Условия стабильности системы с повторными вызовами при регенерирующем входящем потоке

Л. Г. АФАНАСЬЕВА
Московский государственный университет им. М. В. Ломоносова
e-mail: l.g.afanaseva@yandex.ru

УДК 519.218

Ключевые слова: системы с повторным обслуживанием, условие стабильности, классическая политика повторений, постоянная скорость повторений.

Аннотация

Рассматриваются многоканальные системы обслуживания с повторными вызовами двух классов. Для первого класса интенсивность повторения вызовов с орбиты зависит от числа требований, находящихся на ней, а для второго класса она постоянна. Входящий поток регенерирующий, а времена обслуживания имеют произвольное распределение. На основе метода синхронизации входящего потока и специальным образом построенного вспомогательного процесса доказано необходимое и достаточное условие стабильности для систем обоих классов.

Abstract

L. G. Afanaseva, Stability conditions for retrial queueing systems with regenerative input flow, Fundamentalnaya i prikladnaya matematika, vol. 22 (2018), no. 3, pp. 5—18.

We consider two classes of multiserver retrial queueing systems. For the first class, the rate of retrial requests depends on the number of customers on the orbit, and for the second class, the rate is constant. The input flow is supposed to be a regenerative one and service time has an arbitrary distribution. Based on the synchronization of the input flow and an auxiliary service process, we establish necessary and sufficient stability conditions for models of both classes.

1. Введение

Мы рассматриваем многоканальную систему обслуживания, в которую поступает регенерирующий поток требований $X(t)$ с интенсивностью λ [3]. В системе имеется m идентичных приборов. Времена обслуживания требований — независимые одинаково распределённые случайные величины. Если в момент прихода требования в систему имеется хотя бы один свободный прибор, оно немедленно начинает обслуживаться. В противном случае требование направляется на орбиту, с которой осуществляет запросы с целью попасть на обслуживание. Системы такого сорта называются системами с повторными вызовами, и они давно и интенсивно изучаются, так что имеется обширная

Фундаментальная и прикладная математика, 2018, том 22, № 3, с. 5—18.

литература в этом направлении. Изначально системы с повторными вызовами (см., например, [9, 11, 18]) представляли собой альтернативу классическим моделям телефонных систем, а именно систем с отказами, которые не учитывают возможности повторения вызовов от требований, получивших отказ. Если предположить, что каждое блокированное требование независимо от других повторяет запросы через экспоненциально распределённые промежутки времени с параметром ν, то мы получим классическую политику повторов (см. [5]) и суммарный поток запросов с орбиты будет дважды стохастическим пуассоновским потоком со случайной интенсивностью $\nu q(t)$, где $q(t)$ — число требований на орбите в момент t [15]. Если в момент запроса имеется свободный прибор, требование начинает обслуживаться и покидает орбиту. В противном случае остаётся на ней.

Другой класс с повторными вызовами включает в себя системы с постоянной интенсивностью повторов. Эти системы возникли в работе [12] как модели телефонной связи. Далее произошёл быстрый рост литературы (см. [4—8, 10, 14]), посвящённой моделям этого класса, что в первую очередь связано с их успешным применением при анализе коммуникационных и компьютерных сетей, где попытки повторов осуществляются процессором независимо от числа сообщений, хранящихся в узлах сети.

Среди имеющейся обширной литературы отметим несколько работ, посвящённых анализу систем данного класса. Г. И. Фалин и Х. Р. Арталехо [10] изучили многоканальные системы, в которых прибывающие требования либо присоединяются к обычной очереди, либо уходят на орбиту в зависимости от числа требований в очереди. В [7] рассмотрена система типа $M|M|1$ с произвольно распределёнными интервалами между повторами вызовов. Существует большое число работ, посвящённых алгоритмическим методам для систем обоих упомянутых классов (см., например, [4—12, 14, 16]), включая анализ моделей с общим распределением типа BMAP, PH и т. д. интервалов между поступлениями и запросами. Ведущую роль в анализе систем такого рода играет общая теория Ньютса [16] для квазипроцессов рождения и гибели (QBD).

Цель представленной работы — получение необходимых и достаточных условий стабильности систем с повторными вызовами обоих классов при достаточно общих предположениях относительно процессов, определяющих функционирование модели. Для системы второго класса типа $M|M|m$ полученное условие совпадает с установленным в [5]. Отметим также работу [19], в которой рассмотрена система $M|M|1$ с повторениями вызовов и оценена скорость сходимости к предельному распределению.

Применяемый в представленной статье подход, основанный на синхронизации входящего потока и вспомогательного процесса, определяющего число требований на орбите в предположении, что они всё время присутствуют на ней, как мы надеемся, может быть использован и для решения этой проблемы.

В следующем разделе даётся детальное описание изучаемых систем: модели M_1 первого класса и модели M_2 второго класса. В разделе 3 вводятся вспомогательные процессы для модели M_2, а в разделе 4 доказывается необходимое и достаточное условие её стабильности. Оставшаяся часть статьи посвящена доказательству условий стабильности системы M_1 и обсуждению результатов.

2. Описание моделей

Мы исследуем две модели M_1 и M_2 обслуживания с повторными вызовами. Входящий поток требований $\{X(t),\ t \geqslant 0\}$ считается регенерирующим (определения и свойства можно найти в [3, 17]) с моментами регенерации $\{\theta_j\}_{j=1}^{\infty}$ и периодами регенерации $\tau_j = \theta_{j+1} - \theta_j$ $(j = 0, 1, \ldots,\ \theta_0 = 0)$. Пусть $\xi_j = X(\theta_{j+1}) - X(\theta_j)$ — число требований, поступивших в систему на j-м периоде регенерации, и $E\tau_1^{(1)} < \infty$, $E\xi_1 < \infty$. Тогда интенсивность потока равна

$$\lambda_X = \lim_{t\to\infty} \frac{X(t)}{t} = \frac{E\xi_1}{E\tau_1}.$$

Имеется m идентичных приборов, так что времена обслуживания требований $\{\eta_{n)}\}_{n=1}^{\infty}$ — независимые одинаково распределённые случайные величины с функцией распределения $B(x)$ и средним b. Если в момент поступления требования в системе имеется свободный прибор, оно немедленно начинает обслуживаться. В противном случае требование отправляется в накопитель, обычно называемый орбитой, откуда осуществляются повторные запросы на обслуживание. Для модели M_1 мы предполагаем, что поток этих запросов пуассоновский с интенсивностью $\nu(j)$, если на орбите имеется j требований. (На самом деле это дважды стохастический пуассоновский поток [15] со случайной интенсивностью $\nu\big(z(t)\big)$, где $z(t)$ — число требований на орбите в момент t.) При наличии в момент запроса свободного прибора одно из требований на орбите занимает его и начинает обслуживаться. Обычно предполагается [9, 11, 18], что требования посылают запросы независимо друг от друга через экспоненциально распределённые интервалы времени с параметром ν, так что $\nu(j) = \nu j$. В модели M_2 повторы вызовов с орбиты осуществляются через независимые одинаково распределённые промежутки времени $\{\zeta_n\}_{n=1}^{\infty}$ с конечным $E\zeta_n = \nu^{-1}$ независимо от того, есть требования на орбите или нет и сколько их. Таким образом, интенсивность повторения вызовов с орбиты постоянна и равна ν^{-1}. Как и в модели M_1, если в момент запроса с орбиты имеется свободный прибор, а на орбите есть требования, то одно из них покидает орбиту и начинает обслуживаться.

Для моделей M_1 и M_2 мы рассматриваем случайный процесс $q(t)$, представляющий собой число требований в системе в момент t, и хотим найти условия, при которых процесс стабилен. Стабильность означает, что при любом начальном состоянии системы существует предел

$$\lim_{t\to\infty} P(q(t) \leqslant x) = \Phi(x), \tag{1}$$

где $\Phi(x)$ — функция распределения, не зависящая от начального состояния.

Условие 1.

$$P(\xi_1 = 0,\ \tau_1 > 0) + P(\xi_1 = 1,\ \tau_1 - t_1 > \eta_1) > 0,$$

где $\theta_1 + t_1$ — момент поступления единственного требования в интервале (θ_1, θ_2) (первый период регенерации), а η_1 — его время обслуживания.

Заметим, что условие 1 для обеих моделей M_1 и M_2 обеспечивает попадание процесса $q(t)$ в нулевое состояние с положительной вероятностью из любого начального состояния системы.

Условие 2. Для модели M_2 распределение ζ_n между повторами запросов с орбиты имеет вторую экспоненциальную фазу $\zeta_n^{(2)}$, т. е.

$$\zeta_n = \zeta_n^{(1)} + \zeta_n^{(2)},$$

где $\zeta_n^{(1)}$ и $\zeta_n^{(2)}$ независимы и $P(\zeta_n^{(2)} > x) = e^{-\gamma x}$ $(\gamma > 0)$.

Процесс $q(t)$ регенерирующий при условии 1 для модели M_1 и в условиях 1 и 2 для модели M_2. В качестве точек регенерации $\{T_n\}_{n=1}^{\infty}$ для модели M_1 возьмём подпоследовательность $\{\theta_{n_k}\}_{k=1}^{\infty}$ моментов регенерации входящего потока $\{\theta_n\}_{n=1}^{\infty}$, такую что $q(\theta_{n_k} - 0) = 0$, т. е.

$$T_n = \min\left\{\theta_k > T_{n-1} \colon \bigcup_{k=1}^{\infty} \{q(\theta_k - 0) = 0\}\right\}, \quad T_0 = 0. \tag{2}$$

Для модели M_2 дополнительно потребуем, чтобы $\{\theta_{n_k)}\}_{k=1}^{\infty}$ попадали в экспоненциальную фазу интервалов между запросами, т. е.

$$T_n = \min\left\{\theta_k > T_{n-1} \colon \bigcup_{k=1}^{\infty} \{q(\theta_k - 0) = 0\} \cap \bigcup_{l=0}^{\infty} \bigl\{\theta_k \in \bigl(Z_l + \zeta_{l+1}^{(1)},\ Z_{l+1}\bigr)\bigr\}\right\}, \tag{3}$$

где $T_0 = 0$, $Z_l = \zeta_1 + \ldots + \zeta_l$, $Z_0 = 0$. Тогда для регенерирующего процесса $q(t)$ выполнены условия теоремы 1 из [1], так что у процесса $q(t)$ есть две возможности: либо он стабилен, т. е. имеет место (1), либо

$$q(t) \xrightarrow[t \to \infty]{P} \infty. \tag{4}$$

Введём процесс $Y(t)$, представляющий собой число требований, обслуженных в системе к моменту t. Заметим, что это регенерирующий поток и его точки регенерации $\{T_n\}_{n=1}^{\infty}$ определены (2) для модели M_1 и (3) для модели M_2. Пусть $n(t)$ — число занятых приборов в момент t.

Сначала рассмотрим систему M_2, а затем, опираясь на полученные результаты, установим необходимое и достаточное условие стабильности $q(t)$ для модели M_2.

3. Вспомогательные системы и процессы для модели M_2

Пусть $N(t)$ — считающий процес, соответствующий последовательности $\{\zeta_{n)}\}_{n=1}^{\infty}$, т. е.

$$N(t) = \max\{k \geqslant 0 \colon Z_k \leqslant t\},$$
$$Z_0 = 0, \quad Z_k = \zeta_1 + \ldots + \zeta_k, \quad k = 1, 2, \ldots .$$

Обозначим $\tilde{M}_2$ систему с тем же входящим потоком $X(t)$ и процессом повторения вызовов $N(t)$, но всегда имеющую требования на орбите, например, в начальный момент их бесконечно много. Пусть $\tilde{n}(t)$ — число занятых приборов в $\tilde{M}_2$ в момент t, а $\tilde{Y}(t)$ — число обслуженных требований в интервале $(0, t)$. Тогда при одинаковых начальных условиях ($\tilde{n}(0) = n(0)$ и оставшиеся времена обслуживания на занятых приборах в M_2 и $\tilde{M}_2$ одинаковы) очевидно стохастическое неравенство

$$Y(t) \leqslant \tilde{Y}(t). \tag{5}$$

Процессы $X(t)$ и $\tilde{Y}(t)$ зависимы, но при некоторых условиях можно построить для них общие точки регенерации. Например, если выполнены условия 1 и 2, положим

$$\tilde{T}_n^{(0)} = \min\left\{\theta_k > \tilde{T}_{n-1}^{(0)} \colon \bigcup_{k=1}^{\infty}\{\tilde{n}(\theta_k) = 0\} \cap \bigcup_{l=0}^{\infty}\left\{\theta_k \in \left(Z_l + \zeta_{l+1}^{(1)},\ Z_{l+1}\right)\right\}\right\}, \quad \tilde{T}_0^{(0)} = 0. \tag{6}$$

Тогда последовательность $\left\{\tilde{T}_n^{(0)}\right\}$ состоит из точек регенерации $\tilde{X}(t)$ и $\tilde{Y}(t)$. Поскольку $\tilde{Y}(t)$ — регенерирующий поток, существует предел

$$\lim_{t\to\infty} \frac{\tilde{Y}(t)}{t} = \tilde{\lambda}_Y.$$

Пусть $\{\tilde{T}_n\}_{n=1}^{\infty}$ — любая последовательность общих точек регенерации потоков $X(t)$ и $\tilde{Y}(t)$ и $E\tilde{\tau}_n < \infty$, где $\tilde{\tau}_n = \tilde{T}_{n+1} - \tilde{T}_n$. Определим приращения

$$\begin{aligned} \Delta_X(n) &= X(\tilde{T}_{n+1}) - X(\tilde{T}_n), \\ \Delta_Y(n) &= Y(\tilde{T}_{n+1}) - Y(\tilde{T}_n), \\ \tilde{\Delta}_Y(n) &= \tilde{Y}(\tilde{T}_{n+1}) - \tilde{Y}(\tilde{T}_n). \end{aligned} \tag{7}$$

Мы замечаем, что $\{\Delta_X(n), \tilde{\Delta}_Y(n)\}_{n=1}^{\infty}$ — последовательность независимых одинаково распределённых случайных векторов и

$$\lambda_X = \frac{E\Delta_X(1)}{E\tilde{\tau}_1}, \quad \tilde{\lambda}_Y = \frac{E\tilde{\Delta}_Y(1)}{E\tilde{\tau}_1}. \tag{8}$$

Рассмотрим систему $\tilde{M}_0$ с отказами типа $\mathrm{Reg}\,|G|m|0$. Входящий поток $U(t) = X(t) + N(t)$ является регенерирующим (при условии 2) с точками регенерации

$\{\hat{T}_n\}_{n=1}^{\infty}$, определёнными соотношениями

$$\hat{T}_n = \min\left\{\theta_k > \hat{T}_{n-1}\colon \bigcup_{l=1}^{\infty}\{\theta_k \in \big(Z_l + \zeta_{l+1}^{(1)},\ Z_{l+1}\big)\}\right\}, \quad \hat{T}_0 = 0. \tag{9}$$

Тогда поток $\tilde{Y}(t)$ представляет собой число требований, обслуженных за время $(0,t)$, а $\tilde{n}(t)$ — число требований в момент t в системе $\tilde{M}_0$. Пусть $\{t_k\}_{k=1}^{\infty}$ — последовательные моменты скачков потока $U(t)$. Поскольку $\tilde{n}(t)$ — регенерирующий процесс (в условиях 1 и 2), то существуют пределы

$$\lim_{k\to\infty} P(\tilde{n}(t_k) = j) = P_j \quad (j = 0, 1, \ldots, m).$$

Лемма 1. *Если выполнены условия* 1 *и* 2, *то*

$$\tilde{\lambda}_Y = \lim_{t\to\infty} \frac{\tilde{Y}(t)}{t} = (\lambda + \nu)(1 - P_m). \tag{10}$$

Доказательство следует из очевидных неравенств

$$\sum_{k=1}^{U(t)} \mathbb{1}(\tilde{n}(t_k) < m) - m \leqslant \tilde{Y}(t) \leqslant \sum_{k=1}^{U(t)} \mathbb{1}(\tilde{n}(t_k) < m)$$

и предельных теорем теории восстановления. Здесь $\mathbb{1}(A)$ — индикатор события A.

4. Теорема стабильности для системы M_2

Нам понадобится следующее условие.

Условие 3. Распределение времени обслуживания η_n таково, что

$$\eta_n = \eta_n^{(1)} + \eta_n^{(2)},$$

где $\eta_n^{(1)}$ и $\eta_n^{(2)}$ независимы, а $\eta_n^{(2)}$ имеет экспоненциальное распределение, т. е. время обслуживания имеет вторую экспоненциальную фазу.

Мы определим коэффициент загрузки ρ системы M_2 равенством

$$\rho = \frac{\lambda}{\tilde{\lambda}_Y},$$

где $\tilde{\lambda}_Y$ даны (8) или (10).

Теорема 1. *Пусть выполнены условия* 1 *и* 2. *Если* $\rho < 1$, *то процесс* $q(t)$ *стабилен. Если* $\rho > 1$ *или* $\rho = 1$ *и дополнительно выполнено условие* 3, *то имеет место* (4).

Доказательство. При $\rho > 1$ ввиду (5) имеем

$$q(t) = q(0) + X(t) - Y(t) \geqslant q(0) + X(t) - \tilde{Y}(t),$$

следовательно,

$$\lim_{t\to\infty}\frac{q(t)}{t}\geqslant\lambda-\tilde\lambda_Y>0,$$

так что $q(t)\xrightarrow[t\to\infty]{P}\infty$.

Для $\rho=1$ при условии 3 определим последовательность $\{\tilde T_n^{(m)}\}_{n=1}^{\infty}$ общих точек регенерации $X(t)$, $\tilde Y(t)$ и $\tilde n(t)$, предположив, что $\tilde T_n^m$ — момент регенерации $X(t)$, попадающий в экспоненциальную фазу $N(t)$, такой что $\tilde n\big(\tilde T_n^{(m)}\big)=m$, а все времена обслуживания во второй (экспоненциальной) фазе. Тогда в очевидных обозначениях (см. (7)) по распределению

$$\tilde\Delta_Y^{(m)}(n)\geqslant\Delta_Y^{(m)}(n),\tag{11}$$

и в силу (8) коэффициент загрузки равен

$$\rho=\frac{E\Delta_X^{(m)}(1)}{E\tilde\Delta_Y^{(m)}(1)}=1.$$

Из (11) следует, что

$$q\big(\tilde T_{n+1}^{(m)}\big)=q\big(\tilde T_n^{(m)}\big)+\Delta_X^{(m)}(n)-\Delta_Y^{(m)}(n)\geqslant q\big(\tilde T_n^{(m)}\big)+\Delta_X^{(m)}(n)-\tilde\Delta_Y^{(m)}(n).$$

Отсюда следует сходимость

$$q\big(\tilde T_n^{(m)}\big)\xrightarrow[n\to\infty]{P}\infty,$$

поскольку

$$E\Delta_X^{(m)}(n)=E\tilde\Delta_Y^{(m)}(n)$$

(см., например, [13]).

Рассмотрим случай $\rho<1$. Нам достаточно доказать, что для процесса $q(t)$ не выполняется (4). Предположим противное и придём к противоречию.

Рассмотрим мажорирующую систему M_2^β, в которой интервал между запросами с орбиты имеет первую экспоненциальную фазу, т. е. $\zeta_n^{(\beta)}=\varkappa_n+\zeta_n$, где $\varkappa_n$ и ζ_n независимы, а $\varkappa_n$ с параметром β экспоненциально распределена. Тогда для числа требований в системах M_2^β и M_2 имеем стохастическое неравенство $q^\beta(t)\geqslant q(t)$. Кроме того, коэффициент загрузки ρ_β системы M_2^β в силу леммы 1 непрерывно зависит от β и $\lim\limits_{\beta\to\infty}\rho^\beta=\rho$. Поскольку $\rho<1$, можно выбрать β из условия $\rho_\beta<1$. Далее будем рассматривать систему M_2^β, опустив β, что вряд ли вызовет затруднения в понимании.

В качестве общих точек регенерации $\{\hat T_n^{(0)}\}_{n=1}^{\infty}$ потоков $X(t)$ и $\tilde Y(t)$ возьмём подпоследовательность $\{\theta_{n_k}\}_{k=1}^{\infty}$ последовательности моментов регенерации $X(t)$, такую что θ_{n_k} попадает в первую (экспоненциальную) фазу интервалов $\{\zeta_n\}_{n=1}^{\infty}$ и $\tilde n(\theta_{n_k}-0)=0$. По формулам (7) определим приращения $\Delta_X^{(0)}(n)$, $\Delta_Y^{(0)}(n)$ и $\tilde\Delta_Y^{(0)}(n)$ процессов $X(t)$, $Y(t)$, $\tilde Y(t)$ на интервалах $\big(\hat T_n^{(0)},\hat T_{n+1}^{(0)}\big)$, заменив $\tilde T_n$ на $\hat T_n^{(0)}$.

Введём мажорирующую систему M_2^δ, в которой функция распределения времени обслуживания удовлетворяет соотношению $B_\delta(x) = B(x-\delta)$.

Коэффициент загрузки удовлетворяет соотношению $\rho_\delta \underset{\delta\to 0}{\longrightarrow} \rho < 1$, так что можно выбрать δ, для которого $\rho_\delta < 1$. Кроме того, в очевидных обозначениях $q_\delta(t) \geqslant q(t)$, так что стохастическая ограниченность процесса $q_\delta(t)$ влечёт стабильность системы M_2. Далее мы рассматриваем систему M_2^δ, опуская δ.

Лемма 2. *Если* $q(t) \xrightarrow[t\to\infty]{P} \infty$, *то для любого* $\epsilon > 0$ *найдётся* n_1, *такое что при* $n > n_1$

$$E\Delta_Y^{(0)}(n) \geqslant E\tilde{\Delta}_Y^{(0)}(1) - \epsilon.$$

Доказательство. Обозначим $\tilde{M}_2$ систему, аналогичную M_2, но в предположении, что на орбите всё время есть требования, так что $\tilde{Y}(t)$ — это число требований, обслуженных в $\tilde{M}_2$ к моменту t.

Организуем функционирование систем M_2 и $\tilde{M}_2$ на интервалах $\big(\hat{T}_n^{(0)}, \hat{T}_{n+1}^{(0)}\big)$ следующим образом. Если $n\big(\hat{T}_n^{(0)}\big) = j > 0$, то первым j требованиям, поступающим на обслуживание в $\tilde{M}_2$ в этом интервале (если их число не меньше j, а если меньше, то всем поступившим), присваиваем времена обслуживания требований, уже находящихся на приборах в M_2 в момент $\hat{T}_n^{(0)}$. Времена обслуживания следующих поступающих на приборы требований (когда таковые имеются) выбираем для M_2 и $\tilde{M}_2$ соответственно из независимых последовательностей $\{\eta_{nk}\}_{k=1}^\infty$ и $\{\tilde{\eta}_{nk}\}_{k=1}^\infty$ независимых случайных величин с функцией распределения $B(x)$. Положим $\hat{\tau}_n^{(0)} = \hat{T}_{n+1}^{(0)} - \hat{T}_n^{(0)}$ и $q_n = q\big(\hat{T}_n^{(0)}\big)$, так что $q_n \xrightarrow[n\to\infty]{P} \infty$. Для любого $\epsilon > 0$ находим Λ_ϵ из условия

$$E\tilde{\Delta}_Y^{(0)}(1)\mathbb{1}\big(\hat{\tau}_1^{(0)} > \Lambda_\epsilon\big) < \frac{\epsilon}{2},$$

а n_1 из условия

$$E\tilde{\Delta}_Y^{(0)}(n)\mathbb{1}\left(q_n > \frac{\Lambda_\epsilon}{\delta} m\right) < \frac{\epsilon}{2}$$

при $n > n_1$, что возможно в силу конечности $E\tilde{\Delta}_Y^{(0)}(n) = E\tilde{\Delta}_Y^{(0)}(1)$ и сходимости

$$q(t) \xrightarrow[t\to\infty]{P} \infty.$$

Введём события

$$A_n = \left\{q_n > \frac{\Lambda_\epsilon}{\delta} m\right\}, \quad B_n = \{\hat{\tau}_n^{(0)} < \Lambda_\epsilon\}.$$

Согласно выбору времён обслуживания в M_2 и $\tilde{M}_2$ имеет место стохастическое неравенство

$$\tilde{\Delta}_Y^{(0)}(n)\mathbb{1}(A_n B_n) \leqslant \Delta_Y^{(0)}(n)\mathbb{1}(A_n B_n).$$

Отсюда при $n > n_1$ получаем

$$E\Delta_Y^{(0)}(n) \geqslant E\Delta_Y^{(0)}(n)\mathbb{1}(A_nB_n) \geqslant E\tilde{\Delta}_Y^{(0)}(n)\mathbb{1}(A_nB_n) \geqslant \\ \geqslant E\tilde{\Delta}_Y^{(0)}(n) - E\tilde{\Delta}_Y^{(0)}(n)\mathbb{1}(\bar{A}_n) - E\tilde{\Delta}_Y^{(0)}(n)\mathbb{1}(\bar{B}_n) \geqslant E\tilde{\Delta}_Y^{(0)}(1) - \epsilon. \quad \square$$

Теперь предположим, что $\rho < 1$, но система M_2 нестабильна, так что

$$q_n = q\big(\hat{T}_n^{(0)}\big) \xrightarrow[n\to\infty]{P} \infty. \tag{12}$$

В лемме 2 выберем $0 < \varepsilon < E\tilde{\Delta}_Y^{(0)}(1)(1-\rho)$. Переходя в рекуррентных соотношениях

$$q_n = q_{n-1} + \Delta_X^{(0)}(n) - \Delta_Y^{(0)}(n), \quad n = 1, 2, \ldots,$$

к математическим ожиданиям, находим при $n > n_1$, что

$$Eq_n = Eq_{n-1} + E\Delta_X^{(0)}(n) - E\Delta_Y^{(0)}(n) \leqslant \\ \leqslant Eq_{n-1} + E\Delta_X^{(0)}(1) - E\tilde{\Delta}_Y^{(0)}(1) + \varepsilon \leqslant Eq_{n-1},$$

что противоречит (12). $\square$

Следствие 1. *Если $X(t)$ и $N(t)$ — пуассоновские потоки с параметрами λ и ν соответственно, то процесс $q(t)$ стабилен тогда и только тогда, когда*

$$\frac{\lambda}{\lambda+\nu} < \frac{\sum\limits_{j=0}^{m-1} \frac{\alpha^j}{j!}}{\sum\limits_{j=0}^{m} \frac{\alpha^j}{j!}}, \tag{13}$$

где $\alpha = (\lambda+\nu)b$.

Доказательство. Вспомогательная система с отказами $\tilde{M}_0$ в рассматриваемом случае имеет тип $M|G|m|0$, так что

$$P_m = \frac{\alpha^m}{m!}\left(\sum_{j=0}^{m} \frac{\alpha^j}{j!}\right)^{-1}$$

(см., например, [2]).

В соответствии с (10)

$$\tilde{\lambda}_Y = (\lambda+\nu)\sum_{j=0}^{m-1} \frac{\alpha^j}{j!}\left(\sum_{j=0}^{m} \frac{\alpha^j}{j!}\right)^{-1},$$

что согласно теореме 1 доказывает условие (13).

Заметим, что оно совпадает с условием, установленным в [5] для системы типа $M|M|m$, т. е. с экспоненциально распределённым временем обслуживания. $\square$

5. Теорема стабильности для системы M_1

Нам понадобится следующее условие.

Условие 4. Для некоторого $\delta > 0$

$$E\xi_1^{2+\delta} < \infty, \quad E\tau_1^{2+\delta} < \infty, \quad E\eta_1^{2+\delta} < \infty.$$

Как и раньше, мы рассматриваем случайный процесс $q(t)$, представляющий число требований в системе M_2 в момент t.

Теорема 2. *Пусть выполнено условие* 1, *интенсивность запросов с орбиты* $\nu(j)$ *монотонно возрастает по* j *и*

$$\lim_{j\to\infty} \nu(j) = \infty. \tag{14}$$

Если

$$\rho = \frac{\lambda b}{m} < 1,$$

то процесс $q(t)$ *стабилен, т. е. имеет место* (1). *Если* $\rho > 1$ *или* $\rho = 1$ *и дополнительно выполнено условие* 4, *то имеет место* (4), *т. е. система* M_1 *нестабильна.*

Доказательство. Согласно условию 1 процесс $q(t)$ регенерирующий, и из теоремы 1 в [1] следует, что существуют две возможности: либо $q(t)$ стабилен, либо имеет место сходимость (4).

Рассмотрим классическую систему $\mathrm{Reg}\,|G|m|\infty$ с регенерирующим входящим потоком $X(t)$ и обозначим через $q_0(t)$ число требований в ней в момент t. Если $q(0) = q_0(0)$ и остаточные времена обслуживания требований на приборах в системах $\mathrm{Reg}\,|G|m|\infty$ и M_1 одинаковы, то выполняется стохастическое неравенство $q_0(t) \leqslant q(t)$.

В [1] показано, что при $\rho = \lambda b/m > 1$ и при дополнительном условии 4 в случае $\rho = 1$ $q_0(t) \xrightarrow[t\to\infty]{P} \infty$, что означает выполнение (4) для $q(t)$ в системе M_1.

Пусть $\rho < 1$. Рассмотрим мажорирующую систему $M_1^{(1)}$, в которой требования в момент прихода в систему сразу направляются на орбиту и только через запросы с орбиты поступают на обслуживание. Если $q_1(t)$ — число требований в этой системе в момент t, то при одинаковых начальных условиях по распределению $q_1(t) \geqslant q(t)$, так что достаточно доказать стабильность мажорирующей системы, т. е. стохастическую ограниченность процесса $q_1(t)$.

Для системы $M_1^{(1)}$ введём вспомогательную систему $\tilde{M}_1^{(j)}$, предположив, что в $\tilde{M}_1^{(j)}$ на орбите всё время есть требования, а запросы с орбиты поступают с постоянной интенсивностью $\nu(j)$. Пусть $\tilde{Y}_j(t)$ и $Y_1(t)$ — число требований, обслуженных за время $(0,t)$ в системах $\tilde{M}_1^{(j)}$ и $M_1^{(1)}$ соответственно.

В системе $\tilde{M}_1^{(j)}$ $\zeta_l^{(j)}$ — время, через которое после $t_l^{(j)}$ поступит запрос с орбиты. Тогда последовательность $\{\zeta_l^{(j)}\}_{l=1}^{\infty}$ состоит из независимых экспоненциально распределённых случайных величин с параметром $\nu(j)$ и не зависит от

$\{t_l^{(j)}\}_{l=1}^{\infty}$. Определим последовательность $\{\tilde{T}_n^{(j)}\}_{n=1}^{\infty}$ общих точек регенерации потоков $\tilde{Y}(t)$ и $X(t)$ рекуррентными соотношениями

$$\tilde{T}_n^{(j)} = \min\left\{\theta_k > \tilde{T}_{n-1}^{(j)}\colon \bigcup_{l=0}^{\infty}\{t_l^{(j)} < \theta_k \leqslant t_l^{(j)} + \zeta_l^{(j)}\}\right\}, \quad \tilde{T}_0^{(j)} = 0.$$

Тогда $\{\tilde{\tau}_n^{(j)} = \tilde{T}_{n+1}^{(j)} - \tilde{T}_n^{(j)}\}_{n=1}^{\infty}$ — последовательность независимых одинаково распределённых случайных величин и $E\tilde{\tau}_1^{(j)} < \infty$. Для фиксированного j определим приращения $\Delta_X^{(j)}(n)$, $\Delta_{Y_1}^{(j)}(n)$ и $\tilde{\Delta}_Y^{(j)}(n)$ соответственно потоков $X(t)$, $Y_1(t)$ и $\tilde{Y}_j(t)$ на интервалах $(\tilde{T}_n^{(j)}, \tilde{T}_{n+1}^{(j)})$.

Напомним, что $Y_1(t)$ — число требований, обслуженных в системе $M_1^{(1)}$ за время $(0,t)$. Поскольку в $\tilde{M}_1^{(j)}$ на орбите всё время есть требования, а интенсивность запросов с орбиты постоянна и равна $\nu(j)$, то $\tilde{Y}_1^{(j)}(t)$ представляет собой число требований, обслуженных в системе типа $M|G|m|0$ за время $(0,t)$. В соответствии с (10) интенсивность $\tilde{\lambda}_{Y_j}$ потока $\tilde{Y}_j(t)$ определяется равенством

$$\tilde{\lambda}_{Y_j} = \frac{\nu(j)\sum\limits_{k=0}^{m-1}\frac{(\nu(j)b)^k}{k!}}{\sum\limits_{k=0}^{m}\frac{(\nu(j)b)^k}{k!}}.$$

Поскольку $\lim\limits_{j\to\infty}\nu(j) = \infty$, отсюда находим, что

$$\lim_{j\to\infty}\tilde{\lambda}_{Y_j} = \frac{m}{b}. \tag{15}$$

Лемма 2′. *Если*

$$q_1(t) \xrightarrow[t\to\infty]{P} \infty, \tag{16}$$

то для любых $j = 1, 2, \ldots$ *и* $\epsilon > 0$ *найдётся* n_1*, такое что при* $n > n_1$

$$E\Delta_Y^{(j)}(n) \geqslant E\tilde{\Delta}_Y^{(j)}(1) - \epsilon.$$

Доказательство опирается на монотонность $\nu(j)$, а в остальном аналогично доказательству леммы 2 и поэтому здесь не приводится.

Поскольку $\{\tilde{T}_n^{(j)}\}_{n=1}^{\infty}$ — последовательность общих точек регенерации для потоков $\tilde{Y}_j(t)$ и $X(t)$, их интенсивности равны

$$\lambda = \frac{E\Delta_X^{(j)}(1)}{E\tilde{\tau}_j^{(1)}}, \quad \tilde{\lambda}_{Y_j} = \frac{E\tilde{\Delta}_Y^{(j)}(1)}{E\tilde{\tau}_j^{(1)}},$$

и в силу (15)

$$\rho_j = \frac{\lambda}{\tilde{\lambda}_{Y_j}} = \frac{E\Delta_X^{(j)}(1)}{E\tilde{\Delta}_Y^{(j)}(1)} \xrightarrow[j\to\infty]{} \frac{\lambda b}{m} = \rho.$$

Теперь предположим, что $\rho = \lambda b/m < 1$, но имеет место сходимость (16), означающая нестабильность $q_1(t)$, а следовательно, и $q(t)$. Выберем j_0 из условия

$$1 - \rho_{j_0} > 0$$

и возьмём $\varepsilon = (1-\rho_{j_0})E\tilde{\Delta}_Y^{(j_0)}(1)$. Для вложенного процесса $q_1^{(j_0)}(n) = q_1\big(\tilde{T}_n^{(j_0)}\big)$ имеем

$$q_1^{(j_0)}(n+1) = q_1^{(j_0)}(n) + \Delta_X^{(j_0)}(n) - \Delta_Y^{(j_0)}(n).$$

Переходя к математическим ожиданиям из леммы $2'$ для выбранных j_0 и ε при $n > n_1$, получаем

$$\begin{aligned} Eq_1^{(j_0)}(n+1) &= Eq_1^{(j_0)}(n) + E\Delta_X^{(j_0)}(1) - E\Delta_Y^{(j_0)}(n) \leqslant \\ &\leqslant Eq_1^{(j_0)}(n) + E\Delta_X^{(j_0)}(1) - E\tilde{\Delta}_Y^{(j_0)}(1) + \varepsilon = \\ &= Eq_1^{(j_0)}(n) - E\tilde{\Delta}_Y^{(j_0)}(1)(1-\rho_{j_0}) + E\tilde{\Delta}_Y^{(j_0)}(1)(1-\rho_{j_0}) = Eq_1^{(j_0)}(n), \end{aligned}$$

что противоречит (16). □

Возникает вопрос, а каково условие стабильности, когда

$$\lim_{j\to\infty} \nu(j) = \nu < \infty. \tag{17}$$

Использованный подход приводит к следующему результату.

Следствие 2. *Пусть $X(t)$ — пуассоновский поток, а $\nu(j)$ монотонно возрастает и выполнено* (17). *Тогда для системы M_1 с интенсивностью повторений вызовов $\nu(j)$ справедливо утверждение следствия* 1.

6. Заключение

В статье установлено необходимое и достаточное условие стабильности систем обслуживания с повторными вызовами в достаточно общих предположениях относительно процессов, определяющих их функционирование. Рассмотрены два класса моделей: системы, в которых интенсивность потока повторных вызовов зависит от числа требований на орбите (класс I), и системы, в которых она постоянна (класс II).

Для систем первого класса при условии, что интенсивность потока повторных вызовов монотонно стремится к бесконечности, когда число требований на орбите растёт до бесконечности, условие стабильности такое же, как для классических многоканальных систем с общей очередью. Это вполне ожидаемый факт, но, насколько нам известно, до сих пор он не был строго доказан для достаточно общих моделей.

Для систем второго класса условие стабильности выражается через интенсивности входящего потока и потока повторных вызовов с орбиты и вероятность P_m потери требования в системе с отказами и регенерирующим входящим потоком. Явные формулы для P_m известны лишь в случае пуассоновского входящего потока (см., например, [2]). В такой ситуации условие стабильности приводится в следствии 1, и оно совпадает с полученным в [5] для систем с экспоненциально распределённым временем обслуживания.

Тем не менее доказанное в статье условие стабильности может быть использовано для анализа многих прикладных систем. Во-первых, предположение о том, что входящий поток регенерирующий, не является ограничительным, поскольку класс таких потоков весьма широк. Во-вторых, вероятность P_m легко оценивается по наблюдениям за функционированием системы. Пусть, как и раньше, $X(t)$ — число требований, поступивших в систему, $N(t)$ — число вызовов с орбиты, $Z(t)$ — число требований, направленных на орбиту, за время $(0, t)$. Все эти потоки регенерирующие, и почти наверное существуют

$$\lambda_X = \lim_{t\to\infty} \frac{X(t)}{t}, \quad \lambda_N = \lim_{t\to\infty} \frac{N(t)}{\tau(t)}, \quad \lambda_Z = \lim_{t\to\infty} \frac{Z(t)}{t}.$$

Здесь $\tau(t)$ — суммарное время на интервале $(0, t)$, когда на орбите есть требования.

Поскольку в соответствии с леммой 1 коэффициент загрузки равен

$$\rho = \frac{\lambda_X}{(\lambda_X + \lambda_N)(1 - P_m)},$$

в качестве его оценки при достаточно большом T можно взять

$$\rho_T^* = \frac{\hat{\lambda}_X^2}{(\hat{\lambda}_X + \hat{\lambda}_N)(\hat{\lambda}_X - \hat{\lambda}_Z)},$$

где

$$\hat{\lambda}_X = \frac{X(T)}{T}, \quad \hat{\lambda}_N = \frac{N(T)}{\tau(T)}, \quad \hat{\lambda}_Z = \frac{Z(T)}{T}.$$

Тогда ρ_T^* — состоятельная оценка для ρ и можно построить доверительный интервал, а также критерий для проверки гипотезы о стабильности системы.

Автор выражает благодарность гранту 17-01-00468.

Литература

[1] Афанасьева Л. Г., Ткаченко А. Многоканальные системы обслуживания с регенерирующим входящим потоком // Теория вероятн. и её примен. — 2013. — Т. 58, № 2. — С. 210—234.

[2] Севастьянов Б. А. Эргодическая теорема для марковских процессов и её приложение к телефонным системам с отказами // Теория вероятн. и её примен. — 1957. — Т. 2, № 1. — С. 106—116.

[3] Afanaseva L. G., Bashtova E. E. Coupling method for asymptotic analysis of queue with regenerative input and unreliable server // Queueing Systems. — 2014. — Vol. 76, no. 2. — P. 125—147.

[4] Artalejo S. R. Stationary analysis of the characteristics of the $M|M|2$ queue with constant repeated attempts // Opsearch. — 1996. — Vol. 33. — P. 83—95.

[5] Artalejo S. R., Gómez-Corral A., Neuts M. F. Analysis of multiserver queues with constant retrial rate // Europ. J. Operat. Res. — 2001. — Vol. 135. — P. 569—581.

[6] Choi B. D., Rhee K. H., Park K. K. The $M|G|1$ retrial queue with retrial rate control policy // Probab. Eng. Inform. Sci. — 1993. — Vol. 7. — P. 29—46.

[7] Choi B. D., Rhee K. H., Park K. K., Pearce C. E. M. An $M|M|1$ retrial queue with control policy and general retrial times // Queueing Systems. — 1993. — Vol. 14. — P. 175—292.

[8] Choi B. D., Shin J. W., Ahn W. C. Retrial queues with collision arising from unslotted CSMA|CD protocol // Queueing Systems. — 1992. — Vol. 11. — P. 335—356.

[9] Cohen S. W. Basic problems of telephone traffic theory and influence of repeated calls // Philips Telecommunic. Rev. — 1957. — Vol. 18. — P. 49—100.

[10] Falin G. I., Artalejo J. R. Approximation for multiserver queues with bulking/retrial discipline // OR Spectrum. — 1995. — Vol. 17. — P. 239—244.

[11] Falin G. I., Templeton S. G. C. Retrial Queues. — London: Chapman & Hall, 1997.

[12] Fayolle G. A simple telephone exchange with delayed feedback // Teletraffic Analysis in Computer Performance Evaluation / O. S. Boxma, S. W. Cohen, H. C. Tijms, eds. — Amsterdam: Elsevier, 1986.

[13] Feller W. An Introduction to Probability Theory and Its Applications. Vol. II. — Wiley, 1950.

[14] Gómez-Corral A., Ramalhoto M. F. On the stationary distribution of Markovian process arising in the theory of multiserver retrial queueing systems // Math. Comput. Modelling. — 1999. — Vol. 30, — P. 141—158.

[15] Grandell J. Doubly Stochastic Poisson Process. — Berlin: Springer, 1976.

[16] Neuts M. F. Structured Stochastic Matrices of $M|G|1$ Type and Their Applications. — New York: Marcel Dekker, 1989.

[17] Thorisson H. Coupling, Stationarity and Regeneration. — New York: Springer, 2000.

[18] Wilkinson R. I. Theories for toll traffic engineering in the USA // Bell System Tech. J. — 1950. — Vol. 35. — P. 421—514.

[19] Zeifman A., Satin Y., Morozov E., Nekrasova R., Gorshenin A. On the ergodicity bounds for a constant retrial rate queueing model // 8th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). — 2016. — P. 269—272.