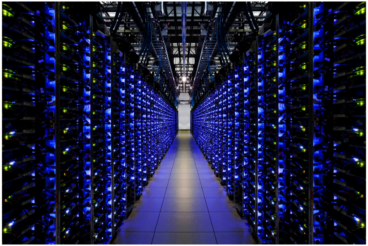
Probabilités et Informatique: du modèle à l'algorithme

Nicolas Gast

Inria, Grenoble, France

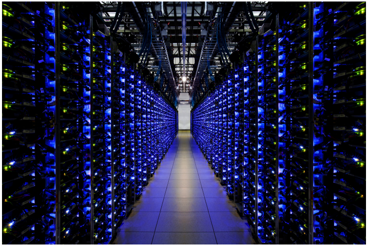
Modèles pour les grands systèmes



Objectifs des modèles:

- Dimensionner le système
- Prédire la performance
- Contrôler le système

Modèles probabilistes pour les grands systèmes



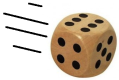
Objectifs des modèles:

- Dimensionner le système
- Prédire la performance
- Contrôler le système

Ces systèmes sont-ils aléatoires?

Les systèmes informatiques sont *a priori* déterministes mais:

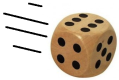
- Parfois le hasard sert à modéliser ce qu'on ne connaît pas exactement.



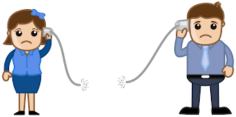
Ces systèmes sont-ils aléatoires?

Les systèmes informatiques sont *a priori* déterministes mais:

- Parfois le hasard sert à modéliser ce qu'on ne connaît pas exactement.

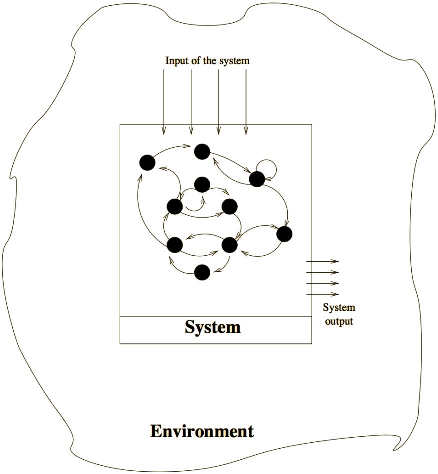


- Parfois, le hasard est introduit intentionnellement



On est amené à construire un modèle de notre système

Modèle = système à événements discrets



- But de ces modèles : comprendre, dimensionner, optimiser.
- Outils : Simulations / Solutions analytiques

Outline

Le modèle d'Erlang [1909]

Modèle des arrivées: Processus de Poisson

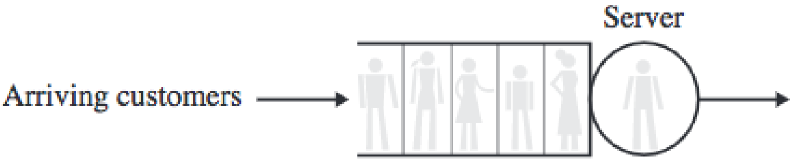
$$\mathbb{P}\left[\text{Un appel pendant } dt\right] = \lambda dt + o(dt).$$

Indépendance temporel.

Temps de traitement d'un appel = 1 unité de temps.

Modélisation par des files d'attente

Exemple: file d'attente.

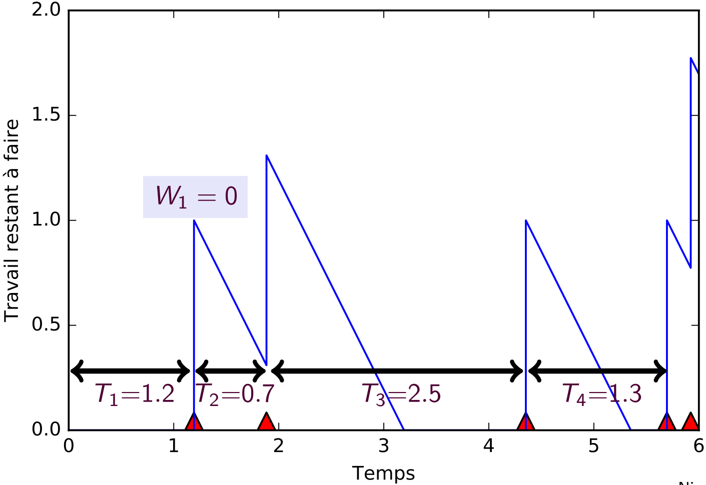


(Source: http://www.cs.cmu.edu/~harchol/PerformanceModeling/)

La formule de Lindley [1952]

- T_n le temps entre les instants d'arrivée des clients $n-1$ et n
- W_n le temps d'attente du $n$ième client.

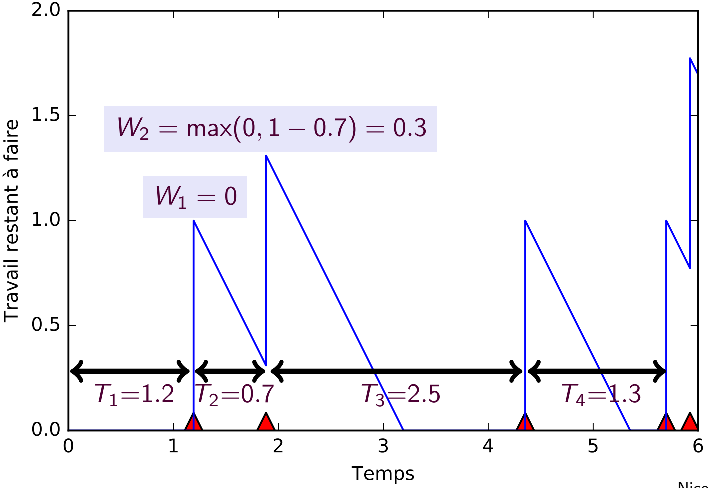
$$W_{n+1} = \max(0, W_n + 1 - T_{n+1})$$



La formule de Lindley [1952]

- T_n le temps entre les instants d'arrivée des clients $n-1$ et n
- W_n le temps d'attente du $n$ième client.

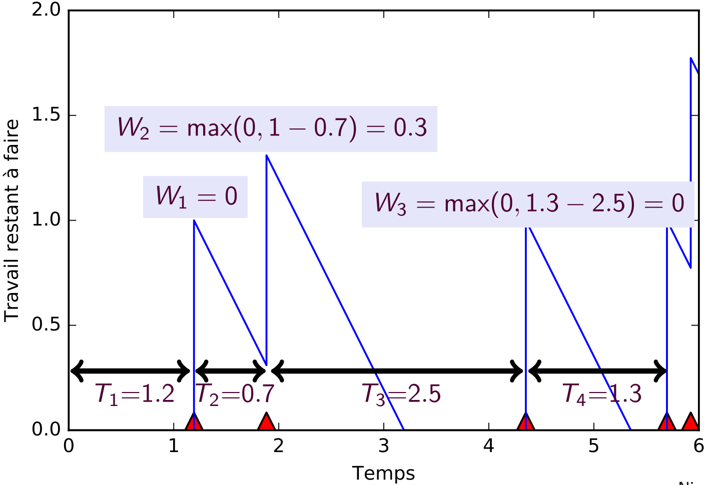
$$W_{n+1} = \max(0, W_n + 1 - T_{n+1})$$



La formule de Lindley [1952]

- T_n le temps entre les instants d'arrivée des clients $n-1$ et n
- W_n le temps d'attente du $n$ième client.

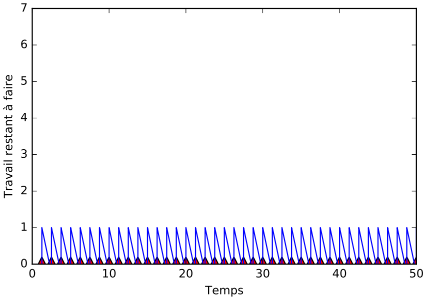
$$W_{n+1} = \max(0, W_n + 1 - T_{n+1})$$



Les modèles déterministes et stochastiques sont différents

Arrivées Déterministe

Une arrivée toutes les 1.25 secondes



Aucune attente : $\mathbb{E}[W_n] = 0$

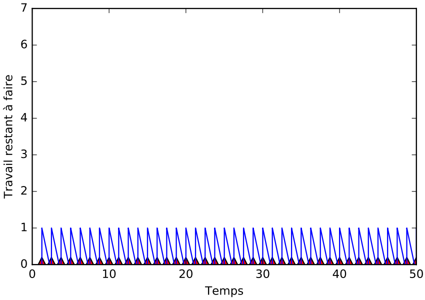
Arrivées selon un proc. de Poisson

Poisson ($\lambda = 0.8$)

Les modèles déterministes et stochastiques sont différents

Arrivées Déterministe

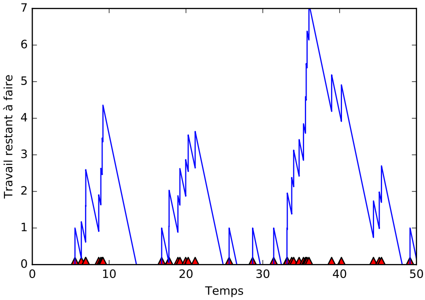
Une arrivée toutes les 1.25 secondes



Aucune attente : $\mathbb{E}[W_n] = 0$

Arrivées selon un proc. de Poisson

Poisson ($\lambda = 0.8$)



$$\mathbb{E}[W_n] = \frac{\lambda}{2(1-\lambda)} = 2$$

Si le temps de service n'est pas déterministe, le temps d'attente est liée à la variabilité

Formule de Pollaczek–Khinchine [1930] https://en.wikipedia.org/wiki/Pollaczek-Khinchine_formula

- Les clients arrivent selon un processus de Poisson d'intensité λ
- Le temps pour servir un client a une variance V (et est de moyenne 1)

On note W le temps d'attente moyen. On a:

$$W = \frac{\lambda}{2(1-\lambda)}(1+V).$$

Remarque: W tend vers l'infini si $\lambda \to 1$ ou $V \to 1$.

Les réseaux de file d'attente

Développements dans les années 50-90

- A.K. Erlang (1909)
- D. Kendall (53) – théorie des files d'attente (notation de Kendall)
- L. Kleinrock (60s-70s)

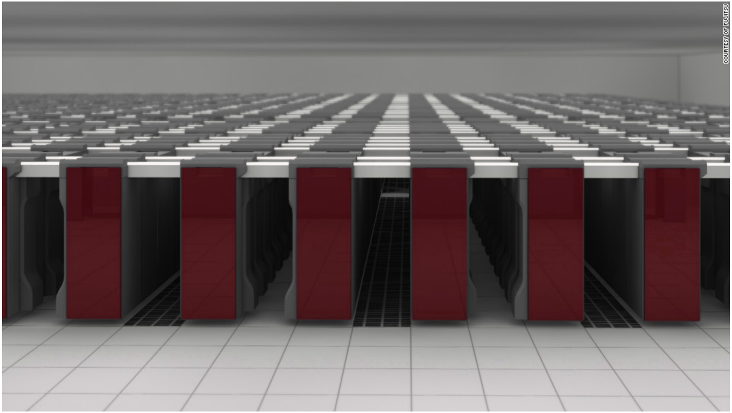
> *"Basically, what I did for my PhD research in 1961–1962 was to establish a mathematical theory of packet networks..."*

(méthodes analytiques, analyse complexe).

Méthodes analytiques (encore aujourd'hui utilisées : formules d'Erlang).

Outline

Problème: passage à l'échelle des modèles



K-Computer, 88 128 processeurs

(crédit : http://edition.cnn.com/2012/03/29/tech/super-computer-exa-flop/)

$$\mathbb{P}\left[X_1 = i_1 \dots X_k = i_k\right] \neq \mathbb{P}\left[X_1 = i_1\right] \dots \mathbb{P}\left[X_k = i_k\right]$$

Certains systèmes se simplifient quand leur taille grandit.

Système stochastique

- N objets
- Comportement aléatoire

Certains systèmes se simplifient quand leur taille grandit.

Système stochastique

- N objets
- Comportement aléatoire

$\xrightarrow{N \to \infty}$

Système dynamique (équation différentielle ordinaire)

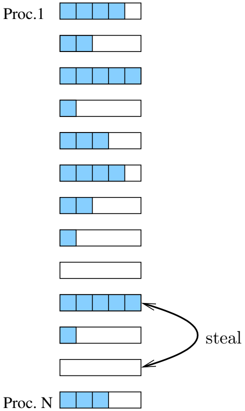
- Comportement déterministe

- Exemples: physique statistique, réactions chimiques, réseaux informatiques, propagation de virus, etc.

https://fr.wikipedia.org/wiki/Algorithme_de_colonies_de_fourmis

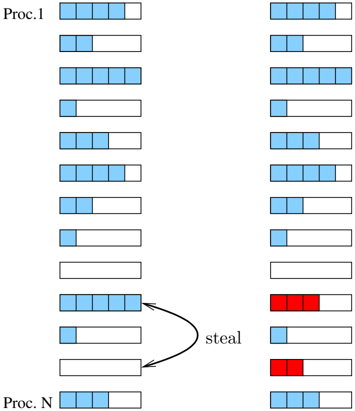
Condition: Interactions localisée et anonymes.

Exemple: modèle de *vol de travail*



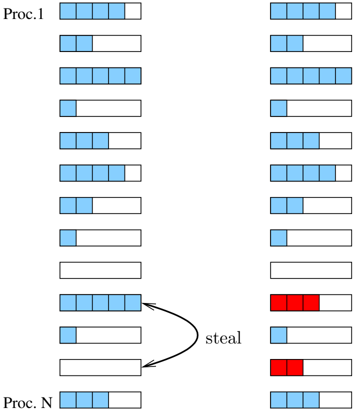
$(4, 2, 5, 1, 3, 4, 2, 1, 0, 5, 1, 0, 3)$

Exemple: modèle de *vol de travail*



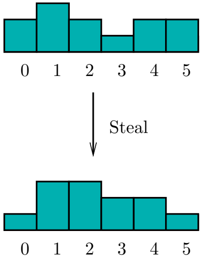
$(4, 2, 5, 1, 3, 4, 2, 1, 0, 5, 1, 0, 3) \to (4, 2, 5, 1, 3, 2, 1, 0, 3, 1, 2, 3)$

Exemple: modèle de *vol de travail*



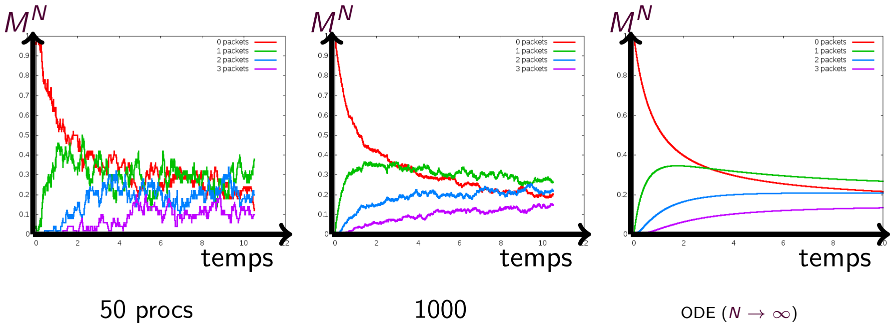
$(4, 2, 5, 1, 3, 4, 2, 1, 0, 5, 1, 0, 3) \to (4, 2, 5, 1, 3, 2, 1, 0, 3, 1, 2, 3)$

Empirical measure : M^N



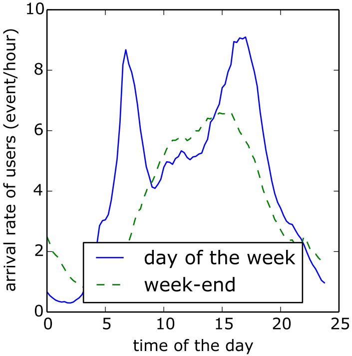
$(2, 3, 2, 1, 2, 2) \to (1, 3, 3, 2, 2, 1)$

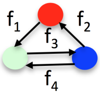
Exemple: modèle de *vol de travail*



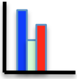
$$\mathbb{E}\left[M^N\right] = m + \frac{c}{N} + O\Big(\frac{1}{N^2}\Big)$$

Autre exemple d'application : Prédiction pour les vélib





$$\frac{dx_1(t)}{dt} = F_1(x_1(t), x_2(t), x_3(t))$$
$$\frac{dx_2(t)}{dt} = F_2(x_1(t), x_2(t), x_3(t))$$
$$\frac{dx_3(t)}{dt} = F_3(x_1(t), x_2(t), x_3(t))$$



$$\mathbb{P}\left[\begin{array}{r} \geq 1 \text{ vélo dispo dans 1h} \\ \text{à la station } x \end{array} \;\middle|\; \#\text{vélos dispo} \right]$$

Outline

Conclusion

- Prendre en compte l'aléatoire est important pour comprendre des systèmes *a priori* déterministes.
- Applications de l'équipe : calcul haute performance, optimisation énergétique, réseaux sans fils.
- Autres exemple d'application : graphes, réseaux sociaux.

https://team.inria.fr/polaris/