ПРИКЛАДНОЙ ВЕРОЯТНОСТНЫЙ АНАЛИЗ И ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

Модель системы обслуживания-запасания с разнотипными заявками

И.А. Алиев

Бакинский государственный университет, Баку, Азербайджан

Последние годы интенсивно исследуются системы обслуживания-запасания (Queuing-Inventory Systems, QIS) [1]. Подробный обзор работ, посвященных различным аспектам изучению таких систем, можно найти в работе [2]. В подавляющем большинстве известных работ считают, что заявки являются идентичными по всем показателям. Однако на практике поставщики товаров различают своих клиентов, и потому для поощрения выгодных клиентов используются различные схемы приоритетов. Подобные модели QIS с разнотипными заявками в доступной литературе мало изучены. Так, в работах [3–6] представлены модели QIS с мгновенным обслуживанием и двумя типами заявок при использовании $(S–1,S)$-политики пополнения запасов; подобные модели при использовании (s,S)-политики изучены в работах [7, 8]. В работе [9] изучена модель QIS с двумя типами заявок и (s,S)-политикой, в которой заявки низкого приоритета уходят в орбиту бесконечного размера, если в моменты их поступления уровень запасов системы меньше, чем s; заявки высокого приоритета принимаются, если уровень запасов системы больше нуля. Дальнейшую библиографию в этом направлении можно найти в списках литературы указанных работ. В работе [10] изучена модель QIS с разнотипными заявками и положительным временем обслуживания. В системе используется рандомизированная политика пополнения запасов, и задача нахождения оптимальных размеров заказов сформулирована как задача марковского программирования. Критерием задачи является минимизация суммарных убытков, связанных с ожиданием заявок в очереди, их потери, доставкой и хранением запасов. Предложены точный и приближенный методы решения поставленной задачи.

Анализ доступных работ показал, что в них дифференциация заявок осуществляется либо с помощью введения критического уровня запасов [3–9] либо по их размеру [10]. В данной работе в отличие от них предлагается другая схема дифференциация: заявки высокого приоритета принимаются при наличии хотя бы одного свободного места в буфере, а заявки низкого приоритета принимаются лишь тогда, когда общая длина очереди заявок меньше заданного порогового значения. Разработан метод расчета характеристик этой системы при использовании предложенной схемы дифференциации заявок.

Склад системы имеет ограниченный размер S, $S < \infty$, и эта система обслуживает пуассоновские потоки заявок двух типов: обычные и приоритетные. Интенсивность обычных заявок (поток заявок первого типа) равна λ_1, а интенсивность приоритетных заявок (поток заявок второго типа) – λ_2. После обслуживания заявки любого типа она с вероятностью $\sigma_1 > 0$ не получает запас и с дополнительной вероятностью $\sigma_2 = 1 - \sigma_1$ получает запас. Если заявка любого типа получает запас, то уровень запасов системы уменьшается на единицу, т.е. заявки являются идентичными по размеру.

Время обслуживания заявок обоих типов зависит от того, получила ли она запас или нет; в обоих случаях эта величина имеет показательную функцию распределения (ф.р.), при этом если заявка не получила запас, то среднее значение времени обслуживания равно μ_1^{-1}, в противном случае оно равно μ_2^{-1}. В реальных системах имеет место соотношение $\mu_1 \neq \mu_2$, так как в случае приобретения запаса выполняются определенные процедуры по его оформлению.

Здесь рассматривается модель QIS с конечной общей очередью разнотипных заявок. Это означает, что заявки обоих типов ожидают в очереди максимальной длины N, $N < \infty$. Предполагается, что если в момент поступления заявки уровень запасов положительный, то заявка первого типа принимается лишь тогда, когда в момент ее поступления суммарная длина очереди меньше, чем заданное пороговое значение r, $1 \leq r \leq N-1$; заявки второго типа теряются лишь тогда, когда очередь полностью заполнена.

Если в момент поступления заявки любого типа уровень запасов равен нулю, то она с вероятностью φ_1 присоединяется к очереди, а с дополнительной вероятностью $\varphi_2 = 1 - \varphi_1$ она уходит из системы не обслуженной. Заявки обоих типов являются нетерпеливыми, если во время их ожидания в очереди уровень запасов системы опускается до нулевого значения, т.е. если при наличии очереди уровень запасов системы опускается до нулевого значения, то заявка каждого типа уходит из системы после случайного времени, которое имеет показательную ф.р. с параметром $\tau > 0$.

В системе используется известная (s, S)-политика, т.е. заказ для поставки запасов делается тогда, когда их уровень опускается до величины s, $s < S/2$, при этом объем заказа равен $S - s$.

Считается, что заказы выполняются с некоторыми случайными задержками, которые имеют показательную ф.р. с параметром $\nu > 0$.

Задача заключается в нахождении совместного распределения уровня запасов системы и числа заявок в системе, и нахождении усредненных характеристик системы: среднего уровня запасов (S_{av}); средней интенсивности заказов (RR); вероятности потери заявок каждого типа (PB_1, PB_2).

Работа системы с конечной очередью описывается двумерной цепью Маркова с состояниями вида (m,n), где m – уровень запасов системы, n – общее число заявок в системе. Пространство состояний (ПС) определяется так:

$$E = \{(m,n): m = \overline{0, S};\ n = \overline{0, N}\}. \quad (1)$$

Переход от состояния (m,n) в состояние (m', n') обозначается через $(m,n) \to (m',n')$, а его интенсивность обозначается через $q((m,n),(m',n'))$. Совокупность этих величин составляет производящую матрицу данной цепи. Рассмотрим задачу ее построения.

Анализируя механизм принятия заявок, схему их обслуживания и поведения в очереди при отсутствии запасов, заключаем, что искомые переходы и интенсивности определяются так:

- интенсивность перехода $(m,n) \to (m,n+1)$ при выполнении условий $m > 0$, $n < r$ равна λ, где $\lambda = \lambda_1 + \lambda_2$;
- интенсивность перехода $(m,n) \to (m, n+1)$ при выполнении условий $m > 0$, $n \geq r$ равна λ_2;
- интенсивность перехода $(0,n) \to (0,n+1)$ равна $\lambda\varphi_1$;
- интенсивность перехода $(m,n) \to (m, n-1)$ при выполнении условий $m > 0$, $n > 0$ равна $\mu_1\sigma_1$;

• интенсивность перехода $(0,n)\to(0,\ n-1)$ при выполнении условия $n>0$, равна $\mu_2\sigma_2$;

• интенсивность перехода $(0,n)\to(0,\ n-1)$ при выполнении условия $n>0$, равна $n\tau$;

• интенсивность перехода $(m,n)\to(m+S-s,\ n)$ при выполнении условия $m\le s$, равна ν.

Все состояния изучаемой конечномерной цепи сообщаются друг с другом, т.е. в ней существует стационарный режим. Стационарная вероятность состояния $(m,n)\in E$ обозначается через $p(m,n)$. Эти вероятности удовлетворяют следующую систему уравнений равновесия (СУР):

$$p(m,n)\sum_{(m',n')\in E_{mn}^{+}} q((m,n),(m',n')) = \sum_{(m',n')\in E_{mn}^{-}} q((m',n'),(m,n))p(m',n'). \qquad (2)$$

Здесь используются следующие обозначения: $q((m,n),(m',n'))$ – интенсивность перехода $(m,n)\to(m',n')$; E_{mn}^{+} – множество тех состояний из (1), в которых можно попасть из состояния (m,n) за один шаг; E_{mn}^{-} – множество тех состояний из (1), из которых можно попасть в состояния (m,n) за один шаг.

Исходя из вышеизложенных фактов об определение интенсивностей переходов между состояниями из ПС (1), заключаем, что СУР (2) в явном виде записывается так:

Случай $s<m\le S$:

$$\begin{aligned}(\lambda I(n<r)+\lambda_2 I(n\ge r)+(\mu_1\sigma_1+\mu_2\sigma_2)I(n>0))p(m,n) &=\\ =\lambda p(m,n-1)\mathrm{I}(n\le r)+\lambda_2 p(m,n-1)I(n>r) &+ \qquad (3)\\ +\mu_1\sigma_1 p(m,n+1)+\mu_2\sigma_2\ p(m+1,n+1)+\nu p(m-S+s,n)&.\end{aligned}$$

Случай $0<m\le s$:

$$\begin{aligned}(\lambda I(n<r)+\lambda_2 I(n\ge r)+(\mu_1\sigma_1+\mu_2\sigma_2)\mathrm{I}(n>0)+\nu)\,p(m,n) &=\\ =\lambda p(m,n-1)\mathrm{I}(n\le r)+\lambda_2 p(m,n-1)\mathrm{I}(n>r) &+ \qquad (4)\\ +\mu_1\sigma_1 p(m,n+1)+\mu_2\sigma_2\ p(m+1,n+1)&.\end{aligned}$$

Случай $m=0$:

$$(\lambda\varphi_1+n\tau+\nu)p(0,n)=\lambda\varphi_1\ p(0,n-1)\mathrm{I}(n>0)+\mu_2\sigma_2\ p(1,n+1). \qquad (5)$$

Здесь $I(A)$ обозначает индикаторную функцию события A. К этой СУР (3) – (5) добавляется еще и условие нормировки

$$\sum_{(m,n)\in E} p\ (m,n)=1. \tag{6}$$

Усредненные характеристики системы определяются с помощью вероятностей состояний. Действительно, средний уровень запасов системы определяется как математическое ожидание соответствующей случайной величины, т.е. имеем

$$S_{av} = \sum_{m=1}^{S} m \sum_{n=0}^{N} p\ (m,n). \tag{7}$$

Заказы на поставку пополнения запасов осуществляются тогда, когда уровень запасов системы опускается от значения s+1 к значению s, т.е. средняя интенсивность заказов определяется следующим образом:

$$RR= \mu_2\sigma_2 \sum_{n=1}^{N} p\ (s+1,n). \tag{8}$$

Заявки первого типа теряются в следующих случаях: 1) в момент поступления заявки уровень запасов системы больше нуля, но общее число заявок в системе превышает пороговое значение r ; 2) заявка теряется из очереди из-за нетерпеливости. Отсюда находим, что вероятность потери заявок первого типа определяется так:

$$PB_1 = \sum_{m=1}^{S}\sum_{n=r}^{N} p(m,n) + \theta \sum_{n=1}^{N} p(0,n) \frac{n\tau}{\lambda\varphi_1 + n\tau}. \tag{9}$$

Заявки второго типа теряются в следующих случаях: 1) если в момент поступления заявки общее число заявок в системе равно N; 2) заявка теряется из очереди из-за нетерпеливости. Следовательно, вероятность потери заявок второго типа определяется так:

$$PB_2 = \sum_{m=0}^{S} p(m,N) + \theta_2 \sum_{n=1}^{N} p(0,n) \frac{n\tau}{\lambda\varphi_1 + n\tau}. \tag{10}$$

В формулах (9), (10) параметры $\theta_1 = \eta_1/(\eta_1+\eta_2)$ и $\theta_2 = 1-\theta_1$ оценивают вероятность того, что в состояниях типа $(m,n), n>0,$ случайно выбранная заявка является заявкой первого и второго типа соответственно, где η_1 и η_2 обозначают интенсивность принятых в систему заявок первого и второго типа соответственно. В частности, величины θ_1 и θ_2 оценивают вероятность того, что покидающая очередь из-за нетерпеливости заявка является заявкой первого и второго типа соответственно. Очевидно, что интенсивность заявок второго типа равна $\eta_2 = \lambda_2$, а поскольку имеется ограничение для заявок первого типа, то величина η_1 вычисляется следующим образом:

$$\eta_1 = \sum_{k=1}^{r} k \frac{\lambda_1^k}{k!} e^{-\lambda_1} + \mathrm{r} \sum_{k=r+1}^{\infty} \frac{\lambda_1^k}{k!} e^{-\lambda_1} = e^{-\lambda_1} \sum_{k=1}^{r} \frac{\lambda_1^k}{(k-1)!} + \mathrm{r}\left(1 - e^{-\lambda_1} \sum_{k=0}^{r} \frac{\lambda_1^k}{k!}\right).$$

Построенная СУР (3) – (6) имеет размерность $(S+1)(N+1)$. Для моделей умеренной размерности вычисления вероятностей состояний из СУР (3) – (6) можно осуществить с помощью известных численных методов линейной алгебры. После нахождения вероятностей состояний из (7) – (10) определяются искомые характеристики системы.

Для моделей большой размерности можно использовать методы спектрального расширения [11], а также метод фазового укрупнения [12]. Разработка соответствующих алгоритмов требует выполнения специальных исследований, и они являются предметами наших дальнейших работ.

ЛИТЕРАТУРА

1. *Schwarz M., Sauer C., Daduna H., Kulik R., Szekli R.* M/M/1 queuing systems with inventory // Queuing Systems. 2006. V. 54. Iss. 1. P. 55–78.
2. *Krishnamoorthy A., Lakshmy B., Manikandan R.* A survey on inventory models with positive service time // OPSEARCH. 2011. V. 48. P. 153–169.
3. *Ha A.Y.* Stock rationing in an M/Ek/1 make-to-stock queue // Management Science. 2000. V. 46. Iss. 1. P. 77–87.
4. *Dekker R., Hill R.M., Kleijn M.J.* On the (*S*–1, *S*) lost sales inventory model with priority demand classes // Naval Research Logistics. 2002. V. 49. Iss. 6. P. 593–610.
5. *Kranenburg A.A., van Houtum G.J.* Cost optimization in the (*S*–1, *S*) lost sales inventory model with multiple demand classes // Operations Research Letters. 2007. V. 35. Iss. 4. P. 493–502.
6. *Isotupa K.P.S.* Cost analysis of an (S-1, S) inventory system with two demand classes and rationing // Annals of Operations Research. 2015. V. 233. P. 411–421.
7. *Isotupa K.P.S.* An (*S*, *Q*) inventory system with two demand classes of customers // Int. J. Operational Research. 2011. V. 12. Iss. 1. P. 12–19.
8. *Isotupa K.P.S.* An (*S*, *Q*) Markovian inventory system with lost sales and two demand classes // Mathematical and Computer Modeling. 2006. V. 43. P. 687–694.

9. *Karthick T., Sivakumar B., Arivarignan G.* An inventory system with two types of customers and retrial demands // Int. J. Systems Science: Operations & Logistics. 2015. V. 2. Iss. 2. P. 90–112.
10. *Melikov A.Z., Fatalieva M.R.* Situational inventory in counter–stream serving systems // Engineering Simulation. 1998. V. 15. P. 839–848.
11. *Chakka R.* Spectral expansion solution for some finite capacity queues // Annals of Operations Research. 1998. V. 79. P. 27–44.
12. *Melikov A.Z., Ponomarenko L.A., Shahmaliyev M.O.* Analysis of perishable queuing-inventory systems with different types of requests // J. Automation and Information Sciences. 2017. V. 49. Iss. 9. P. 42–60.

Алиев Исмаил Алекпер оглы, докторант,isiko94@gmail.com