ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 32 1996 Вып. 4

УДК 621.391.1:621.394.74–503.5

 Б. С. Цыбаков, П. Папантони-Казакос

НАИЛУЧШАЯ И НАИХУДШАЯ ДИСЦИПЛИНЫ ПЕРЕДАЧИ ПАКЕТОВ[1]

Рассматривается задача оптимизации порядка передачи и сброса пакетов в коммуникационных системах с очередями. Эта задача формулируется в терминах системы массового обслуживания (СМО), компонентами которой являются буфер конечной емкости, обслуживающий прибор с постоянным временем обслуживания, и дисциплина обслуживания. Показано, что при любой заданной выборочной функции входного трафика дисциплина LIFO является наилучшей в смысле задержки, а дисциплина FIFO – наихудшей. Проводится сравнение основных характеристик дисциплин LIFO, FIFO и случайной дисциплины и дисциплины d^* с вышибанием требований. Доказательство оптимальности приводится для произвольного входного трафика, функции распределения задержки для дисциплин LIFO, FIFO, случайной дисциплины и d^* найдены для стационарного трафика без памяти, а численное сравнение дисциплин проведено для пуассоновского трафика. Рассматриваются лишь системы с дискретным временем.

§ 1. Введение

Сети с асинхронной передачей (ATM) – это магистрали будущих телекоммуникационных структур, в которых информация передается в форме пакетов (ячеек): До передачи пакеты обычно хранятся в буферах, имеющих ограниченные емкости. Главными характеристиками при этом являются частота отказов передачи пакетов и их задержка. Разработка дисциплин передачи, имеющих наилучшие характеристики, является ключевым элементом при построении высокоэффективных сетей ATM [1–7]. Задача сводится к выбору дисциплины обслуживания в системе массового обслуживания (СМО), имеющей один обслуживающий прибор с постоянным временем обслуживания и буфер с конечной емкостью.

В работе рассматривается СМО, состоящая из конечного буфера с k местами для ожидания и одного обслуживающего прибора с постоянным временем обслуживания, равным единице. В такой системе чаще всего рассматривается дисциплина обслуживания FIFO, в которой вновь прибывающее требование обслуживается по прибытию, если буфер не содержит других требований; помещается в хвост очереди в буфере, если буфер не полный и не пустой; и теряется, если буфер заполнен. Дисциплина FIFO обозначается через d_F.

Сначала в этой работе описывается дисциплина обслуживания с вышибанием из очереди, обозначенная d^* [8, 9]. Она характеризуется задержкой, которая совпадает с задержкой при d_F в периоды, когда потери требований не возникают, и имеет задержку значительно лучшую, чем при d_F, в периоды, когда возникают потери и сильные переполнения буфера. В то же время, дисциплины d_F и d^* дают одинаковые интенсивности потерь. При дисциплине d_F, когда интенсивность входного

[1]Работа выполнена при финансовой поддержке научно-технических комитетов Канады NSERC и OCRI, Миннауки (проект "Доступ") и Нидерландского научного фонда (грант NWO 713-229).

трафика λ велика ($\lambda \to \infty$), задержка обслуженных требований, равная времени, проведенному в буфере до начала обслуживания, стремится к k. В отличие от этого задержка при дисциплине d^* стремится к нулю с ростом λ. Замечательно то, что при дисциплине d^* задержка обслуженных требований улучшается с увеличением интенсивности переполняющего систему входного трафика, так как d^* перекладывает часть системной задержки на теряемые требования, освобождая от нее требования, которые обслуживаются. Мы покажем также, что все моменты распределения задержки для дисциплины d_F не меньше, чем аналогичные моменты для дисциплины d^*. Например, при буфере объема 5 и пуассоновском входном трафике интенсивности λ дисперсия задержки при дисциплине d^* относительно мала (меньше, чем 3) в диапазоне $0 < \lambda < 1,3$. Она меньше, чем дисперсия при дисциплине FIFO, и намного меньше, чем дисперсия при дисциплине LIFO. Это важно для сетей ATM, так как они чувствительны к разбросам задержек.

Основной задачей, рассматриваемой в этой статье, является задача оптимизации – найти оптимальную по задержке дисциплину обслуживания в системе с буфером конечной емкости k. Как хорошо известно, если буфер бесконечный ($k = \infty$) и СМО действует в области стабильности (т.е. в области интенсивности входного трафика, не приводящей к потерям и дающей конечную среднюю задержку), то выполняется так называемый закон сохранения. По этому закону все дисциплины обслуживания (такие как FIFO, LIFO, случайная и др.) эквивалентны в том смысле, что они дают одну и ту же среднюю задержку. Таким образом, задача поиска оптимальной по средней задержке дисциплины становится осмысленной, только если система с бесконечным буфером действует вне области стабильности или если буфер является конечным ($k < \infty$).

Мы покажем, что при конечном буфере и любой заданной выборочной функции входного трафика дисциплина LIFO (обозначаемая через d_L) оказывается оптимальной, дающей минимальную задержку на одно обслуженное требование, а дисциплина FIFO является наихудшей, дающей максимальную задержку на одно обслуженное требование. Удивительно, что столь интенсивно используемая дисциплина обслуживания, каковой является FIFO, в действительности является наихудшей в случае конечного буфера, а в случае бесконечного буфера в области стабильности она эквивалентна любой другой дисциплине по закону сохранения. Более того, мы покажем, что в случае бесконечного буфера и вне области стабильности LIFO опять-таки является оптимальной, а FIFO опять-таки является наихудшей по задержке.

Указанные общие результаты иллюстрируются на примере системы, которая действует в диапазоне $0 < \lambda < \infty$ с одной из рассматриваемых дисциплин d^*, d_F, d_L и d_R, где d_R – случайная дисциплина с вышибанием. Для этих дисциплин и для стационарных входных процессов без памяти даются уравнения для вероятности задержки обслуживания требования. Эти уравнения можно решить численно и по их решению легко найти вероятность потери требования, а также условные первые два момента и дисперсию задержки требования при условии, что требование обслуживается. Проводится сравнение этих характеристик для дисциплин d^*, d_F, d_L и d_R при пуассоновском входном процессе.

Сравнение характеристик различных дисциплин обслуживания, включающих FIFO и LIFO, ранее проводилось в литературе в целях поиска путей улучшения характеристик СМО с конечным буфером. К таким работам относятся интересные работы Доши и Липпер [10], Шансикумар и Сумита [11], Товсли и Баччели [12] и Лиу и Товсли [13], где сравниваются FIFO и LIFO. В [10] и [11] рассматривается СМО с одной очередью, а в [12] и [13] – сети с очередями. Важный случай с конечным буфером рассмотрен Доши и Хеффесом [8] и Яин и Хлачжи [9]. В [8] система $M/M/1$ с непрерывным временем рассмотрена для пяти частных дисциплин, включающих дисциплину FIFO и LIFO с вышибанием. Однако, в ней не поднимается задача об оптимальности. То же самое можно сказать о работе [9], в которой рассматриваются FIFO с вышибанием из начала и из конца очереди и показано, что вышибание из начала очереди лучше в стохастическом смысле. В [9] показано, что для дисциплин d^* и d_F вероятность потери пакета одна и та же. В [9] также показано, что для d^* вероятность того, что задержка пакета больше, чем заданное число, больше или равна аналогичной вероятности для d_F. В работе Печинкина и Соловьева [14] проводится поиск наилучших дисциплин среди консервативных в системе $M/G/1$ с непрерывным временем.

Так как мы в основном интересуемся сетями с АТМ, то рассматриваются СМО, отличные от тех, которые рассматривались в [8, 9]. А именно, рассматриваются системы с дискретным временем, постоянным временем обслуживания и произвольным входным трафиком. Как было сказано, наша главная цель – получить результаты в решении проблемы оптимальности.

§ 2. Система обслуживания с дисциплиной d^*

Рассмотрим СМО, состоящую из буфера с k местами для ожидания и обслуживающего прибора с постоянным временем обслуживания, равным единице. Предположим, что процесс прихода новых требований есть процесс дискретного времени $t \in I_0 \triangleq \{0, 1, 2, \ldots\}$. Число новых требований, приходящих в момент t, обозначается через β_t.

Для представления дисциплины обслуживания d^* и действий системы с этой дисциплиной удобно разбить момент времени t на два t' и t'' так, что $t' < t''$. При этом β_t новых требований приходят в систему в момент t'. Пусть α_t обозначает число требований в буфере в момент t' до прихода β_t новых требований, $0 \leqslant \alpha_t \leqslant k$. Обслуживание действует между моментами t'' и $(t+1)'$. Требование может пойти на обслуживание в момент t'', и если так, то оно должно быть обслужено и уйти из системы к моменту $(t+1)'$ до возможного прихода новых требований. Тогда, если $\alpha_t > 0$, то можно заключить, что одно требование (не из α_t) закончило обслуживание между моментами $(t-1)''$ и t'. Если $\alpha_t = 0$, то либо требование обслужилось между моментами $(t-1)''$ и t', либо прибор был свободным в течение этого интервала. Предположим также, что $\alpha_t = 0$ при $t = 0$, и в этот момент прибор свободный.

Новые требования нумеруются в порядке прибытия номерами $1, 2, \ldots$ Требования, прибывающие в один и тот же момент, при нумерации упорядочиваются случайно с равномерным распределением, т.е. при $\beta_t = \ell$ каждое возможное упорядочение имеет вероятность $(\ell!)^{-1}$. Места в буфере нумеруются $1, \ldots, k$.

Представим теперь дисциплину обслуживания d^*, которая указывает, как размещаются требования в буфере, и дает механизм сбрасывания (потери) требований.

Пусть в момент t' буфер пуст ($\alpha_t = 0$) и приходят β_t, $0 < \beta_t \leqslant k+1$, новых пакетов. Тогда требование с наименьшим номером идет на обслуживание в момент t'', а остающиеся $\beta_t - 1$ требований помещаются в буфер в порядке их номеров. Последнее означает, что если требование i занимает в буфере место j, то требование $i+1$ занимает место $j+1$.

Предположим, что в момент t' буфер пустой, и приходят $\beta_t > k+1$ новых требований. Тогда k поступивших требований с наибольшими номерами помещаются в буфер в порядке своих номеров (требования с наименьшими номерами помещаются на места с наименьшими номерами). Из остающихся $\beta_t - k$ прибывших требований одно (с наибольшим номером) идет на обслуживание в момент t'', а остальные сбрасываются.

Предположим, что в момент t' буфер не пустой, т.е. $\alpha_t > 0$, и приходит $\beta_t > 0$ новых требований. Если $\alpha_t + \beta_t \leqslant k+1$, то требование с наименьшим номером (среди $\alpha_t + \beta_t$) идет на обслуживание в момент t'' (это то требование, которое было на месте 1 в момент t'), а остающиеся $\alpha_t + \beta_t - 1$ требований помещаются в буфер в порядке своих номеров, т.е. в момент t'' очередь остающихся $\alpha_t - 1$ требований сдвигается на одно место вперед (требование с места i переходит на место $i-1$), и

β_t новых требований размещаются за ними в порядке своих номеров. Если $\alpha_t+\beta_t > k+1$, то k требований (из $\alpha_t+\beta_t$) с наибольшими номерами помещаются в буфер в порядке своих номеров, а из остающихся $\alpha_t+\beta_t-k$ требований одно с наибольшим номером идет на обслуживание в момент t'', а остальные сбрасываются.

Из этого следует, что всегда, когда требования пропадают, пропадают те требования, которые либо дольше других находились в буфере, либо имеют номера, меньшие чем требования, остающиеся в буфере в момент пропадания.

В §3 будет найдено стационарное распределение вероятностей для времени, проведенного в буфере теми требованиями, которые получают обслуживание.

В момент $t \in I_0$ пусть одно дополнительное требование (назовем его тест-требованием) прибывает вместе с β_t требованиями трафика. Все требования (тест-требование и β_t новых требований) нумеруются, как сказано выше. Если в момент t тест-требование получает наименьший номер, то ему дополнительно присваивается индекс 0. В общем случае ему дополнительно присваивается индекс j, если оно получает $(j+1)$-й по величине номер (считая от наименьшего), и если $\beta_t = \ell$, то $0 \leqslant j \leqslant \ell$.

Определим $Q_t(x)$ как вероятность того, что тест-требование, возникшее в момент t, не будет потеряно и проведет в буфере время x до своего обслуживания. Стационарное распределение определяется равенством

$$Q(x) \triangleq \lim_{t\to\infty} Q_t(x), \tag{1}$$

если этот предел существует.

Распределение $Q(x)$ не обязательно нормировано, т.е.

$$q \triangleq \sum_{x=0}^{\infty} Q(x) \leqslant 1.$$

Величина $1-q$ есть вероятность того, что тест-требование теряется. Условное распределение того, что тест-требование задерживается на время x до того, как получает обслуживание, при условии, что оно обслуживается, дается выражением $Q(x)/q$.

В §3 рассматривается дисциплина d^* в предположении, что входной трафик является процессом без памяти и стационарным. При этом величины β_t, $t \in I_0$, независимы и одинаково распределены с

$$p_\ell \triangleq \Pr\{\beta_t = \ell\}, \quad \ell \in I_0. \tag{2}$$

В этом случае, если $\lambda \triangleq \sum_{\ell=0}^{\infty} \ell p_\ell$ обозначает интенсивность входного трафика, то разумно ожидать, что $q\lambda \to 1$ при $\lambda \to \infty$ (так как при $\lambda \to \infty$ интенсивность потока потерянных требований стремится к $\lambda - 1$, а вероятность $1-q$ потери тест-требования стремится к $(\lambda-1)/\lambda$).

§ 3. Стационарное распределение при дисциплине d^*

В этом параграфе рассматривается дисциплина d^*, действующая, когда входной трафик без памяти и стационарный, как в (2). Наша цель – найти стационарное распределение $Q(x)$. Для того чтобы выразить $Q(x)$, мы используем переходную вероятность

$$p_{mn} = \Pr\{\alpha_{t+1} = n \mid \alpha_t = m\}, \quad 0 \leqslant m \leqslant k, \quad 0 \leqslant n \leqslant k, \tag{3}$$

стационарную вероятность P_m того, что буфер содержит m требований в момент t' до поступления новых требований, и условную стационарную вероятность $g(x,y,u)$

того, что тест-требование идет на обслуживание в момент $t + x$ при условии, что в момент t тест-требование занимает место y в буфере и всего в буфере находится u требований, включая тест-требование.

Заметим, что последовательность $\{\alpha_t, t \in I_0\}$ является цепью Маркова с переходными вероятностями p_{mn}, которые даются следующими равенствами:

$$
\begin{aligned}
&p_{00} = p_0 + p_1, \\
&p_{mn} = 0, \quad \text{если} \quad m > n + 1, \\
&p_{mn} = p_{n-m+1}, \quad \text{если} \quad 0 \leqslant m \leqslant n + 1, \quad n < k, \\
&p_{mn} = \sum_{\ell=k-m+1}^{\infty} p_\ell, \quad \text{если} \quad n = k, \quad 0 \leqslant m \leqslant k.
\end{aligned}
\tag{4}
$$

Цепь Маркова $\{\alpha_t,\ t \in I_0\}$ имеет стационарное распределение $\{P_m,\ 0 \leqslant m \leqslant k\}$, которое является решением следующей хорошо известной системы уравнений:

$$
P_n = \sum_{m=0}^{k} P_m p_{mn}, \qquad \sum_{m=0}^{k} P_m = 1. \tag{5}
$$

Для того чтобы избежать рассмотрения пределов при $t \to \infty$, мы будем предполагать, что цепь $\{\alpha_t, t \in I_0\}$ обладает стационарным распределением в начальный момент $t = 0$. Стационарное распределение $Q(x)$ представляется следующей теоремой.

Теорема 1. *При дисциплине обслуживания d^* стационарное распределение $Q(x)$ дается следующими уравнениями:*

$$
Q(0) = P_0 \sum_{s=0}^{\infty} \frac{1}{s+1} p_s + (1 - P_0) \sum_{s=k}^{\infty} \frac{1}{s+1} p_s, \tag{6}
$$

$$
\begin{aligned}
Q(x) = &\sum_{(m,\ell,j)\in A} P_m p_\ell \frac{1}{\ell+1} g(x, m+j, \ell+m) + \\
+ &\sum_{(m,\ell,j)\in B} P_m p_\ell \frac{1}{\ell+1} g(x, k-\ell+j, k), \quad x > 0,
\end{aligned}
\tag{7}
$$

где

$$
A = \{(m, \ell, j) : \ell + m \leqslant k, \quad 0 \leqslant m \leqslant k, \quad 0 \leqslant j \leqslant \ell, \quad \ell \geqslant 0\}, \tag{8}
$$

$$
B = \{(m, \ell, j) : \ell + m > k, \quad 0 \leqslant m \leqslant k, \quad 0 \leqslant j \leqslant \ell, \quad \ell \geqslant 0\} \tag{9}
$$

и где $\{g(x, y, u),\ 1 \leqslant x \leqslant k,\ 1 \leqslant y \leqslant k,\ 1 \leqslant u \leqslant k\}$ есть решение уравнения

$$
\begin{aligned}
g(x, y, u) = &\sum_{\ell\in C} p_\ell g(x-1, y-1, \ell+u-1) + \\
+ &\sum_{\ell\in D} p_\ell g(x-1, y-\ell-u+k, k); \quad x > 0,
\end{aligned}
\tag{10}
$$

где

$$
C = \{\ell : \ell + u \leqslant k + 1, \quad \ell \geqslant 0\}, \tag{11}
$$

$$
D = \{\ell : k + 2 \leqslant \ell + u \leqslant y + k + 1, \quad \ell \geqslant 0\} \tag{12}
$$

и

$$
\begin{aligned}
&g(x,y,u)=0, \quad \text{если} \quad y=0 \quad \text{и} \quad x>0,\\
&\text{или если} \quad y>u, \quad \text{или если} \quad x>y, &(13)\\
&g(0,y,u)=\delta_{0y}, &(14)
\end{aligned}
$$

где $\delta_{0y}=1$, *если* $y=0$, *и* $\delta_{0y}=0$, *если* $y>0$.

Доказательство является простым, и оно использует смысл функции $g(x,y,u)$ как условной вероятности того, что тест-требование идет на обслуживание в момент $t+x$ при условии, что оно занимает место y в буфере в момент t, и буфер содержит u требований в этот момент, включая тест-требование.

Рассмотрим момент t, расщепленный на два (t' и t'', как в §2). В момент t' пусть буфер содержит m требований, и пусть прибывают s новых требований плюс тест-требование. Пусть тест-требованию присваивается индекс j, $0\leqslant j\leqslant s$.

Найдем сначала вероятность того, что тест-требование обслуживается без ожидания в буфере, т.е. найдем $Q(x)$ при $x=0$. При этом тест-требование идет на обслуживание в момент t'', что происходит в следующих двух случаях: (а) когда в момент t' буфер пустой ($m=0$) и $j=0$; (б) когда в момент t' буфер не пустой ($m>0$), но $s\geqslant k$ и $j=s-k$. Случай (а) дает первый член в (6), а случай (б) дает второй член в (6).

Найдем теперь $Q(x)$ для $x>0$. Если $s+m\leqslant k$, то тест-требование ждет время x в буфере до начала своего обслуживания с вероятностью $g(x,m+j,s+m)$. Принимая это во внимание и усредняя, получаем первый член в (7). Если $s+m>k$ и $s-j\leqslant k-1$, то тест-требование ждет время x в буфере до начала своего обслуживания с вероятностью $g(x,k-s+j,k)$. В этом случае после усреднения получаем второй член в (7). Если $s+m>k$ и $s-j\geqslant k+1$, то тест-требование теряется в момент t'', и этот случай не дает вклада в $Q(x)$. Если $s+m>k$ и $s-j=k$, то тест-требование получает немедленное обслуживание ($x=0$), и этот случай также не дает вклада в $Q(x)$, $x>0$. Таким образом, равенство (7) доказано.

Остается показать, что $g(x,y,u)$, $x>0$, удовлетворяет (10). С этой целью рассмотрим два момента времени t и $t+1$ и разобъем их на t', t'' и на $(t+1)'$, $(t+1)''$, соответственно, как в §2. Пусть тест-требование приходит в момент t'; пусть в момент t'' оно занимает место y в буфере, а буфер содержит u требований ($u\geqslant y$). Пусть ℓ новых требований прибывают в момент $(t+1)'$.

Если $\ell+u\leqslant k+1$, то в момент $(t+1)''$ тест-требование передвигается на место $y-1$ в буфере (если $y=1$, то оно идет на обслуживание), и в этот момент буфер содержит $\ell+u-1$ требований. Принимая это во внимание и усредняя, получаем первый член в (10). Если $\ell+u>k+1$, то в момент $(t+1)''$ тест-требование передвигается на место $y-\ell-u+k$, и буфер становится полным. В этом случае после усреднения получаем второй член в (10). При выводе (10) мы также использовали равенства (13) и (14). ▲

Уравнение (10) можно решить рекуррентно по x относительно $g(x,y,u)$, начиная с $g(1,y,u)$. В начальной точке имеем

$$
g(1,1,u)=\sum_{0\leqslant\ell\leqslant k-u+1} p_\ell, \tag{15}
$$

$$
g(1,y,u)=p_{k+y-u}, \quad y>1. \tag{16}
$$

В частном случае пуассоновского входного потока

$$
p_\ell=e^{-\lambda}\frac{\lambda^\ell}{\ell!} \tag{17}
$$

и

$$
\sum_{\ell=0}^{\infty} p_\ell\frac{1}{\ell+1}=\frac{1}{\lambda}(1-e^{-\lambda}). \tag{18}
$$

Для пуассоновского входного потока выражение (6) принимает следующий вид:

$$
\begin{aligned}
&Q(0) = \frac{1}{\lambda}(1 - e^{-\lambda}), \quad \text{если} \quad k = 0, \\
&Q(0) = \frac{1}{\lambda}(1 - e^{-\lambda}) - (1 - P_0)e^{-\lambda}\sum_{s=1}^{k}\frac{\lambda^{s-1}}{s!}, \quad \text{если} \quad k \geqslant 1.
\end{aligned}
\tag{19}
$$

Когда $\lambda \to \infty$, (19) дает $Q(0) \to 0$, а $\lambda Q(0) \to 1$. Последний предел означает, что при условии, что требование получает обслуживание, оно получает его без ожидания в буфере. Это свойство является одним из замечательных свойств дисциплины d^*: при увеличении интенсивности трафика время ожидания в буфере для обслуживаемых требований стремится к нулю.

Решение уравнений (5) легко найти в виде функции от P_0 последовательно, начиная с P_1. При этом удобно использовать равенство

$$
P_{n+1} = \frac{1}{p_0}\left[P_n - \sum_{m=0}^{n} P_m p_{mn}\right], \quad 0 \leqslant n \leqslant k-2.
\tag{20}
$$

Вероятность P_0 можно найти с помощью нормировки. Например, при $k = 2$ решение системы (5) имеет вид

$$
\begin{aligned}
&P_0 = \left\{1 + \frac{1 - p_0 - p_1 - p_2}{p_0} - \frac{(1 - p_0 - p_1)(1 - p_1)}{p_0^2}\right\}^{-1}, \\
&P_1 = P_0\frac{1 - p_0 - p_1}{p_0}, \\
&P_2 = P_0\left\{\frac{(1 - p_0 - p_1)(1 - p_1)}{p_0^2} - \frac{p_2}{p_0}\right\}.
\end{aligned}
\tag{21}
$$

§ 4. Сравнение дисциплин d^* и d_F

Здесь будет показано, что дисциплина d^*, рассмотренная в §§ 2, 3, лучше дисциплины d_F по целому ряду параметров при общем входном трафике.

Так же, как в § 2, рассмотрим СМО, состоящую из прибора с постоянным временем обслуживания, равным единице, и буфера с объемом k мест ожидания. В отличие от стационарного трафика без памяти, рассмотренного в § 3, здесь рассматривается произвольный входной трафик β_t, $t \in I_0$, поступления новых требований. В частности, не требуется, чтобы он был стационарным или без памяти. Требования, составляющие трафик, нумеруются так же, как в § 2.

Условимся говорить, что для любого заданного интервала времени $[0, T]$ выборочная функция входного трафика является заданной, если заданы число новых требований, прибывших в $[0, T]$, и моменты прибытия для каждого требования.

Как и ранее, d^* обозначает дисциплину, описанную в §2, а d_F обозначает дисциплину FIFO с потерей новых требований, которые сталкиваются с полным буфером. Поясним, что при d_F требования помещаются в буфер в порядке их номеров и что, когда буфер полный, требования с наибольшими номерами теряются.

Символ d используется для обозначения как d^*, так и d_F. Обозначим через $\delta_t(d)$ время, проведенное в буфере требованием, которое идет на обслуживание в момент t''. Значение $\delta_t(d)$ есть задержка этого требования при дисциплине d. Если буфер пустой в момент t', то считается, что $\delta_t(d) = 0$, хотя никакое требование не идет на обслуживание в момент t'' в этом случае.

Утверждение 1 [9]. *При любой выборочной функции входного трафика и* *любом* $t \in I_0$ *обе дисциплины* d^* *и* d_F *либо начинают обслуживание, либо не имеют требований для обслуживания.*

Доказательство. Это утверждение есть прямое следствие того, что для d^* и d_F

(а) если $\beta_t + \alpha_t \leqslant k+1$, то требования не пропадают в момент t'';

(б) если $\beta_t + \alpha_t > 0$ в момент t', то некоторое требование (из $\beta_t + \alpha_t$) идет на обслуживание в момент t''. ▲

Утверждение 2 [9]. *Для любой выборочной функции входного трафика и любом* $t \in I_0$

$$\delta_t(d^*) \leqslant \delta_t(d_F). \tag{22}$$

Доказательство. При дисциплине d^* требование, которое идет на обслуживание в момент t'', имеет номер, который не меньше, чем номер требования, которое идет на обслуживание в тот же момент при дисциплине d_F. Это следует из того, что всегда, когда происходит пропадание требований при дисциплине d^*, пропадающие требования имеют наименьшие номера среди требований, находящихся в буфере и вновь прибывших. Это дает (22). ▲

При заданных i и T из $I_1 = \{1, 2, \dots\}$ и дисциплине d обозначим сумму задержек, возведенных в степень i, через

$$\Delta_T^{(i)}(d) = \sum_{t=0}^{T} \delta_t^i(d). \tag{23}$$

Утверждение 3. *При любой заданной выборочной функции входного трафика на* $[0, T]$

$$\delta_T^i(d^*) \leqslant \delta_T^i(d_F), \quad \Delta_T^{(i)}(d^*) \leqslant \Delta_T^{(i)}(d_F). \tag{24}$$

Доказательство. Утверждение немедленно следует из утверждения 2. ▲

Если выборочная функция входного трафика не задана, то величины $\delta_t(d)$ и $\Delta_T^i(d)$ являются случайными. В этом случае из (22) и (24) следуют неравенства

$$\Pr\left\{\delta_t^i(d^*) > w\right\} \leqslant \Pr\left\{\delta_t^i(d_F) > w\right\}, \quad w > 0, \tag{25}$$

$$\Pr\left\{\frac{1}{T}\Delta_T^{(i)}(d^*) > w\right\} \leqslant \Pr\left\{\frac{1}{T}\Delta_T^{(i)}(d_F) > w\right\}, \quad i \in I_1, \quad w > 0, \tag{26}$$

$$M_T^{(i)}(d^*) \triangleq \frac{1}{T} E\left\{\Delta_T^{(i)}(d^*)\right\} \leqslant \frac{1}{T} E\left\{\Delta_T^{(i)}(d_F)\right\} \triangleq M_T^{(i)}(d_F), \quad i \in I_1. \tag{27}$$

Если пределы

$$M^{(i)}(d) \triangleq \lim_{T\to\infty} M_T^{(i)}(d) \tag{28}$$

для d^* и d_F существуют, то они называются стационарными моментами задержки порядка i для соответствующей дисциплины.

Неравенства (27) показывают, что стационарная средняя задержка для d^* не больше, чем аналогичная задержка для d_F. То же утверждение верно для любого стационарного момента порядка i.

Приведем теперь теорему, которая выражает стационарную вероятность $Q(x)$ при дисциплине d_F. Эта теорема аналогична теореме 1. Она позволяет провести численное сравнение характеристик дисциплин d^* и d_F.

Теорема 2. *При стационарном входном трафике без памяти (таком, как в (2)) и при дисциплине обслуживания* d_F *стационарное распределение* $Q(x)$ *дается равенством*

$$Q(x)=\sum_{m=0}^{x} P_m \sum_{s=x-m}^{\infty} p_s \frac{1}{s+1}, \quad 0 \leqslant x \leqslant k, \tag{29}$$

в котором $\{P_m,\ 0 \leqslant m \leqslant k\}$ *есть решение системы* (5).

Доказательство. Пусть в момент t' до возможного поступления новых требований буфер содержит m требований, $0 \leqslant m \leqslant k$. Пусть тест-требованию присвоен индекс j, $0 \leqslant j \leqslant s$, в момент его прибытия t' вместе с s новыми требованиями (тест-требование не входит в s). Событие {задержка тест-требования равна x} состоит из следующих непересекающихся событий:

$\{m = x,\ j = 0\}$, имеющего стационарную вероятность

$$P_x \sum_{s=0}^{\infty} p_s \frac{1}{s+1};$$

$\{m = x-1,\ j = 1\}$, имеющего стационарную вероятность

$$P_{x-1} \sum_{s=1}^{\infty} p_s \frac{1}{s+1};$$

$\{m = 0,\ j = x\}$, имеющего стационарную вероятность

$$P_0 \sum_{s=x}^{\infty} p_s \frac{1}{s+1}.$$

Сумма указанных вероятностей дает (29). ▲

§ 5. Оптимальная дисциплина

В этом параграфе доказывается оптимальность дисциплины LIFO в смысле задержки; она дается утверждением 8. Также будет показано, что дисциплина FIFO является наихудшей в том же смысле; это формулируется как утверждение 13.

Так же, как в §4, рассмотрим СМО, состоящую из прибора с постоянным временем обслуживания, равным единице, и буфера с k местами для ожидания. Входной трафик β_t, как и в § 4, считается произвольным, а новые требования нумеруются, как в § 2. Кроме того, мы будем иногда рассматривать ту же самую систему, но с бесконечным буфером, $k = \infty$.

Здесь рассматриваются только те дисциплины обслуживания, которые удовлетворяют следующим двум условиям:

(1°) Если $\beta_t + \alpha_t \leqslant k+1$, то требования не теряются в момент t''. Если $\beta_t + \alpha_t > k + 1$, то в точности $\beta_t + \alpha_t - k - 1$ требование теряется в момент t''. (Выбор $\beta_t + \alpha_t - k - 1$ требований для сброса определяется частной дисциплиной и никак не ограничен.)

(2°) Если в момент t' имеем $\beta_t + \alpha_t > 0$, то некоторое требование (из $\beta_t + \alpha_t$) идет на обслуживание в момент t''. (Выбор этого требования также определяется частной дисциплиной и никак не ограничен.)

Обозначим через $D(k)$ множество дисциплин, удовлетворяющих условиям (1°) и (2°). В случае бесконечного буфера никакие требования не теряются, поэтому при $k = \infty$ дисциплины из $D(k)$ должны удовлетворять только условию (2°). Этот класс дисциплин обозначается через $D(\infty)$.

Представим теперь известную дисциплину LIFO, действующую с буфером емкости k. Предположим, что в момент t' буфер содержит α_t требований и что прибывают β_t новых требований. Тогда в момент t'' требование с наибольшим номером (среди $\alpha_t + \beta_t$ требований) идет на обслуживание и либо k из остающихся $\alpha_t + \beta_t - 1$ требований (если $\alpha_t + \beta_t - 1 > k$) с наибольшими номерами, либо все $\alpha_t + \beta_t - 1$ остающиеся требования (если $\alpha_t + \beta_t - 1 \leqslant k$) размещаются в буфере, а $\alpha_t + \beta_t - k - 1$

требований (если $\alpha_t + \beta_t - 1 > k$) пропадают. Порядок размещения требований в буфере не играет роли. Заметим, что и ранее при описании дисциплины d^* порядок размещения также не играл роли.

Обозначим описанную дисциплину LIFO, действующую с буфером емкости k, через $d_L(k)$. Ясно, что $d_L(k) \in D(k)$.

Рассмотрим интервал времени $[0, t]$ и произвольную, но заданную, выборочную функцию входного трафика. Обозначим через $A_t(d)$ множество номеров тех требований, которые обслуживаются дисциплиной d в $[0, t]$.

Утверждение 4. *При любой заданной выборочной функции входного трафика и любом заданном t, если $A_t(d) = A_t(\widehat{d})$ при некоторых d, $\widehat{d}$ из $D(k)$, то*

$$\Delta_t^{(1)}(d) = \Delta_t^{(1)}(\widehat{d}). \tag{30}$$

Утверждение справедливо как при $k < \infty$, так и при $k = \infty$.

Доказательство. Утверждение следует из того, что сумма задержек требований из $A_t(d)$ не зависит от порядка их обслуживания в $[0, t]$, если этот порядок диктуется некоторой дисциплиной из $D(k)$. ▲

Утверждение 5. *При любой заданной выборочной функции входного трафика, любом моменте t и любых дисциплинах d и $\widehat{d}$ из $D(k)$ обе дисциплины либо начинают обслуживание, либо не имеют требований для обслуживания. Кроме того, для обеих дисциплин число требований в системе одно и то же.*

Доказательство. Так как дисциплины d и $\widehat{d}$ удовлетворяют условиям (1°) и (2°), то можно повторить доказательство утверждения 1. ▲

Утверждение 6. *При любой заданной выборочной функции входного трафика, любом моменте t и любой $d(\infty) \in D(\infty)$*

$$\Delta_t^{(1)}(d_L(\infty)) \leqslant \Delta_t^{(1)}(d(\infty)) \leqslant \Delta_t^{(1)}(d_F(\infty)). \tag{31}$$

Доказательство. Докажем лишь левое неравенство в (31), так как доказательство правого неравенства полностью аналогично.

Пусть заданы выборочная функция входного трафика и момент t. Пусть d – некоторая дисциплина из $D(\infty)$, и пусть $\widehat{d} \in D(\infty)$ – такая дисциплина, что

$$A_t(\widehat{d}) = A_t(d), \tag{32}$$

и такая, что при $\widehat{d}$ каждое требование из множества

$$B(d_L, d) \triangleq A_t(d_L) \cap A_t(d) = A_t(d_L) \cap A_t(\widehat{d}), \quad d_\ell \in D(\infty), \tag{33}$$

идет на обслуживание в тот же самый момент, что и при дисциплине d_L.

Согласно утверждению 4 для дисциплин d и $\widehat{d}$ имеем

$$\Delta_t^{(1)}(d) = \Delta_t^{(1)}(\widehat{d}). \tag{34}$$

Требования из $B(d_L, d)$ дают равные вклады в $\Delta_t^{(1)}(\widehat{d})$ и $\Delta_t^{(1)}(d_L)$. Поэтому для доказательства утверждения остается сравнить вклады в $\Delta_t^{(1)}(\widehat{d})$ и $\Delta_t^{(1)}(d_L)$ от требований, которые не входят в $B(d_L, d)$.

Рассмотрим некоторый момент времени $t_1 \in \{0, \ldots, t\}$, в который начинается обслуживание какого-либо требования, не входящего в $B(d_L, d)$. Согласно утверждению 5 в момент t_1 каждая дисциплина из $D(\infty)$ начинает обслуживание. Требование, которое начинает обслуживаться при дисциплине $\widehat{d}$ в момент t_1, дает вклад в $\Delta_t^{(1)}(\widehat{d})$ не меньший, чем вклад в $\Delta_t^{(1)}(d_L)$ от требования, которое начинает обслуживаться при дисциплине d_L в момент t_1. Это происходит потому, что выбор требования для обслуживания в момент t_1 дисциплиной d_L проводится так, что (1′) она выбирает требование из множества, которое содержит требование, выбираемое дисциплиной $\widehat{d}$ для обслуживания в момент t_1, и (2′) она выбирает требование, имеющее наименьшее время пребывания в буфере.

Это вместе с равенством (34) доказывает левое неравенство в (31). ▲

При любой заданной выборочной функции входного трафика обозначим через E_1 первый момент (после $t = 0$), когда буфер пустой при дисциплине $d \in D(k)$. Имеем $E_1 \geqslant 0$. Легко заметить, что E_1 – вычислимый по выборочной функции момент времени. Кроме того, благодаря утверждению 5 легко заметить, что E_1 не зависит от дисциплины d, если d из $D(k)$.

У т в е р ж д е н и е 7. *При любой заданной выборочной функции входного трафика и любом моменте t из $[0, E_1]$ обе дисциплины $d_L(k)$ и $d_L(\infty)$ либо начинают обслуживание требования с одним и тем же номером, либо вообще не начинают обслуживания.*

Д о к а з а т е л ь с т в о. Случай $E_1 = 0$ тривиален. Пусть $E_1 > 0$. Тогда до момента $E_1 + 1$ дисциплина $d_L(\infty)$ обслуживает те требования, которые не пропадают при дисциплине $d_L(k)$, и $d_L(\infty)$ обслуживает их в том же порядке, что и дисциплина $d_L(k)$. ▲

У т в е р ж д е н и е 8. *При любой заданной выборочной функции входного трафика, любом заданном моменте t и любой $d(k) \in D(k)$*

$$\Delta_t^{(1)}(d_L(k)) \leqslant \Delta_t^{(1)}(d(k)). \tag{35}$$

Д о к а з а т е л ь с т в о. Предположим, что задана выборочная функция входного трафика и что она такая, что число новых требований, прибывших в момент $t = 0$, больше нуля (случай нулевого числа требований при $t = 0$ является тривиальным). По такой выборочной функции можно найти момент E_1, когда буфер в первый раз становится пустым.

В силу утверждения 7 до момента $E_1 + 1$ дисциплина $d_L(k)$ обслуживает те же требования и в те же моменты времени, что и дисциплина $d_L(\infty)$. Поэтому при любом $t \leqslant E_1$

$$\Delta_t^{(1)}(d_L(k)) = \Delta_t^{(1)}(d_L(\infty)). \tag{36}$$

Так как ($а_1$) для любой дисциплины $d(k) \in D(k)$ существует дисциплина $d(\infty) \in \in D(\infty)$, которая на $[0, E_1]$ обслуживает те же требования и в те же моменты, что и $d(k)$, и ($б_1$) выполняется левое неравенство в (31), мы можем использовать (36), чтобы получить

$$\Delta_t^{(1)}(d_L(k)) \leqslant \Delta_t^{(1)}(d(k)), \quad 0 \leqslant t \leqslant E_1. \tag{37}$$

Так как ($а_2$) буфер пустой в момент E_1 и ($б_2$) выполняется левое неравенство в (31) при любой выборочной функции входного трафика, то для рассматриваемой СМО с буфером емкости k момент времени E_1 является рекурсивным моментом в следующем смысле. Начиная с E_1, определим последовательность $\{E_i, i \geqslant 1\}$ последовательных моментов времени, в которые буфер является пустым. Тогда можно применить те же рассуждения, что и выше, чтобы показать, что $\Delta_t^{(1)}(d_L(k))$ вновь минимально в $[0, E_2]$, $[0, E_3]$ и так далее, что дает (35). ▲

Теперь мы переходим к тому, чтобы показать, что $d_F(k)$ – наихудшая дисциплина в смысле задержки.

У т в е р ж д е н и е 9. *При любой заданной выборочной функции входного трафика моменты $T_1, T_2, \ldots$, в которые дисциплина $d(k) \in D(k)$ сбрасывает требования, и числа сброшенных требований в моменты $T_1, T_2, \ldots$ не зависят от $d(k)$.*

Доказательство. Утверждение прямо следует из условий (1°) и (2°) и утверждения 5. ▲

Пусть $\{T_i, i \geqslant 1\}$ – последовательность моментов времени, в которые любая дисциплина из $D(k)$ сбрасывает требования (см. утверждение 9). Определим подмножество $D_1(k)$ множества $D(k)$ следующим образом: в каждый момент времени T_i каждая дисциплина из $D_1(k)$ сбрасывает только новые требования (из β_{T_i}). Без потери общности предположим, что в момент T_i дисциплина $d(k) \in D_1(k)$ сбрасывает новые требования с наивысшими номерами (среди β_{T_i} номеров). Теперь перейдем к следующему утверждению.

Утверждение 10. *При любой выборочной функции входного трафика и любом t существует дисциплина $\widehat{d}(k) \in D_1(k)$ такая, что*

$$\Delta_t^{(1)}(d(k)) \leqslant \Delta_t^{(1)}(\widehat{d}(k)) \tag{38}$$

при любой $d(k) \in D(k)$.

Доказательство. Утверждение следует из утверждений 5 и 9 и того, что в момент T_i любая дисциплина из $D_1(k)$ не сбрасывает $k+1$ наиболее старых требований среди $\beta_{T_i} + \alpha_{T_i}$, $\quad i \in I_1$. ▲

Следствие утверждения 10. *При любой заданной выборочной функции входного трафика и любом t, если дисциплина $\widehat{d}(k) \in D_1(k)$ такая, что*

$$\Delta_t^{(1)}(d(k)) \leqslant \Delta_t^{(1)}(\widehat{d}(k)) \tag{39}$$

при любой $d(k) \in D_1(k)$, то (39) также справедливо при той же $\widehat{d}(k)$ при любой $d(k) \in D(k)$.

Пусть S – множество всех выборочных функций входного трафика и пусть $S_1 \subset \subset S$ – подмножество, построенное следующим образом. Каждой $s \in S$ и связанным с ней моментам $T_1, T_2, \ldots$ сопоставим $s_1(s) \in S$ так, что число новых требований, приходящих в s_1 в момент T_i, $i \in I_0$, меньше, чем соответствующее число в s, точно на число сброшенных в момент T_i требований, когда дисциплина $d(k) \in D_1(k)$ действует с выборочной функцией s. Подмножество S_1 теперь строится как $S_1 = \{s_1 = s_1(s) : s \in S\}$.

Утверждение 11. *Если для любой $s_1 \in S_1$, любого t и любой $d(k) \in D_1(k)$ дисциплина $\widehat{d}(k) \in D_1(k)$ такая, что*

$$\Delta_t^{(1)}(d(k)) \leqslant \Delta_t^{(1)}(\widehat{d}(k)), \tag{40}$$

то (40) также справедливо с той же $\widehat{d}(k)$ при любой $s \in S$, любом t и любой $d(k) \in D(k)$.

Доказательство. Утверждение следует из следствия утверждения 10 и того, что при любой $s \in S$ дисциплина $\widehat{d}(k) \in D_1(k)$ обслуживает те же требования и в те же моменты, что и та же дисциплина, когда она действует с выборочной функцией $s_1(s)$ (единственное отличие в том, что при $s_1(s)$ требования не сбрасываются, а при s сбрасываются). ▲

Утверждение 12. *Неравенство*

$$\Delta_t^{(1)}(d(k)) \leqslant \Delta_t^{(1)}(d_F(k)) \tag{41}$$

справедливо для любой выборочной функции $s_1 \in S_1$, любом t и любой $d(k) \in D_1(k)$.

Доказательство. В соответствии с утверждением 6

$$\Delta_t^{(1)}(d(\infty)) \leqslant \Delta_t^{(1)}(d_F(\infty)) \tag{42}$$

для любой $s_1 \in S_1$, любом t и любой $d(\infty) \in D(\infty)$.

При любой $s_1 \in S_1$ каждая дисциплина $d(\infty)$ идентична некоторой дисциплине $d(k)$ (так как на s_1 дисциплина $d(\infty)$ не задействует больше, чем k мест в буфере), и множество

$$D_2(k) = \{d(k) : d(k) \quad \text{идентичная} \quad d(\infty) \quad \text{на} \quad s_1, s_1 \in S_1, \quad d(\infty) \in D(\infty)\}$$

совпадает с $D_1(k)$, т.е. $D_1(k) = D_2(k)$. При $s_1 \in S_1$ дисциплина $d_F(\infty)$ идентична $d_F(k)$. Таким образом, (41) теперь получается из (42). ▲

Утверждение 13. *При любой выборочной функции $s \in S$, любом t и любой $d(k) \in D(k)$*

$$\Delta_t^{(1)}(d(k)) \leqslant \Delta_t^{(1)}(d_F(k)). \tag{43}$$

Доказательство. Утверждение следует из утверждений 11 и 12. ▲

В заключение приведем теорему, аналогичную теоремам 1 и 2, дающую выражение для $Q(x)$ при дисциплине d_L. Она будет использована при численном сравнении дисциплин d_L, d^* и d_F.

Теорема 3. *При стационарном входном трафике без памяти (таком, как в (2)) и при дисциплине обслуживания d_L стационарное распределение $Q(x)$ дается равенством*

$$Q(x) = \sum_{s=0}^{\infty} \frac{p_s}{s+1} \sum_{y=0}^{\min(s,k)} h(x,y), \tag{44}$$

где $h(x,y)$, $\quad x \geqslant 1$, $\quad 1 \leqslant y \leqslant k$, есть решение уравнения

$$h(x,y) = \sum_{\ell=0}^{k-y+1} p_\ell h(x-1,\ y+\ell-1), \tag{45}$$

$$h(x,y) = 0, \quad \textit{если либо} \quad y \geqslant k+1, \quad \textit{либо} \quad y \geqslant x+1, \quad \textit{либо} \quad y < 0, \tag{46}$$

$$h(x,0) = \delta_{0x}. \tag{47}$$

Доказательство. Обозначим через $h(x,y)$ условную вероятность того, что тест-требование идет на обслуживание в момент $t+x$ при условии, что в момент t оно занимает в буфере место y.

Если в момент t' при прибытии тест-требования общее число новых требований равно s и тест-требованию присваивается индекс y, $\quad 0 \leqslant y \leqslant s$, то в момент t'' тест-требование либо занимает в буфере место y (если $y > 0$), либо идет на обслуживание (если $y = 0$). Поэтому получаем (44). Равенство (45) выводится аналогично тому, как в доказательстве теоремы 1, когда рассматриваются два последовательных момента t и $(t+1)$. ▲

В заключение для полноты приведем теорему, аналогичную теоремам 1, 2 и 3, но относящуюся к случайной дисциплине обслуживания с вышибанием, которая обозначается через d_R. Дисциплина d_R действует следующим образом. Пусть в буфере находятся m требований и дополнительно поступают s новых требований в момент t'. Если $m+s \leqslant k+1$, то требование, идущее на обслуживание в момент t'' выбирается случайно из $s+m$ требований. Если $m+s > k+1$, то в момент t'' сбрасываются $m+s-k-1$ требований (из $m+s$) с наименьшими номерами, а требование для обслуживания в момент t'' выбирается случайно из оставшихся $k+1$ требований.

Теорема 4. *При дисциплине обслуживания d_R стационарное распределение $Q(x)$ дается следующими уравнениями:*

$$Q(0)=\sum_{A}P_m p_s\frac{1}{s+m+1}+\sum_{B}P_m p_s\frac{1}{k+1}+\sum_{s=k+1}^{\infty}p_s\frac{1}{s+1}, \tag{48}$$

где

$$A=\{(m,s): s+m\leqslant k,\quad 0\leqslant m\leqslant k,\quad s\geqslant 0\}, \tag{49}$$
$$B=\{(m,s): s+m\geqslant k+1,\quad s\leqslant k,\quad 0\leqslant m\leqslant k,\quad s\geqslant 0\} \tag{50}$$

и

$$Q(x)=\sum_{C}P_m p_s\frac{1}{s+1}g(x,k-m-s+j,\quad m+s+1)+$$
$$+\sum_{D}P_m p_s\frac{1}{s+1}g(x,j,k+1),\quad x>0, \tag{51}$$

где

$$C=\{(m,s,j): k\geqslant s+m,\quad 0\leqslant m\leqslant k,\quad 0\leqslant j\leqslant s,\quad s\geqslant 0\}, \tag{52}$$
$$D=\{(m,s,j): k\leqslant s+m-1,\quad 0\leqslant m\leqslant k,\quad 0\leqslant j\leqslant s,\quad s\geqslant 0\} \tag{53}$$

и где $\{g(x,y,u),\quad x\geqslant 0,\quad 0\leqslant y\leqslant k,\quad 1\leqslant u\leqslant k+1\}$ *– это решение уравнений,*

$$g(0,y,u)=\frac{1}{u}, \tag{54}$$

$$g(x,y,u)=\sum_{\ell=0}^{\infty}\frac{k-y}{u}p_\ell g(x-1,Y_1,U)+$$
$$+\sum_{\ell=0}^{\infty}\frac{y-k+u-1}{u}p_\ell g(x-1,Y_2,U),\quad x\geqslant 1, \tag{55}$$

где

$$Y_1=\begin{cases} y+1, & \text{если}\quad \ell\leqslant k-u+2,\\ y+\ell-k+u-1, & \text{если}\quad \ell\geqslant k-u+2,\end{cases} \tag{56}$$

$$Y_2=\begin{cases} y, & \text{если}\quad \ell\leqslant k-u+2,\\ y+\ell-k+u-2, & \text{если}\quad \ell\geqslant k-u+2,\end{cases} \tag{57}$$

$$U=\begin{cases} k+1, & \text{если}\quad \ell\geqslant k-u+2,\\ u+\ell-1, & \text{если}\quad \ell\leqslant k-u+2\end{cases} \tag{58}$$

и

$$g(x,y,u)=0,\quad \text{если}\quad y\geqslant k+1\quad \text{или}\quad u>k+1,\quad \text{или}\quad u=0. \tag{59}$$

Вероятности P_m – это решение уравнений (5) *с p_{mn}, заданными равенством* (4).

Доказательство. Отметим сначала, что P_m, как и ранее, является стационарной вероятностью того, что буфер содержит m требований в момент t' до прихода новых требований. Вероятности P_m, $0\leqslant m\leqslant k$ являются решением системы стационарных уравнений (5) с переходными вероятностями (4).

Затем найдем вероятность того, что тест-требование, поступившее в момент t', получит обслуживание без ожидания в буфере, т.е. найдем $Q(0)$. Рассматриваемое событие означает, что тест-требование идет на обслуживание в момент t'', что происходит в следующих трех случаях:

(а) когда в момент t' буфер содержит m, $0\leqslant m\leqslant k$ требований; тест-требование приходит в этот момент t' вместе с s другими новыми требованиями ($s\geqslant 0$) и $m+s+1\leqslant k+1$; тест-требование случайно выбирается для обслуживания в момент t'' из $m+s+1$ кандидатов;

(б) когда в момент t' число $s+1$ новых поступивших требований в сумме с числом требований, находящихся в буфере, больше чем $k+1$, т.е. $m+s \geqslant k+1$, но $s \leqslant k$ и тест-требование случайно выбирается для обслуживания в момент t'' из $k+1$ кандидатов;

(в) когда в момент t' число новых поступивших требований больше чем $k+1$ (т.е. $s > k+1$); тест-требование не теряется и, более того, оно случайно выбирается для обслуживания из $k+1$ кандидатов.

Случай (а) дает первый член, случай (б) дает второй член, а случай (в) дает третий член в (48).

Чтобы получить (51) для $Q(x)$, $x > 0$, удобно ввести в рассмотрение виртуальную память объема $k+1$ с местами (каждое место может хранить одно требование) занумерованными числами $0, \ldots, k$.

Представим себе, что в момент t' новые требования (пусть их число равно ℓ) и требования из буфера (пусть их число равно m) занимают (целиком или частично) эту память так, что

(i) если $\ell + m \geqslant k+2$, то требования с наименьшими номерами теряются, а непотерянные $k+1$ требований занимают места в порядке их (требований) номеров так, что требование с наименьшим номером занимает место k, а требование с наибольшим номером занимает место 0 в памяти.

В момент t'' одно из требований из памяти выбирается случайно для обслуживания, а остальные требования помещаются в буфер.

Оставшаяся часть доказательства использует смысл функции $g(x, y, u)$ как условной стационарной вероятности того, что тест-требование поступает на обслуживание в момент $t+x$ при условии, что в момент t' тест-требование занимает в памяти место y и что в памяти содержится всего u требований (включая тест-требование), $0 \leqslant x < \infty$, $0 \leqslant y \leqslant k$, $1 \leqslant u \leqslant k+1$, x, y, u – целые числа.

Если тест-требование приходит в момент t' вместе с s другими новыми требованиями и оно дополнительно индексируется числом j и если $k \geqslant s+m$, где m – число требований в буфере в момент t' (без новых требований), то с вероятностью $g(x,\ k-m-s+j,\ m+s+1)$ тест-требование ждет время x в буфере до начала своего обслуживания. Принимая это во внимание и проводя усреднение, получаем первый член в (51). Если $k \leqslant s+m-1$, то с вероятностью $g(x, j, k+1)$ тест-требование ждет время x в буфере до начала своего обслуживания. Этот случай после усреднения дает второй член в (51). На этом доказательство (51) завершается.

Равенства (54) и (59) очевидные и остается только показать, что $g(x, y, u)$, $x > 0$, удовлетворяют равенству (55).

Пусть тест-требование приходит в момент t'; пусть в этот момент оно занимает место y в памяти и память содержит u требований. Пусть в момент $(t+1)'$ приходят ℓ новых требований.

Если в момент t'' требование, которое идет на обслуживание, занимает в памяти место с номером, большим (меньшим) чем номер места тест-требования, то в момент $(t+1)'$ тест-требование занимает в памяти место Y_1 (Y_2 соответственно) и в то же время в буфере находится U требований. Поэтому после усреднения получается первый (второй соответственно) член в (55). На этом доказательство заканчивается. ▲

§ 6. Численные результаты

Рассмотрим пуассоновский стационарный трафик без памяти и с интенсивностью λ. Для этого трафика вычислим численно такие характеристики как вероятность $Q(x)$, условное стационарное распределение $Q(x)/q$, среднюю задержку q_1, второй момент q_2 и дисперсию задержки σ^2 для дисциплин d^*, d_F и d_L в зависимости от λ и емкости буфера k. Обозначая

$$q \triangleq \sum_{x=0}^{\infty} Q(x),$$

имеем

$$q_1 \triangleq (q)^{-1} \sum_{x=0}^{\infty} xQ(x),$$

$$q_2 \triangleq (q)^{-1} \sum_{x=0}^{\infty} x^2 Q(x),$$

$$\sigma^2 = q_2 - (q_1)^2.$$

При $k = 5$ условное стационарное распределение $Q(x)/q$ показано на рис. 1 для дисциплины d^*, на рис. 2 – для дисциплины d_F и на рис. 3 – для дисциплины d_L, где $\lambda = 0,5;\ 1;\ 2;\ 5$. На этих рисунках имеют смысл лишь точки $(Q(x)/q, x)$, которые соответствуют целым x. Однако для наглядности эти точки, отно-

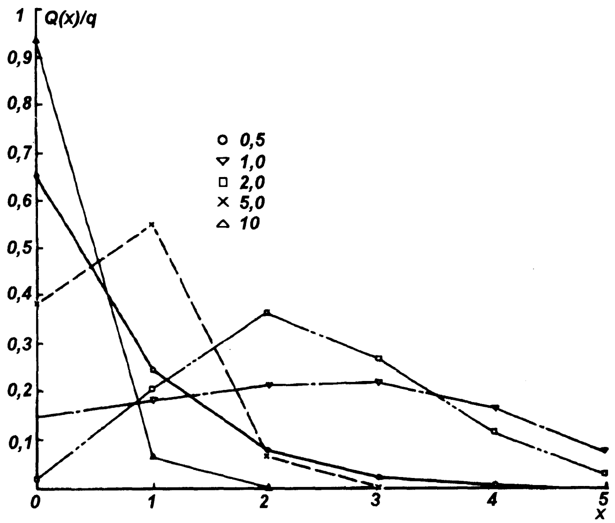

Рис. 1

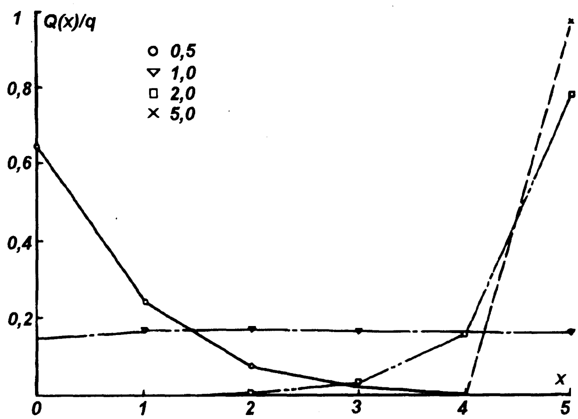

Рис. 2

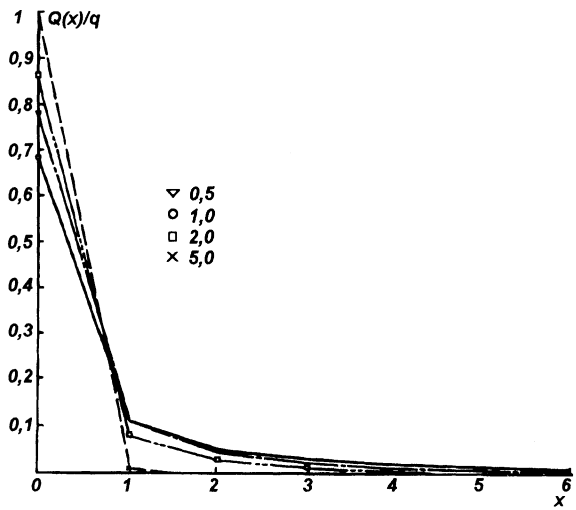

Рис. 3

Рис. 1. Вероятность того, что задержка равна x для требования, которое получает обслуживание. Дисциплина d^*, емкость буфера $k = 5$, интенсивности пуассоновского трафика $\lambda = 0,5; 1; 2; 5; 10$

Рис. 2. Вероятность того, что задержка равна x для требования, которое получает обслуживание. Дисциплина FIFO, емкость буфера $k = 5$, интенсивности пуассоновского трафика $\lambda = 0,5; 1; 2; 5$

Рис. 3. Вероятность того, что задержка равна x для требования, которое получает обслуживание. Дисциплина LIFO, емкость буфера $k = 5$, интенсивности пуассоновского трафика $\lambda = 0,5; 1; 2; 5$

сящиеся к одному и тому же значению λ, соединены прямыми линиями (аналогичное замечание нужно сделать по поводу рис. 5–7 и 10).

При дисциплине d^* условное распределение $Q(x)/q$ концентрируется около нуля при малых λ; оно приблизительно равномерное при средних λ ($\lambda \approx 1$) и оно опять-таки концентрируется около нуля при больших λ.

При дисциплине d_F распределение $Q(x)/q$ ведет себя подобно тому, как при d^* в диапазоне малых и средних λ, но при больших λ оно концентрируется около k.

При дисциплине d_L распределение $Q(x)/q$ имеет пик при $x = 0$ и монотонно падает при $x \to \infty$ при всех λ. Высота пика при $x = 0$ уменьшается от $Q(x)/q = 1$ при $\lambda = 0$ до некоторой величины в диапазоне средних λ, а при увеличении λ снова стремится к единице.

Вероятность того, что требование пропадает, как функция λ показана на рис. 4 для $k = 5$ и всех дисциплин d^*, d_F и d_L.

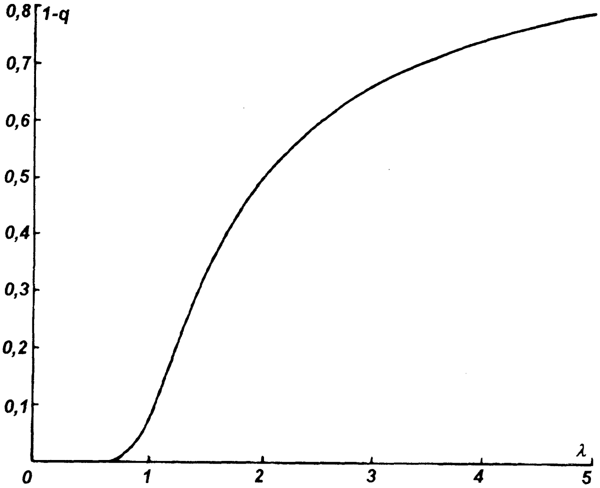

Рис. 4

Рис. 4. Вероятность потери требования $1 - q$ как функция интенсивности пуассоновского трафика λ. Дисциплины FIFO, LIFO, d^*, емкость буфера $k = 5$

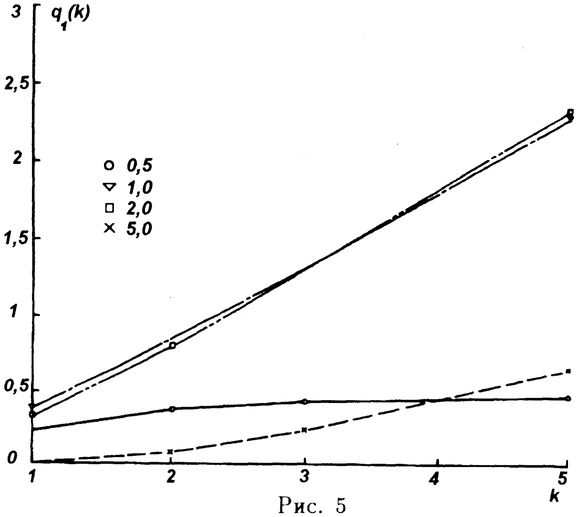

Рис. 5

Рис. 5. Средняя задержка требования, которое получает обслуживание, как функция емкости буфера k. Дисциплина d^*, интенсивности пуассоновского трафика $\lambda = 0,5; 1; 2; 5$

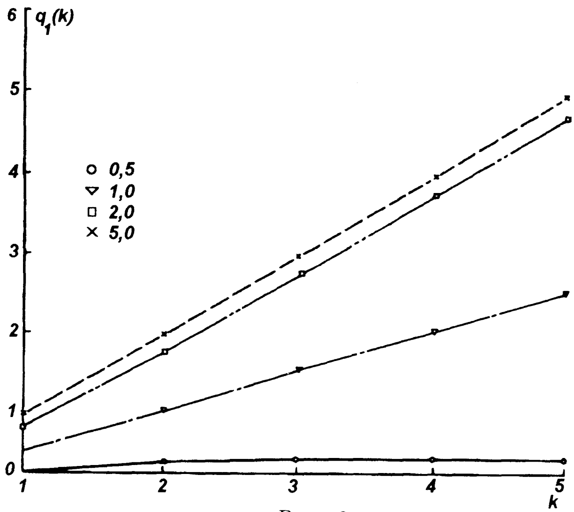

Рис. 6

Рис. 6. Средняя задержка требования, которое получает обслуживание, как функция емкости буфера k. Дисциплина FIFO, интенсивности пуассоновского трафика $\lambda = 0,5; 1; 2; 5$

Условная средняя задержка обслуживания на одно требование $q_1 = q_1(k)$ как функция k, $1 \leqslant k \leqslant 5$, показана на рис. 5 для дисциплины d^*, на рис. 6 – для дисциплины d_F и на рис. 7 – для дисциплины d_L, где $\lambda = 0,5;\ 1;\ 2;\ 5$.

Функция $q_1(k)$ при d^* близка к функции $q_1(k)$ при d_F, когда $\lambda < 1$. Функция $q_1(k)$ при d_F растет почти линейно с ростом k при $\lambda > 1$. Функция $q_1(k)$ при d^* приближается к константе, когда $\lambda > 1$, и эта константа падает с ростом λ. Функции $q_1(k)$ при дисциплинах d_F и d_L стремятся к одному и тому же пределу с ростом k и при $\lambda < 1$.

Зависимость q_1 от λ для $k = 5$ показана на рис. 8 при всех дисциплинах d^*, d_F и d_L.

Средняя задержка q_1 при d_L меньше, чем при d^*, а последняя меньше, чем при d_F, когда $\lambda > 0$. Средняя задержка q_1 при d_F стремится к k, когда $\lambda \to \infty$, а средние задержки q_1 при d^* и d_L стремятся к нулю, когда $\lambda \to \infty$.

Зависимость σ^2 от λ при всех дисциплинах d^*, d_F и d_L

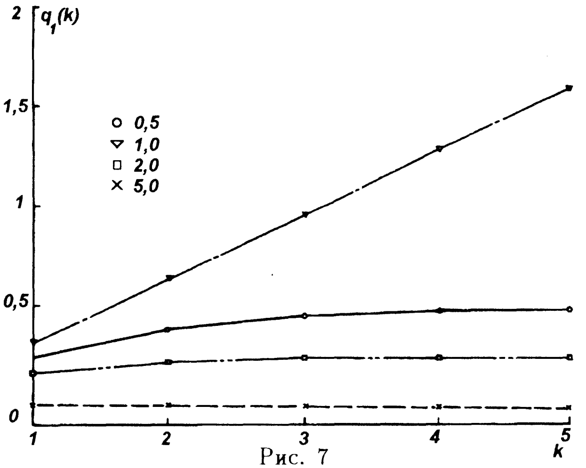

Рис. 7

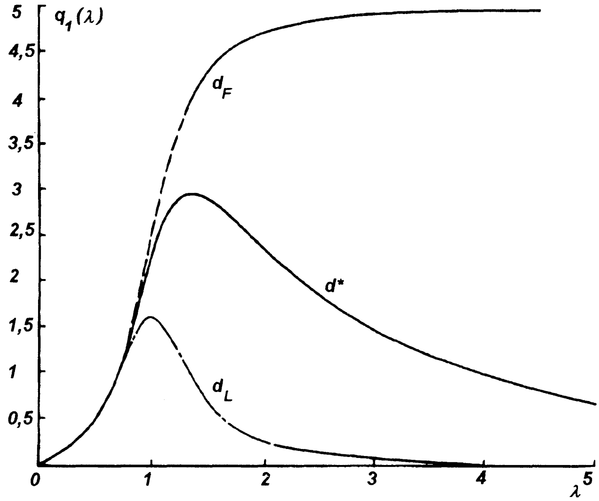

Рис. 8

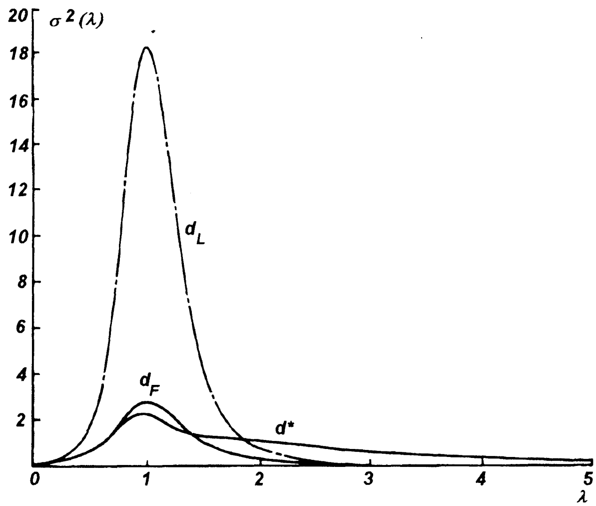

Рис. 9

Рис. 7. Средняя задержка требования, которое получает обслуживание, как функция емкости буфера k. Дисциплина LIFO, интенсивности пуассоновского трафика $\lambda = 0,5; 1; 2; 5$

Рис. 8. Средняя задержка требования, которое получает обслуживание, как функция интенсивности пуассоновского трафика λ. Дисциплины d^*, FIFO, LIFO, емкость буфера $k = 5$

Рис. 9. Дисперсия задержки требования, которое получает обслуживание, как функция интенсивности пуассоновского трафика λ. Дисциплины d^*, FIFO, LIFO, емкость буфера $k = 5$

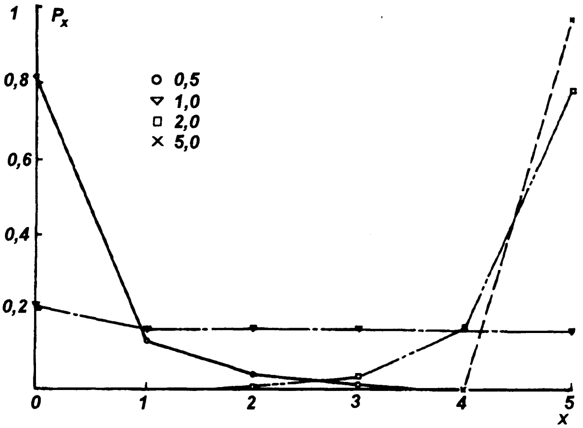

Рис. 10. Вероятность того, что в буфере находятся x требований. Дисциплины d^*, FIFO, интенсивности пуассоновского трафика $\lambda = 0,5; 1; 2; 5$

показана на рис. 9, когда $k = 5$. При d^* и d_F дисперсии σ^2 близки для всех $\lambda > 0$, и они меньше, чем 3. Дисперсия σ^2 при d_F всегда меньше, чем при d_L, когда $\lambda > 0$. Дисперсия σ^2 при d_L значительно выше, чем при d^* и d_F, когда $0,5 < \lambda < 1,6$. В области $0 < \lambda < 1,3$ дисперсия σ^2 при d_F выше, чем при d^*. При d_F, когда $\lambda \to \infty$, первый момент q_1 и второй момент q_2 стремятся к k и k^2 соответственно; таким образом, дисперсия при этом стремится к нулю. При d^* дисперсия σ^2 уменьшается с ростом λ, но она остается выше дисперсий при d_F и d_L.

Стационарные вероятности P_x, $0 \leqslant x \leqslant 5$, показаны на рис. 10 при дисциплинах d^* и d_F для $k = 5$.

Рассчитанные по теореме 4 характеристики для дисциплины d_R на графиках не приведены. Для d_R вероятность $Q(x)/q$ несколько лучше, чем для d^*; функция $q_1(k)$ почти такая же, как для d^*; функция $q_1(\lambda)$ при $\lambda < 1$ почти не отличается от функции, относящейся к d^*, но максимум функции $q_1(\lambda)$, который при $k = 5$ достигается на $\lambda = 1,5$, равен 2,39, а для d^* аналогичный максимум равен 2,9. Для d_R дисперсия $\sigma^2(\lambda)$ идет выше, чем дисперсия для d^* и d_F, но ниже, чем дисперсия для d_L. Для d_R максимум функции $\sigma^2(1)$ при $k = 5$ достигается на $\lambda = 1,1$ и он равен 7,35.

§ 7. Закон сохранения и заключение

Дополнительное качественное понимание поведения задержек требований при различных дисциплинах можно получить с помощью закона сохранения для рассматриваемых СМО с конечным буфером. Чтобы представить этот закон, обозначим через $\Gamma_T(d)$ сумму времен, проведенных в буфере к моменту T всеми требованиями, прибывшими в интервале $[0, T]$ при дисциплине d.

Сумма $\Gamma_T(d)$ связана с ранее использованной суммой $\Delta_T^{(1)}(d)$ равенством

$$\Gamma_T(d) = \Delta_T^{(1)}(d) + \Phi_T(d) + \Theta_T(d),$$

где $\Phi_T(d)$ – сумма времен, проведенных в буфере требованиями, пропавшими на интервале $[0, T]$, а $\Theta_T(d)$ – сумма времен, проведенных в буфере требованиями, которые все еще находятся в буфере к моменту T.

Так же, как в §5, рассмотрим только такие дисциплины обслуживания, которые удовлетворяют условиям (1°) и (2°). Так же, как в §§4 и 5, рассмотрим произвольный входной трафик β_t новых требований, занумерованных, как в §2. Как и ранее, предположим, что в момент $t' = 0$ буфер пустой.

Утверждение 14 (закон сохранения). *При любой заданной выборочной функции входного трафика на* $[0, T]$ *величина* $\Gamma_T(d)$ *не зависит от* $d, \Gamma_T(d) = \Gamma_T$.

Доказательство. Пусть на $[0, T]$ задана выборочная функция входного трафика, и при некотором $t < T$ значение $\Gamma_t(d)$ не зависит от d (а зависит только от выборочной функции). Тогда $\Gamma_{t+1}(d)$ также не зависит то d, так как

$$\Gamma_{t+1}(d) = \Gamma_t(d) + N_t,$$

где N_t – число требований в буфере в момент t'', а согласно утверждению 5 N_t не зависит от d.

Принимая во внимание, что $\Gamma_0(d) = 0$ при любой d, и проводя индукцию по t, мы заканчиваем доказательство. ▲

Если выборочная функция входного трафика не задана, то величины Γ_T, $\Delta_T^{(1)}(d)$, $\Phi_T(d)$, $\Theta_T(d)$ являются случайными. Рассмотрим их математические ожидания. Если при $T \to \infty$ существуют конечные пределы

$$a_\Gamma \triangleq \frac{1}{T} E\,\Gamma_T, \quad a_\Delta(d) \triangleq \frac{1}{T} E\Delta_T^{(1)}(d), \quad a_\Phi(d) \triangleq \frac{1}{T} E\Phi_T(d),$$

то

$$a_r = a_\Delta(d) + a_\Phi(d), \tag{60}$$

так как $\Theta_T(d) \leqslant k$, где k – емкость буфера.

По утверждению 14 величина a_Γ от d не зависит.

Если существует интенсивность входного трафика λ, то $a_\Delta(d)/\lambda$ и $a_\Phi(d)/\lambda$ интерпретируются, соответственно, как средняя задержка требования, которое получает обслуживание при дисциплине d, и среднее время, проводимое в буфере требованием, сброшенным при дисциплине d.

Если дисциплина d такая, что она сбрасывает только новые требования, то $a_\Phi(d) = 0$ и $a_\Delta(d)$ принимает максимально возможное значение a_Γ. Дисциплина d_F является в точности такой, и это есть причина, по которой она наихудшая (теперь в смысле максимума средней задержки требования).

К сожалению, мы не можем столь же просто заключить из (60), что d_L – наилучшая дисциплина. Для этого нужно было бы показать, что $a_\Phi(d)$ минимизируется на d_L.

Дисциплины d^* и d_R являются хорошим компромиссом между d_L и d_F. Как показывают рассмотренные примеры, d^* сильно не отличается от d_L по средней задержке, и в то же время d^* имеет дисперсию задержки, которая меньше дисперсии, даже даваемой дисциплиной d_F в важном диапазоне интенсивностей входного трафика.

Интересной представляется задача поиска дисциплины обслуживания, которая минимизирует среднюю задержку для обслуженных требований при условии заданной дисперсии.

СПИСОК ЛИТЕРАТУРЫ

1. Broadband Aspects of ISDN – Baseline Document (Sinha R., Chief Ed.) T1S1.5/90-001 R1. T1S1 Technical Subcommittee. April, 1990.
2. *Mitra D.* Asymptotically Optimal Design of Congestion Control for High Speed Data Networks // IEEE Trans. Commun. 1992. V. 40. № 2. P. 301–311.
3. *Elwalid A. I., Mitra D.* Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High Speed Networks // IEEE/ACM Trans. Networking. 1993. V. 1. № 3. P. 329–343.
4. *Bertsekas D., Gallager R.* Data networks. Englewoods Cliff: Prentice–Hall, 1992.
5. *Hui J. Y.* Switching and traffic theory for integrated broadband networks. Norwell: Kluwer, 1990.
6. *Wu G. L., Mark J. W.* Discrete Time Analysis of Leaky-Bucket Congestion Control // Computer Networks and ISDN Systems. 1993. V. 26. № 1. P. 79–94.
7. *Sidi M., Liu W. Z., Cidon I., Golal I.* Congestion Control through Input Rate Regulation // IEEE Trans. Commun. 1993. V. 41. № 3. P. 471–477.

8. *Doshi B., Heffes H.* Overload Performance of Several Processor Queueing Disciplines for the M/M/1 Queue // IEEE Trans. Commun. 1986. V. 34. № 6. P. 538–546.
9. *Yin H., Hluchyj H.* Implication of Dropping Packets from the Front of a Queue // Int. Teletraffic Congress. Specialists Seminar. Morristown, 1991.
10. *Doshi B., Lipper.* Comparisons of Service Disciplines in a Queueing System with Delay Dependent Customer Behavior // Applied Probability-Computer Science: The Interface. V. II. R. Disney, T. Ott, Eds. Cambridge: Birkhauser, 1982. P. 269–301.
11. *Shanthikumar J., Sumita U.* Convex Ordering of Sojourn Times in Single-Server Queues: Extremal Properties of FIFO and LIFO Service Disciplines // J. Appl. Prob. 1987. V. 24. № 3. P. 737–748.
12. *Towsley D., Baccelli F.* Comparisons of Service Disciplines in a Tandem Queueing Network with Real-Time Constraints // Operations Research Letters. 1991. V. 10. P. 49–55.
13. *Liu Z., Towsley D.* Stochastic Scheduling in In-Forest Networks // Advances of Applied Probability. 1994. V. 26. № 1. P. 222–241.
14. *Pečinkin A. V., Solovjev A. D.* The analysis and optimization of unichannel queueing systems disciplines // Proc. Third Int. Seminar on Teletraffic Theory. Fundamentals of Teletraffic Theory. Moscow, 1984. P. 342–350.

Поступила в редакцию
29.06.95