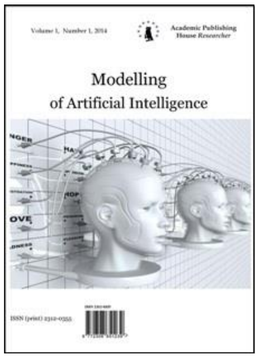
Published in the Russian Federation
Modeling of Artificial Intelligence
Has been issued since 2014.
ISSN: 2312-0355
E-ISSN: 2413-7200
Vol. 12, Is. 4, pp. 194-203, 2016

DOI: 10.13187/mai.2016.12.194
www.ejournal11.com



UDC 519.21

The Employment Periods in the Single-Channel Computing System of Service with the Poisson Entering Flow and the Regulator of Queue

Arsen R. Simonyan [a,*], Elena I. Ulitina [a], Simon Zh. Simavoryan [a], Nadezhda A. Kornienko [a]

[a] Sochi State University, Russian Federation

Abstract
In case of simulation of operations of information systems the mathematical apparatus of the theory of queues is often applied and they are called computing systems with queues. The main characteristics of such models are the periods of zanyatost. In this operation the model with one servicing instrument (server), the Poisson entering flow and the regulator of queue in front of the server is considered. The regulator of queue disperses the logging-in calls on several flows and ranges flows on priorities. In the described computing system the main characteristics are researched by methods of probability theory and a research of operations.
Keywords: information system, computing systems with queues, the theory of queues, the employment period, probability, event streams, calls.

1. Введение

Бурные темпы развития информационно-вычислительных систем диктует необходимость построения математических моделей для их исследования. По мнению известного специалиста по вычислительным систе-мам Л. Клейнрока "применения теории очередей для анализа распределения ресурсов и решения задач о потоках данных в вычислительных системах является, по-видимому, единственным доступным специалистам по вычислительной технике методом, позволяющим понять сложные связи в таких системах" (Клейнрок, 1979).

В этой связи, для изучения вычислительных систем, наиболее подходящими являются одноканальные модели теории очередей с пуассоновскими входящими потоками. Изучению этих моделей посвящены многочисленные работы (Симонян, 2003; Simonyan, 2004; Даниелян, 2001; Бронштейн, 1976; Гнеденко, 1973). В этих работах исследованы и получены результаты для основных характеристик модели при разных (приоритетных, неприоритетных и параметрических дисциплинах).

Вычислительные системы с параметрическими дисциплинами представляют собой особый класс систем с широкими приложениями. Таким системам посвящены работы (Симонян, 2005; Simonyan, 2013; Симонян, 2010; Simonyan, 2016)

* Corresponding author
E-mail addresses: oppm@mail.ru (A.R. Simonyan), ulitinaelena@mail.ru (E.I. Ulitina), simsim58@mail.ru (S.Zh. Simavoryan), kornienko_nadja@mail.ru (N.A. Kornienko)

В данной работе предполагается наличие, перед прибором обслуживания, регулятора очереди, которых входящий поток вызовов распределяет на несколько приоритетных потоков.

2. Обсуждение

2.1. Обозначения и точные определения

В систему обслуживания, состоящую из одного прибора, поступает пуассоновский потоков вызовов с параметром $a > 0$. Перед прибором стоит регулятор очереди, который распределяет вызовы входящего потока по потокам

$$L_1, L_2, \dots, L_r$$

с вероятностями $p_1, p_2, \dots, p_r$ соответственно Длительности обслуживания вызовов в совокупности независимы. Длительность времени обслуживания вызовов потока L_k есть сл. в. с ф. р. $B_k(t)$, $k = \overline{1, r}$. Допускается неограниченная очередь.

Под периодом занятости прибора (или системы) понимается длительность промежутка времени, начинающегося с момента поступления, некоторого вызова, заставшего систему свободной от вызовов до следующего момента освобождения системы от вызовов. Ясно, что если не допускается прерывание обслуживаемого вызова и замещение его на приборе другим, то порядок обслуживания вызовов здесь не имеет значения.

Для образности изложения введем следующую терминологию и установим в связи с этим порядок обслуживания. Вызовы потока еще L_kназовем вызовами приоритета k и будем говорить, что вызовы потока L_i, имеют более высокий приоритет по отношению к вызовам потока L_j, если $i < j$. При этом вызовы приоритета i имеют преимущество перед вызовами приоритета $j(i < j)$. Это преимущество заключается в следующем. Среди вызовов, ожидающих начала обслуживания, вызовы высшего приоритета обслуживаются раньше вызовов низшего приоритета. Для вызовов одного приоритета порядок обслуживания считаем инверсионным. Это значит, что среди вызовов одного приоритета первым из ожидающих обслуживается тот, который поступил позже остальных.

Пример. Имеются r ящиков, занумерованных числами $\overline{1, r}$. Поступающий вызов (изделие) приоритета k помещается в ящик с номером k над имеющимися в нем изделиями. На обслуживание выбирается, начиная с ящика с меньшим номером, изделие, находящееся сверху в этом ящике.

Если во время обслуживания некоторого вызова поступает вызов более высокого приоритета, то можно представить случаи, когда обслуживание прерывается и сразу же начинается обслуживание поступившего вызова более высокого приоритета. В связи с этим мы будем различать следующие схемы обслуживания с преимуществом.

Схема 1. Если во время обслуживания вызова поступает вызов более высокого приоритета, то прерывается обслуживание вызова и начинается обслуживание поступившего вызова; когда система освободится от вызовов более высокого приоритета, чем прерванный вызов, последний дообслуживается оставшееся время обслуживания.

Схема 2. То же, но прерванный вызов «теряется».

Схема 3. То же, но прерванный вызов обслуживается заново (не учитывается время имевшегося обслуживания). При этом будем различать две возможности:

а) неидентичное обслуживание заново, при котором прерванный вызов при возвращении на прибор обслуживается случайное время, не зависящее от предыдущего его (вызова) обслуживания и имеющее ту же ф. р., что и ранее;

б) идентичное обслуживание заново, при котором прежде всего разыгрывается длительность обслуживания вызова. Пусть эта длительность равна t. Если произошло прерывание вызова, то при новом поступлении на прибор на обслуживание вызова следует затратить время, равное t.

Схема 4. Если вызов начал обслуживаться, то он обслуживается до конца, несмотря на поступления вызовов более высокого приоритета (прерывания не происходит).

Введем некоторые предварительные обозначения, годные для всех схем:

$\prod(t)$ — ф. р. периода занятости системы;

$\prod_k(t)$ — ф. р. периода занятости системы обслуживанием вызовов приоритета k и выше, т. е. длительности промежутка времени, начинающегося с момента поступления вызова приоритета k или выше, заставшего систему свободной от вызовов, до следующего момента освобождения системы от вызовов приоритета k и выше; если дополнительно известно, что этот период занятости начинается с обслуживания вызова приоритета $i(i = \overline{1,k})$, то указанную ф. р. будем обозначать через $\prod_{ki}(t)$. Ясно, что $\prod(t) = \prod_r(t)$.

Обозначим еще через $H_k(t)$ ф. р. длины промежутка времени, начинающегося с момента поступления вызова приоритета k, заставшего систему свободной от вызовов, до следующего непосредственно момента освобождения системы от этого вызова и вызовов приоритета выше чем k.

Положим

$$\sigma_i = a(p_1 + \cdots + p_i);\ i = \overline{1,r};\ \ \sigma_0 = 0;\ \ \sigma_r = a;\ \textstyle\prod_0(t) \equiv 0.$$

Системы с прерыванием начатого обслуживания вызова вызовом более высокого приоритета иначе называют системами с *абсолютным* приоритетом. Системы же без прерывания допускают несколько различных дисциплин обслуживания, в зависимости от расположения вызовов разных потоков в очереди. Две такие дисциплины (относительный приоритет, чередование приоритетов) ниже подробно рассмотрены.

2.2. Распределение периода занятости систем без прерывания

Введем следующие обозначения

$$\beta_k(s) = \int_0^\infty e^{-st} dB_k(t)\,, \qquad \pi_k(s) = \int_0^\infty e^{-st} d\Pi_k(t).$$

Для схемы 4 справедлива важная

Теорема.

$$\text{а) } \pi(s) = \sum_{i=1}^{r} p_i \beta_i(s + \sigma - \sigma\pi(s)), \tag{2.1}$$

причем это функциональное уравнение определяет единственную функцию $\pi(s)$, *аналитическую в полуплоскости* $Re\ s > 0$, *в которой* $|\pi(s)| \le 1$;

б) если $\rho_1 = a(p_1\beta_{11} + \cdots + p_r\beta_{r1}) < 1$ *то* $\pi(+0) = 1$; *в противном случае* $0 < \pi(+0) < 1$;

б) *если* $\rho_1 < 1$, то *первые три момента ф.р.* $\prod(t)$ *определяются соотношениями*

$$\sigma\pi_1 = \frac{\rho_1}{\rho} \tag{2.2}$$

$$\sigma\pi_2 = \frac{\rho_2}{\rho^3} \tag{2.3}$$

$$\sigma\pi_3 = \frac{\rho_3}{\rho^4} + 3\frac{\rho_2^2}{\rho^5} \tag{2.4}$$

где

$$\rho_2 = a(p_1\beta_{12} + \cdots + p_r\beta_{r2}),$$
$$\rho_3 = a(p_1\beta_{13} + \cdots + p_r\beta_{r3}), \qquad \rho = 1 - \rho_1.$$

Теорема верна при любом порядке обслуживания вызовов.

Доказательство теоремы. А. Независимо от эволюции системы наступают *катастрофы*, поток которых — пуассоновский с параметром $s > 0$. При перестановке любых двух вызовов в очереди распределение периода занятости не изменится. Поэтому считаем, что вызовы (всех потоков, независимо от принадлежности к тому или иному потоку) обслуживаются в инверсионном порядке, т. е. среди вызовов, ожидающих начала обслуживания, первым обслуживается вызов, поступивший последним.

$\pi(s)$ – вероятность того, что за период занятости *катастрофа* не наступит. Свяжем период занятости с тем вызовом, с обслуживания которого начинается сам период занятости. Обратно, всякому вызову можем сопоставить «период занятости», понимая под этим длину промежутка времени, начинающегося с обслуживания этого вызова до следующего момента освобождения системы от этого вызова и вызовов, поступивших после него.

Периоды занятости, соответствующие вызовам, поступившим в систему во время обслуживания некоторого вызова, не пересекаются, независимы в совокупности и одинаково распределены. Период занятости, соответствующий некоторому вызову, складывается из длительности обслуживания его и длин периодов занятости, соответствующих вызовам, поступившим за время его обслуживания. Пусть за период занятости не произошла *катастрофа,* для этого необходимо и достаточно, чтобы за время обслуживания первого вызова (с которого начинается период занятости и который является вызовом приоритета i с вероятностью p_i) не произошло событие следующего суммарного потока событий: потока *катастроф* и потока вызовов, за соответствующий период занятости которых наступает *катастрофа*. При этом слагаемые потоки независимы и каждый является пуассоновым с параметром s

и $\sigma[1-\pi(s)]$ — соответственно; поэтому суммарный поток — тоже пуассоновый с параметром $s+\sigma-\sigma\pi(s)$. Отсюда следует формула (2.1).

Б. Положим

$$\beta(s)=\sum_{i=1}^{r} p_i\beta_i(s).$$

Так как функции $\beta_i(s)$ вполне монотонны, то и $\beta(s)$ — вполне монотонная функция (см. § 8 доп.). Далее $\beta(0)=1$. Теперь утверждение б) теоремы следует из теоремы 1 § 2 гл. 1(Гнеденко, 1973).

Теорема доказана.

Замечание. Легко видеть, что период занятости в схеме 1 совпадает с периодом занятости в схеме 4.

2.3. Системы с прерыванием. Схема 2 (потеря прерванного вызова).

Теорема. Для схемы 2.

$$h_k(s)=\beta_k(s+\sigma_{k-1})+\frac{\sigma_{k-1}}{s+\sigma_{k-1}}[1-\beta_k(s+\sigma_{k-1})]\pi_{k-1}(s), \tag{3.1}$$

$$\pi_{kk}(s)=h_k\big(s+a_k-a_k\pi_{kk}(s)\big), \tag{3.2}$$

$$\pi_{ki}(s)=\pi_{k-1i}\big(s+a_k-a_k\pi_{kk}(s)\big), \tag{3.3}$$

$$\sigma_k\pi_k(s)=a_1\pi_{k1}(s)+\cdots+a_k\pi_{kk}(s)\qquad (i<k) \tag{3.4}$$

определяет единственные функции $h_k(s_1)\pi_{ki}(s)$, $\pi_k(s)$, $i=\overline{1,k}$, $k=\overline{1,r}$ *аналитические в полуплоскости* $Re\ s>0$, *где* $|h_k(s)|<1$, $|\pi_{ki}(s)|<1$, $|\pi_k(s)|<1$;

б) если, $a_1\beta_{11}+\frac{a_2}{\sigma_1}[1-\beta_2(\sigma_1)]+\cdots+\frac{a_k}{\sigma_{k-1}}[1-\beta_k(\sigma_{k-1})]\le 1,$ (3.5)

то $h_{k+1}(0)=\pi_{ki}(0)=\pi_k(0)=1$; *в противном случае*

$$0<h_{k+1}(0)<1,0<\pi_{ki}(0)<1,0<\pi_k(0)<1;$$

в) положим $\rho_{k1}=a_1\beta_{11}+\frac{a_2}{\sigma_1}[1-\beta_2(\sigma_1)]+\cdots+\frac{a_k}{\sigma_{k-1}}[1-\beta_k(\sigma_{k-1})]$, $\rho_k=1-\rho_{k1}$, $\rho_{k2}=a_1\beta_{12}+2\left[-\frac{a_2}{\sigma_1}C_2\rho_1-\cdots-\frac{a_k}{\sigma_{k-1}}C_k\rho_{k-1}+\frac{\rho_1}{\sigma_1}(\rho_1-\rho_2)++\cdots+\frac{\rho_{k-1}}{\sigma_{k-1}}(\rho_{k-1}-\rho_k)\right]$

Тогда при $p_{k1}<1$ *выполнено*

$$\sigma_k\pi_{k1}=\frac{\rho_{k1}}{\rho_k}, \tag{3.6}$$

$$\sigma_k\pi_{k2}=\frac{\rho_{k2}}{\rho_k^3}, \tag{3.7}$$

$$h_k=\frac{1}{\sigma_{k-1}\rho_{k-1}}[1-\beta_k(\sigma_{k-1})], \tag{3.8}$$

$$h_k=\frac{2}{\sigma_{k-1}\rho_{k-1}}\left[-C+\frac{1-\beta_k(\sigma_{k-1})}{\sigma_{k-1}}\right]+\frac{\rho_{k-12}}{\rho_{k-1}^3}\frac{1-\beta_k(\sigma_{k-1})}{\sigma_{k-1}}, \tag{3.9}$$

Доказательство теоремы. А. Докажем сначала формулы (3.1) —(3.4).

Предполагаем, что независимо от функционирования системы наступают *катастрофы*, поток которых является пуассоновским с параметром $s>0$. Величина

$$\beta_k(s+\sigma_{k-1})=\int_0^\infty e^{-(s+\sigma_{k-1})t}dB_k(t)$$

— вероятность того, что за длительность обслуживания вызова приоритета k не наступали *катастрофы* и вызовы приоритета выше k.

Суммарный поток *катастроф* и вызовов приоритета выше k (см. § 1 доп.) — пуассоновским с параметром $s+\sigma_{k-1}$. Каждый вызов суммарного потока независимо от остальных вызовов с вероятностью $\frac{\sigma_{k-1}}{s+\sigma_{k-1}}$ принадлежит к потоку вызовов приоритета выше k. Следовательно, $\frac{\sigma_{k-1}}{s+\sigma_{k-1}}[1-\beta_k(s+\sigma_{k-1})]$ есть вероятность того, что обслуживание вызова приоритета к прервано, до момента же прерывания *катастрофы* не наступали.

Формула (3.1) вытекает из следующих соображений. Пусть за промежуток времени, отсчитываемый с момента начала обслуживания вызова приоритета k и кончающийся первым моментом освобождения системы от вызовов приоритета выше k и этого вызова приоритета k, не наступали *катастрофы*. Для этого необходимо и достаточно, чтобы

либо вызов приоритета k начал обслуживаться и не был прерван, а за длительность обслуживания *катастрофы* не наступали;

либо обслуживание вызова приоритета k прервано поступлением вызова более высокого приоритета (что означает потерю прерванного вызова); *катастрофы* не наступали до потери вызова и за следующий после потери вызова период занятости обслуживанием вызовов приоритета выше k.

Период занятости обслуживанием вызовов приоритета k и выше, начавшийся с обслуживания вызова приоритета i, назовем ki -периодом.

Прежде всего протекает промежуток ξ начавшийся с обслуживания вызова приоритета k и кончающийся освобождением системы от вызовов приоритета выше k и этого вызова приоритета k. За время ξ вызовы приоритета k могут не поступить, а могут и поступить. Если за время ξ вызовы приоритета k в систему не поступают, то с окончанием ξ кончается и первоначальный -период. Если за время ξ вызовы приоритета k поступали, то с каждым из этих вызовов связан -период, лишь после окончания которых (kk-периодов) первоначальный kk-период кончается.

Вызов приоритета k называем *плохим*, если за период, связанный с ним, наступила *катастрофа*. Каждый вызов приоритета k независимо от остальных вызовов является *плохим* с вероятностью $1-\pi_{kk}(s)$. Поток *плохих* вызовов приоритета k — пуассоновый с параметром $a_k(1-\pi_{kk}(s))$. Поток *плохих* вызовов приоритета k и *катастроф* — пуассоновый с параметром $s+a_k-a_k\pi_{kk}(s)$).

Чтобы вызов приоритета k не оказался *плохим* (вероятность $\pi_{kk}(s)$), необходимо и достаточно, чтобы за время ξ не наступали *катастрофы* и не поступали *плохие* вызовы приоритета k (вероятность $h_k(s+a_k-a_k\pi_{kk}(s))$.Этим доказана формула (3.2). Формулы (3.3) и (3.4) доказываются аналогично.

Формулы (3.1) — (3.4) доказывались при $s>0$. По принципу аналитического продолжения получаем, что они верны при $Re\, s>0$.

Б. Укажем, как для вычисления $\pi_k(s)$ пользоваться системой (3.1) — (3.4) рекуррентных функциональных уравнений.

Пусть $\sigma_0=0$ и $k>1$. Тогда из (3.1) имеем $h_1(s)=\beta_1(s)$. Формула (3.2) дает функциональное уравнение $\pi_{11}(s)=\beta_1(s+a_1-a_1\pi_{11}(s))$, определяющее $\pi_{11}(s)$. Далее, $\pi_1(s)=\pi_{11}(s)$.

Допустим, что мы определили $h_{k-1}(s)$, $\pi_{k-1}(s)$, $\pi_{k-1i}(s)$ $(i=\overline{1,k-1})$. Тогда по формуле (3.1) вычисляется $h_k(s)$, и из (3.2) - $\pi_{kk}(s)$. Используя значения $\pi_{kk}(s)$ и $\pi_{k-1i}(s)$ $(i=\overline{1,k-1})$, по (3.3) вычисляем $\pi_{ki}(s)$ $(i=\overline{1,k-1})$. Теперь осталось подставить значения $\pi_{ki}(s)$ $(i=\overline{1,k})$ в (3.4).

В. Докажем методом математической индукции, что уравнения (3.1) — (3.4) определяют единственные функции $h_k(s)$, $\pi_k(s)$, $\pi_{ki}(s)$ $(i=\overline{1,k})$ при $s>0$, являющиеся вполне монотонными функциями, причем $h_k(s)<1$, $\pi_k(s)<1$, $\pi_{ki}(s)<1$ $(i=\overline{1,k})$. При $k=1$ утверждение верно (см. доказательство теоремы § 2 гл. 1). Предположим, что утверждение верно для всех $j<k$, докажем его справедливость при $j=k$.

По предположению индукции функции $h_{k-1}(s)$, $\pi_{k-1}(s)$, $\pi_{k-1i}(s)$ $(i=\overline{1,k-1})$ —вполне монотонны при $s>0$, где $h_{k-1}(s)<1$, $\pi_{k-1}(s)<1$, $\pi_{k-1i}(s)<1$. Вполне монотонной является также функция

$$\frac{\sigma_{k-1}}{s+\sigma_{k-1}}[1-\beta_k(s+\sigma_{k-1})]=\int_0^\infty \sigma_{k-1}e^{-(\sigma_{k-1}+s)t}[1-B_k(t)]dt.$$

Тогда вполне монотонна функция $-\frac{\sigma_{k-1}}{s+\sigma_{k-1}}[1-\beta_k(s+\sigma_{k-1})]\pi_{k-1}(s)$ (св. 1, п. А, §8 доп.) как произведение двух вполне монотонных сомножителей.

Наконец, из (3.1) следует, что $h_k(s)$ как сумма двух вполне монотонных слагаемых вполне монотонна. При $\pi_{k-1}(s)<1$ из (3.1) имеем $\pi_k(s)<1$. Как и при доказательстве теоремы § 2 гл. 1, показывается, что существует единственное решение $\pi_{kk}(s)$ уравнения (3.2), являющееся вполне монотонной функцией при $s>0$, где $\pi_{kk}(s)<1$.

По св. 2 вполне монотонных функций вполне монотонны функции $\pi_{ki}(s)$ $(i=\overline{1,k-1})$. Действительно, функция $s+a_k-a_k\pi_{kk}(s)$ имеет вполне монотонную производную

$$\pi_{ki}(s)=\pi_{k-1i}(s+a_k-a_k\pi_{kk}(s))<\pi_{k-1i}(s)<1, s>0$$

Для доказательства утверждения а) теоремы осталось показать аналитичность $h_k(s)$, $\pi_k(s)$, $\pi_{ki}(s)$ $(i=\overline{1,k})$ в области $Re\ s>0$, причем

$$|h_k(s)|<1, |\pi_k(s)|<1, |\pi_{ki}(s)|<1.$$

Функция $h_k(s)$ как вполне монотонная функция при $s>0$ представим а в виде

$$h_k(s)=\int_0^\infty e^{-st}dH_k(t), \tag{3.10}$$

где $H_k(t)$ — некоторая мера на $[0,\infty]$. Берем аналитическое продолжение функции $h_k(s)$, заданной в форме (3.10), в области $Re\, s>0$. Из единственности аналитического продолжения $h_k(s)$ в область $Re\ s>0$ и из представления (3.10) из (3.1) имеем $|h_k(s)|<1$.

Аналогично $\pi_{kk}(s)$ представима в виде

$$\pi_{kk}(s)=\int_0^\infty e^{-st}d\prod\nolimits_{kk}(t),$$

где $\prod_{kk}(t)$ — некоторая мера на $[0,\infty]$. Аналитичность $\pi_{kk}(s)$ и $|\pi_k(s)|<1$ при $Re\ s>0$ также следует из теоремы § 2 гл. 1.

Аналитичность же $\pi_{ki}(s)$ $(i=\overline{1,k-1})$ и $\pi_k(s)$ и условия $|\pi_{ki}(s)|<1$, $|\pi_k(s)|<1$ получаются по предположению индукции из (3.3), (3.4).

Г. Пусть $h_k(0)=0$. Тогда по теореме § 2 гл. 1 при выполнении условия $a_kh_{k-1}\le 1$ из (3.2) следует, что $\pi_{kk}(0)=1$; в противном случае $0<\pi_{kk}(0)<1$. Следовательно, при $a_kh_{k-1}\le 1$ на основании (3.1), (3.3), (3.4)

$$\pi_{ki}(0)=\pi_k(0)=h_{k+1}(0)=1\ \big(i=\overline{1,k}\big);$$

в противном случае $0<\pi_{ki}(0)<1$, $i=\overline{1,k}$, $0<\pi_k(0)<1$, $0<h_{k+1}(0)<1$. Пусть $0<h_k(0)<1$. Тогда $0<\pi_{ki}(0)<1$, $i=\overline{1,k}$, k, $0<\pi_k(0)<1$, $0<h_{k+1}(0)<1$. Следовательно, методом математической индукции доказывается, что при замене условия (3.5) на $a_jh_{j1}\le 1$ $(j=\overline{1,k})$ утверждение пункта б) теоремы справедливо. Осталось заметить, что (3.5) и $a_jh_{j1}\le 1$ $(j=\overline{1,k})$ эквивалентны. Действительно, обозначим

$$\rho_{k1}=a_1\beta_{11}+\frac{a_2}{\sigma_1}[1-\beta_2(\sigma_1)]+\cdots+\frac{a_k}{\sigma_{k-1}}[1-\beta_k(\sigma_{k-1})], \tag{3.11}$$

$$\rho_k=1-\rho_{k1}$$

Из соотношений (3.2) — (3.4) следует соотношение

$$\sigma_k\pi_k(s)=\sigma_{k-1}\pi_{k-1}\big(s+a_k\pi_{kk}(s)\big),$$

продифференцировав которое по s и воспользовавшись при $s=0$ значением первого момента π_{kk1}, получаем

$$1+\sigma_k\pi_k(s)=\frac{1}{1-a_kh_{k1}}\big(1+\sigma_{k-1}\pi_{k-1}(s)\big). \tag{3.12}$$

С другой стороны, из (3.1) легко вычислить

$$h_{k1} = \frac{1-\beta_k(\sigma_{k-1})}{\sigma_{k-1}}[1+\sigma_{k-1}\pi_{k-11}]. \tag{3.13}$$

Подставляем значения $1+\sigma_k\pi_{k1}$ и $1+\sigma_{k-1}\pi_{k-11}$ из (3.13) в (3.12) и находим

$$\frac{1-\beta_k(\sigma_{k-1})}{\sigma_{k-1}}\cdot\frac{1}{h_{k1}} = \frac{1-\beta_{k-1}(\sigma_{k-2})}{\sigma_{k-2}}a_{k-1} +$$

$$+\frac{1-\beta_{k-1}(\sigma_{k-2})}{\sigma_{k-2}}\cdot\frac{1}{h_{k-11}} = \cdots = 1-\left[+a_1\beta_{11}+\frac{a_2}{\sigma_1}[1-\beta_2(\sigma_1)]+\cdots\right.$$

$$\left.\ldots+\frac{a_{k-1}}{\sigma_{k-2}}[1-\beta_{k-1}(\sigma_{k-2})]\right] = 1-\rho_{k-11}.$$

Откуда

$$h_{k1} = \frac{1-\beta_k(\sigma_{k-1})}{\sigma_{k-1}}\cdot\frac{1}{1-\rho_{k-11}} \tag{3.14}$$

Теперь, если $a_k h_k \le 1$, то из (3.14)

$$\frac{a_k}{\sigma_{k-1}}[1-\beta_k(\sigma_{k-1})] \le 1-\rho_{k-11} \text{ и } \rho_{k1} \le 1$$

Обратное тоже очевидным образом следует из (3.14).

Мы доказали не только утверждение пункта б), но и получили h_{k1}. Имея h_{k1} легко вычислить по (3.13) π_{k-11}.

2.4. Системы с прерыванием. Схема 3 (обслуживание заново).

Теорема 1. *Для схемы 3 с неидентичным обслуживанием заново*

а) $$h_k(s) = \beta_k(s+\sigma_{k-1})\left\{1-\frac{\sigma_{k-1}}{s+\sigma_{k-1}}[1-\beta_k(s+\sigma_{k-1})]\pi_{k-1}(s)\right\}^{-1}, \tag{4.1}$$

$$\sigma_k\pi_k(s) = a_1\pi_{k1}(s)+\cdots+a_k\pi_{kk}(s), \tag{4.2}$$

$$\pi_{kk}(s) = h_k(s+a_k-a_k\pi_{kk}(s)) \tag{4.3}$$

$$\pi_{ki}(s) = \pi_{k-1i}(s+a_k-a_k\pi_{kk}(s)) \tag{4.4}$$

причем эта система рекуррентных функциональных уравнений определяет единственные функции $h_k(s)$, $\pi_{ki}(s)$, $\pi_k(s)$, $i=\overline{1,k}$, $k=\overline{1,r}$, аналитические в полуплоскости $Re\ s>0$, в которой

$$|h_k(s)|<1, |\pi_{ki}(s)|<1, |\pi_k(s)|<1$$

б) если

$$a_1\beta_{11}+\frac{a_2}{\sigma_1}\left[\frac{1}{\beta_2(\sigma_1)}-1\right]+\cdots+\frac{a_k}{\sigma_{k-1}}\left[\frac{1}{\beta_k(\sigma_{k-1})}-1\right] \le 1, \tag{4.5}$$

то $h_{k+1}(0)=\pi_{ki}(0)=\pi_k(0)=1$; в противном случае

$$0<h_{k+1}(0)<1{,}0<\pi_{ki}(0)<1{,}0<\pi_k(0)<1;$$

в) положим

$$\rho_{k1} = a_1\beta_{11}+\frac{a_2}{\sigma_1}\left[\frac{1}{\beta_2(\sigma_1)}-1\right]+\cdots+\frac{a_k}{\sigma_{k-1}}\left[\frac{1}{\beta_k(\sigma_{k-1})}-1\right],$$

$$\rho_k = 1-\rho_{k1};$$

$$\rho_{k2} = a_1\beta_{12}+2\left\{-\frac{c_2a_2\rho_1}{\sigma_1[\beta_2(\sigma_1)]^2}-\cdots-\frac{c_ka_k\rho_{k-1}}{\sigma_{k-1}[\beta_k(\sigma_{k-1})]^2}+\frac{\rho_1(\rho_1-\rho_2)}{\sigma_1}+\cdots\right.$$

$$\left.\ldots+\frac{\rho_{k-1}(\rho_{k-1}-\rho_k)}{\sigma_{k-1}}+\frac{(\rho_1-\rho_2)^2}{a_2}+\cdots+\frac{(\rho_{k-1}-\rho_k)^2}{a_k}\right\},$$

$$c_i = \int_0^\infty te^{-\sigma_{i-1}t}B_i(s)\,;$$

тогда при $\rho_{k1}<1$

$$\sigma_k\pi_{k1} = \frac{\rho_{k1}}{\rho_k}; \tag{4.6}$$

$$\sigma_k\pi_{k2} = \frac{\rho_{k2}}{\rho_k^3}; \tag{4.7}$$

$$h_{k1} = \frac{1}{\sigma_{k-1}\rho_{k-1}}\left[\frac{1}{\beta_k(\sigma_{k-1})} - 1\right] \tag{4.8}$$

$$h_{k2} = \frac{2}{\sigma_{k-1}\rho_{k-1}}\left\{-\frac{c_k}{[\beta_k(\sigma_{k-1})]^2} + \frac{1}{\sigma_{k-1}}\left[\frac{1}{\beta_k(\sigma_{k-1})} - 1\right] + \right. \tag{4.9}$$

$$\left. + \frac{1}{\sigma_{k-1}\rho_{k-1}}\left[\frac{1}{\beta_k(\sigma_{k-1})} - 1\right]^2\right\} + \frac{\rho_{k-12}}{\rho_{k-1}^3}\cdot\frac{1}{\sigma_{k-1}}\left[\frac{1}{\beta_k(\sigma_{k-1})} - 1\right].$$

Доказательство теоремы 1. Пусть за промежуток времени, начинающийся с момента начала обслуживания вызова приоритета k и оканчивающийся следующим непосредственно моментом освобождения системы от этого вызова и вызовов приоритета выше k (напомним еще, что вызовы одного приоритета обслуживаются в инверсионном порядке), не произошла *катастрофа* (вероятность чего есть $h_k(s)$). Для этого необходимо и достаточно, чтобы

либо за время обслуживания вызова приоритета k не наступило «нежелательное» событие следующего суммарного потока событий: потока *катастроф* и потока вызовов приоритета выше k (вероятность $\beta_k(s + \sigma_{k-1})$);

либо за время обслуживания вызова приоритета k произошло «нежелательное» событие (вероятность $1 - \beta_k(s + \sigma_{k-1})$), причем таким «нежелательным» событием оказалось появление вызова приоритета выше к (вероятность $\frac{\sigma_{k-1}}{s+\sigma_{k-1}}$), и для того чтобы за период занятости системы обслуживанием вызовов приоритета выше k *катастрофа* не наступила (вероятность $\pi_{k-1}(s)$)* и нужно, чтобы за промежуток времени, начинающийся с момента начала (вторичного) обслуживания вызова приоритета k и оканчивающийся следующим моментом освобождения системы от этого вызова и вызовов приоритета выше k, *катастрофа* не произошла (вероятность $h_k(s)$). Следовательно,

$$h_k(s) = \beta_k(s + \sigma_{k-1}) + [1 - \beta_k(s + \sigma_{k-1})]\frac{\sigma_{k-1}}{s + \sigma_{k-1}}\pi_{k-1}(s)h_k(s),$$

т. е. доказана формула (4.1).

Остальные соотношения ничем не отличаются от аналогичных соотношений, приведенных в теореме § 3. Дальнейшие рассуждения проводятся по обычной схеме.

Подобные результаты имеют место и в случае идентичного обслуживания заново. Они приводятся ниже без доказательства.

Теорема 2. Для *схемы 3 с идентичным обслуживанием заново,*

а) рекурентные соотношения

$$h_k(s) = \int_0^\infty e^{-(s+\sigma_{k-1})x}\left[1 - \frac{\sigma_{k-1}}{s + \sigma_{k-1}}(1 - \right. \tag{4.10}$$

$$\left. - e^{-(s+\sigma_{k-1})x})\pi_{k-1}(s)\right]^{-1} dB_k(x)$$

и система (4.2)-(4.4) *определяют единственные функции* $h_k(s)$, $\pi_{ki}(s)$, $\pi_k(s)$, $i = \overline{1,k}$, $k = \overline{1,r}$ *вполне монотонные при* $s > 0$, *для* $|h_k(s)| < 1, |\pi_{ki}(s)| < 1, |\pi_k(s)| < 1$ *при* $Re\ s > 0$;

б) если $\beta_j(s)$ *аналитичны и правой полуплоскости, содержащей прямую* $Re\ s = \sigma_{j-1}$, $j = 2, k$, *и*

$$a_1\beta_{11} + \frac{a_2}{\sigma_1}[\beta_2(-\sigma_1) - 1] + \cdots + \frac{a_k}{\sigma_{k-1}}[\beta_k(-\sigma_k) - 1] \le 1,$$

То $h_{k+1}(0) = \pi_{ki}(0) = \pi_k(0) = 1$*; противном случае*

$$0 < h_{k+1}(0) < 1, 0 < \pi_{ki}(0) < 1; 0 < \pi_k(0) < 1;$$

в) положим

$$\rho_{k1} = a_1\beta_{11} + \frac{a_2}{\sigma_1}[\beta_2(-\sigma_1) - 1] + \cdots + \frac{a_k}{\sigma_{k-1}}[\beta_k(-\sigma_k) - 1];$$

$$\rho_k = 1 - \rho_{k1};$$

* После этого прерванный вызов приоритета k начинает обслуживаться с новой реализацией длительности обслуживания.

$$\rho_{k2} = a_1\beta_{12} + 2\left\{\left(\frac{\rho_1(\rho_1-\rho_2)}{\sigma_1} + \dots + \frac{\rho_{k-1}(\rho_{k-1}-\rho_k)}{\sigma_{k-1}}\right) - \right.$$
$$-\left(\frac{a_2(C_2+2\beta_{21})\rho_1}{\sigma_1} + \dots + \frac{a_k(C_k+2\beta_{k1})\rho_{k-1}}{\sigma_{k-1}}\right) +$$
$$+\left(\frac{a_2}{\sigma_1^2}[\beta_2(-2\sigma_1) - 2\beta_2(-\sigma_1) + 1] + \right.$$
$$\left.\left. + \dots + \frac{a_k}{\sigma_{k-1}^2}[\beta_k(-2\sigma_{k-1}) - 2\beta_k(-\sigma_{k-1}) + 1]\right)\right\};$$
$$C_i = \int_0^\infty te^{\sigma_{i-1}t}B_i(t).$$

Пусть $\beta_j(s)$ *аполитичны в правой полуплоскости, содержащей прямую* $Re\ s = -\sigma_{j-1}$, $\left(j=\overline{2,k}\right)$ и $\rho_{k1} < 1$.

Тогда

$$\sigma_k\pi_{k1} = \frac{\rho_{k1}}{\rho_k}; \tag{4.11}$$
$$\sigma_k\pi_{k2} = \frac{\rho_{k-12}}{\rho_k^3}; \tag{4.12}$$
$$h_{k1} = \frac{1}{\sigma_{k-1}\rho_{k-1}}[\beta_k(-\sigma_{k-1}) - 1]; \tag{4.13}$$
$$h_{k2} = \frac{\rho_{k-12}}{\rho_{k-1}^3}\cdot\frac{1}{\sigma_{k-1}}[\beta_k(-\sigma_{k-1}) - 1] + \tag{4.14}$$
$$+\frac{2}{\sigma_{k-1}\rho_{k-1}}\left\{\frac{1}{\sigma_{k-1}}[\beta_k(-\sigma_{k-1}) - 1] - C_k - 2\beta_{k1} + \right.$$
$$\left.+\frac{1}{\sigma_{k-1}\rho_{k-1}}[\beta_k(-2\sigma_{k-1}) - 2\beta_k(-\sigma_{k-1})] - 1\right\}.$$

3. Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 15-01-99482

Литература

Бронштейн, 1976 – Бронштейн О.И., Духовный И.М. (1976). Модели приоритетного обслуживания в информационно-вычислительных системах. М.: Наука, 220 с.

Гнеденко, 1973 – Гнеденко Б.В., Даниелян Э.А. и др. (1973). Приоритетные системы обслуживания. М.: МГУ, 279 с.

Даниелян, 2001 – Даниелян Э.А., Хачикян Х.З. (2001). Тенденции в изучении модели $M_r|G_r|1|\infty$ теории очередей // Ученые записки ЕГУ, №1, с. 1-19.

Клейнрок, 1979 - Клейнрок Л. (1979). Вычислительные системы с очередями. М.: Мир, 600 с.

Симонян, 2003 - Симонян А.Р., Симонян Э.А. (2003). Оптимальное упорядочение параметров модели Клейнрока // Обозрение прикладной и промышленной математики. Т. 10. С. 23.

Симонян, 2005 – Симонян А.Р., Улитина Е.И. (2005). О параметрических моделях массового обслуживания // Обозрение прикладной и промышленной математики. Т. 12. С. 184.

Симонян, 2010 – Симонян А.Р., Улитина Е.И. (2010). Нестационарные характеристики модели Клейнрока с нелинейной функцией приоритета // Обозрение прикладной и промышленной математики. Т. 17. № 2. С. 57.

Simonyan, 2004 – Simonyan A.R., Ulitina E.I. (2004). A Theorem on the convergence to a stable law in the M|G|1|∞ model // Russian Mathematical Surveys. T. 59. № 3. С. 589-590.

Simonyan, 2013 – Simonyan A.R., Simonyan R.A., Ulitina E.I. (2013). About nested circuits Markov in one parametric queueing model. European researcher. Series A. № 5-1 (48). С. 1119-1130.

Simonyan, 2016 – Simonyan A.R., Ulitina E.I., Simavoryan S.Zh. (2016). New submission of multidimensional limit laws in Prabhu's model // Russian Journal of Mathematical Research. Series A. № 1 (3). C. 38-42.

References

Bronshtein, 1976 – Bronshtein O.I., Dukhovnyi I.M. (1976). Modeli prioritetnogo obsluzhivaniya v informatsionno-vychislitel'nykh sistemakh. M.: Nauka, 220 s.

Gnedenko, 1973 – Gnedenko B.V., Danielyan E.A. i dr. (1973). Prioritetnye sistemy obsluzhivaniya. M.: MGU, 279 s.

Danielyan, 2001 – Danielyan E.A., Khachikyan Kh.Z. (2001). Tendentsii v izuchenii modeli Mr|Gr|1|∞ teorii ocheredei. Uchenye zapiski EGU, №1, s. 1-19.

Kleinrok, 1979 - Kleinrok L. (1979). Vychislitel'nye sistemy s ocheredyami. M.: Mir, 600 s.

Simonyan, 2003 - Simonyan A.R., Simonyan E.A. (2003). Optimal'noe uporyadochenie parametrov modeli Kleinroka. Obozrenie prikladnoi i promyshlennoi matematiki. T. 10. S. 23.

Simonyan, 2005 – Simonyan A.R., Ulitina E.I. (2005). O parametricheskikh modelyakh massovogo obsluzhivaniya. Obozrenie prikladnoi i promyshlennoi matematiki. T. 12. S. 184.

Simonyan, 2010 – Simonyan A.R., Ulitina E.I. (2010). Nestatsionarnye kharakteristiki modeli Kleinroka s nelineinoi funktsiei prioriteta. Obozrenie prikladnoi i promyshlennoi matematiki. T. 17. № 2. S. 57.

Simonyan, 2004 – Simonyan A.R., Ulitina E.I. (2004). A Theorem on the convergence to a stable law in the M|G|1|∞ model. Russian Mathematical Surveys. T. 59. № 3. S. 589-590.

Simonyan, 2013 – Simonyan A.R., Simonyan R.A., Ulitina E.I. (2013). About nested circuits Markov in one parametric queueing model. European researcher. Series A. № 5-1 (48). S. 1119-1130.

Simonyan, 2016 – Simonyan A.R., Ulitina E.I., Simavoryan S.Zh. (2016). New submission of multidimensional limit laws in Prabhu's model. Russian Journal of Mathematical Research. Series A. № 1 (3). S. 38-42.

УДК 519.21

Периоды занятости в одноканальной вычислительной системе обслуживания с пуассоновским входящим потоком и регулятором очереди

Арсен Рафикович Симонян [a,*], Елена Ивановна Улитина [a], Симон Жоржевич Симаворян [a], Надежда Андреевна Корниенко [a]

[a] Сочинский государственный университет, Российская Федерация

Аннотация. При моделировании работ информационно-вычислительных систем часто применяется математический аппарат теории очередей и их называют вычислительными системами с очередями. Основными характеристиками таких моделей являются периоды занятостей. В данной работе рассматривается модель с одним обслуживающим прибором (сервером), пуассоновским входящим потоком и регулятором очереди перед сервером. Регулятор очереди рассеивает на несколько потоков входящие в систему вызовы и ранжирует потоки по приоритетам. В описанной вычислительной системе исследуются основные характеристики методами теории вероятностей и исследования операций.

Ключевые слова: информационно-вычислительная система, вычислительные системы с очередями, теория очередей, период занятости, вероятность, потоки событий, вызовы.

* Корреспондирующий автор
Адреса электронной почты: oppm@mail.ru (А.Р. Симонян), ulitinaelena@mail.ru (Е.И. Улитина), simsim58@mail.ru (С.Ж. Симаворян), kornienko_nadja@mail.ru (Н.А.Корниенко)