Séminaire Mathématique de Béjaïa (LaMOS)
Volume 11, 2012, pp. 53 - 56

12

File d'attente à un seul serveur avec rappels et vacances : Approche régénérative de Markov

Mohamed BOUALEM

Laboratoire de Modélisation et d'Optimisation des Systèmes LAMOS
Université de Béjaïa 06000, Algerie.
email : `robertt15dz@yahoo.fr`

Résumé Dans ce papier, nous considérons l'analyse stationnaire plus détaillée, du modèle d'attente $M/G/1$ avec rappels classiques et vacances du serveur, basée sur les processus régénératifs de Markov. Nous avons obtenu des formules explicites pour la distribution limite de l'état du serveur, la décomposition stochastique et quelques mesures de performance.

Mots clés : Modèles d'attente avec rappels, vacances exhaustives, chaîne de Markov, processus régénératifs.

12.1 Introduction

Les modèles de files d'attente avec rappels et vacances sont caractérisés par la présence simultanée des phénomènes de répétition de demandes et de vacances à la fois. Dans ces modèles, durant la période de vacances, le serveur est occupé par les tâches supplémentaires, ainsi il n'est pas disponible aux nouvelles arrivées de clients primaires ni secondaires. Dans ce cas, tout client qui trouve le serveur non disponible (occupé ou en vacance) est bloqué, alors il quitte la zone de service et rappelle à des intervalles de temps aléatoires, jusqu'à ce qu'il le trouve oisif pour qu'il puisse être servi. La littérature liée à ces systèmes d'attente est vaste et riche, alors il est possible de trouver un grand nombre de variantes et de généralisations. Artalejo [3] a considéré une file $M/G/1$ avec rappels constants et politique de vacances exhaustive. Plus tard, Aissani [1] a considéré une file d'attente avec rappels constants, arrivées par lots et vacances exhaustives du serveur. Les temps de service et les périodes de vacances sont arbitrairement distribués. Il a obtenu les fonctions génératrices du nombre de clients dans le système et dans l'orbite en régime stationnaire. Récemment, Aïssani [2] considère une file d'attente $M/G/1$ avec la politique de rappels constants et vacances du serveur, quand les temps de rappels, les temps de service et les temps de vacances sont distribués arbitrairement. La distribution du nombre de clients dans le système en régime stationnaire est obtenue en termes de fonction génératrice. Par la suite, il donne une approximation pour une telle distribution. Cette étude constitue une analyse complète des rappels constants dans le cas de ces modèles.

Le but de notre travail est de faire une analyse stationnaire du système d'attente $M/G/1$ avec rappels classiques et vacances du serveur. La raison fondamentale pour l'analyse de ce type de modèles est que leur structure apparaît dans plusieurs représentations de systèmes de télécommunication, de production et les problèmes de contrôle de qualité. La majorité des études précédentes donnent seulement des solutions en termes de fonctions génératrices. Dans ce travail, nous présentons une analyse plus détaillée comprenant le calcul récursif des probabilités limites, en utilisant la méthodologie mathématique basée sur le processus régénératif de Markov.

12.2 Description du modèle

On considère un système de files d'attente à un seul serveur où les clients primaires arrivent suivant un flux poissonnien de taux λ ($\lambda > 0$). Un client qui arrive et trouve le serveur non disponible (occupé ou en vacances), quitte l'espace de service pour rejoindre un groupe de clients bloqués appelé "*orbite*". Après un certain temps aléatoire, il renouvelle sa tentative d'entrer en service, une fois, deux fois, ..., jusqu'à ce qu'il le trouve disponible. La discipline d'accès au serveur à partir de l'orbite est gouvernée par une loi exponentielle avec une intensité donnée par $j\alpha$ $(\alpha > 0)$, quand le nombre de clients en orbite est $j \in \mathbb{N}$. Comme cette politique de rappels dépend du nombre de clients dans l'orbite, on l'appelle politique de rappels classiques.
Les temps de service sont supposés d'une loi arbitraire, de fonction de distribution $B(t)$ $(B(0) = 0)$, de transformée de Laplace-Stieltjes $\psi(\omega)$ et des deux premiers moments finis γ_1 et γ_2, respectivement.

Tous les clients entrant dans le système sont servis d'une manière continue et dans un ordre indépendant de leurs temps de service. De plus, on suppose que le serveur prend une vacance chaque fois que le système devient vide (service exhaustif), des deux premiers moments finis $E(X)$ et $E(X^2)$, respectivement.
Les règles qui gouvernent ces périodes de vacances sont :

(1) Le mécanisme qui détermine l'instant de la fin d'une vacance, n'anticipe pas une nouvelle occurrence du processus des arrivées poissonniennes.

(2) Chaque temps de service est indépendant de la séquence des périodes de vacances qui précèdent ce temps de service.

(3) Si aucun client n'arrive durant la période de vacances, on dit qu'il y'a une période d'activité pour le serveur de longueur zéro et le serveur prend une autre vacance.

(4) Juste après la fin des vacances, s'il y a des clients en orbite, le prochain client qui arrive au service est déterminé par une compétition entre deux lois exponentielles de taux λ et α.

Finalement, on suppose que le flux des arrivées primaires, les intervalles entre les rappels successifs et les temps de service sont mutuellement indépendants.
L'état du système à l'instant t peut être décrit par le processus $M(t) = (C(t), N_o(t), \xi(t))_{(t\geq 0)}$, où

$$C(t) = \begin{cases} 0, \text{ si le serveur est oisif,} \\ 1, \text{ si le serveur est occupé,} \\ 2, \text{ si le serveur est en vacance.} \end{cases}$$

$N_o(t)$: le nombre de clients en orbite à l'instant t.
$\xi(t)$: représente le temps de service écoulé du client en service à l'instant t, si $C(t) = 1$, et représente le temps de la vacance écoulé à l'instant t, si $C(t) = 2$.

Plusieurs situations pratiques peuvent être modélisées par les systèmes de files d'attente avec rappels et vacances. Nous présentons ci-dessous un exemple illustratif.

Exemple 1 (Boualem et al. [4]) *Dans le modèle opérationnel du serveur WWW, les requêtes HTTP arrivent au serveur WWW suivant un flux de Poisson et peuvent être interrompues par un utilisateur avant d'arriver au serveur WWW. Lorsque les requêtes arrivent dans le serveur WWW, une requête est sélectionnée pour être servie et les autres entreront dans le tampon situé à l'intérieur du serveur WWW. Dans le tampon, chaque requête attend un certain temps puis demande le service de nouveau. Un programme approprié est mis en application dans le serveur WWW pour diriger les requêtes de service à partir du tampon. À chaque fois qu'elle essaye mais échoue, elle attend un autre moment avant de réessayer de nouveau. Si la page web cible est située dans le même serveur WWW, la requête peut retourner au serveur. Pour garder le serveur WWW en bon fonctionnement, des activités de maintenance telles que scanner les virus peuvent être réalisées lorsque le serveur WWW est inactif. Ce type de maintenance peut être programmé pour fonctionner sur une base régulière. Cependant, ces activités de maintenance ne se répètent pas continuellement. Lorsque ces activités sont achevées, le serveur WWW entrera de nouveau en état d'inactivité et attendra l'arrivée de nouvelles requêtes.*

Dans ce scénario, le tampon dans le serveur WWW, le serveur WWW, la politique de retransmission et les activités de maintenance en période d'inactivité correspondent respectivement à l'orbite, le serveur, la discipline de rappels et la politique de vacances, dans la terminologie de files d'attente.

12.3 Chaîne de Markov induite

L'évolution de notre file d'attente avec rappels et vacances peut être décrite en termes d'une séquence alternée de période d'activité et d'inactivité du serveur.
À la fin de chaque service, le serveur devient libre. La prochaine période d'inactivité du serveur sera de deux types différents :

Type 1 : Si l'orbite devient vide, donc le serveur prend une vacance propre qui est régie suivant les règles de (1) à (4).

Type 2 : Si l'orbite n'est pas vide après la complétion d'un service, alors une compétition entre deux lois exponentielles de taux λ et α déterminera le prochain client qui entrera en service.

Soit $\{\zeta_n, n \in \mathbb{N}\}$ une suite d'instants de la complétion d'un service ou bien de la fin d'une vacance propre.
La séquence des vecteurs aléatoires $Q_n = \{C(\zeta_n), N(\zeta_n)\}$ forme une chaîne de Markov induite pour notre système de files d'attente. Son espace d'état est donné par $S = \{1, 2\} \times \mathbb{N}$.
Les états de la chaîne de Markov $\{Q_n\}_{n=0}^{\infty}$ sont donnés par :

$$(i_{n+1}, j_{n+1}) = \begin{cases} (2, X), & \text{si } \ j_n = 0, \\ (1, j_n - \delta_{j_n} + v_{n+1}), & \text{si } \ j_n > 0, \end{cases} \tag{12.1}$$

où

- X est le nombre de clients primaires qui arrivent vers le système durant une vacance.

- v_{n+1} est la variable représentant le nombre de clients primaires qui arrivent pendant le $(n+1)^{\text{ème}}$ temps de service qui se termine à l'instant ζ_{n+1}, sa distribution est donnée par :

$$K_i = P(v_{n+1} = i) = \int_0^\infty \frac{(\lambda x)^i}{i!} e^{-\lambda x} dB(x), \qquad i \geq 0.$$

- δ_{j_n} est la variable de Bernoulli définie comme suit :

$$\delta_{j_n} = \begin{cases} 1, \text{ si le } (n+1)^{\text{ème}} \text{ client servi provient de l'orbite,} \\ 0, \text{ sinon.} \end{cases}$$

Sa distribution conditionnelle dans le cas de rappels classiques est donnée :

$$P(\delta_{j_n} = 1/j_n = k) = \frac{k\alpha}{\lambda + k\alpha},$$
$$P(\delta_{j_n} = 0/j_n = k) = \frac{\lambda}{\lambda + k\alpha}.$$

Condition d'ergodicité

Théorème 12.1 (Boualem et al. [4]) *La chaîne de Markov induite $\{Q_n\}_{n=0}^\infty$ est ergodique si et seulement si*

$$\rho = \lambda\gamma_1 < 1.$$

12.4 Approche par les processus régénératifs

Plusieurs processus stochastiques survenant, par exemple, dans les systèmes de files d'attente et les systèmes de gestion de stock possèdent la propriété de "régénération" en certains instants, alors le comportement futur du processus après ces instants devient une réplique, c'est-à-dire, le comportement futur du processus après ces instants possède exactement la loi de probabilité qu'il aurait eu s'il avait commencé à l'instant zéro. De tels processus sont appelés "processus régénératifs".

12.4.1 Distributions limites

Dans cette Section, nous employons une approche récursive basée sur la théorie des processus régénératifs de Markov [5], pour calculer les distributions limites :

$$P_{i,j} = \lim_{t\to\infty} P\{Z(t) = (C(t), N(t)) = (i,j)\}, \ (i,j) \in E = \{0,1,2\} \times \mathbb{N}, (P_{00} \equiv 0),$$

sous la condition d'ergodicité $\rho < 1$.
Pour cela, on définit quelques variables aléatoires :
T : la longueur d'un cycle,
T_{ij} : la durée du temps dans un cycle durant lequel le système est à l'état (i,j),
V : la durée d'une vacance propre,
V_j : le nombre de vacances propres dans un cycle pour lesquelles j clients sont laissés en orbite,
N_j : le nombre de fins de service dans un cycle pour lequel j clients sont laissés en orbite.
Alors,

$$P_{i,j} = \frac{E[T_{i,j}]}{E[T]}, \ \forall (i,j) \in E.$$

Le Théorème suivant donne les fonctions génératrices partielles $P_i(z) = \sum_{j=0}^\infty P_{i,j} z^j$, pour $i \in \{0,1,2\}$ des probabilités limites.

Théorème 12.2 (Boualem et al.[4]) *Si $\rho < 1$, alors les fonctions génératrices partielles $P_i(z)$ sont données par :*

$$P_0(z) = \frac{\lambda}{\alpha} \frac{P_{2,\bullet}}{E(X)} \int_1^z \left[\frac{1 - X(t)}{K(t) - t} \exp\left\{ -\frac{\lambda}{\alpha} \int_1^t \frac{1 - K(u)}{K(u) - u} du \right\} \right] dt$$
$$\times \exp\left\{ \frac{\lambda}{\alpha} \int_1^z \frac{1 - K(t)}{K(t) - t} dt \right\}, \tag{12.2}$$

$$P_1(z) = \frac{K(z) - 1}{z - K(z)} \left(P_0(z) + P_2(z) \right), \quad \textit{et} \tag{12.3}$$

$$P_2(z) = \frac{P_{2,\bullet}}{E(X)} \frac{1 - X(z)}{1 - z}, \tag{12.4}$$

où, $P_{2,\bullet} = (1-\rho)\lambda^{-1}$, $K(z) = \psi(\lambda - \lambda z)$ et $X(z)$ est la fonction génératrice de de la variable X.

12.4.2 Décomposition stochastique

Dans cette section, on donne le résultat important concernant la décomposition stochastique de la distribution du nombre de clients dans le système en un point arbitraire au régime stationnaire.

$$P_j = (1-\delta_{0,j})(P_{0,j} + P_{1,j-1}) + P_{2,j}, \quad j \geq 0. \tag{12.5}$$

Théorème 12.3 (Boualem et al. [4]) *Si $\rho < 1$, alors la fonction génératrice du nombre de clients dans le système est donnée par :*

$$P(z) = Q(z)\frac{P_0(z) + P_2(z)}{P_{0,\bullet} + P_{2,\bullet}}, \tag{12.6}$$

où $Q(z)$ est la formule bien connue de Pollaczek-Khintchine pour la file classique ($M/G/1, \infty$), qui est donnée par

$$Q(z) = \frac{(1-\rho)(1-z)K(z)}{K(z)-z}.$$

12.5 Quelques mesures de performance

Notre prochain objectif est de fournir des expressions explicites pour quelques mesures de performance du système $M/G/1$ avec rappels classiques et vacances du serveur. Les résultats sont résumés dans les corollaires suivants.

Corrollaire 12.1 (Boualem et al. [4]). Le nombre moyen de clients dans le système durant une période d'oisiveté, la période d'occupation et la période de vacances sont donnés respectivement par

$$\begin{aligned}
E(N_I) &= P_0^{'}(1) = \frac{\lambda}{\alpha}\frac{\lambda}{1-\rho}P_{2,\bullet} = \frac{\lambda}{\alpha},\\
E(N_B) &= P_1^{'}(1) = \frac{\lambda^2}{2}\frac{\gamma_2}{1-\rho} + \frac{\lambda}{\alpha}\frac{\rho}{(1-\rho)} + \frac{\lambda\rho}{2}\frac{E(X^2)}{E(X)}, \text{ et}\\
E(N_V) &= P_2^{'}(1) = \frac{\lambda^2 E(X^2)}{2}\frac{P_{2,\bullet}}{E(X)} = (1-\rho)\frac{\lambda}{2}\frac{E(X^2)}{E(X)}.
\end{aligned}$$

Corrollaire 12.2 (Boualem et al. [4]). Le nombre moyen de clients dans le système en un point aléatoire est donné par

$$\begin{aligned}
L_s &= E(N_I) + E(N_B) + E(N_V) = P^{'}(1) =\\
&= \frac{1}{1-\rho}\left(\frac{\lambda}{\alpha} + \frac{\lambda^2\gamma_2}{2}\right) + \frac{\lambda}{2}\frac{E(X^2)}{E(X)}.
\end{aligned} \tag{12.7}$$

Le temps moyen d'attente dans le système est obtenu en utilisant les formules de Little et l'équation (12.7). Il est donné par

$$W = \frac{L_s}{\lambda} = \frac{1}{1-\rho}\left(\frac{1}{\alpha} + \frac{\lambda\gamma_2}{2}\right) + \frac{E(X^2)}{2E(X)}.$$

Références

1. A. Aïssani, An $M^X/G/1$ retrial queue with exhaustive vacations, Journal of Statistics and Management Systems, 3(3), 2000, pp.270-286.
2. A. Aïssani, Optimal control of an $M/G/1$ retrial queue with vacations, Journal of Systems Science and Systems Engineering, 17(4), 2008, pp. 487-502.
3. J. R. Artalejo, Analysis of an $M/G/1$ queue with constant repeated attempts and server vacations, Computers and Operations Research, 24(6), 1997, pp. 493-504.
4. Mohamed Boualem, Natalia Djellab and Djamil Aïssani, Approche Régénérative de la File d'Attente $M/G/1$ avec Rappels Classiques et Vacances Exhaustives du Serveur. Journal Européen des Systèmes Automatisés, 45 (1-3), 2011, pp. 253–267.
5. H. C. Tijms, Stochastic models : An algorthmic approach, Wiley, Chichester (ISBN 0-471-95123-4). 385 page, 1994.