

ВЕРОЯТНОСТНАЯ МОДЕЛЬ РАБОТЫ УЗЛА СЕТИ INTERNET

А.Н. ДУДИН, Н.И. ЛИСТОПАД, Г.В. ЦАРЕНКОВ

Проведен анализ вероятностных характеристик работы узлов сети Internet с учетом реального нестационарного характера потоков запросов пользователей. Дана достаточно точная модель вероятности отказа в принятии запроса из-за ограниченности величины входного буфера. Проведенный численный анализ показал, что использование упрощенных моделей, основанных на модели входного потока как стационарного пуассоновского, не всегда оправдано и может привести к грубым погрешностям в расчетах вероятностей отказа.

В [1] предложены стратегии оптимизации функционирования узла сети Internet, позволяющие достичь сбалансированности входных и выходных потоков данных с учетом емкостей буфера маршрутизатора. В данной работе получены результаты, позволяющие применить аппарат [1] в случае, когда потоки информации являются нестационарными и «взрывными», что свойственно сети Internet.

Существенным предположением, сделанным в [1], есть допущение, что входной поток является стационарным пуассоновским, что справедливо при анализе функционирования классических сетей связи. Это объясняется сравнительной однородностью трафика, вследствие чего суперпозиция большого числа независимых рекуррентных потоков образует стационарный пуассоновский поток. Современные же сети передачи данных характеризуются неоднородным характером трафика, наличием всплесков интенсивности запросов, перемежающихся интервалами, когда трафик относительно низок. Трафик в Internet также является неоднородным. Типичный пример реализации входного потока в сервере одной из компьютерных сетей представлен на рис. 1 [2], из которого видно, что интенсивность потока варьируется в широком диапазоне и аппроксимация его стандартным пуассоновским потоком, имеющим постоянную интенсивность, является довольно грубой. Это может привести к значительной погрешности в оценивании характеристик производительности узла и, как следствие, к вынесению неверных решений при решении задачи повышения эффективности работы сети в целом.

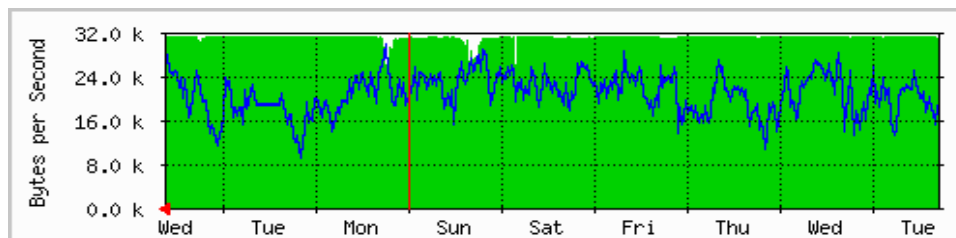


Рис. 1

Из анализа вида траектории входного трафика видно, что ее можно разбить на участки постоянства, где интенсивность потока меняется не

очень значительно, в то время как интенсивности на различных участках отличаются сильно. Это приводит к выводу, что поток принадлежит классу марковских модулированных пуассоновских потоков (ММРР — Markov modulates Poisson Process).

На интуитивном уровне ММРР-поток легко трактуется следующим образом. Имеется W , $w < +\infty$ уровней постоянства потока. На каждом r -м уровне постоянства в систему поступает стационарный пуассоновский поток интенсивности λ_r , $r = \overline{1, N}$. Длительность нахождения потока на r -м уровне постоянства имеет экспоненциальное распределение с параметром φ_r . После завершения пребывания на r -м уровне постоянства поток перескакивает на r' -й уровень постоянства с вероятностью $P_{r,r'}$, $r' = \overline{1, W}$, $r = \overline{1, W}$.

Таким образом, ММРР-поток можно задать числом W уровней постоянства, наборами $\{\lambda_1, \dots, \lambda_W\}$ и $\{\varphi_1, \dots, \varphi_W\}$ интенсивностей и матрицей $P = \|P_{r,r'}\|_{r,r'=\overline{1,W}}$ вероятностей перехода между уровнями. Моделирование реального потока в терминах ММРР-потока сводится к аппроксимации траектории интенсивности реального потока кусочно-постоянной функцией.

Заметим, что для более точной аппроксимации реального потока следует выделить большое число уровней постоянства. Однако следует искать разумный компромисс, поскольку значительное увеличение числа уровней может привести к возникновению проблем при численной реализации изложенных ниже методов.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

В качестве модели процесса обслуживания запросов пользователей в узле сети выберем следующую систему массового обслуживания.

Имеется однолинейная система с конечным буфером емкости, рассчитанной на хранение N запросов. Выходной поток предполагается потоком ММРР-типа. Этот поток полностью описывается в вероятностном смысле путем задания двух матриц: $D_1 = \Lambda$, $D_0 = -\Lambda + \Phi(P - I)$. Здесь Λ и Φ — диагональные матрицы с диагональными элементами λ_r, φ_r , $r = \overline{1, W}$ соответственно; P — матрица переходов между уровнями постоянства, $P = \|P_{r,r'}\|_{r,r'=\overline{1,W}}$; I — тождественная матрица.

Мы не делаем никаких специальных предположений о времени обработки запроса, считая его случайной величиной с функцией распределения $B(t)$ и математическим ожиданием $b_1 = \int_0^{\infty} ((1 - D(t))) dt$.

Основной характеристикой, интересующей нас, является вероятность отказа в обслуживании произвольного запроса.

ВЕРОЯТНОСТЬ ОТКАЗА В ОБСЛУЖИВАНИИ

Расчетные формулы в данном разделе получены на базе теоретических результатов, касающихся более сложной модели системы обслуживания типа ВМАР/SM/1/N. Поток типа ВМАР (batch Markovian arrival process) (см., например, [3]) является подклассом точечных процессов, включающим ММРР-поток как частный случай. Вероятность отказа $P_{\text{отк}}$ вычисляется по следующей формуле:

$$P_{\text{отк}} = 1 + \lambda^{-1} \sum_1^N \vec{P}_i D_0 \vec{1}. \quad (1)$$

Здесь λ — средняя интенсивность потока, получаемая усреднением интенсивностей ММРР-потока по распределению его пребывания на различных уровнях постоянства и вычисляемая по формуле

$$\lambda = \vec{\Theta} D_1 \vec{1}, \quad (2)$$

где $\vec{\Theta}$ — вероятностный вектор-строка, компоненты которого находятся как единственное решение системы линейных алгебраических уравнений

$$\vec{\Theta} = (D_0 + D_1) \vec{\Theta} = \vec{0}, \quad \vec{\Theta} \vec{1} = 1, \quad (3)$$

где $\vec{0}$ — нулевой вектор размерности W , а $\vec{1}$ — единичный вектор-столбец этой же размерности; \vec{P}_i , $i = \overline{0, N+1}$ — вероятностные векторы-строки с компонентами $\vec{P}_i = (p(i, 1), \dots, p(i, w))$, где $p(i, r)$ — стационарная вероятность наличия в системе в произвольный момент времени i запросов и пребывания входного потока на уровне r , $i \geq 0$, $r = \overline{1, w}$.

Эти векторы вычисляются по формулам:

$$\begin{aligned} \vec{P}_0 &= \lambda_1 \vec{\pi}_0 (-D_0)^{-1}, \\ \vec{P}_1 &= \vec{P}_0 D_1 R_0 + \lambda_1 \vec{\pi}_1 R_0, \\ \vec{P}_i &= \lambda_1 \sum_{k=1}^i \vec{\pi}_k R_{i-k}, \quad i = \overline{2, N}, \\ \vec{P}_{N+1} &= \vec{P}_0 D_1 \sum_{l=N}^{\infty} R_l + \lambda_1 \sum_{k=1}^N \vec{\pi}_k \sum_{l=N-k+1}^{\infty} R_l, \end{aligned} \quad (4)$$

где λ_1^{-1} — среднее значение интервалов между моментами ухода из системы обслуженных запросов,

$$\lambda_1^{-1} = \vec{\pi}_0 (-D_0) \vec{1} + b_1 \quad (5)$$

Матрица R_l характеризует переходы между уровнями потока за время обслуживания одного запроса, в течение которого в систему поступило $l \geq 0$ новых запросов, и вычисляется из натурального разложения

$$\sum_{l=0}^{\infty} R_l Z^l = \int_0^{\infty} e^{(D_0 + D_1 Z)t} (1 - B(t)) dt.$$

Вероятностные векторы $\vec{\pi}_k, k = \overline{0, N}$ характеризуют стационарное распределение числа запросов в буфере и уровень ММРР-потока в моменты окончания обслуживания запросов. Процедуры для подсчета этих векторов приведены в [3].

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ВЕРОЯТНОСТЕЙ ОТКАЗА

Алгоритм расчета вероятности отказа (1) для произвольной системы типа ВМАР/SM/1/N реализован в виде соответствующей программы (см. [3]). Построим графики, характеризующие зависимость этой вероятности от различных значений коэффициента загрузки системы $\rho = \lambda b_1$.

Параметры системы обслуживания, соответствующей трафику, реализация которого приведена на рис.1, были следующими.

1. Поток описывается ММРР-поток со значениями параметров $W = 3, \lambda_1 = 14, \lambda_2 = 22, \lambda_3 = 28, \varphi_1 = 0,1, \varphi_2 = 0,05, \varphi_3 = 0,012,$

$$P = \begin{pmatrix} 0 & 0,75 & 0,25 \\ 0,5 & 0 & 0,5 \\ 0 & 1 & 0 \end{pmatrix}.$$

2. Матрицы D_0 и D_1 принимают вид:

$$D_0 = \begin{pmatrix} -14,1 & 0,075 & 0,025 \\ 0,025 & -22,05 & 0,025 \\ 0 & 0,12 & -28,12 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 14 & 0 & 0 \\ 0 & 22 & 0 \\ 0 & 0 & 28 \end{pmatrix}.$$

Среднее время обслуживания выбираем равным $b_1 = 1/32$ с. Проведенный анализ показал, что вероятность отказа в рассматриваемой системе однозначно не определяется средним временем обслуживания, а зависит и от вида функции распределения. Поэтому мы исследовали и зависимость величины $P_{\text{отк}}$ от вида вышеупомянутой функции распределения. В качестве таких функций распределения выбирались следующие:

1) функция распределения вырожденного распределения,

$$B(t) = \begin{cases} 0, & t < b_1, \\ 1, & t \geq b_1; \end{cases}$$

2) функция равномерного распределения в интервале от 0 до $2b_1$,

$$B(t) = \begin{cases} 0, & t < 0, \\ \frac{t}{2b_1}, & 0 \leq t < 2b_1, \\ 1, & t \geq 2b_1; \end{cases}$$

3) функция равномерного распределения в интервале $\left[\frac{3}{112}, \frac{1}{28}\right]$,

$$B(t) = \begin{cases} 0, & t < \frac{3}{112}, \\ \frac{t - \frac{3}{112}}{\frac{1}{28} - \frac{3}{112}}, & \frac{3}{112} \leq t < \frac{1}{28}, \\ 1, & t \geq \frac{1}{28}; \end{cases}$$

4) функция распределения Эрланга k -го порядка,

$$B(t) = \lambda \int_0^t \frac{(\mu\tau)^{k-1}}{(k-1)!} e^{-\mu\tau} d\tau$$

со значениями: $k = 2, \mu = 64$;

5) функция распределения Эрланга k -го порядка,

$$B(t) = \lambda \int_0^t \frac{(\mu\tau)^{k-1}}{(k-1)!} e^{-\mu\tau} d\tau$$

со значениями: $k = 4, \mu = 128$;

6) функция распределения экспоненциального закона,

$$B(t) = 1 - e^{-\mu t}, \quad \mu = 32, \quad t > 0;$$

7) функция распределения гиперэкспоненциального закона,

$$B(t) = \frac{1}{2} \left(1 - e^{-28t} \right) + \frac{1}{2} \left(1 - e^{-\frac{112}{3}t} \right).$$

Нетрудно видеть, что эти распределения имеют математическое ожидание $b_1 = \frac{1}{32}$. В качестве другой важной характеристики распределения часто принимают коэффициент вариации σ (напомним, коэффициентом вариации σ [2] распределения случайной величины ξ называют

величину $\sigma = \frac{\sqrt{D\xi}}{M\xi} = \frac{\sqrt{M\xi^2 - (M\xi)^2}}{M\xi}$.

Для приведенных выше законов распределения коэффициент вариации σ соответственно равен:

- 1) вырожденное распределение — 0;
- 2) равномерное распределение в широком диапазоне — 0,57735;
- 3) равномерное распределение в узком диапазоне — 0,08248;
- 4) эрланговское распределение 2-го порядка — 0,5;
- 5) эрланговское распределение 4-го порядка — 0,25;
- 6) экспоненциальное распределение — 1;
- 7) гиперэкспоненциальное распределение — 1,237.

Результаты расчета вероятности отказа в описанном примере приведены на рис. 2–4, где кривые представлены для различных значений буфера N ($N = 1, 10, 50, 100, 200$) и различных упомянутых видов распределения. Кроме значения вероятности отказа (поток ВМАР) приводятся также значения вероятности отказа, рассчитанной для аналогичной системы со стационарным пуассоновским (а не ВМАР) потоком такой же интенсивности (поток СР), а также значение этой вероятности, рассчитанной для стационарного пуассоновского потока с максимально возможной (в данном примере 28) интенсивностью потока (поток ЧНН — час наибольшей нагрузки). Этот поток соответствует популярной до сих пор при проектировании сетей связи концепции расчета необходимых параметров сети в расчете на худший вариант (ЧНН).

Кривые на рисунках упорядочены для наглядности не в соответствии с приведенной выше нумерацией видов распределений, а в соответствии со значением коэффициента вариации σ распределения (в порядке возрастания).

Значение коэффициента загрузки $\rho = \lambda b_1$ в рассмотренной системе равно 0,678448, а соответствующее значение $\rho_{\text{ЧНН}}$, рассчитанной по ЧНН, равно 0,875. Вариация средней интенсивности потока проводилась путем домножения элементов матрицы D_1 (соответственно интенсивностей $\lambda_1, \lambda_2, \lambda_3$) на масштабирующий коэффициент δ с соответствующей перенормировкой матрицы D_0 для достижения выполнения условия $(D_0 = D_1)\vec{1} = \vec{0}$.

На каждом из рисунков график № 1 соответствует значениям $\delta = 1$, $\rho = 0,678448$, $\rho_{\text{ЧНН}} = 0,875$, а для графиков № 2–6 эти значения следующие:

№	δ	ρ	$\rho_{\text{ЧНН}}$
2	1,2	0,814138	1,05
3	1,4	0,949828	1,225
4	1,5	1,01767	1,3125
5	1,7 _ы	1,15336	1,4875
6	1,78	1,20753	1,5575

Поток ВМАР

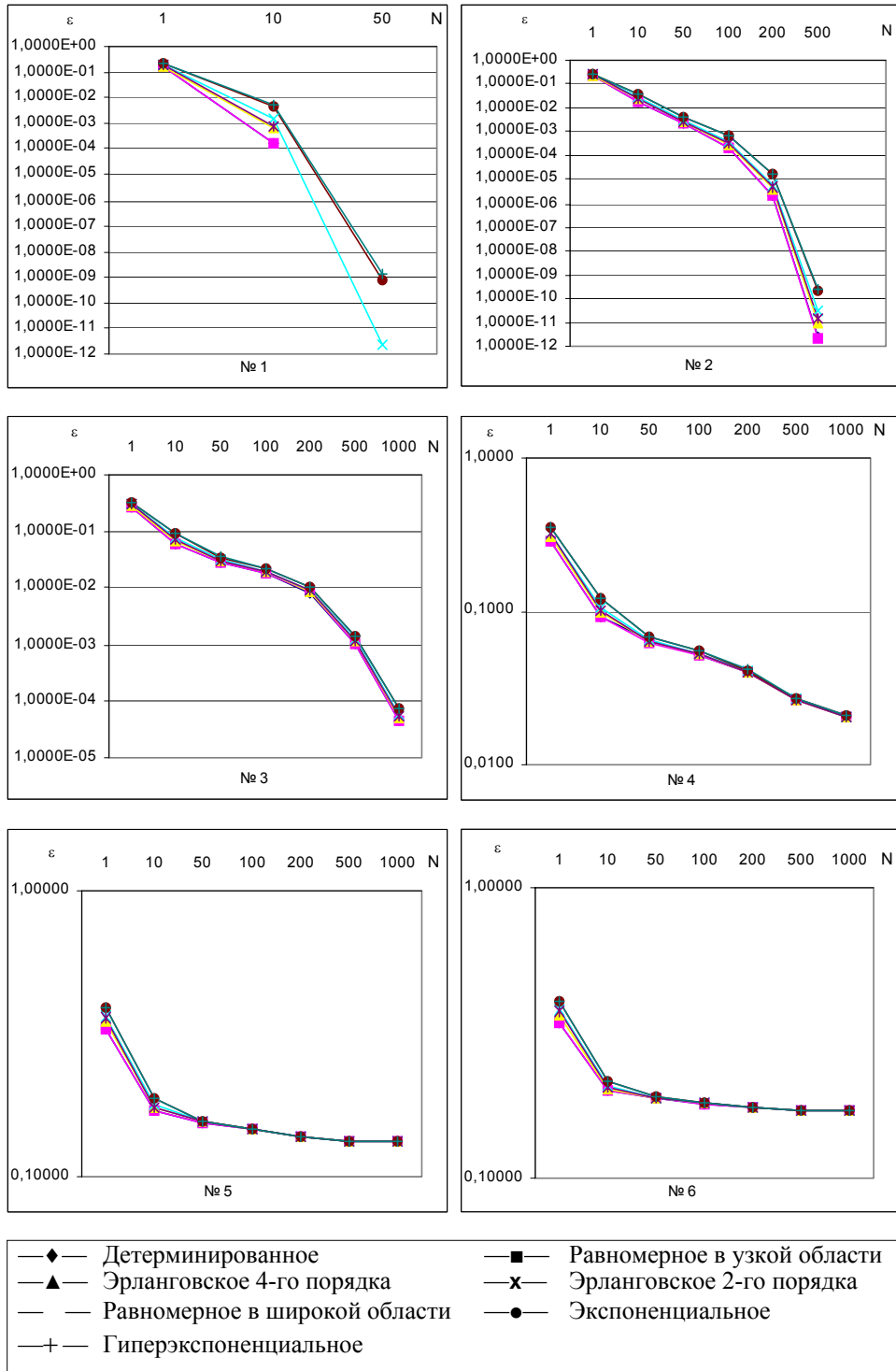


Рис. 2

Поток СП

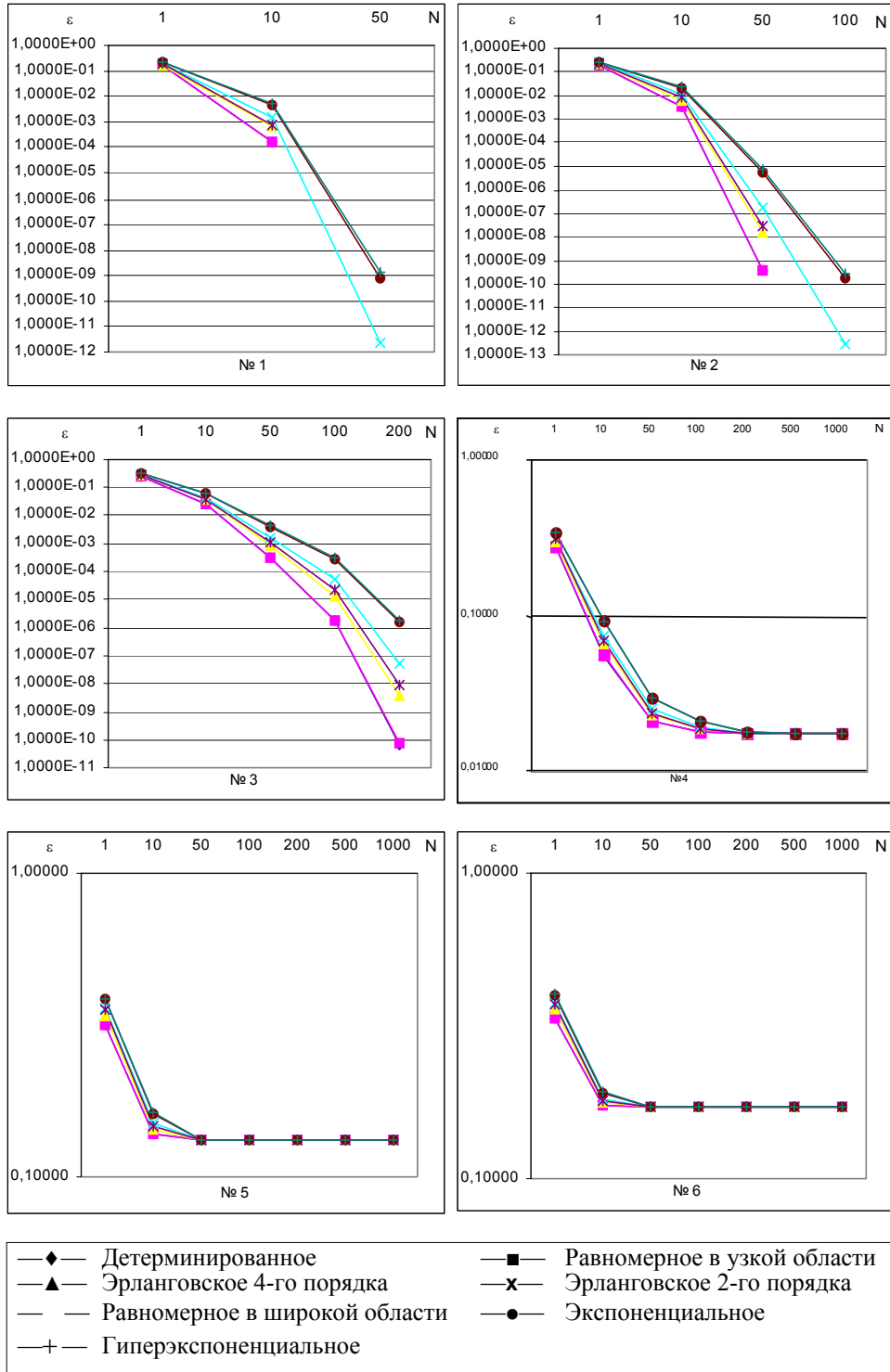


Рис. 3

Поток ЧНН

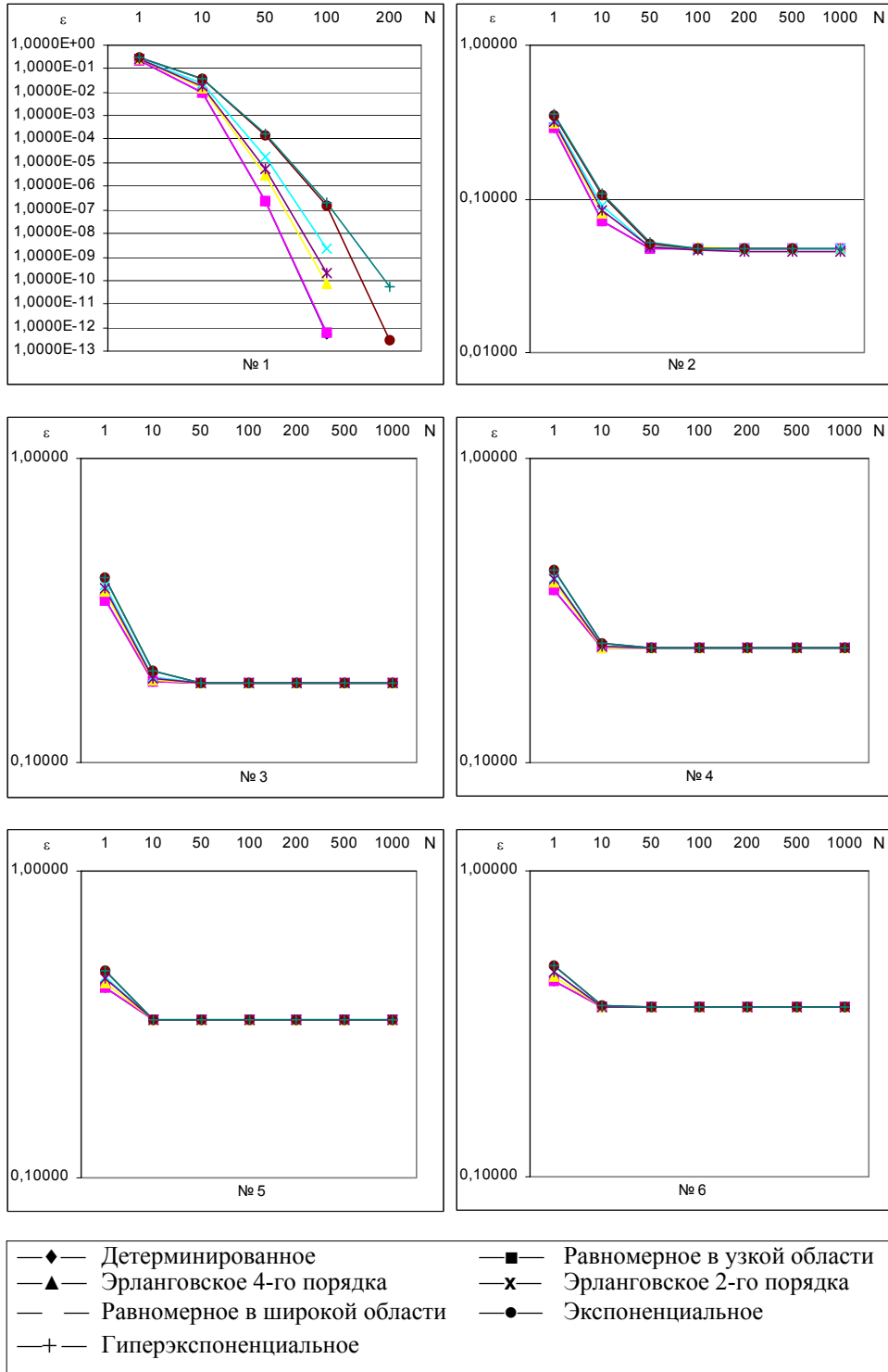


Рис. 4

Анализ рис. 2–4 позволил сделать следующие выводы.

1. Модель вероятности отказа, построенная на основе усредненных стационарных пуассоновских потоков и стационарных пуассоновских потоков с максимально возможной интенсивностью (ЧНН), является достаточно грубой. Из графиков № 1 и 2 видно, что ошибка резко увеличивается с ростом значения буфера N . При $N = 50$ она уже составляет 4–6 порядков.

2. На значение вероятности отказа влияет вид функции распределения. При не критической нагрузке это влияние может быть значительным (см. график № 2). Например, при $N = 500$ вероятность отказа для детерминированного распределения на два порядка отличается от вероятности для экспоненциального распределения. Поэтому применение разработанного математического аппарата должно находиться в строгом соответствии с моделируемым процессом. Так, если на нижних уровнях эталонной модели взаимодействия открытых систем (например, канальном уровне) мы практически безошибочно можем считать время передачи детерминированным, то на более высоких уровнях (например, транспортном, где время передачи должно учитывать получение подтверждения и тайм-ауты) вид функции распределения времени обслуживания будет несколько другим.

При критической нагрузке (значение ρ близко к единице) влияние вида функции распределения существенно только при небольших значениях буфера N .

3. Из проведенного анализа следует, что вероятность отказа, грубо говоря, прямо пропорциональна коэффициенту вариации σ , но не полностью определяется им. Так, сравнение случаев равномерного распределения в узкой области ($\sigma = 0,57735$) и эрланговского σ -го порядка ($\sigma = 0,5$) показывает, что выявленная пропорциональность в этом случае нарушается.

В данной статье проведен анализ вероятностных характеристик работы узлов сети Internet с учетом реального нестационарного характера потоков запросов пользователей. Приведена достаточно точная модель вероятности отказа в принятии запроса из-за ограниченности величины входного буфера. Проведенный численный анализ показал, что применение упрощенных моделей, основанных на модели входного потока как стационарного пуассоновского, не всегда оправдано и может привести к грубым погрешностям в расчетах вероятностей отказа.

ЛИТЕРАТУРА

1. Дудин А.Н., Листопад Н.И. Модели оптимизации работы узла сети Internet. Комплексная защита информации. — 2000. — Вып. 3. — С. 47-53.
2. Дудин А.Н., Листопад Н.И., Царенков Г.В. Улучшенный алгоритм оптимизации работы узла сети Internet // Проблемы проектирования информационно-телекоммуникационных систем: Сб. науч. тр./ Под ред. А.Н. Курбацкого. — Мн.: БГУ, 2001. — С. 28–43.
3. Dudin A.N., Klimenok V.I., Klimenok I.A., Borokhovskiy V.V., Karolik A.V., Tsarenkov G.V. Software «SIRIUS» for evaluation and optimization of queues with the BMAP-input // Advances in Algorithmic Methods for Stochastic Models — Proceedings of the 3-rd International Conference on Matrix Analytic Methods. New Jersey: Notable Publications Inc., 2000. — P. 115–133.

Поступила 5.04.2002