ИНТЕГРАЛЬНАЯ ОЦЕНКА БЛИЗОСТИ НЕСТАЦИОНАРНЫХ И СТАЦИОНАРНЫХ ХАРАКТЕРИСТИК ПРОЦЕССА ОБСЛУЖИВАНИЯ

Боян Н. Димитров

1. Введение. — При исследований систем массового обслуживания обычно сталкиваемся со следующим явлением. Процесс обслуживания начинает в момент $t_0=0$ при некоторых начальных условий (например, для одноканальных систем возможно у прибора ожидает некоторое число вызовов и прибор только начинает очередное (первое) обслуживание), которые очень существенны для определения разного рода нестационарных характеристик обслуживания. Если нагрузка системы $\rho=\frac{a}{b}$ меньше единицы (здесь a есть среднее число поступающих, b — среднее число обслуженных вызовов за единицу времени) и вероятностный режим поступающего потока и дисциплина очереди не меняются со временем, то известно (см. монографию Б. В. Гнеденко и И. Н. Коваленко [2] или книгу А. Обретенова и др. [3]), что характеристики процесса облуживания в момент времени t стремятся при $t \to \infty$ к величинам, независящим от начальных условий. Эти предели принято называть стационарными характеристиками обслуживания и на практике по ним часто судят о работе системы даже на конечном интервале времени. При этом допускается ошибка, которую необходимо учитывать всякий раз, когда нестационарный режим заменяется стационарным. Один из возможных способов вычисления указанной ошибки приводится в настоящей статье.

Рассмотрим однолинейную систему массового обслуживания. Пусть $P_0(t)$ вероятность того, что в момент $t>0$ система свободна от вызовов. Это характеристика абсолютного режима работы системы. Известно [2, 3], что если нагрузка $\rho<1$, тогда при $t \to \infty$ существует предел

$$\lim_{t\to\infty} P_0(t)=p_0.$$

Степень стремления разности

$$\delta(t)=P_0(t)-p_0$$

к нулю при $t \to \infty$ характеризует скорость сходимости абсолютного состояния системы к стационарному режиму. Разность же

$$\Delta(T)=\int_0^T P_0(t)dt-p_0T \tag{1}$$

характеризует ошибку, если мы на период времени $[0, T]$ использовали характеристики стационарного режима вместо абсолютных характеристик.

Действительно, если положить $\chi_{абс}(t)=1$, когда система свободна от вызовов в момент t, и $\chi_{абс}(t)=0$ в противном случае, и образуем такую же величину $\chi_{ст}(t)$ для аналогичной системе (с теми же параметрами обслуживания), работающей в стационарном режиме, то легко понять, что $\int_0^T \chi_{аб}(t)\,dt$ и $\int_0^T \chi_{ст}(t)\,dt$ дают доли времени, которое соответствующая система провела на интервале $[0, T]$ в бездействии. Тогда $\Delta(T)$ есть не что иное, как математическое ожидание разности только что написанных двух величин, т. е.

$$\Delta(T)=E\left[\int_0^T \chi_{абс}(t)dt-\int_0^T \chi_{ст}(t)\,dt\right].$$

Точно сосчитать $\Delta(T)$ не всегда возможно, однако в многих случаях можно определять асимптотическое приближение для $\Delta(T)$ при больших значениях аргумента T. Ниже мы демонстрируем эту возможность на двух частных случаев.

2. Оценка $\Delta(T)$ для системы $M|G|1$ с абсолютно надежным прибором. — Для простоты предположим, что при $t_0=0$ система свободна от вызовов, т. е. $\lim_{t\downarrow 0} P_0(t)=1$. В книге [3] § 7. 4. для преобразования Лапласа

$$p_0(s)=\int_0^\infty e^{-st}P_0(t)\,dt$$

вероятности $P_0(t)$ получено рекуррентное соотношение

$$sp_0(s)=\frac{s}{\lambda+s}+\frac{\lambda}{\lambda+s}\,\pi(s)sp_0(s), \tag{2}$$

где λ — параметр входящего потока вызовов, а $\pi(s)$ — преобразование Лапласа—Стилтьеса функции распределения $\Pi(t)$ периода занятости прибора, определяемое из уравнения

$$\pi(s)=\beta(s+\lambda-\lambda\pi(s)). \tag{3}$$

Здесь обозначено $\beta(s)=\int\limits_0^\infty e^{-st}\,dB(t)$, где $B(t)$ — функция распределения длительности обслуживания отдельного вызова.

Нетрудно записать (2) в виде

$$p_0(s)=\frac{1}{s}-\lambda\left[\frac{1}{s}-\frac{\pi(s)}{s}\right]p_0(s). \tag{4}$$

Однако в правую часть (4) стоят

$$\frac{1}{s}=\int\limits_0^\infty 1\,.\,e^{-st}dt;\qquad \frac{\pi(s)}{s}=\int\limits_0^\infty e^{-st}\,\Pi(t)\,dt,$$

поэтому (4) эквивалентно интегральному уравнению для $P_0(t)$ —

$$P_0(t)=1-\lambda\int\limits_0^t[1-\Pi(t-x)]P_0(x)\,dx. \tag{5}$$

Уравнение (5) представляет собой уравнением Вольтерра с замкнутым циклом (см. Приложение), следовательно его решение $P_0(t)$ может быть записано в виде

$$P_0(t)=\sum_{n=1}^{\infty}(-\lambda)^n[1-\Pi(t)]^{n*}+1.$$

Ввиду того, что если $\lambda\beta_1<1$, то

$$-\lambda\int\limits_0^\infty e^{-st}[1-\Pi(t)]\,dt=-\frac{\lambda}{s}[1-\pi(s)]\neq 1\quad \text{при } \operatorname{Re} s\geqq 0,$$

мы можем применить теорему, цитированную в Приложении, которая даст

$$\lim_{t\to\infty}P_0(t)=\frac{1}{1-\lim\limits_{s\downarrow 0}\left\{\frac{\lambda}{s}[1-\pi(s)]\right\}}=\frac{1}{1-\lambda\pi'(0)}.$$

С помощью (3) подсчитываем последнеее выражение —

$$\lim_{t\to\infty}P_0(t)=1-\rho,$$

где положено

$$\rho=\lambda\beta_1, \quad \beta_1=-\beta'(0)=\int_0^\infty t\,dB(t).$$

Этот результат известен и получен другими способами в [3], гл. 7. Но теорема, приведенная в Приложении, помогает дать ответ на поставленную вначале задачу.

Теорема 1. Если второй момент времени обслуживания конечен, т. е. если

$$\beta_2=\int_0^\infty t^2\,dB(t)<\infty,$$

то

$$\lim_{T\to\infty}\left[\int_0^T P_0(t)dt-(1-\rho)T\right]=\frac{\lambda\beta_2}{2(1-\rho)}, \tag{6}$$

где $\rho=\lambda\beta_1$ предположено меньше единицы.

Доказательство. Образуем функцию

$$F(T)=\int_0^T[P_0(t)-(1-\rho)]dt-C=\Delta(T)-C. \tag{7}$$

Если определить константу C таким образом, чтобы выполнялось $F(T)\to 0$ при $T\to\infty$, то C будет предполагаемый предел разности (1). Однако $P_0(t)$ удовлетворяет уравнению (5), откуда тождественными преобразуваниями получаем

$$P_0(t)-(1-\rho)=1-(1-\rho)-\lambda\int_0^t[1-\Pi(x)].[P_0(t-x)-(1-\rho)]\,dx$$

$$-\lambda(1-\rho)\int_0^t[1-\Pi(x)]dx.$$

Интегрируем это равенство в пределах от 0 до T и получаем

$$\Delta(T)=\rho T-\lambda\int_0^T\left\{\int_0^t[1-\Pi(x)][P_0(t-x)-(1-\rho)]dx\right\}dt \tag{8}$$

$$-\lambda(1-\rho)\int_0^T\left\{\int_0^t[1-\Pi(x)]dx\right\}dt.$$

В первом интеграле правой части меняем порядок интегрирования, а затем производим замену переменной $t-x=y$ во внутренном интеграле, вследствии чего доведем его к виду

$$\int_0^T \left\{ [1-\Pi(x)] \int_0^{T-x} [P_0(y)-(1-\rho)dy \right\} dx = \int_0^T [1-\Pi(x)] \cdot \Delta(T-x)\, dx.$$

Таким образом (8) получает вид

$$\Delta(T)=\rho T-\lambda \int_0^T [1-\Pi(x)]\Delta(T-x)dx-\lambda(1-\rho) \int_0^T \left\{ \int_0^t [1-\Pi(x)]dx \right\} dt.$$

Отсюда с тождественными преобразованиями для определенного равенством (7) $F(T)$ получаем интегральное уравнение

$$F(T)=R(T)-\lambda \int_0^T F(x)[1-\Pi(T-x)]dx, \tag{9}$$

где положено

$$R(T)=\rho T-C-\lambda(1-\rho) \int_0^T \int_0^t [1-\Pi(x)]\, dx\, dt-\lambda C \int_0^T [1-\Pi(t)]\, dt. \tag{10}$$

Так как условия для применимости теоремы из Приложения сводятся к тому, чтобы выполнялось

$$\frac{\lambda}{s}[1-\pi(s)] \neq -1 \quad \text{при } \operatorname{Re} s \geqq 0,$$

которое очевидно имеет место при $\lambda\beta_1<1$, то предел $F(T)$ при $T\to\infty$ будет равняться нулю, если $\lim\limits_{T\to\infty} R(T)=0$. Последнее условие дает уравнение, откуда определяется требуемое значение константы C.

Легко видеть (с помощью (3)), что

$$\int_0^T [1-\Pi(t)]dt \xrightarrow[T\to\infty]{} \pi_1=-\pi'(0)=\frac{\beta_1}{C(1-\rho)}. \tag{11}$$

Далее можем проверить, что

$$\rho T-\lambda(1-\rho) \int_0^T \int_0^t [1-\Pi(x)]dx\, dt \tag{12}$$

$$=\rho\left\{ T-\frac{1}{\pi_1} \int_0^T \int_0^t [1-\Pi(x)]dx\, dt \right\}$$

$$=\rho\int_0^T\left\{1-\frac{1}{\pi_1}\int_0^t[1-\Pi(x)]\,dx\right\}dt$$

$$=\rho\left\{T\left[1-\frac{1}{\pi_1}\int_0^T[1-\Pi(t)]\,dt\right]+\frac{1}{\pi_1}\int_0^T t[1-\Pi(t)]\,dt\right\}.$$

Ввиду конечности второго момента π_2 (это следует из (3), так как $\beta_2<\infty$) верны соотношения

$$\lim_{T\to\infty} T\left\{1-\frac{1}{\pi_1}\int_0^T[1-\Pi(t)]\,dt\right\}=0 \tag{13}$$

и

$$\lim_{T\to\infty}\int_0^T t[1-\Pi(t)]\,dt=\frac{\pi_2}{2}=\frac{\beta_2}{2(1-\rho)^3}. \tag{14}$$

Имея в виду пределов (11)—(14) при переходе к пределу $T\to\infty$ в (9), получим

$$\lim_{T\to\infty} R(T)=C(1+\lambda\pi_1)+\rho\frac{\pi_2}{2\pi_1}=0. \tag{15}$$

Отсюда видно, что если в (7) мы возьмем значение $C=\frac{\lambda\beta_2}{2(1-\rho)}$ — решение уравнения (15), то с ростом T $F(T)$ будет стремиться к нулю. Теорема доказана.

3. Оценка $\Delta(T)$ для системы $M|G|1$ с прибором, ненадежным в свободном состоянии. Во время обслуживания прибор абсолютно надежен. Выходы прибора из строя возможны лишь тогда, когда в системе нет вызовов. Если прибор простаивал время t (из за неимения вызовов в системе), с вероятностью $E(t)$ он мог выйти из строя, после чего сразу началось бы его восстановление, имеющее разпределение $G(t)$. Во время восстановление вызовы могут поступать, но они ждут до окончания восстановления, и лишь потом начинается их обслуживание. Времена „жизни" прибора, его восстановления и длительности обслуживания отдельных вызовов независимы между собой.

Абсолютный режим работы этой системы характеризуется вероятностью $P_0(t)$, что в момент t прибор исправен и в системе вызовов нет. Если в момент $t_0=0$ прибор находится в исправности и свободен от обслуживания, то в книге [3] § 8. 3. доказано, что преобразование Лапласа вероятности $P_0(t)$ задается формулой

$$(16)\qquad p_0(s)=\frac{1}{s+\lambda}\left\{1-\varepsilon(s+\lambda)\gamma(s+\lambda-\lambda\pi(s))-\frac{\lambda}{s+\lambda}[1-\varepsilon(s+\lambda)]\pi(s)\right\}^{-1},$$

где λ — параметр входящего потока, $\varepsilon(s)$, $\gamma(s)$ — преобразования Лапласа-Стильтеса функций $E(t)$ и $G(t)$ соответственно, а $\pi(s)$ определяется из уравнения (3). Для $p_0(s)$ из (16) нетрудно получить соотношение

$$(17)\qquad p_0(s)=\frac{1}{s}-\frac{\varepsilon(s+\lambda)}{s}-\left\{\lambda\left[\frac{1}{s}-\frac{\pi(s)}{s}-\varepsilon(s+\lambda)\frac{\gamma(s+\lambda-\lambda\pi(s))}{s}+\varepsilon(s+\lambda)\frac{\pi(s)}{s}\right]-\varepsilon(s+\lambda)\gamma(s+\lambda-\lambda\pi(s))\right\}p_0(s).$$

Однако функция $\gamma(s+\lambda-\lambda\pi(s))$ есть преобразование Лапласа-Стилтьеса некоторой функции распределения, которую мы обозначим через $\tilde{G}(t)$. В этом легко убедится, применив критерий Бернштейна (см. книгу В. Феллера [4], критерий 2 на стр. 507). Теперь заметим, что в правую часть (17) записаны величины

$$\frac{1}{s}-\frac{\varepsilon(s+\lambda)}{s}=\int_0^\infty e^{-st}\left[1-\int_0^t e^{-\lambda x}\,dE(x)\right]dt;$$

$$\frac{1}{s}-\frac{\pi(s)}{s}=\int_0^\infty e^{-st}[1-\Pi(t)]\,dt;$$

$$\varepsilon(s+\lambda)\left[\frac{\pi(s)}{s}-\frac{\gamma(s+\lambda-\lambda\pi(s))}{s}\right]=\int_0^\infty e^{-st}\left\{\int_0^t e^{-\lambda x}[\Pi(t-x)-\tilde{G}(t-x)]\,dE(x)\right\}dt;$$

$$\varepsilon(s+\lambda)\gamma(s+\lambda-\lambda\pi(s))=\int_0^\infty e^{-st}\int_0^t e^{-\lambda x}\,d_t\,G(t-x)\,dE(x).$$

Поэтому (17) эквивалентно интегральному уравнению

$$P_0(t)=1-\int_0^t e^{-\lambda x}\,dE(x)-\int_0^t K(t-x)\,P_0(x)\,dx\,, \tag{18}$$

где обозначено

$$k(t)=\lambda\left\{1-\Pi(t)+\int_0^t e^{-\lambda x}[\Pi(t-x)-\widetilde{G}(t-x)]\,dE(x)\right\}$$
$$-\int_0^t e^{-\lambda x}\,d_t\,\widetilde{G}(t-x)\,dE(x)\,.$$

Если положить

$$k(s)=\int_0^\infty e^{-st}\,k(t)\,dt\,,$$

тогда легко видеть, что

$$k(s)=\lambda\left[\frac{1}{s}-\frac{\pi(s)}{s}-\varepsilon(s+\lambda)\frac{\gamma(s+\lambda)-\lambda\pi(s))-\pi(s)}{s}\right] \tag{19}$$
$$-\varepsilon(s+\lambda)\,\gamma(s+\lambda-\lambda\pi(s))\,.$$

Отсюда можно показать, что $k(s)\neq-1$ в области $\operatorname{Re} s\geqq 0$. Для этого достаточно доказать неравенство $k(s)\neq-1$ для действительных значений параметра s. В самом деле, пусть s действительное число, $s\geqq 0$. Тогда

$$\frac{1}{s}-\frac{\pi(s)}{s}-\varepsilon(s+\lambda)\frac{\gamma(s+\lambda-\lambda\pi(s))-\pi(s)}{s}\geqq\frac{1}{s}\left[1-\frac{\pi(s)}{s}\right][1-\varepsilon(s+\lambda)]\geqq 0$$

и, так как $\varepsilon(s+\lambda)\gamma(s+\lambda-\lambda\pi(s))<1$, из (19) требуемое неравенство становится очевидным.

Теперь мы сможем воспользоваться уравнением (18) как в случае надежного прибора (см. предыдущий пункт). Это тоже уравнение Вольтерра с замкнутым циклом, а условия теоремы из Приложения выполнены. Проделав все как в предыдущем пункте, мы получим, что для рассматриваемой системы

$$\lim_{t\to\infty} p_0(t)=\frac{[1-\varepsilon(\lambda)](1-\rho)}{1-\varepsilon(\lambda)+\lambda\varepsilon(\lambda)\gamma_1}\,,$$

где положено

$$\rho=\lambda\beta_1\,,\qquad \beta_1=\int_0^\infty t\,dB(t)\,,\qquad \gamma_1=\int_0^\infty t\,dG(t)\,.$$

Затем увидим, что имеет место

Теорема 2. Если вторые моменты времени обслуживания и времени восстановления конечны, т. е. если

$$\beta_2=\int_0^\infty t^2\,dB(t)<\infty \quad \text{и} \quad \gamma_2=\int_0^\infty t^2\,dG(t)<\infty\,,$$

тогда

$$\lim_{T\to\infty}\left[\int_0^T p_0(t)\,dt-\frac{[1-\varepsilon(\lambda)](1-\rho)}{1-\varepsilon(\lambda)+\lambda\varepsilon(\lambda)\gamma_1}\cdot T\right] \tag{20}$$

$$=\frac{1-\varepsilon(\lambda)}{1-\varepsilon(\lambda)+\lambda\varepsilon(\lambda)\gamma_1}\cdot\frac{\lambda\beta_2}{2(1-\rho)}+\frac{\varepsilon(\lambda)[1-\varepsilon(\lambda)][\lambda\gamma_2-(1-\rho)\gamma_1]}{[1-\varepsilon(\lambda)+\lambda\varepsilon(\lambda)\gamma_1]^2}.$$

Эти результаты можно применять для систем обслуживания с r поступающими Пуассоновскими потоками и обслуживания с абсолютным приоритетом. Как показано в [3] § 9. 4., обслуживание вызовов более высокого приоритета (по сравнению с приоритетом вызовов k-ого потока) условно можно считать выходом прибора из строя. Тогда $P_0(t)$ будет вероятностью того, что в момент t прибор доступен для вызова потока с номером k, пришедшего в этот момент. Применяя результатов настоящего пункта к таким системам, нужно учитывать модификации распределения $B(t)$ времени обслуживания, т. е. необходимо брать не $B(t)$, а распределение времени блокировки прибора вызовом потока k; вместо $E(t)$ берется распределение $1-e^{-(\lambda_1+\ldots+\lambda_{k-1})t}$; вместо $G(t)$ — функция распределения периода занятости прибора обслуживанием вызовов более высокого приоритета, чем k — $\Pi_{k-1}(t)$.

Для систем типа $M_r|G_r|1$ и дисциплина обслуживания относительный приоритет (без прерывания начатого обслуживания), вероятность $p_0(t)$ удовлетворяет соотношению (5), где на месте λ стоит $\sigma=\lambda_1+\ldots+\lambda_r$; на месте $\Pi(t)$ необходимо подставить $\Pi_r(t)$ — распределение периода занятости прибора обслуживанием всех r потоков. Соответствующий результат перефразируется очевидным образом, чего мы делать не станем. Если в такой системе прибор ненадежен в свободном состоянии, тогда справедлив результат (20), где вместо $B(t)$ необходимо взять смесь $B(t)=\sum_{i=1}^{r}\frac{\lambda_i}{\sigma}B_i(t)$, $B_i(t)$ — распределение времени обслуживания одного вызова потока i. Порядок обслуживания при дисциплин без прерывания не отражается на значение $P_0(t)$, так что полученные результаты остаются в силе также для инверсионного порядка обслуживания, для случайного выбора, для чередования приоритетов и т. п. (см. [3], гл. 9.).

Приложение

Уравнение

$$F(t)=G(t)+\int_0^t k(t-x)F(x)\,dx, \tag{1}$$

где $G(t)$ и $k(t)$ — известные функции, называется уравнением Вольтера с замкнутым циклом. Его ограниченное и измеримое решение $F(t)$ определяется однозначно и задается формулой

$$F(t)=G(t)+\int_0^t G(t-x)\,Q(x)\,dx\,,$$

где функция $Q(x)$ называется резолвентным ядром уравнения (1) и определяется выражениями

$$Q(x)=\sum_{k=1}^{\infty} K^{n_*}(x);\quad K^{n_*}(x)=\int_0^x K^{(n-1)_*}(x-u)\,K(u)\,du;\quad K_1(x)=K(x)\,.$$

Следующая теорема описывает поведение решения $F(t)$ на бесконечности и является следствием теорем XVII и XVIII из книги Н. Винера и Р. Пэли [1], стр. 92 и стр. 94.

Теорема. Если

$$\int_0^{\infty} e^{-sx}K(x)\,dx\neq 1 \qquad \text{при } \operatorname{Re} s\geq 0 \tag{2}$$

и

$$\lim_{t\to\infty} G(t)=\gamma\,,$$

то

$$\lim_{t\to\infty} F(t)=\frac{\gamma}{1-\int_0^{\infty} K(x)\,dx}\,,$$

причем если выполнено (2), тогда

$$\int_0^{\infty} K(x)\,dx=\lim_{s\downarrow 0}\int_0^{\infty} e^{-sx}K(x)\,dx\,.$$

ЛИТЕРАТУРА

1. Винер, Н., Р. Пэли: Преобразование Фурье в комплексной области. Москва, 1964.
2. Гнеденко, Б. В., И. Н. Коваленко: Введение в теорию массового обслуживания. Москва, 1966.
3. Обретенов, А., Б. Н. Димитров, Е. А. Даниелян: Масово обслужване и приоритетни системи на обслужване. София, 1973 (в печати).
4. Феллер, В.: Введение в теорию вероятностей и ее приложения, т. II, Москва, 1967.

Постъпила на 13. XI. 1972 г.

AN INTEGRAL ESTIMATE OF THE VICINITY OF NONSTATIONARY AND STATIONARY CHARACTERISTICS OF THE QUEUEING PROCESS

B. N. Dimitrov

(SUMMARY)

Let $P_0(t)$ be the probability that in time t (after starting the queueing process) the server is idle and in good repair (when the server could be with breakdowns). $P_0(t)$ characterises the nonstationary (absolutely) regime of the system. If the ratio $\rho=\frac{a}{b}<1$ (a is the average number of arriving customers, b — the average number of served ones in a unit time interval) then $\lim\limits_{t\to\infty} P_0(t)=p_0$ exists and p_0 characterises the stationary regime of the system. The difference

$$\delta(t)=P_0(t)-p_0 \qquad (t\to\infty)$$

gives the speed of convergence of the absolute queueing process to its stationary regime. The quantity

$$\Delta(T)=\int_0^T \delta(t)\,dt=\int_0^T P_0(t)\,dt-p_0 T$$

could be considered as a measure of the error in case we use the stationary characteristics instead of the absolutely ones during the time interval $[0, T]$.

In the paper the behaviour of $\Delta(T)$ when $T\to\infty$ is studied for a system $M|G|1$ in the following cases:

(*i*) of absolutely reliable server (theorem (1);

(*ii*) of server which is nonreliable if being idle (theorem 2).

The method applied make use of a Tauberian theorem for the Volterra's integral equation with closed cycle (see the Appendix).