Modélisation et Évaluation de Performances de Réseaux

Chapitre 3 : Performances des files d'attente

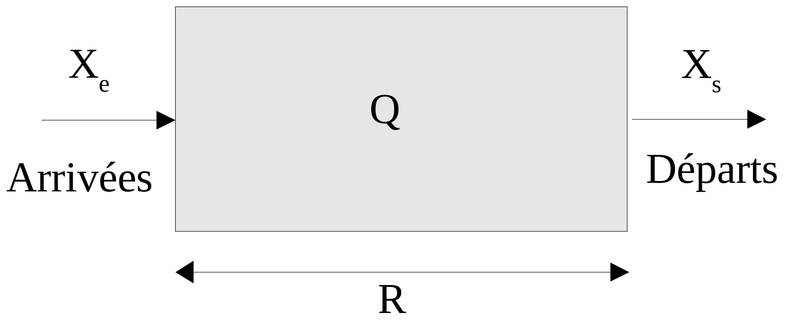
(maj : mars 03)

Fabrice Valois
fabrice.valois@insa-lyon.fr
http://citi.insa-lyon.fr/ ~ fvalois

Paramètres de performances

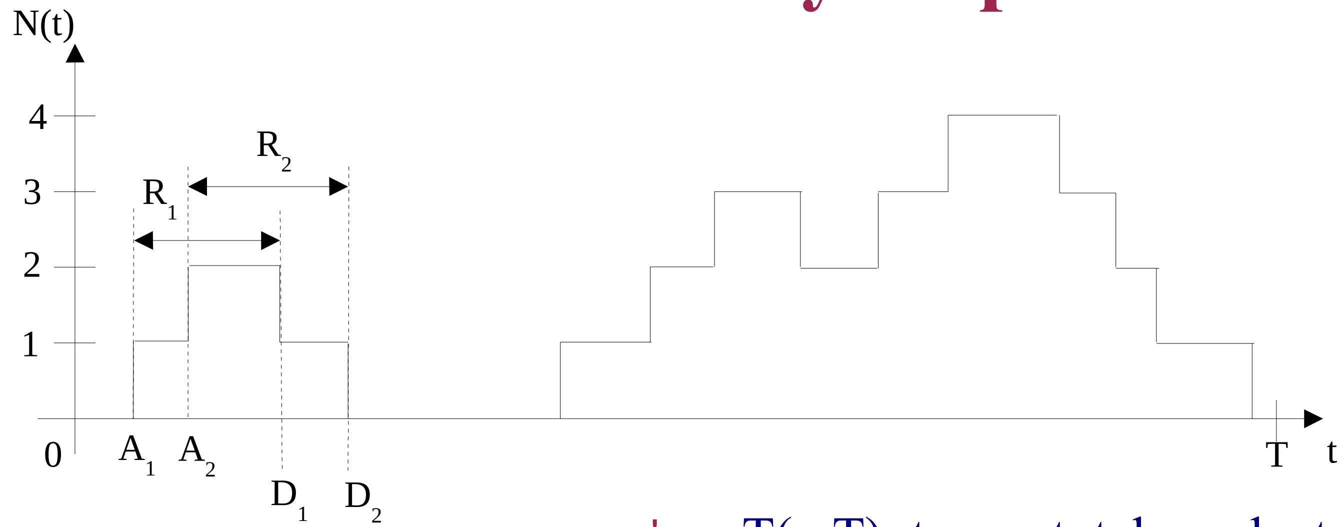
- Analyse opérationnelle
  - En régime transitoire
    - Débit moyen d'entrée
    - Nombre moyen de clients
    - Taux d'utilisation (file d'attente)
    - Débit moyen de sortie
    - Temps de séjour
  - En régime permanent
- Condition de stabilité
- Notion d'ergodicité
- Loi de Little
- Équivalence des instants d'observation d'un système

Analyse opérationnelle



- Analyse opérationnelle = réalisation particulière de l'évolution de ce système pendant une période donnée
  - ⇒ Permet de caractériser le système
- On étudie donc le comportement du système entre t=0 et t=T

Paramètres de l'analyse opérationnelle



- A_k: instant d'arrivée du $k^{ième}$ client
- D_k: instant de départ du $k_{ième}$ client
- R_k: temps de séjour du $k^{ième}$ client dans le système :

$$R_k = D_k - A_k$$

- T: temps total d'observation
- T(n,T): temps total pendant lequel le système contient n clients

$$\sum_{n=0}^{\infty} T(n,T) = T$$

- P(n,T): proportion de temps pendant lequel le système contient n clients

$$P(n,T) = \frac{T(n,T)}{T}$$

- A(T): nombre de clients arrivant dans le système pendant [0,T]
- D(T): nombre de clients quittant le système pendant [0,T]

Paramètres de performances (part. 1)

- Débit moyen d'entrée X_e
  Nombre moyen de clients arrivés dans le système par unité de temps
  $$X_e(T)=\frac{A(T)}{T}$$

- Débit moyen de sortie X_s
  Nombre moyen de clients ayant quitté le système par unité de temps
  $$X_s(T)=\frac{D(T)}{T}$$

Paramètres de performances (part. 2)

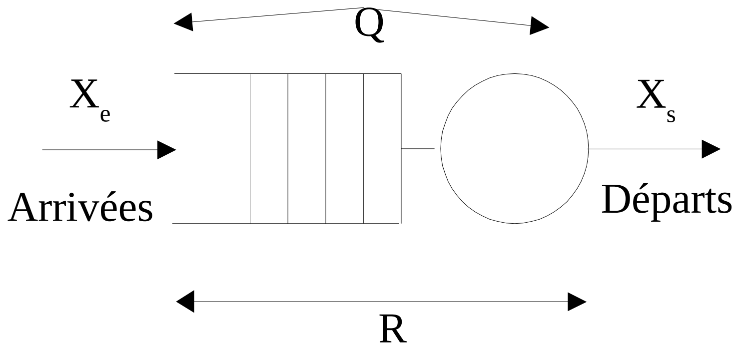
- Nombre moyen de clients Q
  Moyenne temporelle de N(t)

$$Q(T)=\frac{1}{T}\sum_{n=0}^{\infty} n\cdot T(n,T)=\sum_{n=0}^{\infty} n\cdot P(n,T)$$

- Temps moyen de séjour R
  Moyenne (arithmétique) des temps de séjour des clients arrivés dans le système pendant la durée de l'observation

$$R(T)=\frac{1}{A(T)}\sum_{k=1}^{A(T)} R_k$$

Paramètres de performances pour une file simple



- X_e, X_s, Q, R et....
- Taux d'utilisation du serveur U
  Proportion du temps pendant laquelle le serveur est occupé

$$U(T) = \sum_{n=1}^{\infty} P(n, T) = 1 - P(0, T)$$

Paramètres de performances pour un réseau de FA

- On peut considérer plusieurs *niveaux* :
  - Les paramètres de performances du réseau tout entier
  - Les paramètres de performances pour chacune des stations
  - Dans le cas multiclasse :
    - On peut s'intéresser aux paramètres de performances pour chaque classe
    - ... ou toutes classes confondues

Paramètres de performances en régime permanent

- On s'intéresse à l'existence et aux valeurs (éventuelles) des limites lorsque $T \rightarrow \infty$:

$$X_e = \lim_{T \to \infty} X_e(T)$$

$$X_s = \lim_{T \to \infty} X_s(T)$$

$$Q = \lim_{T \to \infty} Q(T)$$

$$R = \lim_{T \to \infty} R(T)$$

$$U = \lim_{T \to \infty} U(T)$$

Stabilité ?

- Notion définie uniquement en régime permanent
- Restrictions : pas de mécanismes de type
  - **Join**
  - **Fork**

- Définition : *Un système est **stable** ssi le débit moyen asymptotique de sortie des clients du système est égal au débit moyen d'entrée des clients dans le système*

$$\lim_{T\to\infty} X_s(T) = \lim_{T\to\infty} X_e(T) \quad \Rightarrow \quad \lim_{T\to\infty} \frac{A(T)}{D(T)}$$

Stabilité d'une file d'attente

- Une file d'attente sera considérée comme stable dès que:

  taux d'arrivées λ < taux de service μ

- i.e. il ne faut pas qu'il arrive, en moyenne, plus de clients dans la file que ce qu 'elle est capable de traiter

Stabilité d'un réseau de FA ouvert

- Propriété : Un réseau de files d'attente monoclasse comportant M stations (chaque station i ayant un taux de service μ_i, C_i serveurs et étant soumise à un taux d'arrivées λ_i d'arrivée des clients) est stable ssi:

$$\lambda_i < C_i.\mu_i \text{ pour tout } i=1, ..., M$$

Notion d'ergodicité

- Paramètres opérationnels ?
  - Issus de l'étude du régime stationnaire
  - Mais cela revient à considérer une évolution *particulière* du système
  - **Question** : toutes les réalisations ont-elles le même comportement asymptotique ?
- Évolution stochastique du système ?

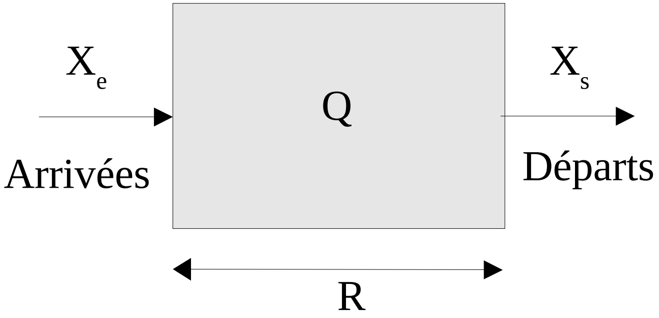
⇒ relation entre paramètres de performances stochastiques et opérationnels ?

Une idée de l'ergodicité

- Système **ergodique** : toutes les réalisations particulières de l'évolution d'un système sont asymptotiquement et statistiquement identiques
- **Ergodicité** ⇔ égalité entre moyennes temporelles et moyennes statistiques
- Pour un système ergodique, les paramètres de performance **opérationnels** sont égaux aux paramètres de performances **stochastiques**
(en régime permanent)
- On peut montrer, qu'un système ergodique tend vers un processus stochastique stationnaire

Loi de Little

- Ne concerne que le régime permanent



- Aucune hypothèse sur la « *boîte noire* »
- Aucune hypothèse sur les variables aléatoires qui caractérisent le système

Loi de Little : énnoncé

- Propriété : *Le nombre moyen de clients Q, le temps moyen de réponse R et le débit moyen X d'un système stable en régime permanent se relient de la façon suivante :*

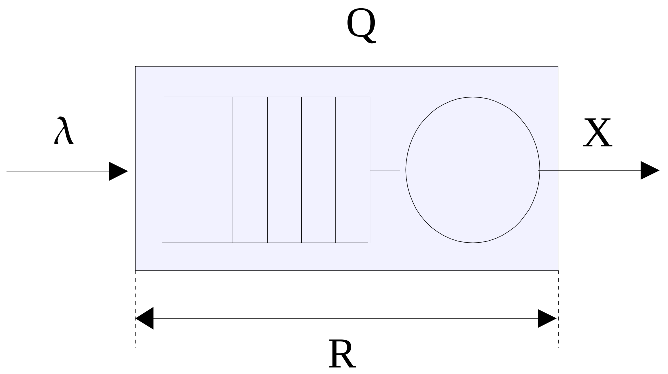
$$Q=R.X$$

- Pseudo-preuve (intuitive mais....)
  - Un client arrivant trouve en moyenne Q clients devant lui
  - Ce client partant laisse derrière lui R.X clients
  - *Donc* dans l'état stationnaire : Q=R.X

Importance de la loi de Little

- Permet de déduire l'une des trois quantités (Q, R, X) en fonction de la connaissance des deux autres
- Peut s'appliquer :
  - Sur une file d'attente (buffer+serveur)
  - Sur la file d'attente (le buffer seulement)
  - Sur le serveur de la file
  - ...

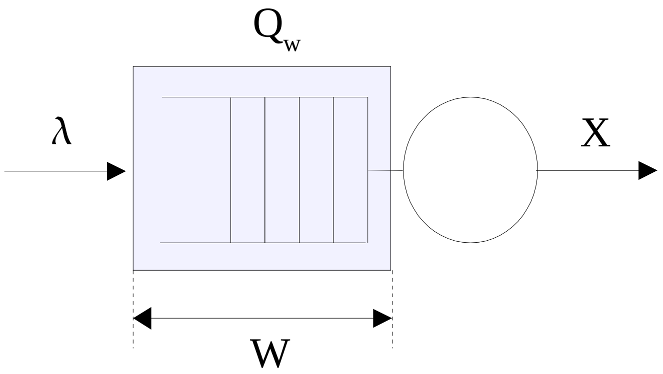
Little sur *file d'attente + serveur*



- Little ⇒relation entre nombre moyen dans la file (en attente ou en service) et le temps moyen total de séjour d'un client dans la file (temps d'attente+temps de service) :

$$Q=R.X=R.\lambda$$

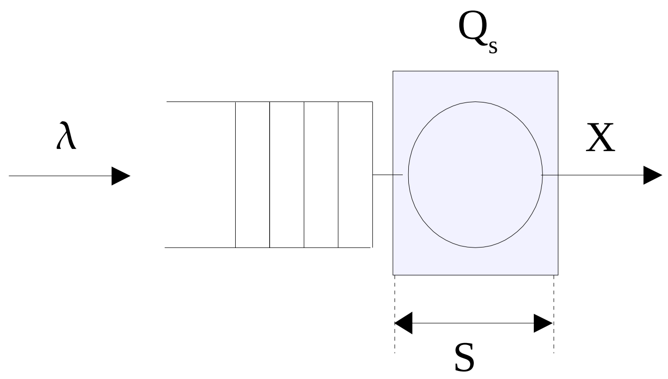
Little sur *file d'attente*



- Ici « *boîte noire* » = buffer
- Little ⇒ relation entre le nombre moyen de clients en attente Q_w et le temps moyen d'attente d'un client avant service W :

$$Q_w = W.X = W\lambda$$

Little sur *serveur de la file*



- Maintenant « *boîte noire* » = serveur
- Little ⇒ relation entre le nombre moyen de clients en service Q_s et le temps moyen de séjour S d'un client dans le serveur :

$$Q_s = S.X = S.\lambda$$

Observations d'un système

- On peut observer un système :
  - À un instant quelconque
  - À un instant d'arrivée d'un client
  - À un instant de départ d'un client

Equivalence des instants d'observations

- Propriété 1 : *Dans un système tel que l'on observe **jamais** l'arrivée simultanée ou le départ simultané de plusieurs clients alors les instants d'arrivée sont équivalents aux instants de départ*
- Propriété 2 : *Dans un système soumis à des arrivées poissoniennes alors les instants d'arrivée sont équivalents aux instants quelconques*