Modèles markoviens de transfert de charges dans les réseaux informatiques.

Maryse Béguin

Le problème du transfert de charge dans un réseau informatique, est un problème complexe ayant donné lieu à de nombreux travaux de recherche. Un tour d'horizon peut être trouvé dans [3]. L'étude présentée ici modélise un transfert de charge sur un réseau de n processeurs totalement connectés. Chaque processeur peut accueillir au plus K tàches. Les performances du système sont évaluées par les indices suivants : (cf. [4]) probabilité stationnaire pour un processeur d'accueillir i tàches, notée $P_i(\lambda)$, probabilité de rejet, notée $Prjt$, nombre moyen de tàches traitées par unité de temps, noté $Ttsk$, temps moyen de réponse noté Mrt. Les derniers indices sont reliés par la formule de Little [2]. Le but de cette étude est de mesurer les répercussions du transfert de charge en comparant les valeurs des indices obtenues avec transfert avec celles obtenues sans transfert. En particulier, le comportement asymptotique pour des systèmes massivement parallèles peut être étudié et interprété. Ces comparaisons permettent d'obtenir des bornes supérieures sur les bénéfices que l'on peut attendre d'un réel transfert. Elles permettent également d'étudier l'opportunité du transfert selon les valeurs des paramètres du système.

Dans cette étude, les hypothèses supplémentaires sont les suivantes.

Chaque processeur génère des tàches, et le temps séparant deux générations est un temps aléatoire modélisé par une distribution exponentielle de paramètre λ. Le temps de service de chaque tàche sur un processeur est modélisé par un temps aléatoire de distribution exponentielle de paramètre μ.

Deux processeurs se transfèrent instantanément des tâches dans les deux situations suivantes. Quand une tâche arrive sur un processeur dont la charge devient j, cette tâche est transférée sur un processeur de charge $j-2$ s'il existe. Parmi les processeurs de charge $j-2$, le processeur qui reÃ§oit la tâche supplémentaire est choisi au hasard. Quand une tâche se termine sur un processeur dont la charge devient j, une tâche en provenance d'un processeur de charge $j-2$ est transférée. Parmi les processeurs de charge $j+2$, celui qui transfère sa dernière tàche arrivée est choisi au hasard.
Il n'y a pas de file d'attente partagée par l'ensemble des processeurs et un processeur de charge K ne génère plus de nouvelles tàches.

Le processus étudié est le processus $\{X_t, t \geq 0\}$, qui à chaque instant fait correspondre le n-uplet dont la i-ème coordonnée représente la charge du i-ème processeur. Ces hypothèses de modélisation conduisent à un processus de Markov traduisant l'évolution de la charge de l'ensemble des processeurs au cours du temps.

Dans la situation sans transfert, les n processeurs se comportent comme n files indépendantes $M/M/1/K$ de taux d'arrivée λ, et de taux de service μ.

Pour étudier le système avec transfert, il est opportun d'utiliser les procédés d'agrégation [5]. L'observation clé est de constater que le nombre total de tâches présentes dans le système évolue comme un processus de naissance et de mort ([1]) sur $\{0, \ldots, Kn\}$ (pour $\mu = 1$) avec
des taux de naissance (de j à $j+1$)

$$\lambda(j) = \begin{cases} n\lambda & \text{pour } j = 0, \cdots, (K-1)n \\ (Kn-j)\lambda & \text{pour } j = (K-1)n, \cdots, Kn-1 \end{cases}$$

et des taux de mort (de j à $j-1$)

$$\mu(j) = \begin{cases} j & \text{pour } j = 1, \cdots, n \\ n & \text{pour } k = n+1, \cdots, Kn \end{cases}$$

Soit $(p_i)_{j=0 \ldots Kn}$ la mesure stationnaire de ce prooessus de naissance et de mort.

Proposition 1 *Les probabilitieś $P_j(\lambda)$ sont reliées à ces quantités de la fagon suivante :*

$$P_0(\lambda) = \sum_{i=0}^{n-1} \frac{n-i}{n} p_i \quad , \quad P_K(\lambda) = \sum_{i=0}^{n-1} \frac{n-i}{n} p_{Kn-i}$$

$$\forall 0 < j < K, \quad P_j(\lambda) = \sum_{i=0}^{n} \frac{i}{n} p_{(j-1)n+i} + \sum_{i=1}^{n-1} \frac{n-i}{n} p_{jn+i}$$

Par suite, tous les indices de performance peuvent s'exprimer en fonction des paramètres λ et n.
Posons

$$\rho = p_{Kn} \frac{1}{\lambda^{K_p}} = \lambda^{Kp} p_0.$$

$$G(n, \lambda) = \sum_{j=0}^{n-2} \frac{n^j}{j!} \left(\frac{1}{\lambda}\right)^{Kp-j},$$

$$F(n, \lambda) = \sum_{j=n-1}^{Kp-1} \frac{1}{\lambda^{Kp-j}},$$

$$H(n, \lambda) = \frac{1}{n} \left(\sum_{j=n}^{K_{p-1}} \frac{j}{\lambda^{K_{p-j}}} + (Kn-j)\lambda^{K_{p-j}} \right).$$

On obtient,

Théorème 1 *Pour le modèle n processeurs avec transfert, les valeurs des autres indices de performance sont les suivantes*

$$
\begin{aligned}
Prjt &= \left[G\left(n, \frac{1}{\lambda}\right)\left(1 - \frac{1}{\lambda}\right) + \frac{n^n}{n!}\lambda^{Kp-n+1} \right]\rho \quad , \\
Ttsk &= n\lambda(1 - Prjt) \quad , \\
Nprt &= \rho n \left[G\left(n, \frac{1}{\lambda}\right)\left(K - \frac{1}{\lambda}\right) + \frac{n^n}{n!}\left(\frac{K}{2} + K\lambda^{Kp-(n-1)}\right) + H(n,\lambda) \right] \quad , \\
Mrt &= \frac{\lambda G(n,\lambda) + G\left(n, \frac{1}{\lambda}\right)\left(K - \frac{1}{\lambda}\right) + \frac{n^n}{n!}\left(K\lambda^{Kp-n-1} + \frac{K}{2} + H(n,\lambda)\right)}{\lambda G(n,\lambda) + G\left(n, \frac{1}{\lambda}\right) + \frac{n^n}{n!}\left(1 + F(n,\lambda) + F\left(n, \frac{1}{\lambda}\right) - \lambda^{Kp-n+1}\right)} \quad .
\end{aligned}
$$

Théorème 2 *Pour n grand, le système avec transfert, présente trois types de comportement*
Système non saturé : $\lambda < 1$

$$\lim_{n \to +\infty} Prjt = 0$$

$$\lim_{n \to +\infty} Mrt = 1$$

Valeur critique : $\lambda = 1$

$$\lim_{n \to +\infty} Prjt = 0$$

$$\lim_{n \to +\infty} Mrt = \frac{K}{2}$$

Système saturé : $\lambda > 1$

$$\lim_{n \to +\infty} Prjt = 1 - \frac{1}{\lambda}$$

$$\lim_{n \to +\infty} Mrt = K - \frac{1}{\lambda}$$

Des preuves numériques montrent que la convergence est extrêmement rapide. Seuls les principaux resultats mathématiques ont été rapportés ici, et il convient de faire une étude comparative plus détaillée des résultats avec et sans transfert pour approfondir le sujet ([?]).

Bibliographie

[1] *Barusha-Reid A. T.*, Elements of the theory of Markov Processes, Mc Graw-Hill, New-York,1960

[2] *Béguin M.*, Transfert de charge dans un réseau de processeurs totalement conectés, Rapport technique MAI-IMAG, Grenoble, no 28, (1996).

[3] *Béguin M. f Vincent J.M. , Ycart B.*, Markovian models for load transferring, Rapport technique MAI-IMAG, Grenoble, no 14, (1995).

[4] *Bertzekas D.P.*; *Tsitsiklis, J.N.*, Parallel and distributed computation, Prentice-Hall,New York,(1989).

[5] *Little J.D.C.*, A proof of the queueing formula $L = \lambda W$, Oper. Res., no 9, 383-387, (1961)

[6] *Rosenblatt M.*, Functions of a Markov process that are Markovian, Journal of Mathematics and Mechanics, 4, 585-596, vol8, (1959).

U.F.R. Mathématiques et Informatique
LMC-IMAG B.P. 53, 38041 Grenoble Cedex
Maryse.Beguin@imag.fr