Сибирский математический журнал
Март—апрель, 2001. Том 42, № 2

УДК 519.2

ОБ ОПТИМАЛЬНОСТИ ДИСЦИПЛИНЫ FCFS В МНОГОКАНАЛЬНЫХ СИСТЕМАХ И СЕТЯХ ОБСЛУЖИВАНИЯ

С. Г. Фосс, Н. И. Чернова

Аннотация: Доказаны два новых свойства оптимальности дисциплины обслуживания FCFS в некотором классе дисциплин, не зависящих от будущего. Эти свойства применяются для изучения асимптотики распределения некоторых характеристик системы с дисциплиной FCFS. Рассматриваются обобщения полученных результатов. Библиогр. 14.

§ 1. Введение

В многоканальных системах используются различные дисциплины обслуживания вызовов, среди них «первый пришел — первый обслуживается» (FCFS), «последний пришел — первый обслуживается» (LCFS), циклическая, дисциплины с разделением времени и т. д. Поэтому вопросы сравнения дисциплин обслуживания естественны и изучались многими авторами.

Мы ограничимся рассмотрением дисциплин, при которых не допускается разделения времени прибора на обслуживание различных вызовов и прерывания обслуживания вызова. Другими словами, будем предполагать, что в любой момент времени на любом приборе может обслуживаться только один вызов, и если обслуживание какого-то вызова началось, то оно продолжается непрерывно, и после его окончания вызов покидает систему.

Обычно поведение системы обслуживания описывается с помощью последовательности конечномерных случайных векторов $W_n = (W_{n,1}, \dots, W_{n,m})$, $n \geq 0$ (здесь m — число приборов), где $W_{n,j}$ — количество остаточной работы в момент прихода вызова с номером n, т. е. количество времени, требуемое j-му прибору для завершения обслуживания всех вызовов (включая n-й), поступивших на этот прибор к данному моменту времени. Мы интересуемся вопросом о минимизации распределения некоторого функционала $\phi(W_n)$ при каждом фиксированном n. Также представляет интерес задача минимизации функционалов от совместных распределений $\phi(W_n, W_{n+1}, \dots, W_{n+k})$. При этом возникают следующие вопросы: (а) среди каких дисциплин обслуживания вести оптимизацию; (б) как определять класс функционалов $\{\phi\}$?

Наряду с последовательностью W_n рассмотрим последовательность d_n времен ожидания (или времен пребывания) клиентов (вызовов) в системе и изучим вопрос о минимизации распределений некоторых функционалов от этой последовательности.

Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований (коды проектов 99–01–00561, 99–01–00504, 00–15–96178), INTAS (грант 97–1625) и Франко-русского центра им. А. М. Ляпунова (грант 97–07).

В дальнейшем будем предполагать, что времена обслуживания вызовов $\{s_n\}$ образуют последовательность независимых одинаково распределенных случайных величин, не зависящую от других входных характеристик системы.

Нам потребуются следующие соглашения. Будем говорить, что случайная величина ξ стохастически не меньше случайной величины ξ^0, и писать $\xi \geq_{st} \xi^0$, если $\mathbf{P}(\xi \geq x) \geq \mathbf{P}(\xi^0 \geq x)$ при любом x. Случайный вектор $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ стохастически не меньше случайного вектора $\boldsymbol{\xi}^0 = \big(\xi_1^0, \ldots, \xi_n^0\big)$,

$$\boldsymbol{\xi} \geq_{st} \boldsymbol{\xi}^0, \tag{1}$$

если можно задать на некотором вероятностном пространстве их копии $\boldsymbol{\eta} =_D \boldsymbol{\xi}$ и $\boldsymbol{\eta}^0 =_D \boldsymbol{\xi}^0$ так, что $\boldsymbol{\eta} \geq \boldsymbol{\xi}$ покоординатно п. н. Отметим, что из (1) следует, что для любого вектора $x = (x_1, \ldots, x_n)$ справедливо неравенство

$$\mathbf{P}(\boldsymbol{\xi} \geq x) \geq \mathbf{P}(\boldsymbol{\xi}^0 \geq x). \tag{2}$$

Однако обратное неверно: вообще говоря, при $n \geq 2$ из (2) не следует (1).

Рассмотрим две дисциплины обслуживания: FCFS (обозначим ее через T^0) и некоторую другую (обозначим ее через T). Все характеристики системы с дисциплиной FCFS будем помечать верхним индексом 0. По-видимому, первые содержательные и достаточно общие результаты по оптимальности дисциплины FCFS получены в [1]. Во-первых, там показано, что при любом $n = 1, 2, \ldots$ для любой выпуклой по Шуру функции $\phi : \mathbb{R}^m \to \mathbb{R}$ и для любой допустимой дисциплины обслуживания T справедливо неравенство

$$\phi(W_n) \geq_{st} \phi\big(W_n^0\big). \tag{3}$$

В частности,

$$\max_j W_{n,j} \geq_{st} \max_j W_{n,j}^0 \tag{4}$$

(определения выпуклости по Шуру и допустимости см. в следующем параграфе). Далее, пусть v_n — момент окончания обслуживания n-го вызова. В [1] также доказано, что для любого натурального n

$$R(v_1, \ldots, v_n) \geq_{st} R\big(v_1^0, \ldots, v_n^0\big), \tag{5}$$

где для вектора $x = (x_1, \ldots, x_n)$ через $R(x)$ обозначается вектор, полученный из x перестановкой координат в порядке неубывания. Из (5) следует, в частности, соотношение

$$\sum_1^n d_j \geq_{st} \sum_1^n d_j^0. \tag{6}$$

Если допустить, что как система с дисциплиной FCFS, так и система с дисциплиной T эргодичны, то

$$\frac{1}{n}\sum_1^n d_j \to \mathbf{E}d, \quad \frac{1}{n}\sum_1^n d_j^0 \to \mathbf{E}d^0 \text{ п. н.},$$

где d и d^0 — соответствующие стационарные времена ожидания, и (6) влечет неравенство

$$\mathbf{E}\, d \geq \mathbf{E}\, d^0. \tag{7}$$

Ранее вопрос о справедливости неравенства (4) обсуждался различными авторами. Так, в [2, с. 220] утверждалось, что это неравенство имеет место почти наверное. Однако Д. Штойян [3] привел пример, показывающий, что это не так, и высказал предположение, что неравенство может быть справедливо стохастически. В [4] было предложено некоторое доказательство (4), но оказалось, что оно по-существу ошибочное (т. е. основано на неверных аргументах).

В работе [5] для систем вида $GI/GI/m$ при выполнении условия эргодичности $\rho \equiv \frac{\mathbf{E}s_1}{m\tau_1} < 1$ доказано более общее, чем (7), неравенство, но только в случае, когда дисциплина T является *циклической* (см. пример 2 в следующем параграфе): для любой возрастающей выпуклой (вниз) функции f

$$\mathbf{E}f(d) \geq \mathbf{E}f(d^0). \tag{8}$$

В [6] (см. также [7]) рассматривается также случай циклической дисциплины T и приводится пример, показывающий, что, вообще говоря, случайные величины d и d^0 стохастически не сравнимы, т. е. неравенство $d \geq_{st} d^0$ не имеет места.

В [8] рассмотрен несколько иной класс дисциплин, в котором порядок поступления вызовов на обслуживание может быть отличным от порядка поступления вызовов в систему (в частности, в этот класс входит дисциплина «последний пришел — первый обслуживается»), и доказаны естественные аналоги утверждений (3) и (5).

В [9] сформулировано утверждение, которое во введенных нами обозначениях имеет вид: при любом $k = 1, 2, \ldots$

$$\min_{0\leq l\leq k} \max_j (W_{n+l,j} - \tau_{n+l+1})^+ \geq_{st} \min_{0\leq l\leq k} \max_j \bigl(W^0_{n+l,j} - \tau_{n+l+1}\bigr)^+, \tag{9}$$

однако корректное его доказательство проведено лишь при $k = 1$. В [10] предлагается несколько иной (по сравнению с [1]) вариант доказательства неравенства (3). В [11] содержится краткое описание результатов работ [1, 8, 9] и приводятся доказательства основных утверждений в частных случаях. Следует также упомянуть обзорную работу [12], где воспроизводится доказательство утверждений из [1, 11] и рассмотрены иные вопросы.

Как показывают недавние работы (см., например, [13] и др.), с использованием теорем сравнения можно получать новые результаты для системы с дисциплиной FCFS, как-то: утверждения о существовании конечных моментов для стационарного времени ожидания, верхние оценки для остаточных распределений различных характеристик и т. д.

В настоящей работе получены два новых результата. Во-первых, доказывается, что имеет место естественное обобщение неравенства (6):

$$h(d_1, \ldots, d_n) \geq_{st} h\bigl(d^0_1, \ldots, d^0_n\bigr) \tag{10}$$

для любой выпуклой по Шуру функции h. В частности, из (10) следует, что (8) справедливо не только для циклической дисциплины, но и для любой дисциплины, не зависящей от будущего. Имеет место и более общее утверждение, относящееся к совместному распределению нескольких выпуклых по Шуру функций (см. замечание 2).

Интересное применение неравенства (10) получено в [14] (со ссылкой на результат настоящей, неопубликованной к тому времени работы) и будет обсуждено в § 4.

Во-вторых, строится пример, показывающий, что при $k \geq 2$ утверждение (9), вообще говоря, неверно. Далее, при $k = 1$ получено следующее обобщение (9): для любых выпуклых по Шуру функций ϕ_1, ϕ_2 и для любых $n = 1, 2, \ldots$, $x_1, x_2 \in \mathbb{R}$

$$\mathbf{P}(\phi_1(W_n) \geq x_1, \phi_2(W_{n+1}) \geq x_2) \geq \mathbf{P}\bigl(\phi_1\bigl(W^0_n\bigr) \geq x_1, \phi_2\bigl(W^0_{n+1}\bigr) \geq x_2\bigr). \tag{11}$$

Работа содержит четыре параграфа. В § 2 вводятся нужные определения и обозначения и формулируются основные утверждения и их следствия, в § 3 приводятся доказательства. В § 4 рассматриваются возможные обобщения и варианты применения полученных результатов.

§ 2. Основные определения и утверждения

2.1. Описание системы. Рассматривается m-канальная система обслуживания, заданная с помощью двух независимых последовательностей случайных величин $\{\tau_i\}_{i\geq 1}$ и $\{s_i\}_{i\geq 1}$, вектора $W_0 = (W_{0,1}, W_{0,2}, \dots, W_{0,m})$ и некоторого класса дисциплин обслуживания. Здесь τ_1 — момент прихода 1-го вызова, τ_i при $i \geq 2$ — длительность интервала времени между моментами прихода $(i-1)$-го и i-го вызовов, s_i — время обслуживания i-го вызова, координата $W_{0,j} \geq 0$ вектора W_0 — момент времени, начиная с которого j-й канал может начать обслуживание вызовов. Под дисциплиной обслуживания понимается какая-либо (случайная) последовательность $T = \{T_n\}_{n\geq 1}$, в которой T_n равняется номеру станции, где должен обслуживаться вызов с номером n.

Пусть $\langle \Omega, \mathscr{F}, \mathbf{P} \rangle$ — вероятностное пространство, на котором заданы все рассматриваемые случайные величины, $\mathbb{N} = \{1, 2, \dots, n, \dots\}$ — натуральный ряд и $t_n = \sum\limits_{i=1}^{n} \tau_i$ — момент прихода вызова с номером n.

Будем предполагать, что времена обслуживания $\{s_i\}_{i\geq 1}$ независимы, одинаково распределены и не зависят от W_0 и $\{\tau_i\}_{i\geq 1}$.

Последовательность $T = \{T_n\}_{n\geq 1}$ задает дисциплину обслуживания вызовов следующим рекуррентным образом.

Пусть $v_{0,j} = W_{0,j}$ для $1 \leq j \leq m$. Вызов с номером 1 начинает обслуживаться на станции с номером T_1, и если $T_1 = k$, то это обслуживание начинается в момент $u_1 = \max\{v_{0,k}, t_1\}$ и продолжается в течение времени s_1. До момента u_1 обслуживание на станции с номером k не ведется. На множестве $\{T_1 = k\}$ положим $v_{1,j} = v_{0,j}$, если $j \neq k$, и $v_1 = v_{1,k} = u_1 + s_1$.

Пусть для $n \in \mathbb{N}$ определены случайные величины: u_n — момент начала обслуживания n-го вызова, $v_{n,j}$ — момент освобождения j-й станции от вызовов с номерами $1, \dots, n$.

На множестве $\{T_{n+1} = k\}$ до момента $u_{n+1} = \max\{v_{n,k}, t_{n+1}\}$ обслуживание вызовов с номерами $n+1, n+2, \dots$ на станции k не ведется, обслуживание вызова с номером $n+1$ начинается на станции с номером $T_{n+1} = k$ в момент времени u_{n+1}. Положим на множестве $\{T_{n+1} = k\}$

$$v_{n+1,j} = v_{n,j}, \quad \text{если } j \neq k, \quad \text{и} \quad v_{n+1} = v_{n+1,k} = u_{n+1} + s_{n+1}.$$

Заметим, что для любых $i < j$ на множестве $\{T_i = T_j\}$ выполняется неравенство $u_i + s_i \leq u_j$.

Под *допустимой* дисциплиной T мы будем понимать дисциплину обслуживания, «не зависящую от будущего».

Определение 1. Назовем дисциплину обслуживания $T = \{T_n\}_{n\geq 1}$ *допустимой в системе* $\Sigma(W_0, \{\tau_i\}, \{s_i\})$, если для любого $n \geq 1$ и для любых наборов натуральных чисел $\{k_1, \dots, k_n\}$ выполняется равенство

$$\begin{aligned}\mathbf{P}\bigl\{T_1 = k_1, \dots, T_n = k_n \mid W_0; \{\tau_i\}_{i=1}^{\infty}; \{s_i\}_{i=1}^{\infty}\bigr\}& \\ = \mathbf{P}\bigl\{T_1 = k_1, \dots, T_n = k_n \mid W_0; \{\tau_i\}_{i=1}^{\infty}; \{s_i\}_{i=1}^{n-1}\bigr\}.& \quad (12)\end{aligned}$$

Количество работы на j-й станции от момента t_n прихода вызова с номером n до окончания обслуживания всех вызовов с номерами $1, \dots, n$ на этой станции равняется $W_{n,j} = (v_{n,j} - t_n)^+$. Векторы W_n связаны очевидным рекуррентным соотношением $W_n = (W_{n-1} - i\tau_n)^+ + s_n e_{T_n}$, где e_k — вектор, k-я координата которого равна единице, а остальные — нулю, а $i = (1, \dots, 1)$ — вектор, составленный из единиц.

Приведем примеры допустимых дисциплин обслуживания.

Пример 1 (дисциплина FCFS). Положим при $n \in \mathbb{N}$

$$T_n = \min_{1 \le j \le m} \{j : v_{n-1,j} = \min_{1 \le k \le m} v_{n-1,k}\} = \arg\min_{1 \le j \le m} \{v_{n-1,j}\}.$$

Здесь под $\arg\min\limits_{1 \le j \le m} \{v_{n-1,j}\}$ понимается наименьшее j такое, что $v_{n-1,j} \le v_{n-1,k}$ для всех $1 \le k \le m$.

Пример 2 (циклическая дисциплина). Положим $T_{nk+j} = j$ при всех $n, k \ge 0$ и $j = 1, \dots, m$.

Пример 3 (случайная дисциплина). Случайные величины $\{T_n\}$ независимы в совокупности, не зависят от $\{W_0, \{\tau_i\}, \{s_i\}\}$ и имеют одно и то же равномерное на множестве $\{1, 2, \dots, m\}$ распределение: $\mathbf{P}(T_n = j) = 1/m$ при всех $j = 1, \dots, m$.

Пример 4 (дисциплина, введенная в работе [13]). Любой из первых $m - 2$ вызовов выбирает станцию с наименьшим временем ожидания из числа станций, не выбранных предыдущими вызовами. Любому из следующих вызовов (с номерами $m-1, m, \dots$) не разрешается обслуживаться на станциях, выбранных предыдущими $m - 2$ вызовами. Из двух оставшихся станций вызов выбирает станцию с наименьшим временем ожидания.

Пусть

$$A_n = \begin{cases} \{1, \dots, m\} \setminus \{T_1, \dots, T_{n-1}\} & \text{для } 1 \le n \le m-2, \\ \{1, \dots, m\} \setminus \{T_{n-m+2}, \dots, T_{n-1}\} & \text{для } n > m-2 \end{cases}$$

— множество «доступных» для вызова n станций. Тогда $T_n = \arg\min\limits_{j \in A_n} \{v_{n-1,j}\}$.

2.2. Выпуклость по Шуру. Для произвольного вектора $x = (x_1, \dots, x_n)$ через $x_{(1)} \le x_{(2)} \le \dots \le x_{(n)}$ обозначим вектор, полученный из x перестановкой координат в неубывающем порядке. Нам понадобится следующее отношение частичного порядка.

Определение 2. Пусть $x = (x_1, \dots, x_l)$ и $y = (y_1, \dots, y_l)$ — два вектора. Будем писать $x \lhd y$, если

$$\sum_{i=k}^{l} x_{(i)} \le \sum_{i=k}^{l} y_{(i)}$$

при всех $1 \le k \le l$.

Определение 3. Назовем функцию $f : \mathbb{R}^n \to \mathbb{R}$ *выпуклой по Шуру*, если для любых векторов $x \lhd y$ выполнено неравенство $f(x) \le f(y)$.

Нетрудно показать, что класс выпуклых по Шуру функций совпадает со следующим классом H_n.

Определение 4. Функция $h : \mathbb{R}_+^n \to \mathbb{R}$ *принадлежит классу* H_n, если для любого $1 \le j < n$ и для любых $(a_j, a_{j+1}) \lhd (b_j, b_{j+1})$ выполнено неравенство

$$h(a_1, \dots, a_j, a_{j+1}, \dots, a_n) \le h(a_1, \dots, b_j, b_{j+1}, \dots, a_n).$$

2.3. Основные утверждения. Пусть $T^{(1)}$ и $T^{(2)}$ — допустимые дисциплины обслуживания. Для $n \in \mathbb{N}$ обозначим через $d_n^{(i)} = u_n^{(i)} - t_n$ время ожидания вызова с номером n до начала обслуживания при дисциплине $T^{(i)}$. Координата $W_{n,j}^{(i)} = (v_{n,j}^{(i)} - t_n)^+$ вектора $W_n^{(i)} = (W_{n,1}^{(i)}, \dots, W_{n,m}^{(i)})$ есть количество работы на j-й станции с момента t_n прихода вызова с номером n до окончания обслуживания на этой станции всех вызовов с номерами $1, \dots, n$.

Определим два способа сравнения дисциплин обслуживания.

Определение 5. Будем писать $T^{(1)} \preccurlyeq_n T^{(2)}$, если для любой функции $h \in H_n$ выполнено неравенство

$$h\big(d_1^{(1)}, \dots, d_n^{(1)}\big) \le_{st} h\big(d_1^{(2)}, \dots, d_n^{(2)}\big). \tag{13}$$

Определение 6. Будем писать $T^{(1)} \curlyeqprec_n T^{(2)}$, если для любых функций $\phi_1, \phi_2 \in H_m$ при всех $x_1, x_2 \in \mathbb{R}$ выполнено неравенство

$$\mathbf{P}\big(\phi_1\big(W_{n-1}^{(1)}\big) \ge x_1,\ \phi_2\big(W_n^{(1)}\big) \ge x_2\big) \le \mathbf{P}\big(\phi_1\big(W_{n-1}^{(2)}\big) \ge x_1,\ \phi_2\big(W_n^{(2)}\big) \ge x_2\big). \tag{14}$$

Теорема 1. *При любом* $n \in \mathbb{N}$ *дисциплина* FCFS *не хуже любой другой допустимой дисциплины* T*:*

$$T^0 \preccurlyeq_n T, \quad T^0 \curlyeqprec_n T.$$

Замечание 1. Ниже приводится пример, показывающий, что неравенство $T^0 \curlyeqprec_n T$ перестает быть верным, если в (14) рассмотреть совместное распределение не двух, а большего числа случайных векторов. Так, например, неравенство

$$\mathbf{P}\big(\phi_1\big(W_{n-2}^0\big) \ge x_1, \phi_2\big(W_{n-1}^0\big) \ge x_2, \phi_3\big(W_n^0\big) \ge x_3\big) \\ \le \mathbf{P}(\phi_1(W_{n-2}) \ge x_1, \phi_2(W_{n-1}) \ge x_2, \phi_3(W_n) \ge x_3),$$

аналогичное (14), для FCFS-дисциплины T^0 и произвольной допустимой дисциплины T, вообще говоря, неверно.

Пример 5. Рассмотрим двухканальную систему обслуживания с $\tau_1 = \tau_2 = \tau_3 = 0$, $W_{0,1} = c > 0 = W_{0,2}$. Пусть s_1, s_2, s_3 — независимые и одинаково распределенные случайные величины, принимающие значения $\alpha < \beta$ с вероятностями $1 - p$ и p соответственно. Рассмотрим в такой системе две дисциплины: FCFS-дисциплину T^0 и дисциплину T такую, что $T_1 = 1$, $T_n = \operatorname*{arg\,min}_{j=1,2} \{W_{n-1,j}\}$ при $n = 2, 3$. Здесь $W_n = W_{n-1} + s_n e_{T_n}$.

Возьмем функции $\phi_i(x_1, x_2) = \max(x_1, x_2)$, $i = 1, 2, 3$, из класса H_2 и покажем, что существуют $c, \alpha, \beta, x_1, x_2, x_3$ такие, что

$$P \equiv \mathbf{P}(\max_j W_{1,j} \ge x_1, \max_j W_{2,j} \ge x_2, \max_j W_{3,j} \ge x_3) \\ \le \mathbf{P}\big(\max_j W_{1,j}^0 \ge x_1, \max_j W_{2,j}^0 \ge x_2, \max_j W_{3,j}^0 \ge x_3\big) \equiv P^0. \tag{15}$$

Заметим, что

$$W_1 = \begin{pmatrix} c + s_1 \\ 0 \end{pmatrix}, \quad W_2 = \begin{pmatrix} c + s_1 \\ s_2 \end{pmatrix}, \quad W_1^0 = \begin{pmatrix} c \\ s_1 \end{pmatrix}.$$

Пусть $\beta \geq x_1 = x_2 > c+\alpha$, $2\beta > c+\alpha+\beta \geq x_3 > \alpha+\beta$, $x_3 > c+\beta$. Тогда $\max\limits_j W_{1,j} \geq x_1$ только при $s_1 = \beta$. На множестве $\{s_1=\beta\}$ выполнены равенства $T_2 = 1$, $T_3 = 2$. Неравенство $\max\limits_j W_{2,j} \geq x_2$ верно всегда, третье неравенство имеет вид

$$\{\max_j W_{3,j} = \max(c+\beta, s_2+s_3) \geq x_3\} = \{s_2 = s_3 = \beta\}.$$

Отсюда $P = p^3$.

Вычислим P^0. Первое событие $\bigl\{\max\limits_j W^0_{1,j} = \max(c, s_1) \geq x_1\bigr\}$ выполнено только при $s_1 = \beta > c$. Неравенство $\bigl\{\max\limits_j W^0_{2,j} = \max(c+s_2, s_1) \geq x_2\bigr\}$ верно всегда.

Рассмотрим множество $\{s_1 = \beta, s_2 = \beta\}$. На нем

$$\bigl\{\max_j W^0_{3,j} = \max(c+\beta, \beta+s_3) \geq x_3\bigr\} = \{s_3 = \beta\}.$$

На множестве $\{s_1 = \beta, s_2 = \alpha\}$ имеем

$$\bigl\{\max_j W^0_{3,j} = \max(c+\alpha+s_3, \beta) \geq x_3\bigr\} = \{s_3 = \beta\}.$$

Отсюда

$$P^0 = \mathbf{P}(s_1=\beta, s_2=\beta, s_3=\beta) + \mathbf{P}(s_1=\beta, s_2=\alpha, s_3=\beta) = p^2 > P.$$

Покажем, что в условиях данного примера неравенство (9), вообще говоря, не выполнено. Возьмем $x = x_1 = x_2 = \beta$, $x_3 = c+\alpha+\beta$, $\tau_1 = \tau_2 = \tau_3 = 0$ и $\tau_4 = c+\alpha$. Тогда

$$\{\max_j (W_{3,j} - \tau_4)^+ \geq x\} = \{\max_j W_{3,j} \geq x_3\}$$

и по-прежнему

$$\mathbf{P}(\min_{1\leq i\leq 3} \max_j (W_{i,j} - \tau_{i+1})^+ \geq x) = p^3 < p^2 = \mathbf{P}\bigl(\min_{1\leq i\leq 3} \max_j \bigl(W^0_{i,j} - \tau_{i+1}\bigr)^+ \geq x\bigr).$$

§ 3. Доказательство теоремы 1

Нам понадобится следующее утверждение, являющееся аналогом леммы 3.1 работы [11].

Лемма 1. *Пусть $W_0 = (W_{0,1}, \dots, W_{0,m})$ — неслучайный вектор и $W_{0,r} \geq W_{0,l}$, где $r \neq l$ — фиксированные числа. Пусть T — допустимая дисциплина обслуживания в системе $\Sigma(W_0, \{\tau_i\}, \{s_i\})$ такая, что $T_1 = r$, $T_2 = l$. Тогда*

(а) *существуют независимые случайные величины s'_1, s'_2, одинаково распределенные с s_1, не зависящие от $\{\tau_i\}$, $\{s_i\}_{i>2}$ и такие, что $s'_1 + s'_2 = s_1 + s_2$ п. н.;*

(б) *существует дисциплина T', допустимая в системе $\Sigma(W_0, \{\tau_i\}, \{s'_1, s'_2, s_3, s_4, \dots\})$, такая, что $T'_1 = l, T'_2 = r$, и для всех $i \geq 3$ выполнены неравенства*

$$W'_{i-1} \lhd W_{i-1} \quad \text{п. н.,} \tag{16}$$

$$d'_i = \bigl(W'_{i-1,T'_i} - \tau_i\bigr)^+ \leq d_i = \bigl(W_{i-1,T_i} - \tau_i\bigr)^+ \quad \text{п. н.,} \tag{17}$$

$$d'_1 \leq d_1, \quad (d'_1, d'_2) \lhd (d_1, d_2) \quad \text{п. н.,} \tag{18}$$

а также

$$\mathbf{P}\bigl(\phi_1\bigl(W'_{i-2}\bigr) \geq x_1,\ \phi_2\bigl(W'_{i-1}\bigr) \geq x_2\bigr) \leq \mathbf{P}\bigl(\phi_1\bigl(W_{i-2}\bigr) \geq x_1,\ \phi_2\bigl(W_{i-1}\bigr) \geq x_2\bigr) \tag{19}$$

для любых функций $\phi_1, \phi_2 \in H_m$ при всех $x_1, x_2 \in \mathbb{R}$.

Доказательство. Положим $A = \{W_{0,r} - \tau_1 - \tau_2 < 0\}$. Легко видеть, что для случайных величин

$$s'_1 = s_1 I(A) + s_2 I(\overline{A}); \quad s'_2 = s_2 I(A) + s_1 I(\overline{A})$$

выполнено утверждение (a) леммы 1. Действительно, независимость и одинаковая распределенность s'_j, $j = 1, 2$, а также их независимость от τ_1, τ_2 следуют из одинаковой распределенности $\{s_i\}$ и их независимости от $\{\tau_i\}$: по формуле полной вероятности

$$\mathbf{P}(s'_1 \in B) = \mathbf{P}(s_1 \in B, A) + \mathbf{P}(s_2 \in B, \overline{A}) \\ = \mathbf{P}(s_1 \in B)\mathbf{P}(A) + \mathbf{P}(s_1 \in B)\mathbf{P}(\overline{A}) = \mathbf{P}(s_1 \in B).$$

Остальные утверждения п. (a) доказываются аналогично.

В [1] (см. также [11, лемма 3.1]) доказано, что неравенство (16) выполнено для следующей дисциплины T' в системе с временами обслуживания $\{s'_1, s'_2, s_3, s_4, \dots\}$: полагаем $T'_1 = l$, $T'_2 = r$ п. н. При $i > 2$ дисциплину T' задаем так:

а) на множестве $\overline{A}$ полагаем $T'_i = T_i$ п. н.;

б) на множестве A полагаем

$$T'_i = \begin{cases} r, & \text{если } T_i = l, \\ l, & \text{если } T_i = r, \\ T_i & \text{в остальных случаях.} \end{cases}$$

Для удобства читателей приведем доказательство (16) из [11], доказав одновременно (17).

Доказательство неравенств (16), (17). Заметим, во-первых, что на множестве $\overline{A}$

$$\begin{gathered} W_{2,l} = (W_{0,l} - \tau_1 - \tau_2)^+ + s_2 \geq W'_{2,l} = ((W_{0,l} - \tau_1)^+ + s_2 - \tau_2)^+, \\ W_{2,r} = W_{0,r} - \tau_1 - \tau_2 + s_1 = W'_{2,r} \quad \text{п. н.} \end{gathered} \tag{20}$$

и $W_{2,j} = W'_{2,j}$ п. н. при всех остальных j. На множестве A имеем

$$W_{2,l} = s_2 = W'_{2,r}, \quad W_{2,r} = ((W_{0,r} - \tau_1)^+ + s_1 - \tau_2)^+ = W'_{2,l} \quad \text{п. н.} \tag{21}$$

и $W_{2,j} = W'_{2,j}$ п. н. при всех остальных j. Легко видеть, что равенства и неравенства (20), (21) сохраняются для соответствующих координат векторов W_i, W'_i при всех $i \geq 2$. Поэтому верно неравенство (16). Более того, $W_{i-1,T_i} \geq W'_{i-1,T'_i}$ п. н. при всех $i > 2$, что влечет (17).

Доказательство неравенства (18). Заметим, что

$$d_1 = \max\{d_1, d_2\} = (W_{0,r} - \tau_1)^+ \geq \max\{d'_1, d'_2\} \\ = \max\{(W_{0,l} - \tau_1)^+, (W_{0,r} - \tau_1 - \tau_2)^+\} \text{ п. н.}$$

На множестве $\overline{A}$

$$d_1 + d_2 = W_{0,r} - \tau_1 + (W_{0,l} - \tau_1 - \tau_2)^+ \geq W_{0,r} - \tau_1 + (W_{0,l} - \tau_1)^+ - \tau_2 = d'_1 + d'_2 \quad \text{п. н.}$$

На множестве A имеем $d_2 = 0 = d'_2$, следовательно, $d_1 + d_2 = d_1 \geq d'_1 = d'_1 + d'_2$ п. н. Поэтому $(d'_1, d'_2) \lhd (d_1, d_2)$ п. н. согласно определению 2, что доказывает неравенство (18).

Доказательство неравенства (19). Заметим, что при $i > 3$ неравенство (19) непосредственно следует из (16). Докажем (19) при $i = 3$. По формуле полной вероятности достаточно доказать (19) на событиях A и $\overline{A}$ в отдельности.

На множестве A

$$\begin{pmatrix} W_{1,l} \\ W_{1,r} \end{pmatrix} = \begin{pmatrix} (W_{0,l} - \tau_1)^+ \\ (W_{0,r} - \tau_1)^+ + s_1 \end{pmatrix} \lhd \begin{pmatrix} (W_{0,l} - \tau_1)^+ + s_1 \\ (W_{0,r} - \tau_1)^+ \end{pmatrix} = \begin{pmatrix} W'_{1,l} \\ W'_{1,r} \end{pmatrix} \quad \text{п. н.} \tag{22}$$

и $W'_2 \lhd W_2$ п. н. в силу (16), что доказывает (19) на множестве A.

Вследствие независимости s_1, s_2, s'_1, s'_2 от τ_1, τ_2 достаточно доказать требуемое неравенство при любых фиксированных τ_1, τ_2, удовлетворяющих $\overline{A}$.

Обозначим для краткости $b = W_{0,r} - \tau_1 \geq a = (W_{0,l} - \tau_1)^+ \geq 0$, $\tau = \tau_2$. В силу $\overline{A}$ имеем $b > \tau$. Пусть $a < b$. Поскольку векторы W_i и W'_i, $i = 1, 2$, отличаются только двумя координатами, докажем неравенство (19) для функций $\phi_1, \phi_2 \in H_2$.

Координаты векторов W_i, W'_i имеют вид

$$\begin{pmatrix} W_{1,r} \\ W_{1,l} \end{pmatrix} = \begin{pmatrix} b + s_1 \\ a \end{pmatrix}, \quad \begin{pmatrix} W_{2,r} \\ W_{2,l} \end{pmatrix} = \begin{pmatrix} b + s_1 - \tau \\ (a - \tau)^+ + s_2 \end{pmatrix},$$

$$\begin{pmatrix} W'_{1,r} \\ W'_{1,l} \end{pmatrix} = \begin{pmatrix} b \\ a + s_2 \end{pmatrix}, \quad \begin{pmatrix} W'_{2,r} \\ W'_{2,l} \end{pmatrix} = \begin{pmatrix} b + s_1 - \tau \\ (a + s_2 - \tau)^+ \end{pmatrix}.$$

При $x_1, x_2 \in \mathbb{R}$ введем события

$$A_0 \equiv A_0(x_1, x_2, \phi_1, \phi_2) = \left\{ \phi_1 \begin{pmatrix} b + s_1 \\ a \end{pmatrix} \geq x_1,\ \phi_2 \begin{pmatrix} b + s_1 - \tau \\ (a - \tau)^+ + s_2 \end{pmatrix} \geq x_2 \right\},$$

$$A_1 \equiv A_1(x_1, x_2, \phi_1, \phi_2) = \left\{ \phi_1 \begin{pmatrix} b + s_1 \\ a \end{pmatrix} \geq x_1, \quad \phi_2 \begin{pmatrix} b + s_1 - \tau \\ (a + s_2 - \tau)^+ \end{pmatrix} \geq x_2 \right\},$$

$$A_2 \equiv A_2(x_1, x_2, \phi_1, \phi_2) = \left\{ \phi_1 \begin{pmatrix} b \\ a + s_2 \end{pmatrix} \geq x_1,\ \phi_2 \begin{pmatrix} b + s_1 - \tau \\ (a + s_2 - \tau)^+ \end{pmatrix} \geq x_2 \right\}.$$

Для доказательства (19) достаточно показать, что

$$\mathbf{P}(A_0) \geq \mathbf{P}(A_1) \geq \mathbf{P}(A_2). \tag{23}$$

Первое неравенство в (23) очевидно в силу монотонности $\phi_2 \in H_2$ по второму аргументу. Докажем второе неравенство. Обозначим $u_1 = \min\{s_1, s_2\}$, $u_2 = \max\{s_1, s_2\}$ и рассмотрим полную группу событий C_1, C_2, C_3:

$$C_1 = \left\{ x_1 \leq \phi_1 \begin{pmatrix} b + u_1 \\ a \end{pmatrix} \right\}, \quad C_2 = \left\{ x_1 > \phi_1 \begin{pmatrix} b + u_2 \\ a \end{pmatrix} \right\},$$

$$C_3 = \left\{ \phi_1 \begin{pmatrix} b + u_1 \\ a \end{pmatrix} < x_1 \leq \phi_1 \begin{pmatrix} b + u_2 \\ a \end{pmatrix} \right\}.$$

Воспользуемся формулой полной вероятности: при $j = 1, 2$

$$\mathbf{P}(A_j) = \mathbf{P}(A_j C_1) + \mathbf{P}(A_j C_2) + \mathbf{P}(A_j C_3) \equiv P_{j1} + P_{j2} + P_{j3}. \tag{24}$$

Покажем, что $P_{1i} \geq P_{2i}$ при каждом $i = 1, 2, 3$.

1. Заметим, что $C_1 \subseteq \left\{ x_1 \le \phi_1 \begin{pmatrix} b+s_1 \\ a \end{pmatrix} \right\}$. Отсюда

$$\begin{aligned} P_{11} = \mathbf{P} \left\{ \phi_1 \begin{pmatrix} b+s_1 \\ a \end{pmatrix} \ge x_1,\ \phi_2 \begin{pmatrix} b+s_1-\tau \\ (a+s_2-\tau)^+ \end{pmatrix} \ge x_2,\ C_1 \right\} \\ = \mathbf{P} \left\{ \phi_2 \begin{pmatrix} b+s_1-\tau \\ (a+s_2-\tau)^+ \end{pmatrix} \ge x_2,\ C_1 \right\} \\ \ge \mathbf{P} \left\{ \phi_1 \begin{pmatrix} b \\ a+s_2 \end{pmatrix} \ge x_1,\ \phi_2 \begin{pmatrix} b+s_1-\tau \\ (a+s_2-\tau)^+ \end{pmatrix} \ge x_2,\ C_1 \right\} = P_{21}. \end{aligned} \tag{25}$$

2. Отметим, что

$$C_2 \subseteq \left\{ x_1 > \phi_1 \begin{pmatrix} b+s_1 \\ a \end{pmatrix} \right\}.$$

Значит,

$$C_2 \cap \left\{ \phi_1 \begin{pmatrix} b+s_1 \\ a \end{pmatrix} \ge x_1 \right\} = \varnothing \text{ и } P_{12} = 0.$$

Кроме того,

$$C_2 \subseteq \left\{ x_1 > \phi_1 \begin{pmatrix} b+s_2 \\ a \end{pmatrix} \right\} \subseteq \left\{ x_1 > \phi_1 \begin{pmatrix} b \\ a+s_2 \end{pmatrix} \right\},$$

поэтому

$$C_2 \cap \left\{ \phi_1 \begin{pmatrix} b \\ a+s_2 \end{pmatrix} \ge x_1 \right\} = \varnothing,$$

и, следовательно, $P_{22} = 0 = P_{12}$.

3. Введем полную группу событий

$$D_0 = \{s_1 = s_2\}, \quad D_1 = \{s_1 < s_2\}, \quad D_2 = \{s_1 > s_2\}.$$

Очевидно, $D_0 \cap C_3 = \varnothing$. Поэтому при $j = 1, 2$

$$P_{j3} = \mathbf{P}(A_j C_3 D_1) + \mathbf{P}(A_j C_3 D_2). \tag{26}$$

Рассмотрим P_{13}. Так как

$$C_3 \cap D_1 = \left\{ s_1 < s_2,\ \phi_1 \begin{pmatrix} b+s_1 \\ a \end{pmatrix} < x_1 \le \phi_1 \begin{pmatrix} b+s_2 \\ a \end{pmatrix} \right\},$$

то $\mathbf{P}(A_1 C_3 D_1) = 0$. Заметим, что

$$C_3 \cap D_2 = \left\{ s_2 < s_1,\ \phi_1 \begin{pmatrix} b+s_2 \\ a \end{pmatrix} < x_1 \le \phi_1 \begin{pmatrix} b+s_1 \\ a \end{pmatrix} \right\} \subseteq \left\{ \phi_1 \begin{pmatrix} b+s_1 \\ a \end{pmatrix} \ge x_1 \right\},$$

поэтому

$$P_{13} = \mathbf{P} \left\{ \phi_2 \begin{pmatrix} b+s_1-\tau \\ (a+s_2-\tau)^+ \end{pmatrix} \ge x_2,\ C_3,\ D_2 \right\}. \tag{27}$$

Рассмотрим P_{23}. Поскольку

$$C_3 \cap D_2 \subseteq \left\{ \phi_1 \begin{pmatrix} b \\ a+s_2 \end{pmatrix} < x_1 \right\},$$

то $\mathbf{P}(A_2 C_3 D_2) = 0$ и

$$\begin{aligned} P_{23} = \mathbf{P}(A_2 C_3 D_1) \le \mathbf{P} \left\{ \phi_2 \begin{pmatrix} b+s_1-\tau \\ (a+s_2-\tau)^+ \end{pmatrix} \ge x_2,\ C_3,\ D_1 \right\} \\ = \mathbf{P} \left\{ \phi_2 \begin{pmatrix} b+s_2-\tau \\ (a+s_1-\tau)^+ \end{pmatrix} \ge x_2,\ C_3,\ D_2 \right\}. \end{aligned} \tag{28}$$

Последнее равенство имеет место, так как s_1 и s_2 независимы и одинаково распределены.

Отметим, что если $s_2 < s_1$, то по определению 2

$$\begin{pmatrix} b + s_2 - \tau \\ (a + s_1 - \tau)^+ \end{pmatrix} \lhd \begin{pmatrix} b + s_1 - \tau \\ (a + s_2 - \tau)^+ \end{pmatrix}.$$

Поэтому из (27) и (28) следует неравенство $P_{13} \geq P_{23}$.

Тем самым доказано неравенство (23), а вместе с ним неравенство (19) и лемма 1. $\square$

Зафиксируем произвольное $n \in \mathbb{N}$. Будем при $0 \leq k \leq n$ использовать запись $T^{(1)}_{(1,n)} = T_{(1,k)} \cup T^0_{(k+1,n)}$, если $T^{(1)}_i = T_i$ для $1 \leq i \leq k$ и

$$T^{(1)}_i = T^0_i = \arg\min_{1 \leq j \leq m} \{v_{i-1,j}\}$$

для $k+1 \leq i \leq n$, т. е. первые k вызовов обслуживаются в соответствии с дисциплиной T, а вызовы с номерами $k+1 \leq i \leq n$ — в соответствии с дисциплиной FCFS. Вместо $T^0_{(n,n)}$ будем писать T^0_n. Будем для краткости полагать $T_{(1,0)} \cup T^0_{(1,n)} = T^0_{(1,n)}$.

Доказательство теоремы проводится с использованием принципа обратной индукции, предложенного в [1, лемма 2]. Шаг индукции требует доказательства следующего утверждения.

Лемма 2. *Для произвольных целых чисел $1 \leq k \leq n$ и дисциплины $T^{(1)}$ вида $T^{(1)}_{(1,n)} = T_{(1,k)} \cup T^0_{(k+1,n)}$ существует дисциплина $T^{(2)}$ такая, что*

$$T^{(2)}_{(1,k)} = T_{(1,k-1)} \cup T^0_k \quad \text{и} \quad T^{(2)} \preccurlyeq_n T^{(1)}, \quad T^{(2)} \preccurlyeq_n T^{(1)}.$$

Доказательство. При $k = n$ утверждение леммы очевидно. Действительно, $T^{(1)}_{(1,n)} = T_{(1,n)}$ и $T^{(2)}_{(1,n)} = T_{(1,n-1)} \cup T^0_n$. По определению T^0 имеем $d^{(2)}_n \leq d_n$ п. н., так что $T^{(2)} \preccurlyeq_n T$ в силу монотонности $h \in H_n$ по последнему аргументу.

Векторы $W^{(2)}_n$ и W_n отличаются только r-й и l-й координатами в случае $T_n = r \neq l = T^{(0)}_n$ (и совпадают при $T_n = T^{(0)}_n$), причем

$$W^{(2)}_n \lhd W_n \quad \text{п. н.},$$

поэтому $\phi_2\big(W^{(2)}_n\big) \leq \phi_2\left(W_n\right)$ п. н. для любой $\phi_2 \in H_m$ и, следовательно, $T^{(2)} \preccurlyeq_n T$.

При любом фиксированном k, $1 \leq k \leq n-1$, на множестве элементарных исходов $\big\{T_k = T^0_k\big\}$ дисциплина $T^{(2)} \equiv T^{(1)}$ уже имеет вид $T^{(2)}_{(1,n)} = T_{(1,k-1)} \cup T^0_{(k,n)}$.

Пусть $T_k = r$, а $l = \arg\min_j \{v_{k-1,j}\} = \arg\min_j \{W_{k-1,j}\} \neq r$. По формуле полной вероятности достаточно доказать утверждение леммы на множестве $\{r \neq l\}$.

Будем далее считать, что $T_k = r \neq l = T^0_k$.

По условию дисциплина $T^{(1)}$ на $(k+1)$-м шаге совпадает с FCFS, т. е. $T^{(1)}_{k+1} = T^0_{k+1} = \arg\min_j \{W_{k,j}\}$. Поскольку $W_{k-1,l} \leq W_{k-1,j}$ для всех j и $W_{k,j} = (W_{k-1,j} - \tau_k)^+ + s_k I(j = r)$, то $T^{(1)}_{k+1} = j \neq l$ возможно лишь при $W_{k,l} = W_{k,j} = 0$.

Поэтому при любом фиксированном j на множестве $\{T_k = r \neq l = T_k^0\} \cap \{T_{k+1}^0 = j \neq l\}$ можно вместо $T^{(1)}$ рассмотреть дисциплину $T^{(3)}$, полагая $T_{k+i}^{(3)} = l$, если $T_{k+i}^{(1)} = j$ и $T_{k+i}^{(3)} = j$, если $T_{k+i}^{(1)} = l$ при всех $i \geq 1$. Очевидно, $d_{k+i}^{(1)} = d_{k+i}^{(3)}$, $W_{k+i}^{(1)} = R(W_{k+i}^{(3)})$ п. н.

Другими словами, можно заранее предполагать, что на множестве $\{T_k = r \neq l = T_k^0\}$ выполняется равенство $T_{k+1}^{(1)} = l$.

Рассмотрим $1 \leq k \leq n-1$ и воспользуемся леммой 1, взяв в качестве W_0 вектор W_{k-1}, а в качестве последовательностей $\{\tau_i\}_{i\geq 1}$ и $\{s_i\}_{i\geq 1}$ — последовательности $\{\tau_i\}_{i\geq k}$ и $\{s_i\}_{i\geq k}$. Точнее, применим лемму 1 на множестве $\{W_{k-1} = \mathrm{const}\}$.

Построим согласно лемме 1 дисциплину T', совпадающую с T для первых $k-1$ вызовов и с T^0 для k-го вызова. В силу неравенств (17), (18) для любой функции $h \in H_n$

$$h\big(d_1^{(1)}, \dots, d_{k-1}^{(1)}, d_k', d_{k+1}', \dots, d_n'\big) \leq h\big(d_1^{(1)}, \dots, d_{k-1}^{(1)}, d_k^{(1)}, d_{k+1}^{(1)}, \dots, d_n^{(1)}\big) \quad \text{п. н.} \tag{29}$$

Из неравенства (19) следует, что при $1 \leq k \leq n-1$

$$\mathbf{P}\big(\phi_1\big(W_{n-1}'\big) \geq x_1,\ \phi_2\big(W_n'\big) \geq x_2\big) \leq \mathbf{P}\big(\phi_1\big(W_{n-1}^{(1)}\big) \geq x_1,\ \phi_2\big(W_n^{(1)}\big) \geq x_2\big). \tag{30}$$

Заменим в системе с дисциплиной T' последовательность $s_1, \dots, s_{k-1}, s_k'$, $s_{k+1}', s_{k+2}, \dots$ на $\{s_i\}$ и обозначим через $T^{(2)}$ дисциплину в такой системе. Неравенство (29) станет верно «по распределению», неравенство (30) останется справедливым.

Итак, в первоначальной системе построена дисциплина обслуживания $T^{(2)}$ такая, что $T^{(2)} \preccurlyeq_n T^{(1)}$, $T^{(2)} \preccurlyeq_n T^{(1)}$. Более того, $T_{(1,k)}^{(2)} = T_{(1,k)}' = T_{(1,k-1)}^{(1)} \cup T_k^0$ по построению. $\square$

Пусть $n \in \mathbb{N}$. Докажем с помощью принципа «обратной индукции» следующее утверждение.

(I) Для любого $0 \leq k \leq n-1$ и любой дисциплины T дисциплина $T_{(1,n)}^{(1)} = T_{(1,k)} \cup T_{(k+1,n)}^0$ не хуже, чем T, т. е. выполнены неравенства

$$T^{(1)} \preccurlyeq_n T, \quad T^{(1)} \preccurlyeq_n T. \tag{31}$$

Действительно, при $k = n-1$ утверждение (I) следует из леммы 2. Предположим, что утверждение (I) верно при всех $k \geq k_0 + 1$, меньших n. Докажем, что оно верно и для $k = k_0$.

По предположению индукции для любой дисциплины T дисциплина $T_{(1,n)}^{(1)} = T_{(1,k_0+1)} \cup T_{(k_0+2,n)}^0$ не хуже T в смысле неравенств (31). По лемме 2, найдется дисциплина $T^{(2)}$ такая, что $T_{(1,k_0+1)}^{(2)} = T_{(1,k_0)} \cup T_{k_0+1}^0$ и дисциплина $T^{(2)}$ не хуже, чем $T^{(1)}$. По предположению индукции

$$T_{(1,n)}^{(3)} = T_{(1,k_0+1)}^{(2)} \cup T_{(k_0+2,n)}^0 = T_{(1,k_0)} \cup T_{k_0+1}^0 \cup T_{(k_0+2,n)}^0 \equiv T_{(1,k_0)} \cup T_{(k_0+1,n)}^0$$

не хуже, чем $T^{(2)}$. Итак, (I) верно для $k = k_0$.

Теорема 1 следует из утверждения (I) при $k = 0$. $\square$

Замечание 2. Используя неравенства (17) и (18) леммы 2, нетрудно показать, что наряду с неравенством $T^0 \preccurlyeq_n T$ (см. теорему 1) имеет место и более общее утверждение: для любого $k = 1, 2, \dots$, для любых натуральных $n_1, \dots, n_k$ и для любых функций $h_i \in H_{n_i}$, $i = 1, \dots, k$, справедливо неравенство

$$(h_1(d_1, \dots, d_{n_1}), \dots, h_k(d_1, \dots, d_{n_k})) \geq_{st} \big(h_1\big(d_1^0, \dots, d_{n_1}^0\big), \dots, h_k\big(d_1^0, \dots, d_{n_k}^0\big)\big). \tag{32}$$

§ 4. Обобщения и применения полученных результатов

4.1. Возможные обобщения. Утверждения теоремы 1 естественным образом обобщаются на сети обслуживания с многоканальными станциями (системами), в которых маршрутизация вызовов по станциям проводится независимо от состояния системы (например, на сети джексоновского типа). Доказательства достаточно проводить для сетей с неслучайной маршрутизацией вызовов, а затем воспользоваться формулой полной вероятности.

Ниже мы сформулируем в виде предположения 1 одно из возможных обобщений теоремы 1, не доказывая его, а лишь ограничиваясь рядом рассуждений, поясняющих ход возможного доказательства. Предварительно нам потребуется несколько расширить класс рассматриваемых дисциплин обслуживания, введя так называемые «задержки» начала обслуживания.

Будем теперь понимать под дисциплиной обслуживания T последовательность двумерных векторов:

$$T = \{T_n, \Delta_n\}, \tag{33}$$

где $\Delta_n \geq 0$ — случайные задержки, смысл которых в том, что если $T_1 = k$, то вызов с номером 1 начинает обслуживаться не в момент времени $\max\{v_{0,k}, t_1\}$, а несколько позже, в момент $u_1 = \max\{v_{0,k}, t_1\} + \Delta_1$, т. е. с задержкой Δ_1. При этом, как и ранее, полагаем $v_{1,j} = v_{0,j}$ при $j \neq k$ и $v_1 = v_{1,k} = u_1 + s_1$. Далее, если $v_{n,j}$ и u_n определены, то на множестве $T_{n+1} = k$ вызов с номером $n+1$ начинает обслуживание в момент $u_{n+1} = \max\{v_{n,k}, t_{n+1}\} + \Delta_{n+1}$ (т. е. с задержкой Δ_{n+1}), причем $v_{n+1,j} = v_{n,j}$ при $j \neq k$ и $v_{n+1} = v_{n+1,k} = u_{n+1} + s_{n+1}$.

При этом разумно назвать дисциплину *допустимой*, если для любых натуральных чисел n, k_1, … , k_n и любых измеримых множеств B_1, … , B_n

$$\begin{aligned}\mathbf{P}\big\{T_1 = k_1, \Delta_1 \in B_1, \dots, T_n = k_n, \Delta_n \in B_n \mid W_0; \{\tau_i\}_{i=1}^{\infty}; \{s_i\}_{i=1}^{\infty}\big\}\\ = \mathbf{P}\big\{T_1 = k_1, \Delta_1 \in B_1, \dots, T_n = k_n, \Delta_n \in B_n \mid W_0; \{\tau_i\}_{i=1}^{\infty}; \{s_i\}_{i=1}^{n-1}\big\}.\end{aligned} \tag{34}$$

Так как выпуклые по Шуру функции покоординатно не убывают, дисциплина обслуживания с задержками «хуже» соответствующей дисциплины без задержек, являющейся допустимой в смысле определения 1. Если применить к последней теорему 1, то можно получить, что утверждение теоремы 1 останется справедливым и в классе дисциплин (33), допустимых в смысле (34).

Замечание 3. Можно рассматривать и дальнейшие обобщения понятия дисциплины обслуживания, разрешая вызовам обслуживаться на каждом приборе в порядке, отличном от порядка поступления вызовов в систему (см., например, [8]). При этом можно доказать оптимальность дисциплины FCFS в некотором еще более широком классе дисциплин обслуживания.

Рассмотрим теперь сеть обслуживания, состоящую из конечного числа (скажем, K) станций, каждая из которых представляет собой многоканальную систему. Вызовы поступают извне на станцию с номером 1. После окончания обслуживания на какой-то станции вызов либо переходит на другую станцию, либо покидает сеть по следующему правилу. Для каждой станции $k = 1, \dots, K$ задана последовательность чисел $l_{k,1}, l_{k,2}, \dots$, принимающих значения $1, 2, \dots, K+$

1 и имеющих такой смысл: вызов, закончивший обслуживание на k-й станции j-м по счету, направляется на станцию с номером $l_{k,j}$, если $l_{k,j} \leq K$, и покидает сеть, если $l_{k,j} = K + 1$. Времена обслуживания на каждой станции k образуют последовательность независимых одинаково распределенных случайных величин $\{s_j^{(k)}\}$, не зависящую ни от времен прихода, ни от других времен обслуживания. Будем предполагать, что числа $l_{k,j}$ подобраны так, что никакой вызов не может блуждать бесконечно долго в сети. На каждой станции k рассматриваются дисциплины обслуживания $T^{(k)} = \{T_j^{(k)}, \Delta_j^{(k)}\}$, являющиеся *допустимыми* в следующем смысле: при любых k, j случайные величины $T_{j+1}^{(k)}, \Delta_{j+1}^{(k)}$ могут зависеть только от локальной истории станции, т. е. от предыдущих времен поступления и времен обслуживания вызовов.

Предположим, что в сеть извне поступает лишь конечное число (скажем, n) вызовов, и через z_j, $j = 1, \ldots, n$, обозначим время пребывания в сети j-го вызова. Сформулируем следующее

Предположение 1. *Для любой выпуклой по Шуру функции h справедливо неравенство*

$$h(z_1^0, \ldots, z_n^0) \leq_{st} h(z_1, \ldots, z_n). \tag{35}$$

Здесь под $(z_1^0, \ldots, z_n^0)$ понимается вектор, полученный при использовании дисциплины FCFS на каждой станции.

Как уже говорилось, мы не будем приводить доказательство (35). Отметим лишь, что оно, по сути, может быть основано на тех же идеях, что и доказательство теоремы 1, но является существенно более громоздким (отметим также, что для таких сетей справедливы естественные аналоги неравенства (3)). Проиллюстрируем схему возможного доказательства на примере тандема из двух многоканальных систем обслуживания.

А именно, рассмотрим сеть из двух станций (систем). После окончания обслуживания на первой станции каждый вызов направляется на вторую, а после окончания обслуживания на второй станции покидает сеть. Схема доказательства неравенства (35) в этом частном случае может выглядеть так. Во-первых, если заменить на первой станции дисциплину $T^{(1)}$ на FCFS, то на этой станции все «будет лучше», в том числе вызовы будут и покидать станцию раньше (см. неравенство (5)). Одновременно введем на второй станции дополнительные задержки, чтобы вызовы не могли начать обслуживаться ранее, чем поступали при дисциплине $T^{(1)}$. При этом на второй станции ничего не изменится. Затем заменим на второй станции дисциплину $T^{(2)}$ на FCFS — станет «лучше». Наконец, уберем введенные ранее задержки на входе на вторую станцию.

Замечание 4. Следует пояснить, почему неравенство (35) может иметь место лишь для совместных распределений времен пребывания всех n вызовов, поступающих в систему. Дело в том, что одним из аргументов, используемых в предложенной схеме доказательства, является следующий: на любой станции суммарное количество обслуживаний всех вызовов с номерами $1, \ldots, n$ является постоянным числом, одним и тем же для всех дисциплин обслуживания и на всех элементарных исходах. Если же это условие нарушено, то утверждение, вообще говоря, может быть неверно.

Например, рассмотрим тандем из двух станций, первая из которых является двухканальной, а вторая — одноканальной. На первую станцию поступает два вызова; первый из закончивших обслуживание переходит на вторую станцию, обслуживается там и затем покидает сеть; закончивший обслуживаться последним на первой станции покидает сеть сразу же. Пусть $W_0 = (W_{0,1}, W_{0,2})$ — начальный вектор на первой станции (где $0 \le W_{0,1} < W_{0,2}$), $t_1 \le t_2$ — моменты прихода и s_1, s_2 — времена обслуживания вызовов на первой станции, s — время обслуживания на второй станции. Предположим, что $t_1 = t_2 = W_{0,1}$, $s_1 \le W_{0,2} - W_{0,1}$ и $s > W_{0,2} - W_{0,1}$ п. н. Тогда при дисциплине FCFS $z_1^0 = s_1 + s$ п. н., а если направить 1-й вызов на второй прибор, то $z_1 = s_1 + W_{0,2} - W_{0,1}$. Следовательно, если взять $n = 1$ (т. е. n меньше числа поступающих в сеть вызовов), то неравенство (35) перестает быть верным.

Рассмотрим теперь сеть обслуживания, в которую поступает бесконечное число вызовов (j-й вызов — в момент $\tau_1 + \dots + \tau_j$). Предположим, что эта сеть является эргодичной как при дисциплинах FCFS на всех станциях, так и при дисциплинах $T^{(k)}$, $k = 1\dots, K$. Более того, предположим, что $\mathbf{E}f(z^0)$ и $\mathbf{E}f(z)$ конечны для некоторой неубывающей выпуклой (вниз) функции f, где z^0 и z — соответствующие стационарные времена пребывания вызова в сети. Тогда

$$\frac{1}{n}\sum_{j=1}^{n} f(z_j^0) \to \mathbf{E}f(z^0) \quad \text{и} \quad \frac{1}{n}\sum_{j=1}^{n} f(z_j) \to \mathbf{E}f(z) \quad \text{п. н.}$$

При каждом фиксированном n обозначим через $z_{n,j}^0$, $j = 1, \dots, n$ (соответственно $z_{n,j}$), времена пребывания вызовов во вспомогательной системе, в которую поступают лишь первые n вызовов. Нетрудно показать, что нормированные суммы

$$\frac{1}{n}\sum_{j=1}^{n} f(z_{n,j}^0)$$

также сходятся п. н. к $\mathbf{E}f(z^0)$ (соответственно $\frac{1}{n}\sum\limits_{j=1}^{n} f(z_{n,j})$ сходятся к $\mathbf{E}f(z)$). Поэтому если использовать в неравенстве (35) функции вида

$$h(z_1, \dots, z_n) = \frac{1}{n}\sum_{1}^{n} f(z_j),$$

то можно получить

Следствие 1. *Пусть z — стационарное время пребывания в сети обслуживания и f — неубывающая выпуклая функция. Тогда*

$$\mathbf{E}f(z^0) \le \mathbf{E}f(z).$$

4.2. Применения результатов. Предложим некоторые применения результатов § 2. Символами $GI/GI/m$ будем обозначать многоканальную систему, в которой интервалы τ_n образуют последовательность независимых одинаково распределенных случайных величин, не зависящую от $\{s_n\}$. Положим $a = \mathbf{E}\tau_1 > 0$, $b = \mathbf{E}s_1$ и будем предполагать, что $\rho \equiv \frac{b}{ma} < 1$.

1. В [14] для системы $GI/GI/m$ с дисциплиной FCFS доказано следующее утверждение.

Теорема 2. *Пусть $(m-1)a > b$. Тогда из $\mathbf{E}s_1^c < \infty$ при $3/2 \le c \le 2$ следует $\mathbf{E}(d^0)^{2c-2} < \infty$, из $\mathbf{E}s_1^c < \infty$ при $c > 2$ — $\mathbf{E}(d^0)^c < \infty$.*

При этом в [14] сначала конечность соответствующих математических ожиданий доказывается для дисциплины T из примера 4, а затем используется неравенство $\mathbf{E}d^c \geq \mathbf{E}(d^0)^c$, вытекающее из (10), если рассмотреть функции h вида

$$\frac{1}{n}\sum_{i=1}^{n}(d_i)^c, \quad c \geq 1,$$

и устремить n к бесконечности.

2. Приведем примеры получения верхних оценок на распределение $\mathbf{P}(d^0 > x)$ в системах $GI/GI/m$.

2.1. Воспользуемся сравнением со случайной дисциплиной обслуживания (см. пример 3).

Пусть $x = y + z$, где $0 \leq y = y(x) \to \infty$ и $0 \leq z = z(x) \to \infty$ при $x \to \infty$. Для любого $\alpha > 1$

$$\mathbf{P}(d^0 \geq x) = \mathbf{P}((d^0 - z)^+ \geq y) \leq \frac{\mathbf{E}((d^0 - z)^+)^\alpha}{y^\alpha} \leq \frac{\mathbf{E}((d - z)^+)^\alpha}{y^\alpha}.$$

Здесь d — стационарное время ожидания в многоканальной системе со случайной дисциплиной обслуживания. Так как на всех приборах стационарное остаточное время работы имеет одно и то же распределение и один из приборов выбирается случайно и независимо от состояния системы, то распределение d совпадает с распределением стационарного времени ожидания в одноканальной системе обслуживания $GI/GI/1$, в которой времена между моментами прихода вызовов равны τ_n и времена обслуживания распределены как $\sigma_n = s_n I(T_n = 1)$, т. е.

$$\mathbf{P}(\sigma_n > x) = \frac{1}{m}\mathbf{P}(s_1 > x) \quad \text{при } x \geq 0.$$

Если, к примеру, предположить, что случайные величины s_n имеют так называемое субэкспоненциальное распределение, то, как хорошо известно,

$$\mathbf{P}(d > x) \sim \frac{\rho}{1-\rho}\frac{1}{m}\int_x^\infty \mathbf{P}(s_1 > t)\,dt$$

при $x \to \infty$. Значит,

$$\mathbf{P}(d^0 \geq x) \leq (1 + o(1))\frac{\rho}{1-\rho}\frac{1}{my^\alpha}\int_z^\infty (t - z)^\alpha \mathbf{P}(s_1 > t)\,dt. \tag{36}$$

Минимизацию последнего выражения по y можно проводить для конкретных распределений случайной величины s_1.

2.2. Воспользуемся сравнением с циклической дисциплиной обслуживания (см. пример 2). Справедливы неравенства

$$\mathbf{P}(d^0 \geq x) \leq \mathbf{P}\big(W_1^0 + \dots + W_m^0 \geq mx\big) \leq \mathbf{P}(W_1 + \dots + W_m \geq mx),$$

где $(W_1, \dots, W_m)$ — стационарный вектор виртуальных времен ожидания. Предположим, что времена обслуживания имеют субэкспоненциальное распределение. Тогда можно показать, что

$$\mathbf{P}(W_1 + \dots + W_m \geq mx) \sim m\mathbf{P}(W > mx),$$

где W распределена как стационарное время ожидания в одноканальной системе обслуживания $GI/GI/1$ с временами обслуживания s_n и интервалами между моментами прихода вызовов, распределенными как $\tau_1 + \dots + \tau_m$. Поэтому

$$\mathbf{P}(d^0 \geq x) \leq (1 + o(1))\frac{\rho m}{1-\rho}\int_{mx}^\infty \mathbf{P}(s_1 > t)\,dt. \tag{37}$$

Примеры показывают, что на различных распределениях как первая оценка (36) может быть лучше второй (37), так и вторая может быть лучше первой.

Авторы выражают благодарность рецензенту за полезные замечания.

ЛИТЕРАТУРА

1. *Фосс С. Г.* Об аппроксимации многоканальных систем обслуживания // Сиб. мат. журн. 1980. Т. 21, № 6. С. 132–140.

2. *Боровков А. А.* Вероятностные процессы в теории массового обслуживания. М.: Наука, 1972.

3. *Stoyan D.* A critical remark on a system approximation in queueing theory // Math. Operationsforsh. Statist. 1976. V. 7, N 6. P. 953–956.

4. *Gittins J. C.* A comparison of service disciplines for $GI/G/n$ queues // Math. Operationsforsh. Statist. 1978. V. 9, N 2. P. 255–260.

5. *Wolff R.* An upper bound for multi-channel queues // J. Appl. Probab. 1977. V. 14. P. 884–888.

6. *Whitt W.* On stochastic bounds for the delay distribution in the GI/GI/s queues // Oper. Res. 1981. V. 29, N 3. P. 604–608.

7. *Whitt W.* Deciding which queue to join // Oper. Res. 1986. V. 34, N 1. P. 55–62.

8. *Фосс С. Г.* Сравнение дисциплин обслуживания в многоканальных системах с ожиданием // Сиб. мат. журн. 1981. Т. 22, № 1. С. 190–197.

9. *Фосс С. Г.* Об одном способе оценивания скорости сходимости в теоремах эргодичности и непрерывности для многоканальных систем обслуживания // Тр. Ин-та математики / АН СССР. Сиб. отд-ние. 1985. Т. 5. С. 126–137.

10. *Daley D. J.* Certain optimality properties of the first-come first-served discipline for G/G/s queues // Stochastic Proc. Appl. 1987. V. 25, N 2. P. 301–308.

11. *Foss S. G.* Comparison of service disciplines in $G/GI/m$ queues. INRIA. France, 1989. (Research Report N 1097).

12. *Liu Z., Nain Ph., Towsley D.* Sample path methods in the control of queues // Queueing Systems. 1995. V. 21, N 1–2. P. 293–335.

13. *Sigman K., Scheller-Wolf A.* New bounds for expected delay in FIFO $GI/GI/c$ queues // Queueing Systems. 1997. V. 25. P. 77–96.

14. *Scheller-Wolf A.* Further delay moment results for FIFO multiserver queues // Queueing Systems. 2000. V. 34, N 1–4. P. 387–400.

Статья поступила 1 ноября 2000 г.

Фосс Сергей Георгиевич,
Институт математики им. С. Л. Соболева СО РАН, Новосибирск 630090
foss@math.nsc.ru

Чернова Наталья Исааковна
Новосибирский гос. университет, Новосибирск 630090
cher@nsu.ru