ԵՐԵՎԱՆԻ ՊԵՏԱԿԱՆ ՀԱՄԱԼՍԱՐԱՆԻ ԳԻՏԱԿԱՆ ՏԵՂԵԿԱԳԻՐ
УЧЕНЫЕ ЗАПИСКИ ЕРЕВАНСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

Բնական գիտություններ 3, 1980 Естественные науки

Математика

УДК 519.217

Э. А. ДАНИЕЛЯН

ОБ ОДНОЙ СИСТЕМЕ С ДИНАМИЧЕСКИМИ ПРИОРИТЕТАМИ

В работе рассматривается система $\vec{M}_r|\vec{C}_r|1|\infty$ массового обслуживания. Вызовы разных типов в очереди с течением времени с разными скоростями линейно наращивают свой приоритет. На обслуживание берется вызов с наивысшим в момент освобождения прибора приоритетом.

Установлена связь между безусловными и условными (при условии прекращения доступа вызовов в систему с момента t) виртуальными временами ожидания в момент t, получено распределение времени ожидания для вызовов r-го потока, при $r=2$ вычислены стационарные средние времени ожидания вызовов первого и второго потоков.

§ 1. Введение

Простейшие приоритетные дисциплины относительного, абсолютного, чередующегося приоритетов возникают как оптимальные дисциплины, минимизирующие в некоторых классах дисциплин линейную стоимостную функцию [1]. Однако последняя не всегда отражает правильно реальную ситуацию, при которой часто фактическая стоимостная функция бывает неизвестна.

Поэтому нередко определяют не стоимостную функцию, а относительное время ожидания вызовов разных категорий срочности.

Перед проектировщиком системы может быть поставлена задача: при фиксированных исходных данных системы, т. е. при фиксированных распределениях входных потоков программ и длительностей их счета сконструировать дисциплину, для которой отношения средних времен ожидания программ различных категорий срочности будут равны заранее заданным числам.

Но для дисциплин относительного, абсолютного, чередующегося и т. д. приоритетов при фиксации исходных данных указанные отношения сами фиксированы.

Следовательно, при конструировании требуемой дисциплины необходимо предусмотреть некоторые дополнительные степени свободы, позволяющие проектировщику путем их варьирования достичь поставленной цели. Такие дисциплины носят название динамических дисциплин.

Одна из возможных динамических дисциплин описана и в случае системы $\vec{M}_r|\vec{M}_r|1|\infty$ изучена Л. Клейнроком [1].

Эта же дисциплина в более общем случае системы $\vec{M}_r|\vec{G}_r|1|\infty$ служит предметом нашего исследования.

В одноканальную систему с ожиданием поступают независимые пуассоновские потоки 1-вызовов, ..., r—вызовов с параметрами $a_1, \ldots, a_r$ соответственно. Длительности обслуживания вызовов независимы в совокупности, не зависят от процесса поступления и для i—вызовов имеют функцию распределения $B_i(t)$, $B_i(+0)=0$.

В момент $t=0$ в системе отсутствуют вызовы.

k-вызов, поступивший в момент τ в систему и не попавший к моменту $t>\tau$ на прибор, в момент t получает приоритет

$$q_k(t)=b_k\cdot(t-\tau) \quad (b_1>\cdots>b_r>0),$$

т. е. при $i<j$ i-вызов «наращивает» с течением времени приоритет быстрее, чем j-вызов.

В момент t окончания обслуживания на прибор выбирается вызов с максимальным в момент t приоритетом. При наличии нескольких таковых вызовов выбирается прибывший раньше.

Описанная модель обладает r степенями свободы—константы $b_1, \ldots, b_r$.

Очевидно, что вызовы одного и того же индекса обслуживаются в порядке поступления.

Обозначим через $w_k(t)$ виртуальное время ожидания k-вызова в момент t, т. е. время, которое пришлось бы ждать до начала своего обслуживания k-вызову, если бы он поступил в систему в момент t. Пусть $\overline{w}_k(t)$—условное виртуальное время ожидания k-вызова в момент t при условии прекращения с момента t доступа вызовов в систему.

В настоящей заметке устанавливается связь между $w_k(t)$ $(k=\overline{1, r})$ и $\overline{w}_k(t)$, находится $w_r(t)$ и в случае $r=2$ вычисляются стационарные средние времени ожидания 1-вызовов и 2-вызовов.

§ 2. Анализ времени ожидания

Введем обозначения $(i=\overline{1, k};\ k=\overline{1, r-1})$

$$\hat{a}_{ik}=a_i\left(1-\frac{b_{k+1}}{b_i}\right), \quad \hat{\sigma}_k=\hat{a}_{1k}+\cdots+\hat{a}_{kk},$$

$$\sigma_k=a_1+\cdots+a_k.$$

Пусть функции $\hat{\pi}_k(s)(k=\overline{1, r-1};\ \operatorname{Re}s\geqslant 0)$ есть решения функциональных уравнений

$$\hat{\sigma}_k\hat{\pi}_k(s)=\sum_{i=1}^{k}\hat{a}_{ik}\beta_i(s+\hat{\sigma}_k-\hat{\sigma}_k\hat{\pi}_k(s)), \tag{1}$$

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

Вопросы существования и единственности решений уравнений типа (1) изучены, например, в [2].

Положим ($k=\overline{1, r-1}$; $j=\overline{1, r}$; $\mathrm{Re}\, s \geqslant 0$):

$$\hat{\mu}_{k+1}(s)=s+\hat{\sigma}_k-\hat{\sigma}_k\hat{\pi}_k(s),$$

$$\beta_{jl}=\int_0^\infty t^l dB_j(t) \quad (l=1,\ 2) \quad \rho_{jl}=\sum_{m=1}^{j}\frac{1}{l!}a_m\beta_{ml},$$

$$\rho_j=1-\rho_{j1},\quad \hat{\rho}_{r-1}=1-\sum_{l=1}^{r-1}\hat{a}_{lr-1}\beta_{l1}.$$

Результаты формулируются в терминах преобразований Лапласа-Стильтьеса:

$$\bar{\omega}_k(s,\ t)=Me^{-s\bar{w}_k(t)},\quad \omega_k(s,\ t)=Me^{-sw_k(t)},$$

где M—знак математического ожидания.

Теорема. 1) Верны соотношения ($k=\overline{1, r}$; $\mathrm{Re}\, s\geqslant 0$; $t>0$):

$$\omega_k(s,\ t)=\bar{\omega}_k(\hat{\mu}_k(s),\ t).$$

2) Условные виртуальные времена ожидания r-вызова в момент t для рассматриваемой динамической дисциплины и дисциплины относительного приоритета при одинаковых исходных данных совпадают.

Прежде чем приступить к доказательству теоремы, выведем из ее утверждения одно интересное следствие.

В силу утверждения теоремы при $\rho_{r1}<1$ существует стационарное распределение времени ожидания r-вызова, определяемое своим преобразованием Лапласа-Стильтьеса

$$\omega_r(s)=\lim_{t\to+\infty}\omega_r(s,\ t)=\frac{\rho_r\hat{\mu}_r(s)}{\hat{\mu}_r(s)-\sigma_r+\sigma_r\beta_{(r)}(\hat{\mu}_r(s))}, \tag{3}$$

где

$$\beta_{(r)}(s)=\sum_{l=1}^{r}\frac{a_l}{\sigma_r}\beta_l(s),$$

а выражение для $\omega_r(s,\ t)$ (см. (2) и утверждение 2) теоремы) взято из [2].

Из (3) вычисляется первый момент ω_{r1} стационарного времени ожидания r-вызова

$$\omega_{r1}=\rho_{r2}/(\hat{\rho}_{r-1}\rho_r). \tag{4}$$

Если ω_{k1} ($k=\overline{1, r}$)—среднее стационарного времени ожидания k-вызова, то в силу «закона сохранения» Клейнрока (3)

$$\sum_{l=1}^{r}a_l\beta_{l1}\omega_{l1}=(\rho_{r1}\rho_{r2})/\rho_r$$

и из (4) получаем

$$\sum_{i=1}^{r-1} a_i\beta_{i1}\omega_{i1}=\left(\frac{\rho_r-\rho_{r-1}}{\hat{\rho}_{r-1}}+\rho_{r1}\right)\frac{\rho_{r2}}{\rho_r},$$

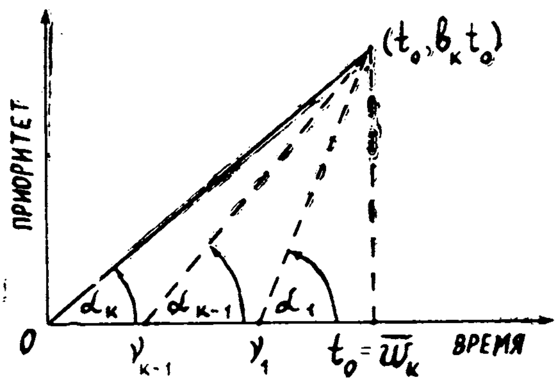
в частности, при $r=2$

$$\omega_{11}=-\frac{\rho_{22}}{a_1\beta_{11}\rho_2}\left(\frac{\rho_2-\rho_1}{\hat{\rho}_1}+\rho_{21}\right). \tag{5}$$

Таким образом, доказано

Следствие. Для рассматриваемой динамической дисциплины стационарные средние ω_{21} и ω_{11} времени ожидания в случае двух входящих потоков находятся по формулам (4), (5).

Доказательство теоремы. Установим связь между $w_k(t)$ и $\overline{w}_k(t)$. Очевидно, что k-вызовы, ..., r-вызовы, поступающие после момента t, не могут опередить в обслуживании виртуальный в момент t k-вызов, поскольку они «наращивают» свой приоритет не быстрее упомянутого виртуального k-вызова. Но некоторые $\overline{1,\ k-1}$-вызовы, поступающие после момента t, могут опередить в обслуживании виртуальный в момент t к-вызов.



Зафиксируем t и k. На рис. 1 начало координат перенесено в момент t ($t_0=\overline{w}_k(t)$). Углы $\alpha_1, \ldots, \alpha_k$ определяются равенствами

$$b_i=\mathrm{tg}\,\alpha_i\left(i=\overline{1,\ k};\ 0<\alpha_i<\frac{\pi}{2}\right).$$

Из точки $(t_0,\ b_k\cdot t_0)$ проводятся $k-1$ прямых, образующих с осью времен углы $\alpha_1, \ldots, \alpha_{k-1}$. Прямые отсекают от оси времени отрезки $[0,\ v_1^0), \ldots, [0,\ v_{k-1}^0)$, где $v_{k-1}^0<v_{k-2}^0<\cdots<v_1^0$, а величины $v_i^0(i=\overline{1,\ k-1})$ находятся из соотношений

$$b_k\cdot t_0=b_i(t_0-v_i^0),$$

или

$$v_i^0=t_0\left(1-\frac{b_k}{b_i}\right)\quad(i=\overline{1,\ k-1}). \tag{6}$$

Смысл величин $v_i^0(i=\overline{1,\ k-1})$ заключается в следующем: i-вызовы, поступающие в промежутке $[0,\ v_i^0)$, наверняка обслуживаются раньше виртуального k-вызова.

Определим процессы с независимыми приращениями (по z)

$$\tilde{\beta}_i(z)=\beta_i^1+\cdots+\beta_i^{\nu_i(z)}\quad(i=\overline{1,\ k-1}), \tag{7}$$

где $\{\beta_i^j\}_{j\geq 1}$ —последовательность независимых, одинаково распределенных случайных величин с функцией распределения $B_i(t)$, а $\nu_i(z)$ — однородный пуассоновский процесс с мгновенной интенсивностью $a_i>0$.

Фактически $\tilde{\beta}_i(v_i^0)$ представляет собой суммарное время обслуживания i-вызовов, поступающих за промежуток $[0, v_i^0)$.

Следовательно, в первом приближении виртуальное время ожидания k-вызова в момент t не меньше величины (t_0—нулевое приближение)

$$t_1=t_0+\tilde{\beta}_1(v_1^0)+\dots+\tilde{\beta}_{k-1}(v_{k-1}^0). \tag{8}$$

Укажем, как строится второе приближение. Из точки $(t_1, b_k\cdot t_1)$ проводятся k—1 прямых, образующих с осью времени углы $\alpha_1, \dots, \alpha_{k+1}$. Прямые отсекают от оси времени отрезки $[0, v_1^1), \dots, [0, v_{k-1}^1)$, где $v_{k-1}^1<v_{k-2}^1< \dots <v_1^1$, а величины $v_i^1 (i=\overline{1, k-1})$ находятся из соотношений

$$v_i^1=t_1\cdot\left(1-\frac{b_k}{b_i}\right) \quad (i=\overline{1, k-1})$$

или на основе (8)

$$v_i^1=(t_0+\tilde{\beta}_1(v_1^0)+\dots+\tilde{\beta}_{k-1}(v_{k-1}^0))\left(1-\frac{b_k}{b_i}\right). \tag{9}$$

Во втором приближении виртуальное время ожидания k-вызова в момент t не меньше величины

$$t_2=t_1+\tilde{\beta}_1(v_1^1-v_1^0)+\dots+\tilde{\beta}_{k-1}(v_{k-1}^1-v_{k-1}^0)=$$

$$=t_0+\tilde{\beta}_1(v_1^1)+\dots+\tilde{\beta}_{k-1}(v_{k-1}^1),$$

при выводе выражения для которой использованы независимость приращений процессов $\tilde{\beta}_i(z)$ $(i=\overline{1, k-1})$ и формула (9).

Продолжение приведенной рекуррентной процедуры на n-ом шаге приводит к следующему n-ому приближению:

$$\begin{cases} v_i^{n-1}=t_{n-1}\left(1-\dfrac{b_k}{b_i}\right) \quad (i=\overline{1, k-1};\ n\geq 1), \\ t_n=t_0+\tilde{\beta}_1(v_1^{n-1})+\dots+\tilde{\beta}_{k-1}(v_{k-1}^{n-1}). \end{cases} \tag{10}$$

При этом $(n>1;\ i=\overline{1, k-1})$: $0<t_0\leq t_1\leq\dots$ и, очевидно, с вероятностью единица

$$\lim_{n\to\infty} t_n=w_k(t). \tag{11}$$

Перейдем к подсчету распределения виртуального времени ожидания в момент t. Вычислим условное математическое ожидание ($\operatorname{Re} s\geq 0$):

$$M\{e^{-st_n}|t_0=u\}=e^{-su}\, M\left\{\exp\left[-s\sum_{i=1}^{k-1}\tilde{\beta}_i\left(t_{n-1}\left(1-\frac{b_k}{b_i}\right)\right)\right]\Big|t_0=u\right\}=$$

$$e^{-su}\int_{u}^{\infty}\exp\left\{-\sum_{i=1}^{k-1}\hat{a}_{ik-1}(1-\beta_i(s))v\right\}d_vP\{t_{n-1}<v/t_0=u\}=$$

$$e^{-su}M\{e^{-\hat{\tau}_{k-1}(s)t_{n-1}}/t_0=u\}=$$

$$e^{-(s+\hat{\tau}_{k-1}(s))u}M\{e^{-\hat{\tau}_{k-1}(\hat{\tau}_{k-1}(s))t_{n-2}}/t_0=u\}=$$

$$\exp\{-[s+\hat{\tau}_{k-1}(s)+\dots+\hat{\tau}_{k-1}(\dots(\hat{\tau}_{k-1}(s))\dots)]u\},$$

где обозначено

$$\hat{\tau}_{k-1}(s)=\sum_{i=1}^{k-1}\hat{a}_{ik-1}(1-\beta_i(s)). \qquad (12)$$

Следовательно, в силу формулы для условного математического ожидания и (12) имеем

$$\omega_k(s,\ t)=\bar{\omega}_k\,(s+\hat{\tau}_{k-1}(s)+\hat{\tau}_{k-1}(\hat{\tau}_{k-1}(s))+\dots,\ t). \qquad (13)$$

Чтобы упростить правую часть (13), заметим, что в обычной системе с относительным приоритетом при тех же исходных данных процедура ($t_0=\bar{w}_k(t)$; k—фиксировано)

$$t_1=t_0+\tilde{\beta}_1(t_0)+\dots+\tilde{\beta}_{k-1}(t_0),$$

$$\dots\dots\dots\dots$$

$$t_{n+1}=t_{n-1}+\tilde{\beta}_1(t_n-t_{n-1})+\dots+\tilde{\beta}_{k-1}(t_n-t_{n-1})= \qquad (14)$$

$$=t_0+\tilde{\beta}_1(t_n)+\dots+\tilde{\beta}_{k-1}(t_n)\quad(n\geqslant 1)$$

$$\dots\dots\dots\dots\dots\dots$$

приводит к безусловному виртуальному времени ожидания k-вызова в момент t:

$\lim\limits_{n\to\infty} t_n=w_k(t)$ (уже для относительного приоритета).

Отличие этой процедуры от предыдущей заключается в справедливости в данном случае формулы (13) с единственной заменой $\hat{\tau}_{k-1}(s)$ на

$$\tau_{k-1}(s)=\sum_{i=1}^{k-1}a_i(1-\beta_i(s)), \qquad (15)$$

что следует из сравнения (10) с (14). Но нам известна следующая связь в случае дисциплины относительного приоритета (см. [2]):

$$\omega_k(s,\ t)=\bar{\omega}_k(\mu_k(s),\ t),$$

т. е.

$$\mu_k(s)=s+\tau_{k-1}(s)+\tau_{k-1}(\tau_{k-1}(s))+\dots. \qquad (16)$$

Сравнивая (12) с (15), заключаем на основе (16) и (1)

$$\hat{\mu}_k(s) = s + \hat{\tau}_{k-1}(s) + \hat{\tau}_{k-1}(\hat{\tau}_{k-1}(s)) + \cdots . \tag{17}$$

Из (17) и (13) следует (2).

Поскольку все 1-вызовы, ..., r-вызовы, поступающие до момента t, «наращивают» свой приоритет не медленнее виртуального в момент t r-вызова, то последний пропускает вперед себя все вызовы, накопившиеся в системе в момент времени t. Наконец, при отсутствии прерываний накопленная (и не выполненная) к моменту t «работа» не зависит от дисциплины обслуживания, следовательно, справедливо утверждение 2) теоремы.

Кафедра теории вероятностей и математической статистики

Поступила 17.04.1980

ЛИТЕРАТУРА

1. **Клейнрок Л.**, Вычислительные системы с очередями, изд-во «Мир», М., 1979.
2. **Гнеденко Б. В. и др.**, Приоритетные системы обслуживания, изд-во МГУ, М., 1973.
3. **Бронштейн О. И., Духовный И. М.**, Модели приоритетного обслуживания в информационно-вычислительных системах, изд-во «Наука», М., 1976.

Է. Ա. ԴԱՆԻԵԼՅԱՆ

ԴԻՆԱՄԻԿ ՆԱԽԱՊԱՏՎՈՒԹՅՈՒՆՆԵՐՈՎ ՄԻ ՀԱՄԱԿԱՐԳԻ ՄԱՍԻՆ

Ամփոփում

Սպասումով միալար զանգվածային սպասարկման համակարգ են ներմտնում r *անկախ Պուասոնի հոսքեր։* i*-րդ* $(i=\overline{1,r})$ *հոսքի պահանջները կոչվում են* i*-պահանջներ։* i*-պահանջների* $(i=\overline{1,r})$ *սպասարկման ժամանակները ունեն* $B_i(t)$, $B_i(+0)=0$, *բաշխման ֆունկցիա։ Սպասարկման ժամանակները անկախ են, կախված չեն ներմտնող հոսքերից։* $t=0$ *պահին համակարգը ազատ է պահանջներից։*

i*-պահանջը* $(i=\overline{1,r})$ *ընկնելով համակարգ* u *պահին և մնալով հերթում մինչև* $t>u$ *պահը,* t *պահին ձեռք է բերում* $b(t-u)$ *նախապատվություն* $(b_1>b_2>\ldots>b_r>0)$*։*

Սպասարկման ավարտին հերթից սարքի վրա է ընտրվում մաքսիմալ նախապատվություն ունեցող պահանջը։

Դիցուք $w_k(t)$*-ն* k *պահանջի հնարավոր սպասման ժամանակն է* t *պահին, իսկ* $\overline{w}_k(t)$*-ն պայմանական հնարավոր սպասման ժամանակը այն պայմանի դեպքում, երբ* t *պահից փակվում է պահանջների մուտքը համակարգ։*

Աշխատանքում գտնված է $w_k(t)$*-ի և* $\overline{w}_k(t)$*-ի միջև եղած կապը* $(k=\overline{1,r})$*, գտնված է* $w_r(t)$*-ի բաշխման ֆունկցիան։* $r=2$ *դեպքում ստացված են 1 և 2-պահանջների ստացիոնար սպասման ժամանակների մաթեմատիկական սպասումները։*