Réseaux Avancés

Cours 9: *Introduction à la théorie des files d'attente*

Osman SALEM
osman.salem@parisdescartes.fr
Maître de Conférences

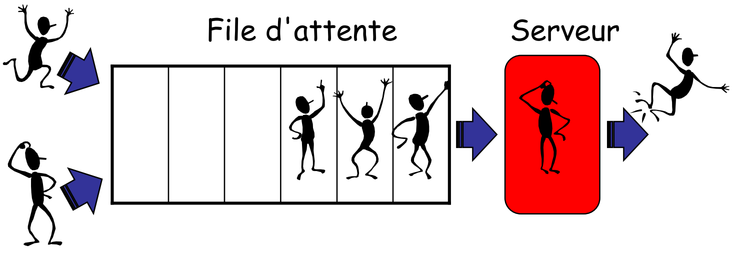
UNIVERSITÉ PARIS DESCARTES
UFR DE MATHÉMATIQUES ET INFORMATIQUE

Théorie des files d'attente

- L'échange d'information entre 2 applications est généralement variable
  - Volume à échanger, fréquences des échanges, etc.
- Ces informations transitent dans différents systèmes, et ne pourront être transmis instantanément
- Elles seront mis dans des files d'attente (queue) avant d'être traitées par ces systèmes
- La théorie des files d'attente permet de modéliser les processus dans lesquels les clients arrivent, attendent leur tour pour être servis, sont servis, et partent

Théorie des files d'attente

- Une file d'attente simple peut être représentée par le schéma suivant :



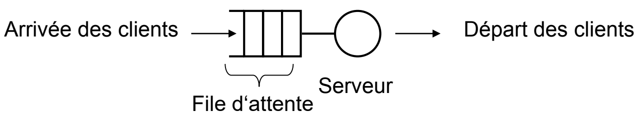
Les files d’attente

- Quelques exemples d’application:

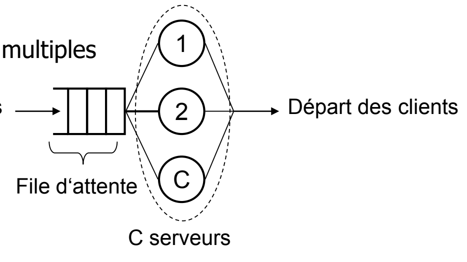
	Clients	Ressource ou serveur	Activité ou service
Systèmes informatiques	processus	processeur	temps de traitement
	demande d'E/S	disque dur	lecture ou écriture
Réseaux de communication	message	réseau	temps de transmission
Système de production	pièce	machine	temps d'usinage
	palette	station de chargement	temps de chargement d'une pièce brute
Guichet SNCF	usager	employé	réservation du billet

Les files d’attente

- La file simple
  - Une file simple (ou une station) est constituée d’une file d’attente (ou buffer) et d’un ou plusieurs serveurs
    - File avec un seul serveur



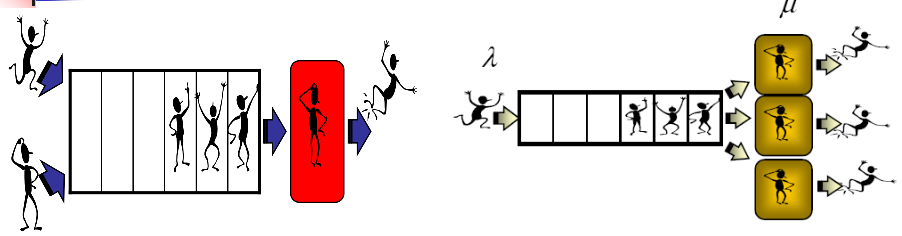
    - File à serveurs multiples



Modèle de Files d’attente

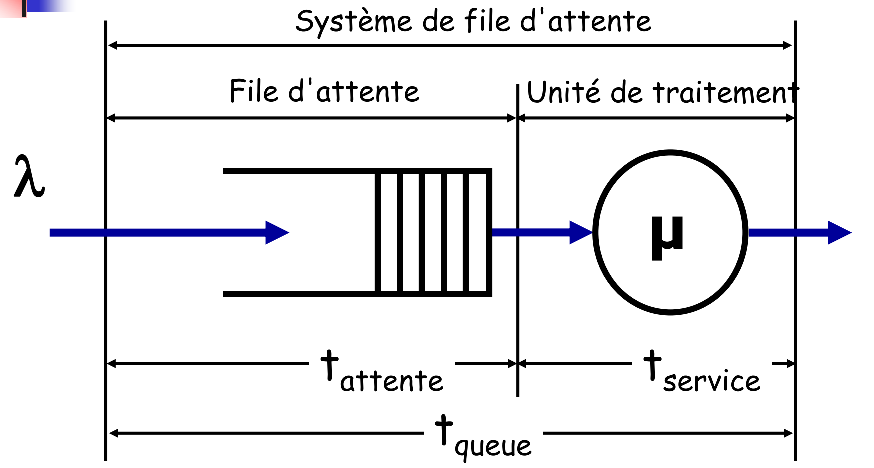
- Une file d’attente se compose
  - Clients (qui demandent un service)
  - Un ou plusieurs serveurs (qui fournissent le service)
  - Une salle d’attente (quand aucun des serveurs n’est disponible)
- Caractéristiques d’une file d’attente
  - Le processus d’arrivées
  - Le processus de service
  - La discipline de service (ordonnancement)
  - La capacité du système
  - Le nombre de serveurs

Caractérisation



- Nombre de serveurs (mono/multiserveurs)
- Arrivées des clients (Processus de Poisson)
- Durée de service (loi exponentielle)
- Discipline de service (FIFO, LIFO, etc.)
- Classes de services (priorité)

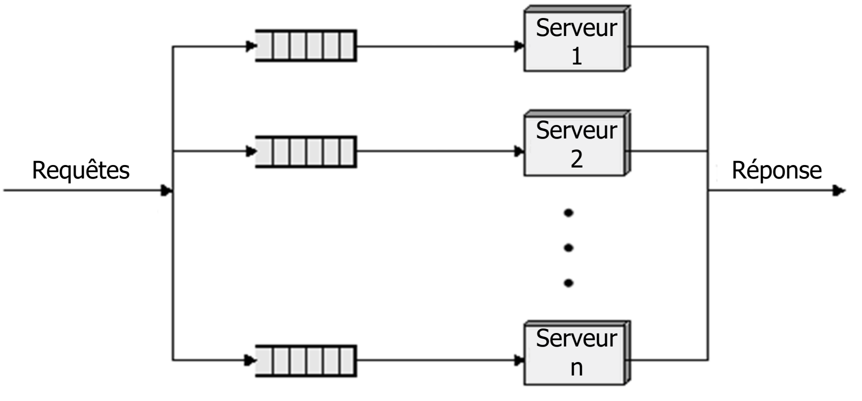
Théorie des files d'attente



$$t_q = t_a + t_s$$

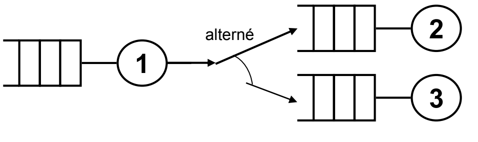
Exemple

- Après un service RR du serveur DNS



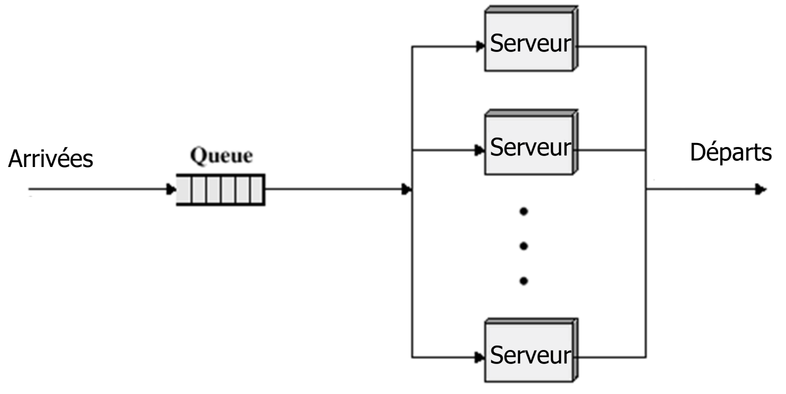
Exemple

- Un service RR du serveur DNS



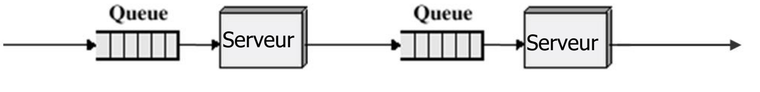
Serveurs parallèles

- Le client passe un seul serveur
- Tous les serveurs fournissent le même service



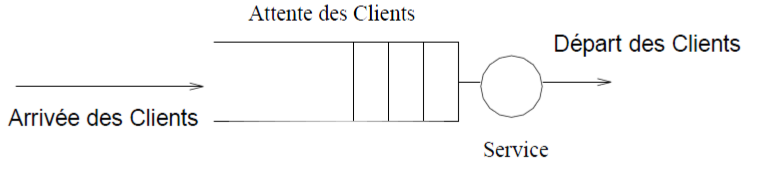
Serveurs en série

- Le client doit visiter plusieurs serveurs successifs dans un ordre fixe
- les serveurs ne fournissent pas le même service



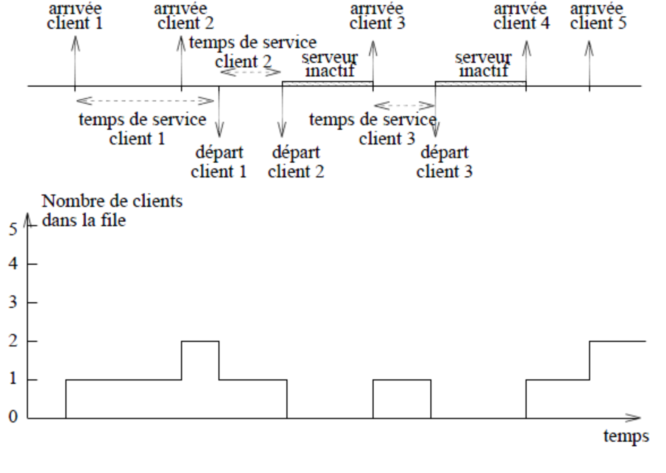
Modélisation

- Le nombre de clients dans la file est modélisé par un processus aléatoire
  - Le plus souvent par une chaîne de Markov
- On cherche le régime stationnaire
  - Pour déduire les performances



Modélisation

- Nombre de Clients dans le système

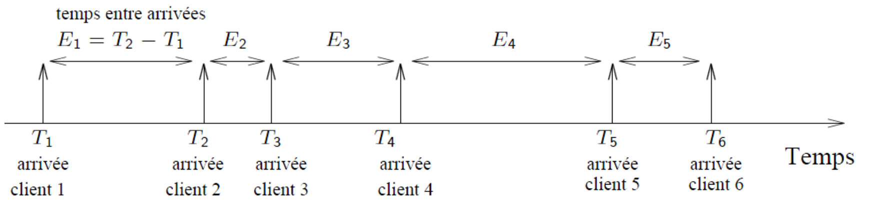


Caractérisation du système

- Arrivées de clients
- Durées de services
- Nombre de serveurs
- Nombre maximum de clients dans le système
- Discipline de service
- Notation de Kendall

Arrivées des Clients

- Processus des arrivées des clients



- On suppose que les temps d'interarrivées successifs E_1, E_2, E_3, etc...
  - sont indépendants entre eux
  - ont même loi de probabilité
- Les arrivées de clients sont alors
  - entièrement caractérisées par la loi des interarrivées

Arrivées des Clients

- Cas particulier important : arrivées poissoniennes
- Les inter-arrivées successives E_1, E_2, E_3, etc... sont indépendantes entre elles et distribuées selon une **loi exponentielle** de paramètre λ
- Le processus des arrivées est une de Poisson de paramètre λ
- λ : nombre moyen des clients/seconde
- La durée moyenne des interarrivées vaut $1/\lambda$ (seconde)

Processus de Poisson

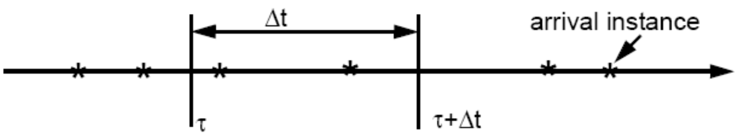
- Scientifique français : Simeon Denis Poisson (1781-1840)
- Pour un nombre d'arrivée Poissonien avec un taux λ, la probabilité de voir ***n*** arrivées dans un intervalle de temps Δt

$$P[A(t+\Delta t) - A(t) = n] = e^{-\lambda \Delta t} \frac{(\lambda \Delta t)^n}{n!} \quad , \qquad n = 0, 1, 2, ...$$

- C'est la distribution de Poisson de paramètre λ
- λ est le taux d'arrivée des clients à la station

Processus du Poisson

- Utilisé pour représenter le nombre des arrivées



- Pour un nombre d'arrivée poissonnien avec un taux λ, la probabilité de voir ***n*** arrivées dans un intervalle de temps **Δt** (très ***petit***)

$$\Pr(n) = \frac{e^{-\lambda\Delta t}(\lambda\Delta t)^n}{n!} \qquad E(n) = \lambda\Delta t$$

$$\Pr(0) = e^{-\lambda\Delta t} = 1 - \lambda\Delta t + \frac{(\lambda\Delta t)^2}{2!} \ldots = 1 - \lambda\Delta t + o(\Delta t) \to \Pr(0) = 1 - \lambda\Delta t$$

$$\Pr(1) = \lambda\Delta t e^{-\lambda\Delta t} = \lambda\Delta t[1 - \lambda\Delta t + \frac{(\lambda\Delta t)^2}{2!} \ldots] = \lambda\Delta t + o(\Delta t) \to \Pr(1) = \lambda\Delta t$$

$$\Pr(>= 2) = \ldots = 0$$

Processus de Poisson & distribution exponentielle

- Les temps inter-arrivées t (temps entre les arrivées) dans le processus du poisson suivent une distribution exponentielle

$$\Pr_0(X \le t) = 1 - e^{-\lambda t}$$

- Intervalle *[0,T]* est divisé en *n* sous intervalle
  - Quand *n* ->∞, *T/n -> 0*

$$\Pr(n) = \frac{e^{-\lambda\Delta t}(\lambda\Delta t)^n}{n!}$$

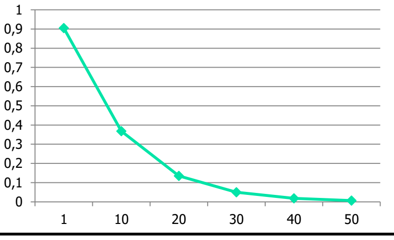
$$\Pr_0(X > t) = P\left[0\ arrivee\ dans\ (0,t)\right] = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}$$

Example

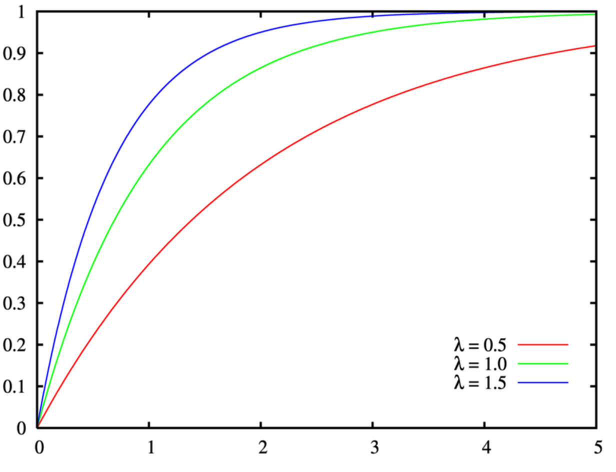
- Pour n=0

$$\Pr_0(t) = P[0\ \ dans\ t] = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}$$

  - 0 arrivée dans l'intervalle [0,t]
- 1 paquet/10sec (0.1 p/s)
- Après une 1 sec, probabilité de 90% pour voir 0 paquet
  - 10% de voir deux paquets dans un intervalle < 1sec
- Après 20sec, probabilité de 10% pour voir 0 paquets
  - 90% de voir deux paquets dans un intervalle < 20sec



Distribution Exponentielle



$$P(T < t) = 1 - e^{-\lambda t} \qquad E[T] = 1/\lambda$$

Processus de Poisson et loi exponentielle

- A partir de la distribution de Poisson, on peut déduire la distribution exponentielle

$$\Pr_0(X \leq t) = 1 - e^{-\lambda t}$$

- Si le processus d'arrivée est poissonnien, le temps inter-arrivée suit la loi exponentielle
- La loi exponentielle est sans mémoire

Propriété de superposition

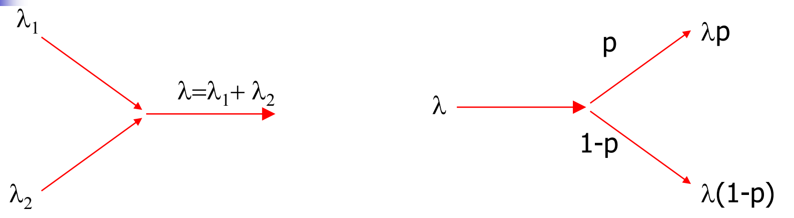
- *Si* $X_1, X_2, \ldots, X_n$ sont des processus de poisson
  - Avec de taux respectifs $\lambda_1, \lambda_2, \ldots \lambda_n$
- Leur superposition

$$S_n = X_1 + X_2 + \ldots + X_n$$

- S_n est un processus de poisson de taux λ:

$$\lambda = \lambda_1 + \lambda_2 + \ldots + \lambda_n$$

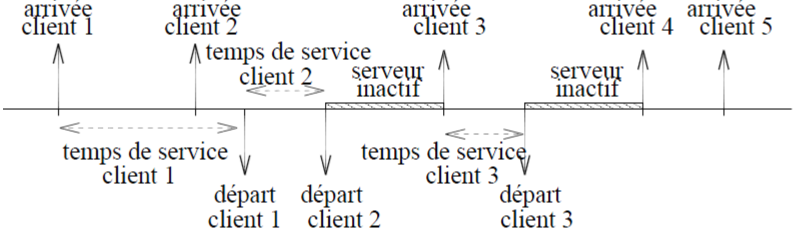
Groupement et dissociation de processus de Poisson



- $X_1,\ldots, X_k$ sont de processus de Poisson avec un taux $\lambda_1,\ldots, \lambda_k$
- Le groupement est un processus de Poisson
  $X= X_1+\ldots+ X_k$
- Avec un taux $\lambda= \lambda_1+\ldots+ \lambda_k$

- X: un processus de poisson de taux λ
- Dissociation en X_1 et X_2 avec des probabilités p et 1-p
- X_1 est Poisson avec un taux $\lambda_1= \lambda p$
  X_2 est Poisson avec un taux $\lambda_2= \lambda(1-p)$

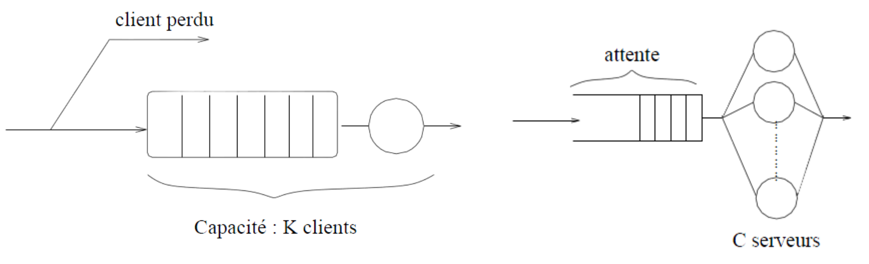
Durées de Services



- Durées de service aléatoire
  - Caractérisées par une loi de probabilité
- Cas particulier
  - Durée de service exponentielle de paramètre µ
  - Durée moyenne de service: 1/µ (seconde)

Nombre de clients dans le système

- Nombre Maximum de Clients dans le système
  - nombre maxi. de clients dans le système = nombre de serveurs + nombre de positions dans le buffer d’attente



Nombre de clients dans le système

- File d’attente fini
  - Si le buffer est plein, le client est perdu
  - Système stable
- File d’attente infini
  - Système à attente pure
  - Condition de stabilité : $\lambda < \mu$ (taux d’arrivée < vitesse de service)

Discipline de service

- D: la discipline de service
  - FIFO: First In First Out ⇔ FCFS (First Come First Serve)
  - LIFO: Last In First Out
  - Priorité: avec différente priorité
  - RSS: Random Selection for Service (aléatoire)
  - Round Robin: QUANTUM ou un slice de temps successivement à chaque client
  - Par défaut, D= FIFO

Notation de Kendall

- David George Kendall (1918 – 2007)
- Notation pour représenter une file d'attente en 1953
- Elle se présente sous la forme de: ***A/S/N/C/P/D***
- A: loi des inter-arrivées (temps entre deux arrivées successives)
  - M: distribution exponentielle (M= Markov)
    - λ est le nombre moyen des arrivées par unité de temps
    - μ est le nombre moyen de départs par unité de temps
  - E_n: distribution d'Erlang à ***n*** phases
  - D: distribution déterministe (constante)
  - U: distribution Uniforme
  - G: distribution quelconque (G= General)
  - GI: distribution quelconque avec inter-arrivées indépendantes

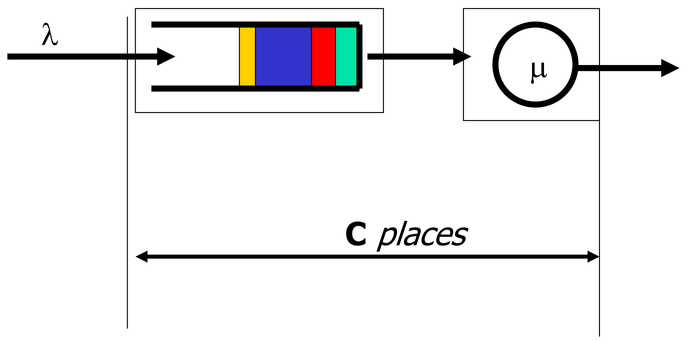
Notation de Kendall

- Elle se présente sous la forme de: ***A/S/N/C/P/D***
- **S**: loi de Service (idem à l'arrivée)
- **N**: nombre de serveurs {1,2,3..,∞}
- **C**: capacité du système (nb max de clients) - Defaults C = ∞
- **P**:la population des usagers
- **D**: la discipline de service
- **C**, **P** et **D** ont une valeur par défaut ∞, ∞ et FIFO

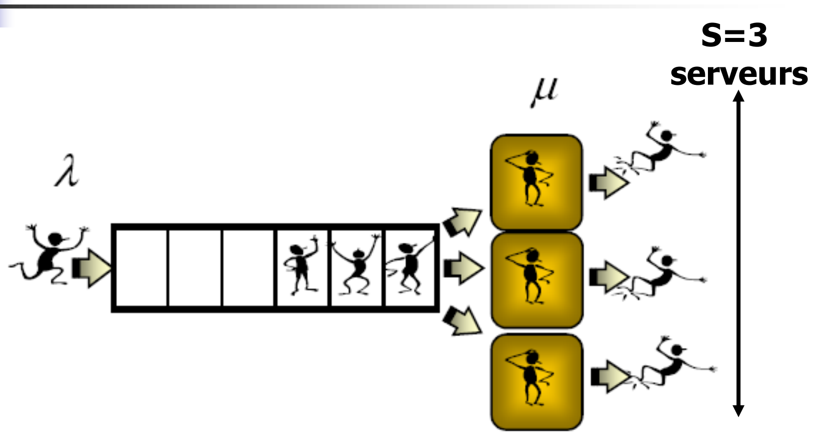
Notation de Kendall

- M/M/1 ⇔ M/M/1/∞/FIFO
  - M: Les arrivées sont Markovien (Distribution exponentielle λ)
  - M: Départs sont Markovien (Distribution exponentielle μ)
  - 1 seule file d'attente de capacité infini et un service PAPS (FIFO ou FCFS)
- M/M/1/*N* : *N* est la capacité du système
  - Nb de clients simultanément présents dans le système
  - S'il y a déjà *N* clients dans le système, les nouveaux clients quittent sans obtenir le service désiré
- M/M/S : S serveurs, capacité infinie, FIFO
- M/M/N/N : C serveurs, pas de buffer d'attente, FIFO
- M/M/N/K (K > N): N serveurs, buffer de taille ***K-N***, FIFO
- M/G/1: arrivées poissonniennes, loi de service quelconque (générale), 1 serveur, capacité infinie, FIFO
- G/G/1 : arrivée quelconque, service quelconque, 1 serveur, capacité infinie, FIFO
- D/M/1 : arrivée déterministe

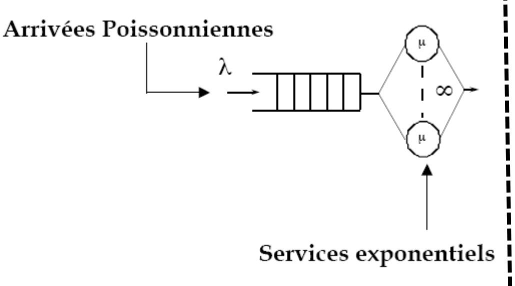
M/M/1/C

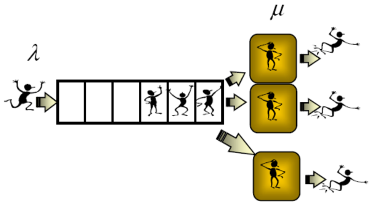


M/M/S (S =3)



M/M/∞

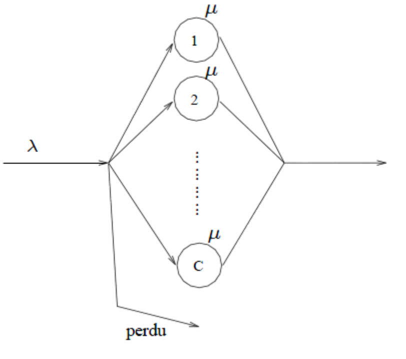




Files Markoviennes

- M/M/1
- M/M/1/K
- M/M/C
- M/M/C/C
- M/M/C/K
- etc.

File M/M/C/C



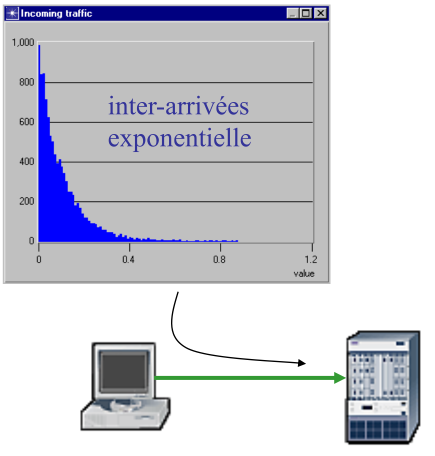
Loi exponentielle et processus de Poisson

- Durée de service exponentielle (paramètre μ)
- Arrivées de clients
  - Inter-arrivées indépendantes et exponentielles (paramètre λ)
  - Arrivées: processus de Poisson (paramètre λ)
- Pourquoi ces cas sont-ils importants ?
  - Propriété sans mémoire (loi exponentielle)
  - C'est la base de représentation markovienne

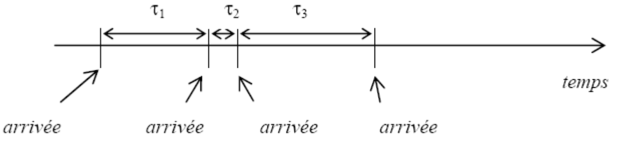
Distribution exponentielle

- La fonction de répartition de la loi exponentielle
  - $F(t) = 1 - e^{-\lambda t}$
- La fonction de densité de la loi exponentielle
  - $f(t) = \lambda e^{-\lambda t}$
- λ est l'inverse de la moyenne
  - $E(x) = \int_0^\infty f(x).dx = \frac{1}{\lambda}$
- Propriété
  - Distribution sans mémoire
    - Le temps restant avant la prochaine arrivée est indépendante du temps écoulé depuis la dernière arrivée

Densité de la loi exponentielle



Distribution exponentielle



- Sans mémoire

$$p(X > t + s \mid X > t) = \frac{p(X > t + s, X > t)}{p(X > t)} =$$

$$\frac{p(X > t + s)}{p(X > t)} = \frac{1 - 1 + e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = p(X > s)$$

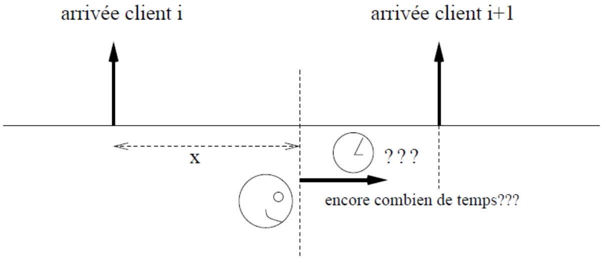
- La distribution exponentielle est la seule distribution continue sans mémoire

Distribution exponentielle

Propriété sans mémoire :

$$\mathbb{P}(X \geq x + x_0 \mid X \geq x) = \mathbb{P}(X \geq x_0) \quad \text{[ne dépend pas de } x\text{]}$$

(i) inter-arrivées exponentielles : le temps qui nous sépare de l'arrivée du prochain client ne dépend pas du temps écoulé depuis l'arrivée du client précédent :

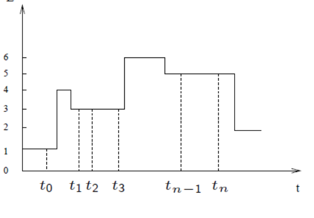


(ii) durées de service exponentielle : le temps qui reste avant que le serveur ne soit libéré ne dépend pas du temps que le client a déjà passé en service

Chaîne de Markov : Définition

- Définition : une chaîne de Markov à temps continu ou à état discret (CMTCED) est un processus $(X_t)_{t\in R}$ qui vérifie la *propriété de Markov* (***sans mémoire***)*:*

$$\Pr\left\{X(t_{k+1}) = x_{k+1} \mid X(t_k) = x_k, \ldots, X(t_0) = x_0\right\}$$
$$= \Pr\left\{X(t_{k+1}) = x_{k+1} \mid X(t_k) = x_k\right\}$$



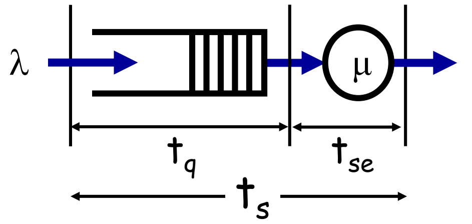
- Le futur est indépendant du passé conditionnellement au présent …

Théorie des files d'attente

- Le modèle ***M/M/1*** est l'un des modèles de file d'attente retenu pour la modélisation des systèmes en télécommunication :
  - M = processus Markovien en entrée (distribution exponentielle des arrivées)
  - M = processus Markovien en sortie
  - 1 = il n'y a qu'un seul processeur (système mono serveur)

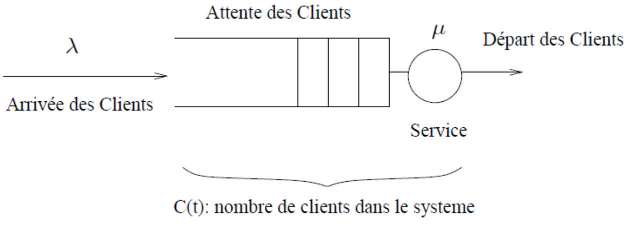
Théorie des files d'attente

- Ce modèle (***M/M/1***) suppose que :
  - Les clients désirant un service arrivent dans une file d'attente qui n'est pas limitée en taille. La file d'attente ne peut pas déborder.
  - Un serveur est placé devant cette file d'attente
  - La discipline de service est FIFO
  - Comment calculer le temps d'attente (T_q) dans la file ?



Premier exemple : file M/M/1

- Premier exemple : cas de la file M/M/1



- File M/M/1 :
  - arrivées Poisson
  - services exponentiel
  - 1 seul serveur
  - buffer d'attente infini
  - discipline FIFO

Le calcul des files d'attente

- Taux d'arrivée des clients = λ
(nombre moyen de clients qui arrive par période de temps)

- Taux de service = μ
(nombre de clients qui peut être servi pendant une période de temps)

- La charge du système $\rho = \lambda / \mu$ (Taux d'utilisation d'un système)
  - Pourcentage de temps où le serveur est occupé

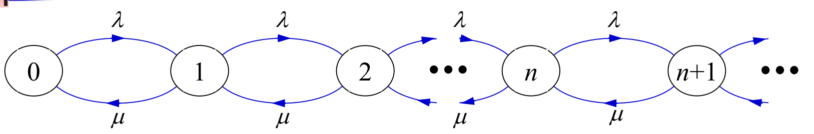
- Longueur moyenne de la file = L_q

- Longueur moyenne du système= L_s

- Temps moyen d'attente dans la file = W_q

- Temps moyen d'attente dans le système = W_s

Processus de naissance et de mort



- Les transitions sont seulement vers les états voisins
  - L'arrivé d'un client (naissance) est représenté par une transition de l'état *n* vers l'état *n+1*
  - La fin de service (mort) d'un client est représenté par une transition de l'état *n* vers l'état *n-1*

Evolution du nombre de clients

Evolution du nbre $C(t)$ de clients

$\pi_n(t) = \mathbb{P}(C(t) = n)$, proba. qu'il y ait n clients dans le système à l'instant t

$*$ proba. arrivée entre t et $t + \Delta t$: $\lambda \Delta t + o(\Delta t), \quad n \to n+1$
$*$ proba. départ entre t et $t + \Delta t$: $\mu \Delta t + o(\Delta t), \quad n \to n-1$
$*$ proba. qu'il ne se passe rien entre t et $t + \Delta t$: $1 - (\lambda + \mu)\Delta t + o(\Delta t), \quad n \to n$

Equations de Chapman Kolmogorov

$$\left\{ \begin{array}{lcl} \pi_n(t + \Delta t) & = & \pi_{n-1}(t)\, \lambda\, \Delta t + \pi_n(t)\, (1 - (\lambda + \mu)\Delta t) \\ & & + \pi_{n+1}(t)\, \mu\, \Delta t + o(\Delta t), \quad n \geq 1 \\ & & \\ \pi_0(t + \Delta t) & = & \pi_0(t)\, (1 - \lambda \Delta t) + \pi_1(t)\, \mu\, \Delta t + o(\Delta t) \end{array} \right.$$

Equations de Chapman Kolmogorov

$$\begin{cases} \pi_n(t+\Delta t) &= \pi_{n-1}(t)\,\lambda\,\Delta t + \pi_n(t)\,(1-(\lambda+\mu)\Delta t) \\ & \quad +\pi_{n+1}(t)\,\mu\,\Delta t + o(\Delta t),\ \ n \geq 1 \\ & \\ \pi_0(t+\Delta t) &= \pi_0(t)\,(1-\lambda\Delta t) + \pi_1(t)\,\mu\,\Delta t + o(\Delta t) \end{cases}$$

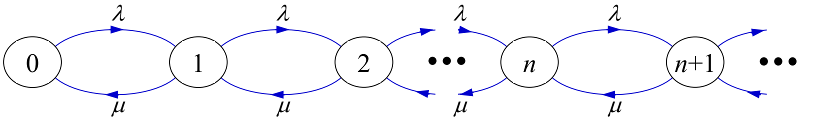
$$\begin{cases} \lim_{\Delta t \to 0} \frac{\pi_n(t+\Delta t)-\pi_n(t)}{\Delta t} &= \pi_{n-1}(t)\,\lambda - \pi_n(t)\,(\lambda+\mu) \\ & \quad +\pi_{n+1}(t)\,\mu,\ \ n \geq 1 \\ & \\ \lim_{\Delta t \to 0} \frac{\pi_0(t+\Delta t)-\pi_0(t)}{\Delta t} &= -\lambda\,\pi_0(t) + \mu\,\pi_1(t) \end{cases}$$

$$\boxed{\begin{cases} \frac{d}{dt}\pi_n(t) = \pi_{n-1}(t)\,\lambda - \pi_n(t)\,(\lambda+\mu) + \pi_{n+1}(t)\,\mu,\ \ n \geq 1 \\ \\ \frac{d}{dt}\pi_0(t) = -\lambda\,\pi_0(t) + \mu\,\pi_1(t) \end{cases}}$$

Etude du régime stationnaire

- Sous une condition que nous avions précisée ($\lambda<\mu$) la distribution du nombre de clients dans la file converge (lorsque *t* tend vers l'infini):

$$\boxed{\pi_n(t) \overset{t\to\infty}{\longrightarrow} \pi_n}$$



Etude du régime stationnaire

Distribution en régime stationnaire :

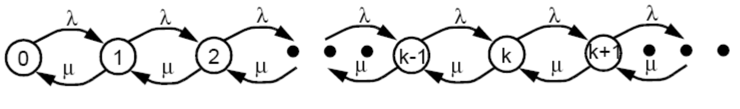
La distribution π_n du système en régime stationnaire est obtenue en résolvant $\boxed{\frac{d}{dt}\pi_n(t) = 0, n = 0, 1, 2, \ldots}$

$$\begin{cases} -\lambda\pi_0 + \mu\pi_1 = 0 \\ \lambda\pi_{n-1} - (\lambda + \mu)\pi_n + \mu\pi_{n+1} = 0 \end{cases}$$

Solution : distribution stationnaire de la M/M/1

$$\boxed{\pi_n = \pi_0\,(\lambda/\mu)^n, n = 0, 1, 2, \ldots}$$

Etude du régime stationnaire



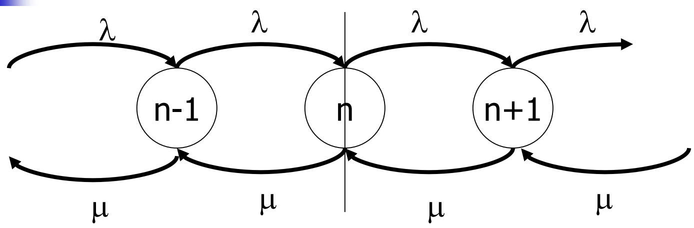
$\text{Taux de } \textit{départ} \text{ n} = P_n(\lambda + \mu)$

$\text{Taux d'arriv}\textit{ées} \text{ n} = \lambda P_{n-1} + \mu P_{n+1}$

$\text{R}\textit{égime permanent } n: \; P_n(\lambda + \mu) = \lambda P_{n-1} + \mu P_{n+1}$

$\text{R}\textit{égime permanent } 0: \; \lambda P_0 = \mu P_1$

Etude du régime stationnaire



$$\lambda P_0 = \mu P_1$$

$$(\lambda + \mu) P_n = \lambda P_{n-1} + \mu P_{n+1}$$

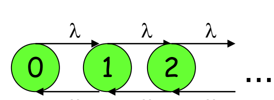
Solution de P_0 et P_n

- Step 1

$$P_1 = \frac{\lambda}{\mu} P_0, \quad P_2 = \left(\frac{\lambda}{\mu}\right)^2 P_{0,} \quad P_n = \left(\frac{\lambda}{\mu}\right)^n P_0$$

- Step 2

$$\sum_{n=0}^{\infty} P_n = 1, \quad then \quad P_0 \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n = 1, \quad \Rightarrow \quad P_0 = \frac{1}{\sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n}$$



Calcul des P_0 et P_n

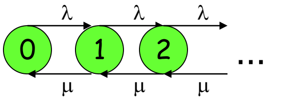
- Step 3

$$\rho = \frac{\lambda}{\mu},\ then\ \sum_{n=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^n = \sum_{n=0}^{\infty}\rho^n = \frac{1-\rho^{\infty}}{1-\rho} = \frac{1}{1-\rho}\ \{\rho < 1\}$$

- Step 4

$$P_0 = \frac{1}{\sum_{n=0}^{\infty}\rho^n} = 1-\rho \quad \text{and} \quad P_n = \rho^n(1-\rho)$$

$$P_n = \rho^n P_0$$



Etude du régime stationnaire

Condition de Stabilité du Système

- sous quelle condition un régime stationnaire s'établit-il ?
- condition de stabilité :

λ	$<$	μ
Taux d'Arrivée	$<$	Taux de Service

- $\rho = \lambda/\mu$ est appelé *trafic offert* ou *facteur de charge*(unité : Erlang) ; le système est stable si $\rho < 1$.

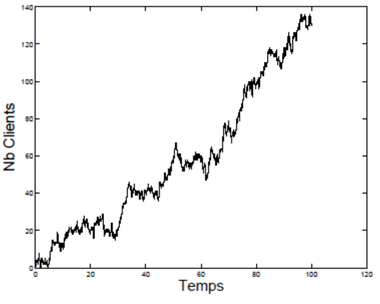
Distribution stationnaire de la M/M/1

- si $\rho < 1$ alors la somme $\sum_{n=1}^{+\infty}\rho^n$ converge $\sum_{n=1}^{\infty}\rho^n = 1/(1-\rho)$;
- distribution du système en régime stationnaire

$$\boxed{\pi_n = (1-\rho)\rho^n, n = 0, 1, 2, \ldots}$$

Etude du régime stationnaire

- Système instable
  - cas d'une file M/M/1 avec ρ≥ 1
  - les clients arrivent plus vite que la capacité de traitement du serveur (λ≥ μ)
  - la file est instable, le nombre de clients dans le buffer d'attente augmente indéfiniment, aucun régime stationnaire ne peut s'établir



$\lambda = 10, \mu = 9$. File instable.

Mesures de performance

- L_s = nb moyen de clients dans le système
- L_q = nb moyen de clients dans la file d'attente
- W_s= temps moyen passé par le client dans le système
- W_q= temps moyen passé par le client dans la file d'attente
- Par définition:

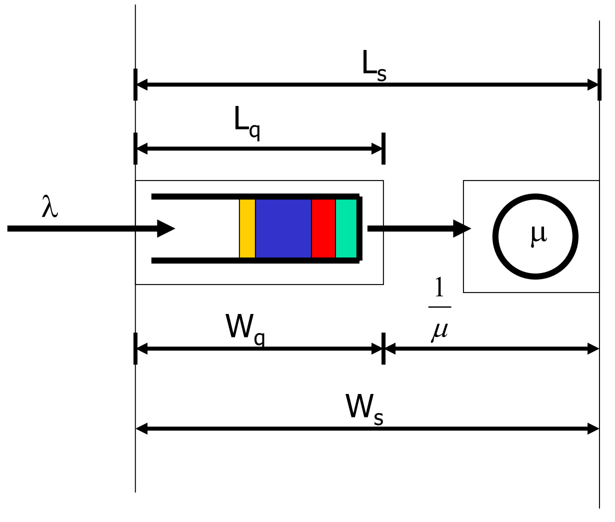
$$L_s = \sum_{n=0}^{\infty} n.p_n$$

$$et$$

$$L_q = \sum_{n=s+1}^{\infty} (n-s).p_n$$

- s est le nombre de serveurs

M/M/1



Calcul de L_s

- L_S: nombre moyen de clients dans le système

$$L_s = \sum_{n=0}^{\infty} nP_n = \sum_{n=0}^{\infty} n\rho^n (1-\rho) = (1-\rho)\rho \sum_{n=1}^{\infty} n\rho^{n-1}$$

$$(1-\rho)\rho \tfrac{d}{d\rho}\left(\sum_{n=0}^{\infty} \rho^n\right) = (1-\rho)\rho \tfrac{d}{d\rho}\left(\tfrac{1}{1-\rho}\right)$$

$$(1-\rho)\rho\left(\tfrac{1}{(1-\rho)^2}\right) = \frac{\rho}{(1-\rho)} = \frac{\lambda}{\mu-\lambda}$$

Performances Moyennes

- Performances Moyennes en Régime Stationnaire
  - Performances moyennes : taux moyen d'utilisation du serveur, délai moyen (attente et service), temps moyen d'attente, etc.
  - La notion de performances *moyennes n'a de sens que* dans le cas d'un système stable
- Taux Moyen d'Utilisation du Serveur
  - Représente la proportion du temps pendant lequel le serveur est actif (système non vide)
  - Taux moyen d'utilisation:

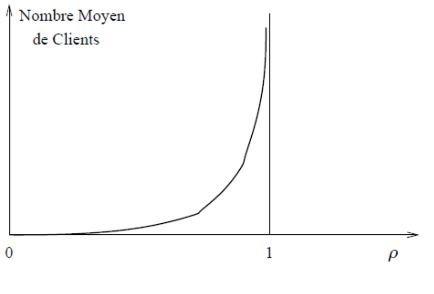
$$\boxed{1 - \pi_0 = \rho = \lambda / \mu}$$

Performances Moyennes

- Nombre Moyen L_S de Clients
  - nombre moyen L_S de clients dans le système

$$\boxed{\text{Nb Moyen de Clients} \qquad L_s = \rho \,/\, (1-\rho)}$$

  - influence du facteur de charge ρ



Performances Moyennes

- Débit Moyen X
- la M/M/1 est un *système sans perte ou système à attente pure*
- pas de blocage des clients (à l'entrée du système) puisque le buffer d'attente est infini
- débit moyen d'entrée/sortie

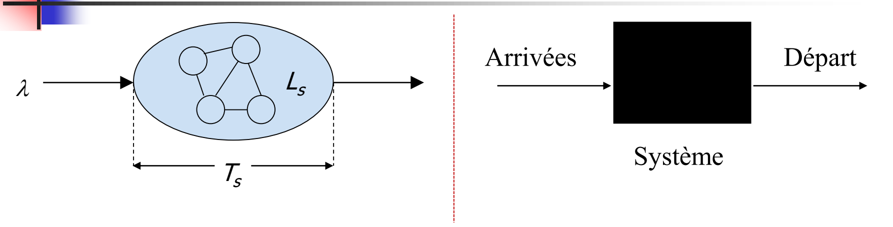
$$\boxed{\text{débit moyen : } \quad X = \lambda}$$

Performances Moyennes

- Avant de donner l'expression du *temps de séjour moyen W_s dans le système nous allons énoncer une formule* importante.
- Une formule importante : formule de Little
  - Le nombre moyen des clients et les temps d'attente sont liées par la célèbre formule de little:
  - $L_x = \lambda * W_x$

$$\underbrace{\text{Débit Moyen}}_{\text{unité : clients/sec.}} = \frac{\overbrace{\text{Nombre Moyen de Clients dans le Système}}^{\text{unité : nombre de clients}}}{\underbrace{\text{Temps de Séjour Moyen dans le Système}}_{\text{unité : sec.}}}$$

Formule de Little



- $W_s = W_q + (1/\mu)$
- Formule de Little:
  nb moyen de clients dans le système = Taux d'entrée x Temps de réponse
- Pour tout système de files d'attentes:

$$L = \lambda_{eff} W$$

$$\Rightarrow L_s = \lambda_{eff} W_s \qquad et \qquad L_q = \lambda_{eff} W_q$$

Ws, Wq et Lq

$$W_s = \tfrac{L_s}{\lambda} = \left(\tfrac{\lambda}{\mu-\lambda}\right)\left(\tfrac{1}{\lambda}\right) = \tfrac{1}{\mu-\lambda}$$

$$W_q = W_s - \tfrac{1}{\mu} = \left(\tfrac{1}{\mu-\lambda}\right) - \left(\tfrac{1}{\mu}\right) = \tfrac{\lambda}{\mu(\mu-\lambda)}$$

Taux d'utilisation $= 1 - p_0 = \rho = \dfrac{\lambda}{\mu}$

Ws, Wq et Lq

$$L_q = \lambda W_q = \lambda \frac{\lambda}{\mu(\mu-\lambda)} = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

$$= \sum_{i=0}^{\infty} (i-1)\, p_i = \frac{\rho}{1-\rho} - (1-p_0)$$

$$= \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho}$$

Autres performances

$$P[\text{client doit attendre}] = 1 - p_0 = \rho$$

$$P[\text{serveur occupé}] = 1 - p_0 = \rho$$

$$P[N > K] = 1 - P[N \le k] = 1 - \sum_{i=0}^{k} P_i = 1 - P_0 \frac{1-\rho^{k+1}}{1-\rho} = \rho^{k+1}$$

- Débit
  - Le service s'effectue avec un taux dans chaque état où le système contient au moins un client

$$X = P[N > 0]\mu = \sum_{i=1}^{\infty} P_i.\mu = [1 - P_0].\mu = \rho.\mu = \lambda$$