П. 2 Комплекс моделей процессов обработки запросов в системе

Для оценки своевременности представления запрашиваемой информации рассмотрим существующие подходы к формализации технологий информационного обслуживания пользователей.

Как показывают практические исследования, весьма высокую степень адекватности при формализации процесса обработки запросов пользователей обеспечивают различные методы теории массового обслуживания. В реальности технологий обслуживания может быть очень много. Это бесприоритетное и приоритетное обслуживание одним или несколькими приборами, многофазное обслуживание, обслуживание в режиме разделения времени и т.п. Для оценки некоторых из этих технологий при различного рода ограничениях уже существуют методические разработки, в том числе в приложении к анализу вычислительных систем и сетей. Применительно к системам массового обслуживания с ожиданием термин «технология обслуживания» совпадает с термином «дисциплина обслуживания», определяющим порядок выборки очередного запроса из буфера для обработки на приборе. Под запросами понимаются не только запросы пользователей на получение выходных документов, но и задачи на пересылку файлов или ввод информации в базу данных (БД), а также некоторые технологические заявки по управлению вычислительным процессом, администрированию доступа к передающей среде в сетях ЭВМ, обеспечению безопасности информации и пр.

В общем виде процессы информационного обслуживания пользователей предлагается формализовать как процессы массового обслуживания пуассоновских потоков запросов в надежно функционирующих одно- и многолинейных системах с ожиданием и буфером бесконечного объема (см. рис.П.2.1). Надежность может быть оценена с помощью модели предыдущего раздела П.1. Предположение о пуассоновости потоков запросов на обработку в систему может быть обосновано тем, что среди потоков типа Пальма [19] пуассоновский поток ставит систему обслуживания в наиболее жесткие условия функционирования и для показателей времени ожидания заявок в очередях дает верхние оценки. Более того, потоки заявок одного типа представляют собой, как правило, сумму большого числа потоков от различных источников. Интенсивность каждого из слагаемых потоков мала по сравнению с интенсивностью суммарного потока – в такой ситуации действует предельная теорема В. Григолиониса [21], согласно которой суммарный поток будет пуассоновским. Все приведенные соображения, а также результаты статистических исследований, проводимые в ходе испытаний реальных ИС, свидетельствует о возможности использования допущения о пуассоновости потоков заявок на обслуживание.

Предположение о бесконечности числа мест для ожидания означает на практике выделение для хранения запросов, входной и выходной информации таких объемов памяти буферов и базы данных, которые при правильной эксплуатации гарантируют отсутствие информационных потерь в системе вследствие их возможного переполнения. Поскольку в последние годы прослеживается весьма устойчивая тенденция к существенному увеличению объемов оперативной, встроенной и внешней памяти и ее удешевлению в современных средствах электронно-вычислительной техники, проблемы с недостатком памяти возникают все реже, и в ближайшем будущем, по-видимому, перестанут вызывать практические затруднения. С учетом изложенного введенное предположение о бесконечности числа мест для ожидания в системе представляется вполне обоснованным.

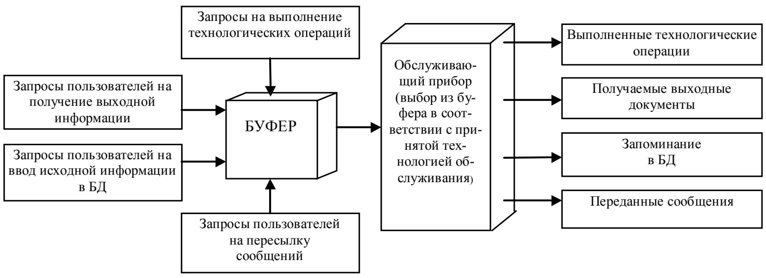

Рис. П.2.1 Формализация процессов обработки запросов

Наконец, предположение о надежности (безотказности) системы объясняется тем, что при оценке характеристик своевременности речь будет вестись лишь об условной вероятности своевременного представления информации пользователям при условии, что надежность выполнения функций ИС обеспечена.

В настоящее время существует несколько подходов к аналитической оценке своевременности обработки запросов в системах массового обслуживания. Наиболее простым является подход, позволяющий получить явное выражение стационарной ФР времени реакции системы на запрос $P_{св.i}(t)$. В этом случае, подставляя вместо t задаваемое допустимое значение $T_{зад.i}$, получаем искомую вероятность своевременного представления выходной информации за время $T_{зад.i}$. Однако, необходимо отметить, что явный вид ФР $P_{св.i}(t)$ получен лишь для наиболее простых систем без приоритетов, например, для систем $M/M/1/\infty$ [8-22].

Другой подход применим к оценке тех систем, для которых ФР времени получения ответа на запрос выражается в терминах различных преобразований Лапласа-Стилтьеса. Так, для широкого класса приоритетных систем $M/G/1/\infty$ с различными технологиями обслуживания запросов вероятностно-временные характеристики получены как раз в такой форме. В предлагаемых ниже утверждениях П.2.2-П.2.3 искомые выражения совместного распределения длины очереди и времени ожидания начала обслуживания получены именно в виде функциональной зависимости в терминах преобразований Лапласа-Стилтьеса и производящих функций. Они дают достаточное представление о математической сложности моделей. В этом случае искомая вероятность может быть рассчитана на основе обратного преобразования Лапласа-Стилтьеса. И, хотя существует несколько прикладных способов такого обратного преобразования, практические расчеты потребуют не только дополнительного программирования на высококвалифицированном уровне, но и существенных временных затрат. Такого рода дополнительные условия заметно усложняют работу системного аналитика. Поэтому на практике нередко используются подходы, обеспечивающие приближенную оценку искомой вероятности. Наиболее широко распространенный способ приближенной оценки состоит в аппроксимации ФР времени реакции системы на запрос с помощью неполной гамма-функции. Исследования Дж. Мартина [13] на ряде приоритетных технологий обслуживания показали достаточно высокую инженерную точность такой аппроксимации. Именно этот подход и реализован в инструментарии КОК.

Определение 1. Запрашиваемая информация i–го типа считается представленной своевременно в соответствии с критерием среднего времени обработки, если $T_{полн.i1} \le T_{зад.i.}$ (критерий 1).

Определение 2. Запрашиваемая информация i–го типа считается представленной своевременно в соответствии с вероятностным критерием, если $P_{св.i}=P(t_{полн.i} \le T_{зад.i}) \ge P_{зад.i}$, где $t_{полн.i}$ — реальное время реакции системы на запросы i–го типа (критерий 2).

Суть формализации заключается в моделировании процессов обработки с помощью бесприоритетных и приоритетных систем M/G/1/∞ [9-19 и др.] — см. рис. П. 2.1.

Вероятность своевременной обработки для всех технологий аппроксимируется с помощью неполной гамма-функции:

$$P_{св.i} = P\left(t_{полн.i} \le T_{зад.i}\right) = \frac{\int\limits_0^{\gamma_i^2 T_{зад.i}/T_{полн.i}} t^{\gamma_i - 1} e^{-t}\,dt}{\int\limits_0^{\infty} t^{\gamma_i - 1} e^{-t}\,dt}, \text{ где } \gamma_i = \frac{T_{полн.i}}{\sqrt{T_{полн.i2} - T_{полн.i}^2}}.$$

$$S = \frac{\sum\limits_{i=1}^{I} \lambda_i P_{св.i}}{\sum\limits_{i=1}^{I} \lambda_i}, C = \frac{\sum\limits_{i=1}^{I} \lambda_i P_{св.i}\left[Ind(\boldsymbol{\alpha}_1) + Ind(\boldsymbol{\alpha}_2)\right]}{\sum\limits_{i=1}^{I} \lambda_i}, \; Ind(\boldsymbol{\alpha}) = \begin{cases} 0, & \text{если } \boldsymbol{\alpha} = истина, \\ 1, & \text{если } \boldsymbol{\alpha} = ложь \end{cases},$$

α_1 = (для i-го типа используется критерий 1 и $T_{полн.i} \le T_{зад.i}$);
α_2 = (для i-го типа используется критерий 2 и $P_{св.i} \ge P_{зад.i}$).

Обозначения исходных данных λ_i, β_{i1}, β_{i2}, β_{i3}, ρ и описания технологий обработки запросов см. в разделе 2.2. В инструментально-моделирующих комплексах, описанных в монографии, исследованы следующие технологии обработки.

Технология 1, режим R=«однозадачный» (БПО) позволяет моделировать такую технологию обработки, когда все запросы имеют одинаковый приоритет и обрабатываются последовательно в порядке поступления.

Технология 1, режим R=«многозадачный» (БПМ) состоит в следующем: если на обработке находится n запросов, то все они обслуживаются одновременно в режиме разделения времени, но каждый в n раз медленнее, чем если бы он обслуживался один в свободной системе.

Технология 2 (ОП). Запросы более высокого приоритета имеют преимущество перед запросами низшего приоритета, а именно: среди запросов, ожидающих начала обработки, запросы более высокого приоритета обрабатываются впереди запросов низшего приоритета. Из запросов одного приоритета следующим выбирается на обработку запрос, который поступил в систему раньше. Если во время обработки некоторого запроса поступает запрос более высокого приоритета, то прерывания не происходит.

Технология 3 (АП). В отличии от технологии 2 запросы более высокого приоритета абсолютно прерывают обработку запроса с низким приоритетом. Прерванный запрос дообслуживается с прерванного места.

Технология 4 (Пак.). Первый поступивший в свободную систему запрос образует первый пакет. Очередной пакет формируется из запросов, поступивших в систему за время обработки предыдущего пакета, а если таковых нет, то из следующего же запроса, поступившего в свободную систему. Очередной пакет начинает обрабатываться сразу же после обработки всех заявок предыдущего пакета. В пакете, поступившем на обработку, выбираются запросы наивысшего приоритета и обрабатываются последовательно в порядке их поступления в систему. Пакет запросов обрабатывается без прерываний независимо от других поступающих в систему запросов.

Технология 5 (Ком.) обобщает варианты обработки запросов, описанные в технологиях 2–4. Все поступающие запросы разбиваются на n групп (рис. 2.3.4). Запросы e-й группы имеют более высокий приоритет над запросами g-й группы при $e<g$ ($e, g=1, n$) Внутри каждой группы приоритеты запросов относительные. Система обладает возможностью обработки запросов в соответствии с различными базовыми технологиями. В качестве базовых выступают вышеописанные технологии 2, 3 и 4. Для обработки запросов g-й группы назначается какая-либо одна из технологий (2 или 4). Между запросами e-й и g-й групп назначаются либо относительные, либо абсолютные приоритеты.

Для технологий 1 (R=<однозадачный режим>), 2 и 3 использованы классические результаты теории массового обслуживания [9-19, 97-98].

Для технологии 4 в отличие от технологии 5 (см.рис.2.2.3) имеем всего одну группу, пусть в ней n типов запросов. Описание технологии 4 приведено в разделе 3. Докажем для технологии 4 одно важное Утверждение 2.1.

Прежде, чем его сформулировать, примем следующие обозначения: все векторы являются n-мерными; если опущены пределы суммирования, то оно производится от 1 до n. Положим $\boldsymbol{yz}=(y_1z_1, \ldots, y_nz_n)$, $\boldsymbol{y}^{\boldsymbol{z}}=y_1^{z1} \ldots y_n^{zn}$, $\boldsymbol{B}^{*\boldsymbol{i}}= B_1^{*i1} \times \ldots \times B_n^{*in}$, где B_n^{*in} – i_k-я свертка ФР B_k $\left(B_k^{*0} = \begin{cases} 0, t \le 0, \\ 1, t > 0 \end{cases}\right)$, $\boldsymbol{0}=(0,\ldots,\ 0)$, $\sum\limits_{k=1}^{0} y_k = 0.$ Неравенство $\boldsymbol{y} \ge \boldsymbol{z}$ означает соответствующее неравенство между всеми компонентами, $\beta_k(s) = \int\limits_0^{\infty} \exp(-s\,t)\partial B_k(t)$ — преобразование Лапласа — Стилтьеса ФР B_k, а β_{kj} – ее j-й момент, $\lambda_n = \sum\limits_{k=1}^{n} \lambda_k$.

Обозначим через $\overrightarrow{\boldsymbol{\zeta}}(\boldsymbol{t})$ – вектор, характеризующий размер последнего пакета в технологии 4, поступившего до t, т.е. $\zeta_k(t)$ – число запросов k-го типа в нем $\left(k = \overline{1,n}\right)$, t_m – момент начала обслуживания m-го пакета, $\vec{\boldsymbol{\zeta}}_{\boldsymbol{m}} = \vec{\boldsymbol{\zeta}}\left(t_m + 0\right)$. Последовательность $\{\vec{\boldsymbol{\zeta}}_{\boldsymbol{m}}\}$ образует цепь Маркова. Пусть $\{\pi_i\}$ – ее стационарное распределение (если оно существует), $\boldsymbol{\pi(z)} = \sum\limits_{i \ge 0} \boldsymbol{\pi}_{\boldsymbol{i}} \boldsymbol{z}^{\boldsymbol{i}}$, $W_l(t)$ – виртуальное время ожидания начала обработки в системе запроса приоритета l, поступившего в момент t, W_{lj} – j-й момент виртуального времени ожидания начала обработки этого запроса.

Утверждение П.2.1 [130-132]. При условии $\rho = \sum\limits_{l} \lambda_l \beta_{l1} < 1$ существует стационарное распределение $\{\boldsymbol{\pi_i}\}$ цепи $\{\vec{\boldsymbol{\zeta}}_{\boldsymbol{m}}\}$ и предел $\lim\limits_{t \to \infty} P\{\, W_l(t) < w,\ \ \boldsymbol{\zeta}(\boldsymbol{t}) = \boldsymbol{i}\,\} = Q_l(w,\boldsymbol{i})$, причем

$$q_l(s,\boldsymbol{z}) = \sum_{i \ge 0} \boldsymbol{z}^{\boldsymbol{i}} \int\limits_0^{\infty} exp(-s\,w)\partial Q_l(w,\boldsymbol{i}) =$$

$$= [\,\boldsymbol{\pi}'_{z'}(1)]^{-1} \lambda_1 \{[\,s - \lambda_l + \lambda_l \beta_l(s)]^{-1} [\boldsymbol{\pi}(\boldsymbol{z\beta}(\sum_{k=1}^{l}(\lambda_k - \lambda_k \beta_k(s)))\ -$$

$$- \boldsymbol{\pi}(\,\boldsymbol{z}\,\boldsymbol{\beta}(\,s + \sum_{k=1}^{l-1}(\,\lambda_k - \lambda_k \beta_k(s))))\,] + \lambda^{-1} \boldsymbol{\pi}(\,\boldsymbol{z}\,\boldsymbol{\beta}(\lambda))\},$$

где производящая функция $\pi(z)$ удовлетворяет функциональному уравнению

$$\boldsymbol{\pi}(z)=\boldsymbol{\pi}(\boldsymbol{\beta}(\lambda-\sum_k \lambda_k z_k)\text{-}(1\text{-}\lambda^{-1}\sum_k \lambda_k z_k)\boldsymbol{\pi}(\boldsymbol{\beta}(\lambda)). \quad (\text{П.2.1})$$

Доказательство

1. Существование стационарного распределения. Пусть t_m – момент поступления на обслуживание m-го пакета. Положим $t_0 = 0$, $u_m = t_m - t_{m-1}$, $m\geq 1$. Определим при $0\leq t<u_m$ случайную функцию $\xi_m(t)$ равенством $\xi_m(t)=W_l(t_{m-1}+t)$. Величина $W_l(t)$ образует регенерирующий процесс с зависимыми циклами регенерации марковского типа [15, 16, 20]. Для него

$$P\{\vec{\zeta}_{m+1}=j,\ u_{m+1}>\tau,\ \xi_{m+1}(\tau)<w|\vec{\zeta}_m=i,\ u_m=x\}=P\{\vec{\zeta}_{m+1}=\boldsymbol{j},\ u_{m+1}>\tau,\ \xi_{m+1}(\tau)<w|\vec{\zeta}_m=\boldsymbol{i}\}.$$

С последовательностью $\{(\vec{\zeta}_m,\ \boldsymbol{u_m})\}$ связаны функции

$$F_{ij}(\tau)=P\{u_{m+1}<\tau,\ \ \vec{\zeta}_{m+1}=j|\vec{\zeta}_m=\boldsymbol{i}\},\quad p_{ij}=F_{ij}(+\infty),$$

$$A_{ij}(\tau)=F_{ij}(\tau)/p_{ij},\quad p_{ij}\neq 0.$$

Кроме того, $\{\vec{\zeta}_m\}$ образует цепь Маркова и справедливы следующие свойства.

1. Цепь Маркова $\{\vec{\zeta}_m\}$ неприводимая, непериодическая и при $\rho<1$ все ее состояния эргодические [12]. Эргодичность проверяется с помощью критерия Мустафы [9, 10].

2. $A_{ij}(+0)=0$, так как $B_k(+0), k=\overline{1,n}.$

3. Распределение A_{ij} нерешетчатое [15].

4. Среднее время пребывания в состоянии $\boldsymbol{i}$

$$E(u_{m+1}|\vec{\zeta}_m=\boldsymbol{i})=\sum_k i_k\beta_{k1}+\lambda^{-1}[\boldsymbol{\beta}(\lambda)]^{\boldsymbol{i}}<\infty.$$

5. Функция $\mu_i(t, w) = P\{u_{m+1}>t,\ \xi_{m+1}(t)\ |\ \vec{\zeta}_m = \boldsymbol{i}\}$, интегрируема по t по Риману на любом конечном промежутке.

Условия 1 – 5 и условие Утверждения П.2.1 являются достаточными [15] для существования предела

$$\lim_{t\to\infty} P\{W_l(t)<w,\ \vec{\zeta}(\mathrm{t})=\mathrm{i}\}=Q_l(w,\mathrm{i})=r_i\int_0^\infty \mu_i(t,w)\,dt,$$

$$r_i=\pi_i(\sum_{j\geq 0}\pi_j E(u_{m+1}|\vec{\zeta}_m=\boldsymbol{j}))^{-1}.$$

2. Получение основных формул. Найдем $q_l(s,z)$. Используя предыдущее выражение, получаем

$$q_l(s,\mathrm{z})=\sum_{i\geq 0}\mathrm{z}^{\mathrm{i}} r_i\int_0^\infty E_i\{\exp[-s\xi_{m+1}(t)]\}\,dt.$$

Здесь $E_i(\delta)=E\{\delta I(u_{m+1}>t|\vec{\zeta}_m=\boldsymbol{i})\}$ – условное математическое ожидание, $I(A)$ – индикатор события A. Обратимся к методу введения дополнительного события. Считаем, что независимо от функционирования нашей системы наступают катастрофы, поток которых — пуассоновский с параметром $s>0$. Тогда $E_i=\{exp[-s\xi_{m+1}(t)]\}$ – вероятность того, что $u_{m+1}>t$ и за время ожидания в системе запроса l-го приоритета, поступившего в момент t_m+t не произойдет катастрофы при условии, что m-й пакет имеет размер $\boldsymbol{i}$. Если m-й пакет обслуживался меньше t и до момента t не поступало ни одного запроса (вероятность чего $\boldsymbol{B}^{*i}(t)exp(-\lambda t)$), то этот запрос образует очередной пакет и сразу поступает на обслуживание. Если он обслуживался время $\tau>t$ и за это время поступило $r_k\geq 0$ запросов приоритета k, выше $l\left(\frac{(\lambda_k\tau)^{r_k}}{r_k}exp(-\lambda_k\tau)\right)$, а до момента t поступило $r_l\geq 0$ запросов приоритета $l\left(\frac{(\lambda_l t)^{r_l}}{r_l}exp(-\lambda_l\tau)\right)$, то катастрофы не должно произойти в интервале $[t_m+t,\ t_m+\tau]$ (вероятность чего равна $exp(-s(\tau-t))$), а также за время обслуживания запросов приоритета выше l (вероятность этого $\beta_1^{r_1}(s)\dots\beta_{l-1}^{r_{l-1}}(s)$) и за время обслуживания запросов приоритета l, поступивших до момента t (вероятность $\beta_l^{r_l}(s)$). Таким образом,

$$E_i\{\exp[-s\xi_{m+1}(t)]\}=\boldsymbol{B}^{*i}(t)\exp(-\lambda t)+\int_t^\infty d\,\boldsymbol{B}^{*i}(\tau)\times$$

$$\times\exp[-s(\tau-t)]\sum_{\substack{r_k\geq 0\\ k=\overline{1,l}}} P_r^{(l-1)}(\tau)\frac{(\lambda_l t)^{r_l}}{r_l!}\exp(-\lambda_l t)\beta_1^{r_1}(s)\dots\beta_l^{r_l}(s)=$$

$$=\boldsymbol{B}^{*i}(t)\exp(-\lambda t)+\int_t^\infty d\,\boldsymbol{B}^{*i}(\tau)\exp[-s(\tau-t)-\tau\sum_{k=1}^{l-1}(\lambda_k(1-\beta_k(s)))-\lambda_l t(1-\beta_l(s))],$$

где $P_r^{(l-1)}(\tau)=\prod_{k=1}^{l-1}\frac{(\lambda_k\tau)^{r_k}}{r_k!}\exp(-\lambda_k\tau).$

Осталось правую часть проинтегрировать по t и, умножив на $d_i z^i$, просуммировать по i. *Доказательство утверждения П.2.1 завершено.*

При этом функция $q_l(s,z)$ дает возможность вычислить основные расчетные характеристики технологии 4:

$$T_{ож.l} = -\frac{\partial\ q_l(s,1)}{\partial s}\bigg|_{s=0},\ T_{ож.l\,2} = \frac{\partial^2 q_l(s,1)}{\partial s^2}\bigg|_{s=0}. \qquad (П.2.2)$$

Полученные выражения реализованы в инструментариях, описанных в монографии.

Далее перейдем к получению ФР и первых двух моментов времен ожидания в очереди и пребывания запросов в системе (в терминах преобразования Лапласа-Стилтьеса) для технологии 5. Уже имеем n групп, K_n приоритетов запросов (описание технологии – см. раздел 2). Справедливо следующее Утверждение П.2.2.

Утверждение П.2.2 [131-133]. Для случаев, когда параметры технологии 5 принимают значения из таблицы П.2.1, при условии $\rho_{1K_n} = \sum_{l=1}^{K_n} \lambda_l \beta_{l1} < 1$ для преобразований Лапласа-Стилтьеса времени ожидания в очереди запросов r-го приоритета j-й группы $W_r^{(j)}(s)$ и полного времени пребывания на обработке $V_r^{(j)}(s)$ получаем:

$$W_r^{(j)}(s) = \begin{cases} [\,s - \lambda_r + \lambda_r \beta_r(\,\mu_r(s)\,)\,]^{-1}[\,(1-\rho_{1K_n})\,\mu_r(s) + \\ + \sum\limits_{l=r+1}^{K_n} \lambda_l I_{rl}^{(j)}(1-\beta_l(\,\mu_l(s)\,)\,)\,] \text{ для случая } \mathrm{D}(\mathrm{j}) = \text{технология } 2, \\ \{\,\rho_{1K_{j-1}}(1-\sum\limits_{l=K_{j-1}+1}^{K_j} \lambda_l f_{l1})\,/\,[\,\pi_0(1+\rho_{1K_{j-1}} h_{K_{j-1}1})\,]\,\}\{\,[\,s-\lambda_r(1-f_r(s))\,]^{-1} \times \\ \times[\,\pi_0[\,h_{K_{j-1}}(\,g_r(s))-h_{K_{j-1}}(\,s+g_{r-1}(s))\,] + \\ +\pi(\,f_{K_{j-1}}(\,g_r(s)),\dots,f_{K_j}(\,g_r(s))\,)- \\ -\pi(\,f_{K_{j-1}+1}(\,s+g_{r-1}(s)),\dots,f_{K_j}(\,s+g_{r-1}(s))\,)\,]+ \\ +\Lambda_{1K_j}^{-1}[\,\pi(\,f_{K_{j-1}+1}(\,\Lambda_{K_{j-1}+1,K_j}),\dots,f_{K_j}(\,\Lambda_{K_{j-1}+1,K_j}\,))\,] \\ +\pi_0\,[\,h_{K_{j-1}}(\,\Lambda_{K_{j-1}+1,K_j})-1\,]\,\} \text{ для случая } \mathrm{D}(\mathrm{j}) = \text{технология } 4, \end{cases} \qquad (П.2.3)$$

$$V_r^{(j)}(s) = W_r^{(j)}(s)\beta_r(\,s+\Lambda_{1K_{j-1}}-\Lambda_{1K_{j-1}}h_{1K_{j-1}}(s)) \qquad (П.2.4)$$

для случаев D(j) = технология 2 или 4,

где $\beta_r(s)$ – преобразование Лапласа-Стилтьеса ФР времени обслуживания в системе заявки r-го приоритета,

β_{re} – e-й момент ФР $B_r(\,\beta_{re} = \int_0^\infty t^e\, d\,B_r(t)\,),\quad e=1,2,3;$

$$\Lambda_{mi} = \sum_{l=m}^{i} \lambda_l,\ h_i(s) = \sum_{l=1}^{i}\lambda_l \beta_l\left(s+\Lambda_{1i}-\Lambda_{1i}h_i(s)\right)\Big/\Lambda_{1i},\quad m,i=\overline{1,K_n},$$

$$\mu_r(s) = s+\Lambda_{1_{r-1}}-\Lambda_{1_{r-1}}h_{1_{r-1}}(s);$$

$$f_i(s) = (\,s+\Lambda_{1K_{j-1}}-\Lambda_{1K_{j-1}}h_{K_{j-1}}(s)),\ g_i(s) = \sum_{l=K_{j-1}+1}^{r}\lambda_l\,(\,1-f_l(s)),\ K_{j-1}+1 \le m \le K_j\,;$$

$$\begin{aligned} &\pi(\,z_{K_{j-1}+1},\dots,z_{K_j}) = \pi(\,f_{K_{j-1}+1}(\sum_{l=K_{j-1}+1}^{K_j}(\lambda_l-\lambda_l z_l)),\dots,f_{K_j}(\sum_{l=K_{j-1}+1}^{K_j}(\lambda_l-\lambda_l z_l)))- \\ &-\left[1-\left(\Lambda_{1K_{j-1}}+\sum_{l=K_{j-1}+1}^{K_j}\lambda_l z_l\right)\Big/\Lambda_{1K_j}\right]\times\pi\left(f_{K_{j-1}+1}(\,\Lambda_{K_{j-1}+1K_j}),\dots,f_{K_j}(\,\Lambda_{K_{j-1}+1K_j})\right)\times \\ &\times\pi_0\left\{h_{K_{j-1}}\left(\sum_{l=K_{j-1}+1}^{K_j}(\lambda_l-\lambda_l z_l)\right),\dots,\ h_{K_j}(\,\Lambda_{K_{j-1}+1K_j})\times\left[1-\left(\Lambda_{1K_{j-1}}+\sum_{l=K_{j-1}+1}^{K_j}\lambda_l z_l\right)\Big/\Lambda_{1K_j}\right]-\right. \\ &\left.-\left(\Lambda_{1K_{j-1}}+\sum_{l=K_{j-1}+1}^{K_j}\lambda_l z_l\right)\Big/\Lambda_{1K_j}\right\},\quad \pi_0 = \pi(\,z_{K_{j-1}+1},\dots,z_{K_j})\Big|_{z_{K_{j-1}+1}=0,\dots,z_{K_j}=0}; \\ &\hat{\Lambda}_{1K_{j-1}} = \sum_{l=1}^{j-1}(1-y_{1j})\sum_{l=K_{j-1}+1}^{K_j}\lambda_i\,;\ \hat{h}_{K_{j-1}}(s) = \sum_{l=1}^{j-1}(1-y_{1j})\sum_{l=K_{j-1}+1}^{K_j}\lambda_i\beta_i\left(s+\hat{\Lambda}_{1K_{j-1}}\ \hat{h}_{K_{j-1}}(s)\right)\Big/\hat{\Lambda}_{1K_{j-1}}. \end{aligned} \qquad (П.2.5)$$

Таблица П.2.1.

Варианты параметров технологии 5, при которых справедливы предложенные для $W_r^{(j)}(s)$, $V_r^{(j)}(s)$ формулы (y_{1j} – это значение, задаваемое параметром *Abs* в инструментариях)

Варианты базовых дисциплин для j-й группы	$y_{1j},\ l < j$		$y_{jm},\ j < m$	
	$D(l_j) = T2$	$D(l_j) = T4$	$D(m_j) = T2$	$D(m_j)=T4$
$D(j_j)$ = *Технология 2 (Т2)*	0; 1	0	0	0; 1
$D(j_j)$ = *Технология 4 (Т4)*	0; 1	0	0	0

Доказательство утверждения П.2.2 полностью аналогично доказательству Утверждения П.2.1.

Следствие из Утверждений П.2.1 и П.2.2 [36]. Для технологии 5 при условии $\rho_{1Kn}<1$ справедливы следующие оценки для вероятностно-временных характеристик:

$$T_{полн.r} = V_{r1}^{(g)} = W_{r1}^{(g)} + \beta_{r1}\left(1-R_{K_{g-1}}\right)^{-1}, \qquad (П.2.6)$$

$$T_{полн.r2} = V_{r2}^{(g)} = W_{r2}^{(g)} + 2V_{r1}^{(g)}\beta_{r1}\left(1 - R_{K_{g-1}}\right)^{-1} + \beta_{r2}\left(1 - R_{K_{g-1}}\right)^{-2} + \beta_{r1}\left(1 - R_{K_{g-1}}\right)^{-3} R_{K_{g-1}}^{(2)}, \quad (П.2.7)$$

где $\sum_{e=m}^{u} = 0,$ при $u > m,\ u,m = \overline{1,K_n}$, $\rho_{mu}^{(e)} = \sum_{d=m}^{\min(r,n)} \lambda_d \beta_{de} + \sum_{d=r+1}^{u} \lambda_d \beta_{de} I_{rd}^{(g)},$

$$I_{rd}^{(g)} = \begin{cases} 1, \text{ при } K_{g-1} + 1 \le m \le K_g, i = \overline{g+1,n} \\ y_{de}, \text{ при } K_g < K_{i-1} + 1 \le m \le K_1 \end{cases}$$

$$R_{K_{g-1}}^{(e)} = \sum_{d=1}^{g-1}(1 - y_{dg}) \sum_{m=K_{g-1}+1}^{K_g} \lambda_m \beta_{me}, R_{K_{g-1}} = R_{K_{g-1}}^{(1)}, \ \tilde{\rho}_{mu}^{(e)} = 1 - \rho_{mu}^{(e)}, \rho_{mu} = \rho_{mu}^{(1)},$$

$$E_r = \rho_{K_{g-1}+1r}^{(2)} \tilde{\rho}_{1K_{g-1}} + \rho_{K_{g-1}+1r}\, \rho_{1K_{g-1}}^{(2)}, \ C_r = \left(\tilde{\rho}_{1K_{g-1}}^{3} - \rho_{K_{g-1}+1K_g}^{3}\right)^{-1}\left[\left(\rho_{1K_{g-1}} + \rho_{K_{g-1}+1r-1}\right)^{3} - \rho_{K_{g-1}+1r}^{3}\right]$$

Доказательство. Выражения для $T_{ож.r}$, $T_{ож.r2}$, $T_{полн.r}$, $T_{полн.r2}$ получаются путем соответственно одно- и двукратного дифференцирования преобразований Лапласа-Стилтьеса $W_r^{(j)}(s)$ и $V_r^{(j)}(s)$ (т.е. выражений (П.2.3) и (П.2.4)) в точке $s=0$ аналогично (П.2.2). Математические детали дифференцирования опущены. Полученные выражения реализованы в инструментариях, описанных в монографии.

Доказательство следствия завершено.

Для технологии (R=<многозадачный режим>) результаты работы [129] для $i=1, 2$ обобщены на случай $i\ge 1$:

$T_{полн.i1} = \beta_{i1}/(1-\rho)$, $\quad T_{полн.i2} = \int_0^\infty \frac{2t^2}{(1-\rho)^2} d B_i(t) - \frac{2}{1-\rho}\int_0^\infty d B_i(t) \int_0^t \delta(u) du,$ (П.2.10)

где $B_i(t)$ — ФР времени обработки запросов i-го типа в свободной системе,

$$B_i(t) = 1 - \exp\left(-t/\beta_{i1}\right), \ \ B(t) = \sum_{i=1}^{I} \lambda_i B_i(t) \Big/ \sum_{i=1}^{I} \lambda_i; \ \ \delta(u) = u + \int_0^u (u-z) \sum_{n=1}^{\infty} \rho^n d F^{*n}(z),$$

$F(z) = \dfrac{\int_0^z [1 - B(t)]dt}{\int_0^\infty t dB(t)}$, $\quad F^{*n}(z)$ — n-я свертка ФР $F(z)$,

Доказательство см. в работе [129]. Используемое в инструментариях выражение $T_{полн.i2}$ получено путем обычного интегрирования выражения (П.2.10).

В итоге для моделирования получаем все необходимые выражения, в т.ч. для расчета $P_{св.r}(T_{зад.})$.

Таким образом, на этом математические доказательства для «Комплекса моделей процессов обработки запросов в системе» можно считать полностью завершенным.

Необходимые для моделирования пределы исходных значений λ_i, $T_{зад\ i}$, $P_{св\ зад\ i}$ задают в ТЗ или в постановках функциональных задач, технологию обработки и распределение запросов по приоритетам – в конструкторской документации, а значения среднего времени обработки запроса i-го приоритета в свободной системе β_{i1} (а при необходимости более точных оценок – еще 2-й β_{i2} и 3-й β_{i3} моменты) устанавливают в результате натурных испытаний, экспериментов, дополнительного моделирования или сравнения с аналогами.