Н. Д. Введенская, Большая система обслуживания с передачей сообщения по нескольким путям,
Пробл. передачи информ., 1998, том 34, выпуск 2, 98–108

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 34 1998 Вып. 2

УДК 621.394:519.2

 Н.Д. Введенская

БОЛЬШАЯ СИСТЕМА ОБСЛУЖИВАНИЯ С ПЕРЕДАЧЕЙ СООБЩЕНИЯ ПО НЕСКОЛЬКИМ ПУТЯМ[1]

Пусть на систему с N приборами поступает пуассоновский поток сообщений интенсивности λN. Поступившее сообщение разбивается на n пакетов, каждый из которых независимо от остальных посылается на случайно выбранный прибор. Время обслуживания пакета распределено экспоненциально со средним значением единица. Показывается, что если $\rho = \lambda n < 1$, то при $N \to \infty$ распределение длин очередей на приборах стремится к распределению длин очередей в системе $M|M|1$ с входным потоком интенсивности ρ. Отсюда выводится, что при таком методе передачи сообщений и при небольших значениях ρ кодирование может ускорить доставку сообщений. Кратко рассматривается также случай, когда пакет состоит из M мини-пакетов, время обслуживания которых распределено экспоненциально со средним значением $1/M$. При $M \to \infty$ распределение времени ожидания в такой системе стремится к распределению времени ожидания в системе $M|D|1$.

§ 1. Введение

Рассматривается система обслуживания, состоящая из N приборов. Поступившее в систему сообщение разбивается на несколько частей (пакетов), каждая из которых независимо от остальных посылается на случайно выбранный прибор. При этом распределения длин очередей на приборах оказываются зависимыми. Однако можно ожидать, что при $N \to \infty$ распределения длин очередей делаются независимыми, и их просто описать аналитически. Сформулируем условия, при которых доказывается асимптотическая независимость очередей.

Пусть на систему поступает пуассоновский поток сообщений интенсивности λN. Поступившее сообщение мгновенно разбивается на n пакетов. Каждый пакет независимо выбирает один из приборов и мгновенно направляется на него. Все приборы равновероятны. Несколько пакетов одного сообщения могут выбрать один и тот же прибор. Время обслуживания пакета не зависит от времени обслуживания других пакетов и распределено экспоненциально, его среднее значение равно единице. Эту систему мы обозначим S_N.

Кроме системы S_N мы рассмотрим систему $S_{N,M}$, которая отличается от S_N тем, что в ней каждый пакет состоит из M мини-пакетов, и время обслуживания мини-пакетов распределено экспоненциально со средним $1/M$. Мини-пакеты одного пакета направляются на один прибор и обслуживаются друг за другом, так что время обслуживания пакета распределено по закону Эрланга M-й степени со средним единица.

[1]Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований (номер проекта 96-01-0050).

Остановимся сначала на системе S_N.

Мы показываем, что при $\rho = \lambda n < 1$ система S_N описывается эргодической цепью Маркова, так что существует стационарное распределение вероятностей для состояний S_N. Пусть r_i, $i = 0, 1, \ldots,$ – число приборов, в которых длина очереди равна i. Положим

$$h_i = r_i/N, \quad u_i = \sum_{j=i}^{\infty} h_j. \tag{1}$$

Здесь u_i – доля приборов, у которых очереди не меньше, чем i (пакет, который обслуживается, также учитывается). Пусть $\mathbf{E}_N h_i$, $\mathbf{E}_N u_i$ – средние значения случайных величин h_i, u_i относительно стационарного распределения в системе S_N. Показывается, что при $\rho < 1$

$$\lim_{N\to\infty} \mathbf{E}_N h_i = (1-\rho)\rho^i, \quad \lim_{N\to\infty} \mathbf{E}_N u_i = \rho^i. \tag{2}$$

Из соотношений (2) вытекает, что для системы S_N справедливо сформулированное в [1] утверждение: кодирование может ускорить передачу сообщения. Именно, пусть сообщение, состоящее первоначально из k пакетов, закодировано в сообщение из n пакетов так, что его можно восстановить по любым k пакетам из n (код с максимальным расстоянием). Временем передачи пакета назовем интервал времени между моментом поступления пакета в систему и моментом окончания его обслуживания. Временем передачи $T(n,k,\lambda)$ закодированного n-пакетного сообщения назовем время передачи последнего из переданных быстрее всего k пакетов. Конечно, передача n вместо k пакетов увеличивает нагрузку на систему в n/k раз. Однако, используя формулы (2), можно показать, что при $N \to \infty$ и $\lambda < \lambda_{\text{cr}}(n,k)$ среднее время передачи n-пакетного сообщения $T(n,k,\lambda)$ меньше, чем среднее время передачи k-пакетного $(k=n)$ сообщения $T(k,k,\lambda)$.

Аналогичные утверждения справедливы для системы $S_{N,M}$.

Отметим, что при небольших значениях N ускорение передачи сообщения при кодировании показано методом моделирования в [2–5], для конечной системы ускорение передачи сообщений имеет место при $\lambda < \lambda_{\text{cr}}(k,n,N)$ (в [4–5] время передачи пакета фиксировано).

Возможность исследования систем S_N и $S_{N,M}$ при $N \to \infty$ связана с их симметричностью. Используемый ниже подход аналогичен методу среднего поля статистической физики. Этот метод неоднократно применялся при изучении систем обслуживания (см., например, [6–10]). Приводимые ниже доказательства аналогичны доказательствам в [9].

Следуя [9], обозначим через $\overline{\mathcal{U}}$ пространство последовательностей

$$u = \{u_i\}_{i=0}^{\infty}, \quad 1 = u_0 \geqslant u_1 \geqslant u_2 \geqslant \ldots \geqslant 0, \tag{3}$$

с нормой

$$\rho(u^{(1)}, u^{(2)}) = \sup_{i\geqslant 1} |u_i^{(1)} - u_i^{(2)}|/i;$$

здесь $\overline{\mathcal{U}}$ замкнуто относительно этой нормы. Положим

$$\begin{gathered}
\mathcal{U} \subset \overline{\mathcal{U}}, \quad \mathcal{U} = \left\{ u \in \overline{\mathcal{U}},\ \sum_{i=1}^{\infty} u_i < \infty \right\}, \\
v_i = v_i(u) = \sum_{k=i}^{\infty} u_k, \quad u \in \mathcal{U}, \qquad (4) \\
\mathcal{U}_N \subset \mathcal{U}, \quad \mathcal{U}_N = \{u \in \mathcal{U},\ u_i = r_i/N\},
\end{gathered}$$

где r_i – целые неотрицательные числа, $r_i \leqslant N$.

Работа системы S_N описывается однородной по времени цепью Маркова, состояниями которой являются векторы $(x_1, x_2, \dots, x_N)$, где x_j, $j = 1, \dots, N$, – число пакетов в очереди на j-м приборе. От этой цепи Маркова мы перейдем к фактор-цепи, состояниями которой являются векторы $u = (u_0, u_1, \dots) \in \mathcal{U}_N$, заданные в (1).

Производящий оператор фактор-цепи Маркова имеет вид

$$A_N f(u) = N \sum_{i=1}^{\infty} \left[f\left(u - \frac{e_i}{N}\right) - f(u) \right] (u_i - u_{i+1}) +$$
$$+ \lambda N \sum_{i_1=1}^{\infty} \dots \sum_{i_n=1}^{\infty} \left[f\left(u + \frac{e_{i_1} + \dots + e_{i_n}}{N}\right) - f(u) \right] \times$$
$$\times (u_{i_1-1} - u_{i_1}) \dots (u_{i_n-1} - u_{i_n}) + O\left(\frac{1}{N}\right). \tag{5}$$

Здесь $u \in \mathcal{U}_N$, e_i – вектор, у которого все компоненты равны нулю, а i-я компонента равна единице.

Поясним формулу (5). Первая сумма соответствует тому, что окончилось обслуживание пакета на одном из приборов, очередь в котором имела длину i. Вторая сумма соответствует тому, что вновь поступившие в систему пакеты попали на приборы с длинами очередей $i_\ell - 1$, $\ell = 1, \dots, n$. Член $O(1/N)$ – поправка ко второй сумме – соответствует тому, что когда ℓ пакетов попадают на один и тот же прибор с очередью длины $i - 1$, то длина очереди на этом приборе становится $i + \ell - 1$. Например, при $n = 2$ поправка имеет вид

$$\lambda \sum_{i=1}^{\infty} \left[\left(f\left(u + \frac{e_i + e_{i+1}}{N}\right) - f(u) \right) - \left(f\left(u + \frac{2e_i}{N}\right) - f(u) \right) \right] (u_{i-1} - u_i).$$

Наша цель – показать, что при $N \to \infty$ эволюция значений u_i становится детерминированной, и фактор-цепь Маркова сходится к динамической системе, эволюция которой описывается системой дифференциально-разностных уравнений вида

$$\frac{du_i(t)}{dt} = \rho[u_{i-1}(t) - u_i(t)] - u_i(t) + u_{i+1}(t), \quad i = 1, 2, \dots, \tag{6}$$

с начально-краевыми условиями

$$u_0(t) = 1, \tag{7}$$
$$u_i(0) = g_i. \tag{8}$$

В дальнейшем рассматриваются свойства решений $u^\varepsilon = \{u_i^\varepsilon\}_{i=0}^{\infty}$ более общей краевой задачи

$$\frac{du_i^\varepsilon(t)}{dt} = \rho \sum_{\ell=1}^{n} \varepsilon_\ell [u_{i-\ell}^\varepsilon(t) - u_i^\varepsilon(t)] - u_i^\varepsilon(t) + u_{i+1}^\varepsilon(t), \quad i = 1, 2, \dots, \tag{9}$$

$$u_m^\varepsilon(t) = 1, \quad m \leqslant 0, \tag{10}$$

$$\varepsilon_1 = 1 - \sum_{\ell=2}^{n} \ell \varepsilon_\ell, \quad \sum_{\ell=2}^{n} \ell \varepsilon_\ell < 1; \quad \varepsilon_1 \to 1, \quad \varepsilon_\ell = O(1/N^{\ell-1}), \quad \ell > 1, \quad N \to \infty.$$

Решение этой задачи иногда будет обозначаться через $u^\varepsilon(t, g)$. При $\varepsilon_\ell \equiv 0$, $\ell = 2, \dots, n$, задача (9), (10), (8) переходит в задачу (6)–(8).

Рассматривая ниже систему $S_{N,M}$, мы обозначаем через u_i долю приборов в системе, у которых очереди состоят не менее чем из i мини-пакетов. Работа системы $S_{N,M}$ тоже описывается однородной фактор-цепью Маркова, состояниями которой являются векторы $u = (u_0, u_1, \ldots) \in \mathcal{U}_N$. Мы не будем здесь выписывать производящий оператор этой фактор-цепи, так как его вид мало отличается от (5). Нам надо показать, что при $N \to \infty$ эволюция значений u_i становится детерминированной, и фактор-цепь Маркова сходится к динамической системе, эволюция которой описывается системой дифференциально-разностных уравнений (она будет выписана ниже).

В § 2 рассматриваются свойства решений краевых задач для уравнений (6) и (9). В § 3 доказываются эргодичность фактор-цепи Маркова, описывающей работу системы S_N, и ее сходимость при $N \to \infty$ к динамической системе, заданной уравнениями (6). Аналогичные утверждения формулируются для системы $S_{N,M}$. В § 4 показывается, как при $N \to \infty$ кодирование и рассылка пакетов по разным серверам может ускорить доставку сообщения.

§ 2. О некоторых дифференциально-разностных уравнениях

Рассмотрим сначала краевую задачу для конечной системы дифференциальных уравнений, полученной урезанием системы (9),

$$\frac{du_i^{\varepsilon,k}(t)}{dt} = \rho \sum_{\ell=1}^{n} \varepsilon_\ell [u_{i-\ell}^{\varepsilon,k}(t) - u_i^{\varepsilon,k}(t)] - u_i^{\varepsilon,k}(t) + u_{i+1}^{\varepsilon,k}(t), \quad i = 1, 2, \ldots, k-1, \tag{11}$$

$$u_m^{\varepsilon,k}(t) = 1, \quad m \leqslant 0, \quad u_k^{\varepsilon,k} = \text{const} \geqslant 0, \tag{12}$$

$$u_i^{\varepsilon,k}(0) = g_i, \quad i = 1, 2, \ldots, k-1. \tag{13}$$

Лемма 1. *Пусть $u_i^{\varepsilon,k}(0) = g_i$, $i = 1, \ldots, k-1$, $\varepsilon_\ell \geqslant 0$, удовлетворяют неравенствам* (3). *Тогда $u_i^{\varepsilon,k}(t)$ удовлетворяют этим неравенствам при всех t.*

Доказательство. В силу непрерывной зависимости от начальных условий достаточно рассмотреть случай, когда в (3) стоят строгие неравенства. Предположим противное: пусть $u_i^{\varepsilon,k}(t) > u_{i+1}^{\varepsilon,k}(t)$, $i = 0, \ldots, k-1$, при $t < t_0$, а при $t = t_0$ в этих соотношениях появляются равенства. Тогда найдется такое i, что либо $u_{i-1}^{\varepsilon,k}(t_0) > u_i^{\varepsilon,k}(t_0) = u_{i+1}^{\varepsilon,k}(t_0)$, либо $u_{i-1}^{\varepsilon,k}(t_0) = u_i^{\varepsilon,k}(t_0) > u_{i+1}^{\varepsilon,k}(t_0)$. Из уравнений (11) видно, что в первом случае $du_i^{\varepsilon,k}(t_0)/dt > du_{i+1}^{\varepsilon,k}(t_0)/dt$, а во втором случае $du_{i-1}^{\varepsilon,k}(t_0)/dt > du_i^{\varepsilon,k}(t_0)/dt$. Это противоречит предположению о том, что $u_{i-1}^{\varepsilon,k}(t) > u_i^{\varepsilon,k}(t) > u_{i+1}^{\varepsilon,k}(t)$ при $t < t_0$. ▲

Лемма 2. *Пусть $g_{i,1}$, $g_{i,2}$ удовлетворяют неравенствам* (3), *и $g_{i,1} \geqslant g_{i,2}$, $i = 1, \ldots, k-1$. Тогда для решений $u_{i,1}^{\varepsilon,k}$, $u_{i,2}^{\varepsilon,k}$ задачи* (11)–(13), *отвечающих начальным условиям $g_{i,1}$, $g_{i,2}$ соответственно, справедливы неравенства $u_{i,1}^{\varepsilon,k}(t) \geqslant u_{i,2}^{\varepsilon,k}(t)$, $i = 0, \ldots, k-1$.*

Доказательство. Снова достаточно рассмотреть случай, когда в условиях леммы имеют место строгие неравенства. Пусть эти неравенства выполнены при $t < t_0$, и $u_{i,1}^{\varepsilon,k}(t_0) = u_{i,2}^{\varepsilon,k}(t_0)$ для некоторых i. Пусть j – наибольший из индексов, для которых имеет место равенство. Из уравнений (11) и монотонности $u_{i,1}^{\varepsilon,k}$, $u_{i,2}^{\varepsilon,k}$ по i видно, что $du_{j,1}^{\varepsilon,k}(t_0)/dt > du_{j,2}^{\varepsilon,k}(t_0)/dt$. Это противоречит тому, что $u_{j,1}^{\varepsilon,k}(t) > u_{j,2}^{\varepsilon,k}(t)$ при $t < t_0$. ▲

Лемма 3. *Пусть $g = \{g_i\}_{i=0}^{\infty} \in \overline{\mathcal{U}}$. Тогда задача* (9), (10), (8) *имеет решение $u^\varepsilon = \{u_i^\varepsilon\}_{i=0}^{\infty} \in \overline{\mathcal{U}}$. Это решение единственно и может быть получено как предел при $k \to \infty$ решений $u^{\varepsilon,k} = \{u_i^{\varepsilon,k}\}_{i=0}^{k}$ с начальными значениями $g_i(t)$, $i = 1, \ldots, k-1$.*

Доказательство. Рассмотрим последовательность решений $u^{\varepsilon,k}$, $k = 2, 3, \ldots$, при $u_k^{\varepsilon,k} = 0$. В силу леммы 2 из того, что $u_i^{\varepsilon,k+1} \geqslant 0$ при $i \leqslant k$, вытекает, что $u_i^{\varepsilon,k} \leqslant u_i^{\varepsilon,k+1}$, $i < k$. Кроме того, в силу леммы 1 $u_i^{\varepsilon,k} \leqslant 1$. Следовательно, при любом i существует $u_i^\varepsilon = \lim\limits_{k\to\infty} u_i^{\varepsilon,k}$. Переходя от дифференциальных уравнений к интегральным, проверяем, что u_i, $i = 1, 2, \ldots$, удовлетворяют уравнениям (9). Единственность доказывается методом последовательных приближений Пикара. ▲

Лемма 4. *Пусть* $g = \{g_i\}_{i=1}^{\infty} \in \mathcal{U}$. *Тогда для решений задачи* (9), (10), (8) $u^\varepsilon(t, g) \in \mathcal{U}$ *при* $0 \leqslant t < \infty$.

Доказательство. Из уравнений (9) следует, что для $v_1^\varepsilon(t) = v_1(u^\varepsilon(t)) = \sum\limits_{i=1}^{\infty} u_i^\varepsilon(t)$ имеет место уравнение $dv_1^\varepsilon(t)/dt = \rho - u_1^\varepsilon(t)$, и значит, v_1^ε ограничено при любом $0 \leqslant t < \infty$. ▲

Лемма 5. *При* $\rho < 1$ *задача* (9), (10) *обладает в* $\mathcal{U}$ *стационарным решением* $\nu^\varepsilon = \{\nu_i^\varepsilon\}_{i=0}^{\infty}$, *и*

$$\nu_i^\varepsilon = O\big((\rho_1^\varepsilon)^i\big), \quad i \to \infty,$$

где ρ_1^ε, $\rho < \rho_1^\varepsilon < 1$, *– корень уравнения*

$$y(x) = \rho \sum_{\ell=2}^{n} \varepsilon_\ell [x^{n-\ell} - x^n] - x^n + x^{n+1}. \tag{14}$$

Доказательство. Непосредственно проверяется, что стационарное решение имеет вид

$$\nu_i^\varepsilon = \rho \sum_{\ell=1}^{n} \sum_{k=\ell}^{n} \varepsilon_k \nu_{i-\ell}^\varepsilon,$$

и ν_i монотонны по i.

Далее, ρ является корнем уравнения (14) при $\varepsilon_\ell = 0$, $\ell > 1$, а при $\varepsilon_\ell > 0$, $\ell > 1$, у (14) есть один действительный корень ρ_1^ε, $\rho < \rho_1^\varepsilon$, $\rho_1^\varepsilon < 1$ ($\rho_1^\varepsilon \to \rho$ при $\varepsilon_\ell \to 0$, $\ell > 1$).

Если у уравнения (14) n не равных единице корней ρ_m^ε, $m = 1, \ldots, n$, различны, то стационарное решение задачи имеет вид

$$\nu_i^\varepsilon = \sum_{m=1}^{n} c_m (\rho_m^\varepsilon)^i, \quad \sum_{i=1}^{\infty} \nu_i^\varepsilon < \infty, \tag{15}$$

где c_m находятся из условий $\nu_1^\varepsilon = \rho$, $\nu_\ell^\varepsilon = 1$, $\ell = 0, \ldots, -n+2$. Поскольку ν_i^ε монотонно убывают, главным членом в (15) является член $c_1\rho_1$. При кратных корнях в выражении для ν_i^ε появляются слагаемые вида $i^k(\rho_m^\varepsilon)^i$, $m > 1$, не влияющие на асимптотику ν_i^ε. ▲

Лемма 6. *Пусть* $g \in \mathcal{U}$. *Тогда для решения задачи* (9), (10), (8) $v_1^\varepsilon(t) = v_1(u^\varepsilon(t, g))$ *ограничено равномерно по* t, *и для* $u_i^\varepsilon(t)$ *справедливы соотношения*

$$\lim_{t\to\infty} (u_i^\varepsilon(t) - \nu_i^\varepsilon) = 0, \quad i = 1, 2, \ldots \tag{16}$$

Доказательство. Из уравнений (9) видно, что $|du_k^\varepsilon/dt| < \mathrm{const}$. Рассмотрим сначала случаи, когда $g_i \geqslant \nu_i^\varepsilon$ (или $g_i \leqslant \nu_i^\varepsilon$). Из лемм 2, 4 вытекает, что $u_i^\varepsilon(t) \geqslant \nu_i^\varepsilon$ (или $u_i^\varepsilon(t) \leqslant \nu_i^\varepsilon$). Покажем, что при этом

$$\left| \int_0^\infty (u_i^\varepsilon(t) - \nu_i^\varepsilon)\, dt \right| < \infty, \quad i = 1, 2, \ldots \tag{17}$$

Соотношение (16) будет тогда следствием ограниченности интеграла от знакопостоянной величины с ограниченной по t производной. Из уравнений (9) получаем, что

$$\frac{d(v_1(u^\varepsilon) - v_1(\nu^\varepsilon))}{dt} = \frac{dv_1(u^\varepsilon)}{dt} = -u_1^\varepsilon(t) + \nu_1^\varepsilon,$$

и значит, $|v_1(u^\varepsilon) - v_1(\nu^\varepsilon)|$ убывает по t. Тем самым доказана равномерная по t ограниченность $v_1(u^\varepsilon)$.

Пусть (16) доказано для u_k^ε, $k = 1, 2, \ldots, i-1$. Имеем

$$\frac{dv_i(u^\varepsilon(t))}{dt} = \rho \sum_{\ell=1}^{n} \sum_{k=\ell}^{n} \varepsilon_k u_{i-\ell}^\varepsilon(t) - u_i^\varepsilon(t).$$

Следовательно,

$$v_i(u^\varepsilon(t)) - v_i(\nu^\varepsilon) - \rho \int_0^t dt \sum_{\ell=1}^{n} \sum_{k=\ell}^{n} \varepsilon_k [u_{i-\ell}^\varepsilon(t) - \nu_{i-\ell}^\varepsilon] = - \int_0^t (u_i^\varepsilon(s) - \nu_i^\varepsilon)\, ds.$$

Из уже доказанной ограниченности левой части этого равенства следует справедливость утверждения для i.

В случае произвольных начальных данных $g \in \mathcal{U}$ решение $u^\varepsilon(t, g)$ лежит между решениями $u^\varepsilon(t, g^+)$ и $u^\varepsilon(t, g^-)$, где $g_i^+ = \max(g_i, \nu_i^\varepsilon)$, $g_i^- = \min(g_i, \nu_i^\varepsilon)$, а $u_i^\varepsilon(t, g^+)$ и $u_i^\varepsilon(t, g^-)$ сходится к ν_i^ε при $t \to \infty$, $i = 1, 2, \ldots$ ▲

Лемма 7. *У решения задачи* (9), (10), (8) *существуют производные* $\partial u_i^\varepsilon / \partial g_j$, $\partial^2 u_i^\varepsilon / \partial g_j \partial g_k$; $0 \leqslant \partial u_i^\varepsilon / \partial g_j \leqslant 1$, *сумма* $\sum\limits_i \partial u_i^\varepsilon / \partial y_j$ *ограничена равномерно по* j *и* t, *и* $\partial^2 u_i^\varepsilon / \partial g_j \partial g_k = 0$, $i < \infty$, $j, k = 1, 2, \ldots$, $t \geqslant 0$.

Доказательство. В силу линейности уравнений (9) производные по g_j от u_i^ε удовлетворяют уравнениям (9) с начальными и краевыми условиями $\partial u_i^\varepsilon(0)/\partial g_j = \delta_{ij}$, $\partial^2 u_i^\varepsilon(0)/\partial g_j \partial g_k = 0$, $i < \infty$, $j, k = 1, 2, \ldots$ Повторяя рассуждения, аналогичные приведенным в леммах 1–3 и 6, убеждаемся в том, что эти производные существуют при $t \geqslant 0$, причем $0 < \partial u_i^\varepsilon / \partial g_j \leqslant 1$ и $\sum\limits_i \partial u_i^\varepsilon / \partial g_j \leqslant 1$, а $\partial^2 u_i^\varepsilon / \partial g_j \partial g_k = 0$. ▲

Рассмотрим краевую задачу, аналогичную задаче (6)–(8):

$$\frac{du_{i,M}(t)}{dt} = \rho(u_{i-M,M}(t) - u_{i,M}(t)) - M(u_{i,M}(t) + u_{i+1,M}(t)), \quad i = 1, 2, \ldots, \tag{18}$$

$$u_{m,M}(t) = 1, \quad m \leqslant 0, \tag{19}$$

$$u_{i,M}(0) = g_{i,M}, \quad i = 1, 2, \ldots, \tag{20}$$

и аналог уравнений (9), (10):

$$\frac{du_{i,M}^\varepsilon(t)}{dt} = \rho \sum_{\ell=1}^{n} \varepsilon_\ell [u_{i-\ell M,M}^\varepsilon(t) - u_{i,M}^\varepsilon(t)] -$$

$$- M(u_{i,M}^\varepsilon(t) - u_{i+1,M}^\varepsilon(t)), \quad i = 1, 2, \ldots, \tag{21}$$

$$u_{m,M}^\varepsilon(t) = 1, \quad m \leqslant 0. \tag{22}$$

Так же, как и выше, показывается, что для решения задачи (18)–(20) и задачи (21), (22), (20) справедливы леммы, аналогичные леммам 1–7.

Обозначим через $\nu_M = \{\nu_{i,M}\}$ стационарное решение уравнений (18), (19).

Лемма 8. *Справедливы неравенства*

$$\nu_{i,M} \leqslant \rho \nu_{i-M,M}. \tag{23}$$

Доказательство. Просуммировав уравнения (18) по i, в силу монотонности $\nu_{i,M}$ по i получаем

$$\nu_{i,M} = \rho \sum_{\ell=1}^{M} \nu_{i-\ell,M}/M \leqslant \rho \nu_{i-M,M}.$$

Отметим здесь естественное соотношение между величиной индекса $i^{(1)}$ у решения $\{u_{i^{(1)}}\}_{i^{(1)}=0}^{\infty}$ задачи (8)–(10) ($M = 1$) и индекса $i^{(M)}$ у решения $\{u_{i^{(M)}}\}_{i^{(M)}=0}^{\infty}$ задачи (18)–(20) ($M > 1$): $i^{(1)} \leftrightarrow i^{(M)}M$. Поэтому можно сказать, что $\nu_{i,M}$ убывает по $i^{(M)}$ быстрее, чем ν_i убывает по i.

Задача (18)–(20) является разностной аппроксимацией задачи

$$\frac{\partial u(x,t)}{\partial t} = \rho[u(x-1,t) - u(x,t)] + \frac{\partial u(x,t)}{\partial x}, \tag{24}$$

$$u(x,0) = g(x), \quad x > 0, \quad u(x,t) \equiv 1, \quad x \leqslant 0. \tag{25}$$

Здесь $g(x)$, $0 < g(x) \leqslant 1$, $x \geqslant 0$, – невозрастающая функция. Эта начально-краевая задача описывает распределение времени ожидания в системе $M|D|1$.

Обозначим через $\overline{g}_M(x)$, $0 \leqslant x$, монотонное восполнение последовательности $g_{i,M}$, т.е. монотонную функцию, у которой $\overline{g}_M(i/M) = g_{i,M}$. Справедливо

Предложение 1. *Пусть при $M \to \infty$ имеется сходимость $\overline{g}_M(x) \to g(x)$ в пространстве $\mathbf{L}^1_{(x)}$. Тогда для любого $0 \leqslant t < \infty$ при $M \to \infty$ монотонное восполнение решения задачи* (18)–(20) *сходится в пространстве $\mathbf{L}^1_{(x,t)}$ к решению задачи* (24), (25).

При $M \to \infty$ монотонное восполнение стационарного решения задачи (18)–(20) *сходится в пространстве $\mathbf{L}^1_{(x)}$ к стационарному решению задачи* (24), (25).

§ 3. Свойства систем S_N и $S_{N,M}$

Рассмотрим пространство $L = C(\overline{\mathcal{U}})$ непрерывных функций $f : \overline{\mathcal{U}} \to \mathbb{R}$ с метрикой $\|f\| = \max_{u \in \overline{\mathcal{U}}} |f(u)|$ и пространство $L_N = C(\mathcal{U}_N)$. Вложение $\mathcal{U}_N \in \overline{\mathcal{U}}$ индуцирует отображение $L \to L_N$.

Система обслуживания S_N определяет марковский процесс $U_N(t)$, который задан оператором (5). Пространством состояний этого процесса является $\mathcal{U}_N$. В пространстве функций $f : \mathcal{U}_N \to \mathbb{R}$ процессу $U_N(t)$ соответствует полугруппа $T_N(t)$:

$$T_N(t)f(g) = \mathbf{E}(f(U_N(t)) \,|\, U_N(0) = g), \quad g \in \mathcal{U}_N, \quad f \in L_N,$$

которая действует в пространстве L_N. Отображение $g \to u(t,g)$, где $u(t,g)$ – решение задач (6), (7), (8), задает в пространстве $\overline{\mathcal{U}}$ динамическую систему, ей отвечает полугруппа $T(t)$, действующая в пространстве L,

$$T(t)f(g) = f(u(t,g)), \quad g \in \overline{\mathcal{U}}, \quad f \in L.$$

Полугруппы T_N и T являются сильно непрерывными и сжимающими (см., например, [11, гл. 1, п. 1.1]).

Теорема 1. *Пусть $\rho < 1$. Тогда процесс U_N эргодичен, т.е. существует единственная стационарная вероятностная мера μ_N, к которой при $t \to \infty$ сходятся распределения в момент t для любого начального распределения, и $\mathbf{E}_N v_1 < \infty$, где $\mathbf{E}_N$-математическое ожидание по стационарной мере. Кроме того,*

$$\lim_{t\to\infty} |T_N(t)u_i - \nu_i^\varepsilon| = 0. \tag{26}$$

Доказательство. Процесс U_N действует в пространстве $\mathcal{U}_N$. Из формулы (5) вытекает, что действие оператора A_N на функции $f_i(u) = u_i$ имеет вид

$$A_N f_i(u) = \rho \sum_{\ell=1}^{n} \varepsilon_\ell [u_{i-\ell}(t) - u_i(t)] - u_i(t) + u_{i+1}(t).$$

Отсюда следует, что для $w_{i,N}(t,u) = T_N(t)u_i$ справедливы уравнения

$$\frac{dw_{i,N}(t,u)}{dt} = \rho \sum_{\ell=1}^{n} \varepsilon_\ell [w_{i-\ell,N}(t) - w_{i,N}(t)] - w_{i,N}(t) + w_{i+1,N}(t),$$

$$T_N(t)v_1(u) = \sum_{i=1}^{\infty} w_{i,N}(t,u).$$

Из леммы 6 вытекает, что справедливо соотношение (26), и что $T_N(t)v_1(u) < \infty$ равномерно по t. Все состояния системы S_N сообщаются. Для доказательства эргодичности S_N достаточно поэтому заметить, что всякое подмножество u в $\mathcal{U}$ с ограниченным $v_1(u)$ конечно. ▲

Следствие 1. *Соотношение (26) и лемма 5 показывают, $\mathbf{E}_N u_i = O((\rho_1)^i)$, т.е. $\mathbf{E}_N u_i$ стремятся к нулю медленнее, чем ρ^i.*

Обозначим через A генератор полугруппы T, а через $\mathcal{D}(A)$ – область определения A. Если функция $f \in L$ обладает частными произведениями $\partial f/\partial u_i \in L$, $\sup_i \|\partial f/\partial u_i\| < \infty$, то $f \in \mathcal{D}(A)$,

$$Af(u) = \sum_{i=1}^{\infty} \frac{\partial f}{\partial u_i}\frac{du_i}{dt} = \sum_{i=1}^{\infty} \big\{\rho[u_{i-1}(t) - u_i(t)] - u_i(t) + u_{i+1}(t)\big\}\frac{\partial f}{\partial u_i}.$$

Обозначим через D множество всех функций $f \in L$, обладающих производными $\partial f/\partial u_i$, $\partial^2 f/\partial u_i \partial u_j$, для которых существует такая константа $C = C(f) < \infty$, что

$$\sup_{u,i} \left|\frac{\partial f(u)}{\partial u_i}\right| \leqslant C, \quad \sup_{u,i,j} \left|\frac{\partial^2 f}{\partial u_i \partial u_j}\right| \leqslant C. \tag{27}$$

Напомним определение: подпространство D', $D' \subset \mathcal{D}(A)$, является существенным подпространством для замкнутого линейного оператора A, если $\overline{A|_{D'}} = A$ (см., например, [11, гл. 3, п. 3]).

Лемма 9. *Множество D является существенным подпространством для оператора A.*

Доказательство. Очевидно, что $D \subset \mathcal{D}(A)$, и D плотно в $\mathcal{D}(A)$. В силу [11, гл. 1, п. 3.3] достаточно доказать, что оператор $T(t)$ не выводит D из себя при любом t. Пусть $f(g) \in D$, $g \in \overline{\mathcal{U}}$. Из леммы 7 вытекает, что $f(u)$ обладает производными $\partial f/\partial g_i$, $\partial^2 f/\partial g_j \partial g_k$ и для них выполняются условия (27), а значит, $f(u(t,g)) \in D$. ▲

Теорема 2. *Пусть f – произвольная непрерывная функция $f : \overline{\mathcal{U}} \to \mathbb{R}$. Тогда справедливо соотношение*

$$\lim_{N\to\infty} \sup_{g\in\mathcal{U}_N} |T_N(t)f(g) - f(u(t,g))| = 0, \quad t \geqslant 0. \tag{28}$$

Сходимость в (28) равномерна по t на любом конечном интервале изменения t.

Доказательство. Воспользуемся результатом, связывающим сходимость полугрупп со сходимостью их генераторов [11, гл. 1, п. 6.1]. В силу этого результата и леммы 9 достаточно доказать, что для любой функции $f \in D$

$$\lim_{N\to\infty} \sup_{u\in\mathcal{U}_N} |A_N f(u) - Af(u)| = 0. \tag{29}$$

Рассмотрим первые две суммы в (5):

$$N\sum_{i=1}^{\infty}\left(f\left(u-\frac{e_i}{N}\right)-f(u)\right)(u_i-u_{i+1})+$$

$$+N\lambda\sum_{i=1}^{\infty}\left(f\left(u+\frac{\sum_{j=1}^{n}e_{i_j}}{N}\right)-f(u)\right)(u_{i_1-1}-u_{i_1})\dots(u_{i_n-1}-u_{i_n})=$$

$$=N\sum_{i=1}^{\infty}\left(f\left(u-\frac{e_i}{N}\right)-f(u)\right)(u_i-u_{i+1})+$$

$$+N\lambda\sum_{i=1}^{\infty}\sum_{k=1}^{n}\left(f\left(u+\frac{\sum_{\ell=k}^{n}e_{i_\ell}}{N}\right)-f\left(u+\frac{\sum_{\ell=k+1}^{n}e_{i_\ell}}{N}\right)\right)(u_{i_1-1}-u_{i_1})\dots(u_{i_n-1}-u_{i_n})=$$

$$=\sum_{i=1}^{\infty}\frac{\partial f(u)}{\partial u_i}(u_{i+1}-u_i)+\lambda n\sum_{i=1}^{\infty}\frac{\partial f(u)}{\partial u_i}(u_{i-1}-u_i)+O\left(\frac{1}{N}\right)$$

Третья сумма в (5), очевидно, также имеет порядок $O(1/N)$ при $N \to \infty$. Отсюда следует (29). ▲

Теорема 3. *Пусть $\rho < 1$. Тогда на множестве $\mathcal{U}$ существует единственная вероятностная мера μ, которая инвариантна относительно динамической системы $g \to u(t,g)$, $g \in \mathcal{U}$. Эта мера сосредоточена в неподвижной точке ν динамической системы, $\nu = \{\nu_i\}_{i=0}^{\infty}$, $\nu_i = \rho^i$, и $\lim\limits_{N\to\infty} E_N u_i = \rho^i$.*

Доказательство. Множество $\overline{\mathcal{U}}$ компактно, поэтому множество вероятностных мер на $\overline{\mathcal{U}}$ компактно относительно слабой сходимости. Из теоремы 2 следует, что всякая мера μ, которая является предельной точкой для последовательности мер μ_N, инвариантна относительно полугруппы T. В лемме 6 доказано, что единственная вероятностная мера на $\mathcal{U}$, инвариантная относительно динамической системы $g \to u(t,g)$, $g \in \mathcal{U}$, сосредоточена в точке ν. Поэтому для доказательства теоремы надо доказать, что $\mu(\mathcal{U}) = 1$ для всякой предельной меры. Это следует из того, что для всякого N в силу леммы 5 $u_i < c\rho_1^i$, $c = \mathrm{const}$, и $\mathbf{E}_N v_i = \mathbf{E}_N \sum\limits_{j=i}^{\infty} u_j < \infty$. ▲

Так же, как выше, показывается, что для системы $S_{N,M}$ справедливы теоремы, аналогичные теоремам 1–3 (с заменой динамической системы (6)–(8) на систему (18)–(20) и динамической системы (9), (10), (8) на систему (21), (22), (20)).

При $M \to \infty$ распределение времени обслуживания в системе $S_{N,M}$ сходится к детерминированному распределению, сосредоточенному в единице, а распределение времени ожидания в стационарном случае при $M \to \infty$, $N \to \infty$, $N \gg M$, сходится к распределению, которое задается стационарным решением задачи (24), (25).

§ 4. Передача закодированных сообщений

Остановимся на системе S_N. Сравним среднее время передачи k-пакетного сообщения $T(k,k,\lambda)$ и среднее время передачи наиболее быстро переданных k пакетов из n-пакетного сообщения $T(n,k,\lambda)$. Все выкладки проводятся для предельного случая $N=\infty$, когда времена передачи пакетов независимы, $T(1,1,\lambda)=1/(1-\lambda)$. В этом случае можно воспользоваться формулами для порядковых статистик независимых экспоненциально распределенных случайных величин [12]:

$$T(n,k,\lambda)=\frac{1}{(1-n\lambda)}\sum_{i=n-k+1}^{n}\frac{1}{i}. \tag{30}$$

Кодирование уменьшит среднее время передачи k пакетов (по которым сообщение декодируется), если

$$T(k,k,\lambda)=\frac{1}{(1-k\lambda)}\sum_{i=1}^{k}\frac{1}{i}>T(n,k,\lambda)=\frac{1}{(1-n\lambda)}\sum_{i=n-k+1}^{n}\frac{1}{i}. \tag{31}$$

Как видно из формул (30)–(31), при любых k, n, $n>k$, найдется такое λ_{cr}, что при $\lambda<\lambda_{\text{cr}}$ справедливо неравенство (31). И для любого $\rho<1$ найдутся столь большие k, n, $\rho n/k<1$, что для $\lambda=\rho/k$ неравенство (31) справедливо.

Приведем два простых примера того, как кодирование может ускорить доставку сообщения.

1. Пусть $k=1$, $n=2$, т.е. вместо каждого однопакетного сообщения в систему посылается два однопакетных сообщения. Тогда

$$T(1,1,\lambda)=\frac{1}{(1-\lambda)}>T(2,1,\lambda)=\frac{1}{2(1-2\lambda)}$$

при $\lambda<\lambda_{\text{cr}}=1/3$.

2. Пусть $k=2$, $n=3$. Тогда

$$T(2,2,\lambda)=\frac{3}{2(1-2\lambda)}>T(3,2,\lambda)=\frac{5}{6(1-3\lambda)}$$

при $\lambda<\lambda_{\text{cr}}=4/17$.

Численные расчеты показывают, что при $N=\infty$ и постоянной длине пакетов (т.е. для системы $S_{N,M}$ при $N,M=\infty$) кодирование также уменьшает среднее время передачи сообщения.

Автор выражает благодарность Г.А. Кабатянскому, Ф.И. Карпелевичу, Е.А. Круку и Ю.М. Сухову за многочисленные полезные обсуждения работы.

СПИСОК ЛИТЕРАТУРЫ

1. *Кабатянский Г.А., Крук Е.А.* Об избыточном кодировании на транспортном уровне сети передачи данных // Помехоустойчивое кодирование и надежность ЭВМ. М.: Наука, 1987. С. 143–150.

2. *Семенов С.В., Крук Е.А.* К вопросу эффективности кодирования на транспортном уровне сети передачи данных // Тр. IX Всесоюз. симпоз. по проблеме избыточности в информационных системах. Тез. докл. Л., 1986. С. 77–80.
3. *Семенов С.В., Семенов А.В.* Моделирование сети передачи данных с коммутацией пакетов в условиях сетевого кодирования // Проблемы передачи и обработки информации. Л., 1991. С. 80–94.
4. *Vvedenskaya N.D.* Delay of a Message in Packet Switched Network with Unreliable Elements // Proc. Sixth Joint Swedish-Russian Int. Workshop on Information Theory. Molle, Swedan, 1993. P. 76–80.
5. *Vvedenskaya N.D., Weber J.H.* Transmission of Messages Using Few Parallel Servers // Proc. Conference on Distributed Computer Communication Networks. Tel-Aviv, Israel, 1996. P. 241–245.
6. *Dobrushin R.L.* Switching Networks, Gibbson Fields–Interconnection // Proc. 1st World Congress of the Bernoulli Society. Tashkent, 1986. Utrecht: VNS sci. press, 1987. V. 1. P. 377–393.
7. *Столяр А.Л.* Асимптотика стационарного распределения для одной замкнутой системы обслуживания // Пробл. передачи информ. 1989. Т. 25. № 4. С. 80–92.
8. *Baccelli F., Karpelevich F.I., Kelbert M.Ya., Puhalski A.A., Rybko A.N., Suhov Ya.M.* A Mean Field Limit for a Class of Queueing Networks // J. Statist. Phys. 1992. V. 6. № 3/4. P. 803–825.
9. *Введенская Н.Д., Добрушин Р.Л., Карпелевич Ф.И.* Система обслуживания с выбором наименьшей из двух очередей – асимптотический подход // Пробл. передачи информ. 1996. Т. 32. № 1. С. 20–34.
10. *Vvedenskaya N.D.* Transmission of Messages Using Parallel Servers: Asymptotic Approach // Proc. 1997 IEEE Int. Sympos. on Information Theory. Ulm, 1997. P. 32.
11. *Ethier S.N., Kurtz T.G.* Markov procceses characterization and convergence. N.Y.: John Willey and Sons, 1986.
12. *Дэйвид Г.* Порядковые статистики. М.: Наука, 1978.

Поступила в редакцию
05.03.97
После переработки
07.10.97