

НЕПРЕРЫВНЫЕ ДИФFUЗИОННЫЕ МОДЕЛИ МАССОВОГО ОБСЛУЖИВАНИЯ И МЕТОДИКА РАСЧЕТА ИХ ХАРАКТЕРИСТИК

В статье предлагается методика расчета характеристик стохастических сетей на основе новых диффузионных моделей массового обслуживания. Рассмотрены системы массового обслуживания с бесконечной очередью, с конечной очередью и потерями, с переменными параметрами поступления и обслуживания заявок.

Для расчета характеристик систем массового обслуживания (СМО) в предположении, что времена между заявками во входном потоке и обслуживания распределены по произвольному закону, не существует точных методов. Использование же традиционных методов, основанных на пуассоновских входных потоках и экспоненциальности времени обслуживания, не всегда оправдано. Существующие аппроксимационные методы, использующие два первых момента распределений входного потока и времени обслуживания, имеют большую погрешность в случае отклонения коэффициентов вариаций этих распределений от единицы.

Для расчета характеристик сетевых моделей систем сеть декомпозируется на отдельные узлы (рисунок 1).

В этой статье излагается новый метод декомпозиции разомкнутых однородных сетевых моделей систем управления на отдельные узлы на уровне двух моментов распределений параметров потоков. С целью определения параметров выходного потока, а также основных харак-

теристик отдельного узла вводится двумерная диффузионная аппроксимация процессов функционирования узла. Указанная методика также обобщается на сетевые модели с неоднородными потоками заявок.

Для определения среднего значения и дисперсии распределения времени между соседними заявками на входе и выходе каждого узла сети необходимо выразить среднее и дисперсию времени в выходном потоке отдельного узла через параметры входного потока и времени обслуживания. Для решения этой задачи выведены точные формулы:

$$\bar{\tau}_{вых} = \tau_{\mu} + p'_0 \tau'_{\lambda}, \quad (1)$$

$$D_{вых} = D_{\mu} + p'_0 D'_{\lambda} + p'_0 (1 - p'_0) (\bar{\tau}'_{\lambda})^2, \quad (2)$$

где $\bar{\tau}_{вых}$, $\bar{\tau}_{\mu}$, $D_{вых}$, D_{μ} – соответственно средние и дисперсии времени между заявками выходного потока и времени обслуживания, $\bar{\tau}'_{\lambda}$, D'_{λ} – среднее и дисперсия остаточного времени (времени простоя узла), а p'_0 – вероятность того, что обслуженная заявка оставляет узел пустым. Для системы без потерь средние времена между за-

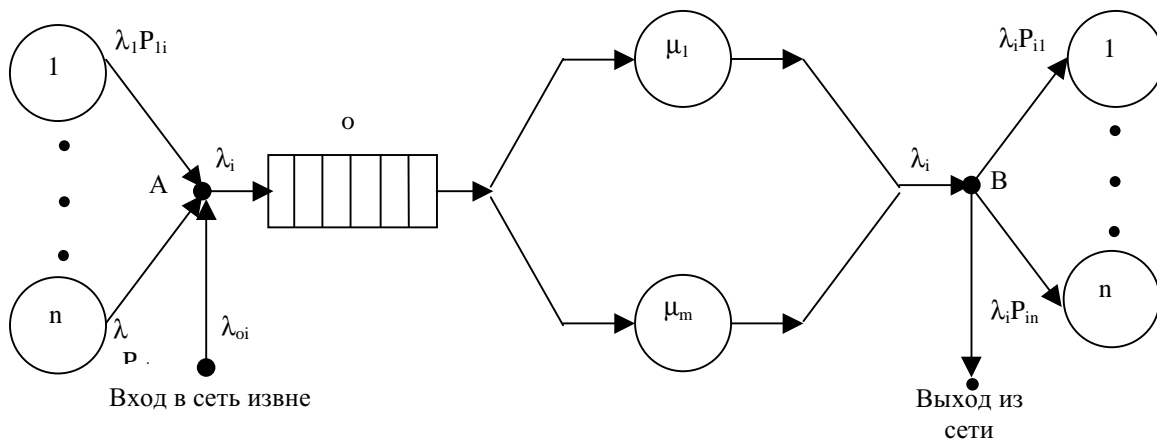


Рисунок 1. Структура I-й СМО сети (А – узел композиции, В – узел декомпозиции)

явками в выходном и входном потоках совпадают. Уравнения (2) решаются совместно со следующими уравнениями преобразования дисперсии времени между заявками:

$$D_{\Sigma} = (\lambda_1 / \lambda_{\Sigma})^3 D_1 + (\lambda_2 / \lambda_{\Sigma})^3 D_2 \quad (3)$$

– для суммы двух независимых потоков с интенсивностями λ_1, λ_2 и дисперсиями D_1, D_2 (узел композиции потоков) и

$$Dp = \frac{1}{p} \left(D + \frac{1-p}{p^2} \right) \quad (4)$$

– для потока с интенсивностью λ и дисперсией времени между заявками D , в котором заявки с вероятностью p уходят из потока (узел декомпозиции потоков). Из уравнений (2)-(4) следуют основные уравнения декомпозиции сети, так называемые уравнения баланса для средних и дисперсий времени между заявками на входе и выходе i -го узла сети ($i=1, \dots, n$)

$$\lambda_i = \bar{\tau}_{oi}^{-1} + \sum_{j=1}^n p_{ji} \lambda_j, \quad (5)$$

$$D_{ex.i} = \left(\frac{\bar{\tau}_{oi}^{-1}}{\lambda_i} \right)^3 D_{oi} + \sum_{j=1}^n \left(\frac{p_{ji} \lambda_j}{\lambda_i} \right)^3 \frac{1}{p_{ji}} \left(D_{выхj} + \frac{1-p_{ji}}{p_{ji} \lambda_j^2} \right) \quad (6)$$

где $\bar{\tau}_{oi}$ и D_{oi} – среднее и дисперсия времени между заявками в потоке, идущем от внешнего источника на вход i -го узла, λ_i – интенсивность

потока на входе и выходе i -го узла, $D_{ex.i}, D_{выхi}$ – дисперсии времени между заявками во входном и выходном потоках i -го узла, p_{ji} – вероятность передачи заявки от j -го узла к i -му, а n – количество узлов в сети. Для решения системы (6) совместно с (5) и (2) предложена итерационная процедура, где в качестве первого приближения в уравнении (2) используется замена $p'_0 = p_0, \bar{\tau}'_{\lambda} = \bar{\tau}_{\lambda}, D_{\lambda} = D_{\lambda}$. Здесь $p_0 = 1 - \rho$ – вероятность простоя узла, $\bar{\tau}_{\lambda}$ и D_{λ} – среднее и дисперсия времени между заявками во входном потоке. Это приближение в случае экспоненциальной сети не вносит погрешности, а уравнения (6) становятся линейными относительно искомым дисперсий. На последующих этапах значения дисперсий входных и выходных потоков уточняются с использованием уравнений (2) и (6), одновременно вычисляются основные характеристики отдельных узлов и всей сети в целом.

При приближенном расчете замкнутых сетевых моделей систем в выражениях (5) и (6) первые слагаемые исчезают. В этом случае в итерационной процедуре значения интенсивностей λ_i подбираются так, чтобы с заданной точностью ϵ выполнялось условие

$$\sum_{i=1}^n N_i = N,$$

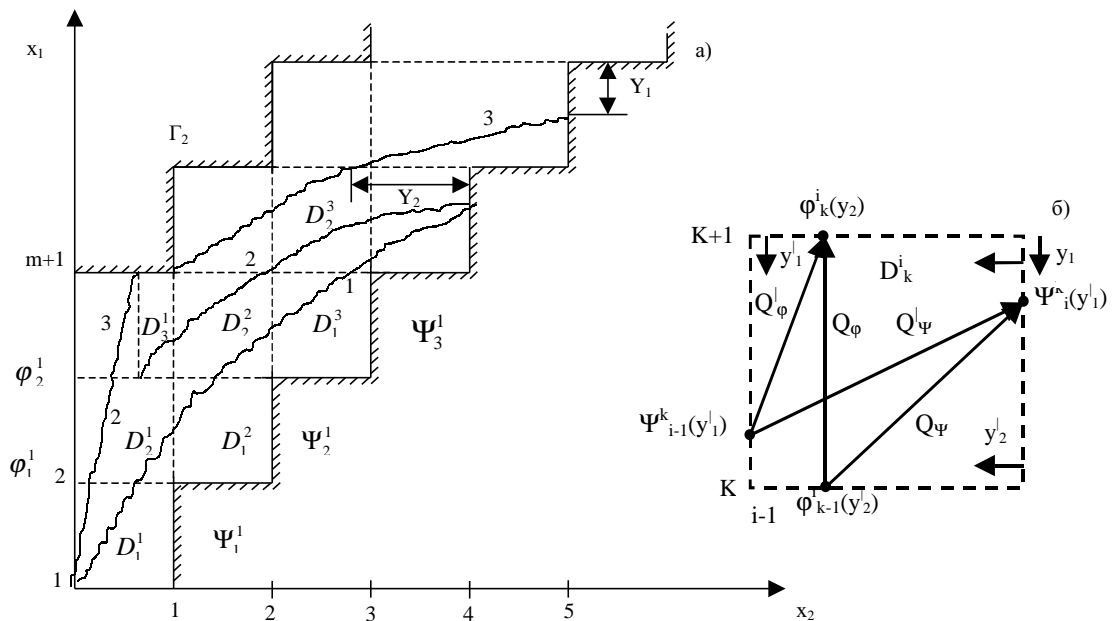


Рисунок 2. Аппроксимация СМО GI/G/1/m с потерями, а также с переменными параметрами двумерным диффузионным процессом (x_1, x_2) (а) и выделенная область D_k^i (б) (индекс k определяет уровень по оси x_1 , а i – по x_2). Траектория 1 – для СМО GI/G/1/∞ при $m \rightarrow \infty$, траектория 2 – для СМО с потерями, траектория 3 – для СМО с переменными параметрами.

где через N_i обозначено среднее количество заявок в i -м узле, а N – количество заявок, постоянно циркулирующих в замкнутой сети.

Для определения параметров $\rho'_0, \tau'_\lambda, D'_\lambda$, входящих в формулы (1) и (2), а также основных характеристик узла процесс его функционирования на периоде занятости аппроксимируется двумерным диффузионным процессом (x_1, x_2) . Процесс $x_1(t)$ аппроксимирует число заявок $N_1(t)$, поступивших в узел к моменту времени t , а $x_2(t)$ – число заявок $N_2(t)$, покинувших узел к тому же времени. Текущее значение N числа заявок в узле определяется разностью $N = [x_1] - [x_2]$, где $[x]$ – целая часть от x (рисунок 2а, траектория 1).

Потребуем, чтобы компоненты двумерного процесса (x_1, x_2) в моменты времени первого прохождения целочисленного уровня (моменты поступления и ухода заявок) имели средние значения и дисперсии, совпадающие со средними и дисперсиями компонент дискретного процесса (N_1, N_2) . Тогда коэффициенты сноса $a_i = \bar{\tau}_i^{-1}$ и диффузии $\sigma_i = D_i \bar{\tau}_i^{-3}$ процесса x_i можно выразить через средние значения $\bar{\tau}_i$ и дисперсии D_i интервалов времени между скачками дискретного процесса $N_i (i=1,2)$. В области D , определенной условием $N \geq 0$ и приведенной на рисунке 2а, плотность распределения $\omega(t, x_1, x_2)$ векторного диффузионного процесса удовлетворяет уравнению Колмогорова

$$\frac{\partial \omega}{\partial t} = \sum_{i=1}^2 \left(b_i \frac{\partial^2 \omega}{\partial x_i^2} - a_i \frac{\partial \omega}{\partial x_i} \right). \quad (7)$$

Так как период занятости начинается с уровня $x_1=1$, то начальным условием для (7) будет $\omega(0, x_1, x_2) = \delta(x_1 - 1)\delta(x_2)$. Рассматривая функционирование узла на периоде занятости, добавим к уравнению (7) граничное условие поглощения $\omega|_{\Gamma_1} = 0$. Граница Γ_1 , определенная условием $[N]=0$, имеет ступенчатый характер, а попадание ординаты процесса на границу физически означает завершение периода занятости. Вследствие сложного характера границы, решение уравнения (7) будем искать в виде совокупности решений в подобластях $D_\kappa = (x_1 \leq \kappa + 1, x_2 \leq \kappa)$, ($\kappa = 1, 2, \dots$) и «сшивая» их на границах $x_1 = \kappa$, ($\kappa = 2, 3, \dots$). Введем обозначения $\Psi_\kappa(y_1)$ для распределения ординаты процесса x_1 в момент достижения уровня $x_2 = \kappa$ (границы Γ) и $\varphi_\kappa(y_2)$ для распределения ординаты процесса x_2 в момент прохождения процессом (x_1, x_2) уровня $x_1 = \kappa + 1$ (границы области D_κ). Эти распределения позволяют определить все основные характеристики функционирования от-

дельного узла. Указанные распределения можно представить через решение уравнения (7), т. е. плотность распределения $\omega_k(t, x_1, x_2)$. Для этого в статье с использованием аппарата теории функций Грина $Q_k(t, x_1, x_2 | k, y'_2)$ получены рекуррентные формулы:

$$\omega_k(t, x_1, x_2) = \int_0^\infty \varphi_{k-1}(y'_2) Q_k(t, x_1, x_2 | k, y'_2) dy'_2, \quad (8)$$

$$\varphi_k(y_2) =$$

$$= \int_0^\infty \varphi_{k-1}(y'_2) Q_\varphi(y_2 | y'_2) dy'_2 (\varphi_1(y_2) = Q_\varphi(y_2 | 0)), \quad (9)$$

$$Q_\varphi(y_2 | y'_2) = \frac{1}{\pi \sqrt{b_1 b_2}} \cdot \exp\left[\frac{a_1}{b_1} + \frac{a_2}{b_2} (y'_2 - y_2 + 1)\right] \times$$

$$\times \left[\sqrt{\frac{\gamma}{\beta_1}} K_1(2\sqrt{\beta_1 \cdot \gamma}) - \sqrt{\frac{\gamma}{\beta_2}} K_1(2\sqrt{\beta_2 \cdot \gamma}) \right];$$

$$\beta_1 = \frac{1}{2b_1} + \frac{(y'_2 - y_2 + 1)^2}{2b_2}; \quad \beta_2 = \frac{1}{2b_1} + \frac{(y'_2 + y_2 + 1)^2}{2b_2};$$

$$\gamma = \frac{a_1^2}{2b_1} + \frac{a_2^2}{2b_2}; \quad y_2 \in [0, \infty];$$

$K_1(\bullet)$ – функция Макдональда;

$$\Psi_k(y_1) =$$

$$= \int_0^\infty \varphi_{k-1}(y'_2) Q_\Psi(y_1, y'_2) dy'_2 (\Psi_1(y_1) = Q_\Psi(y_1 | 0)), \quad (10)$$

$$Q_\Psi(y_1 | y'_2) = \frac{1 + y'_2}{\pi \sqrt{b_1 b_2}} \cdot \exp\left[\frac{a_1}{b_1} (1 - y_1) + \frac{a_2}{b_2} (1 + y'_2)\right] \times$$

$$\times \left[\sqrt{\frac{\gamma}{\beta_3}} K_1(2\sqrt{\beta_3 \cdot \gamma}) - \sqrt{\frac{\gamma}{\beta_4}} K_1(2\sqrt{\beta_4 \cdot \gamma}) \right];$$

$$\beta_3 = \frac{(1 - y_1)^2}{2b_1} + \frac{(1 + y'_2)^2}{2b_2}; \quad \beta_4 = \frac{(1 + y_1)^2}{2b_1} + \frac{(1 + y'_2)^2}{2b_2};$$

$$y_2 \in [0, \infty].$$

Если m_Ψ и D_Ψ – среднее и дисперсия распределения $\Psi(y_1) = \sum_{\kappa=1}^\infty \Psi_\kappa(y_1)$, то искомые параметры двумерного диффузионного приближения $\rho'_0, \tau'_\lambda, D'_\lambda$, входящие в уравнение (2) и формулы для основных характеристик узла, можно выразить через эти величины m_Ψ, D_Ψ и извест-

ные параметры входного потока: среднее $\bar{\tau}_\lambda$ и дисперсию D_λ времени между заявками. Решение этой задачи связано с использованием функции плотности времени первого достижения марковским процессом заданного уровня $g(t|y_1)$. Окончательные формулы имеют следующий вид:

$$\bar{\tau}'_\lambda = \bar{\tau}_\lambda \cdot m_\psi, \quad (11)$$

$$D'_\lambda = D_\lambda m_\psi + \bar{\tau}_\lambda^2 D_\psi \quad (12)$$

– для среднего и дисперсии времени простоя узла и

$$p'_0 = 1 / \sum_{k=1}^{\infty} k p_k \quad \left(p_k = \int_0^{\infty} \psi_k(y_1) dy_1 \right) \quad (13)$$

– для вероятности того, что обслуженная заявка оставляет узел пустым. Подставив последние формулы в (1) и (2) получим окончательные формулы для определения параметров выходного потока.

В случае сетевых моделей с неоднородными потоками, в которых заявки различаются маршрутами и законами поступления и обслуживания, параметры потоков различных классов ($m=1, \dots, M$) усредняются с целью приведения неоднородного потока к однородному. Эти параметры будут описывать так называемую обобщенную заявку. При этом соблюдается условие, чтобы однородный поток создавал такую же нагрузку в каждом узле сети, что и неоднородный поток. Для определения параметров потоков обобщенных заявок получены следующие формулы:

$$\lambda_i^{ob} = \sum_{m=1}^M \lambda_i^{(m)} \quad (14)$$

– для интенсивности поступления потока обобщенных заявок на вход i – го узла ($i=1, \dots, n$);

$$\bar{\tau}_{\mu i}^{ob} = \sum_{m=1}^M \left(\lambda_i^{(m)} / \mu_i^{(m)} \right) / \lambda_i^{ob} \quad \left(\mu_i^{(m)} = 1 / \bar{\tau}_{\mu i}^{(m)} \right) \quad (15)$$

$$D_{\mu i}^{ob} = \sum_{m=1}^M \left[D_{\mu i}^{(m)} \lambda_i^{(m)} + (\bar{\tau}_{\mu i}^{(m)} - \bar{\tau}_{\mu i}^{ob})^2 \lambda_i^{(m)} \right] / \lambda_i^{ob}$$

– для среднего и дисперсии времени обслуживания;

$$\lambda_{0i}^{ob} = \sum_{m=1}^M \lambda_{0i}^{(m)} \quad (16)$$

$$D_{0i}^{ob} = \sum_{m=1}^M \left(\lambda_{ji}^{(m)} / \lambda_{0i}^{(m)} \right)^3 D_{0i}^{(m)}$$

– для интенсивности и дисперсии потока от внешнего источника;

$$p_{ji}^{ob} = \sum_{m=1}^M p_{ji}^{(m)} \lambda_j^{(m)} / \lambda_j^{ob} \quad (17)$$

– для значений обобщенной матрицы вероятностей передач.

Итерационная процедура расчета сети с неоднородными потоками будет такая же, что и для однородного потока с приведенными параметрами, т. е. заключается в решении уравнений (5) и (6) совместно с (2). Используя полученные характеристики для отдельного узла сети, определяются характеристики сети для каждого типа заявок, аналогично сетям с однородными потоками.

Далее рассмотрим модели систем с ограниченной емкостью накопителя и с потерями (рисунок 2а траектория 2), а также с переменными параметрами законов поступления и обслуживания заявок (рисунок 2а траектория 3). Такие модели отражают процесс прохождения задач в отдельной ЭВМ и в более сложных комплексах – сетях передачи данных, имеющих буферные накопители ограниченной емкости. Последняя модель также отражает работу саморегулирующихся динамических систем.

Для определения характеристик моделей систем, содержащих узлы с потерями и ограничениями на длину очереди, сеть также декомпозируется на отдельные узлы на уровне двух моментов. При этом уравнения баланса (5) и (6) модифицируются: учитываются потоки отказов, идущие от узлов с потерями к другим узлам. Вводится матрица $Q = \{q_{ij}\}$ вероятностей передач заявок потоков отказов от i -го узла на вход j -го ($i, j=1, \dots, n$), а в системы (5) и (6) добавляются соответствующие слагаемые, учитывающие потоки отказов. Тогда уравнения баланса примут вид:

$$\lambda_{ibx} = \bar{\tau}_{oi}^{-1} + \sum_{j=1}^n P_{ij} \lambda_{jvbx} + \sum_{j=1}^n q_{ji} \lambda_{jotk} \quad (18)$$

$$D_{ibx} = \left(\frac{\bar{\tau}_{oi}^{-1}}{\lambda_{ibx}} \right)^3 D_{oi} + \sum_{j=1}^n \left(\frac{\lambda_{jvbx} P_{ji}}{\lambda_{ibx}} \right)^3 \frac{1}{P_{ji}} \left(D_{jvbx} + \frac{1 - P_{ji}}{\lambda_{jvbx}^2 P_{ji}} \right) + \sum_{j=1}^n \left(\frac{\lambda_{jotk} q_{ji}}{\lambda_{ibx}} \right)^3 \frac{1}{q_{ji}} \left(D_{jotk} + \frac{1 - q_{ji}}{\lambda_{jotk}^2 q_{ji}} \right), \quad (19)$$

Для определения среднего $\bar{\tau}_{отк}$ и дисперсии $D_{отк}$ времени между заявками в потоке отказов получены формулы:

$$\bar{\tau}_{отк} = \bar{\tau}_{ex} \bar{\tau}_{бвх} / (\bar{\tau}_{бвх} - \bar{\tau}_{ex}), \quad (20)$$

где $\bar{\tau}_{вх}$, и $\bar{\tau}_{вых}$ – средние времена между последующими заявками во входном и выходном потоках, и

$$D_{\text{поток}} = D_{N_{\text{отк}}} \bar{\tau}_{\text{отк}}^3 / T_{\text{ц}}, \quad (21)$$

где $T_{\text{ц}}$ – среднее время цикла занятости, $D_{N_{\text{отк}}}$ – дисперсия числа потерянных заявок на цикле занятости. Численное определение $D_{N_{\text{отк}}}$ выполняется аналогично вычислению распределения $\varphi_k(y_2)$: вычисляются условные распределения числа и квадрата числа потерянных заявок.

Основные характеристики для узла с потерями определяются с учетом того, что уравнение Колмогорова решается в области D , определенной двухсторонними условиями $N \geq 0$ и $N \leq m$, m – максимально допустимое число заявок в узле (рисунок 2а траектория 2). Граничные условия для уравнения (7) физически эквивалентны условиям поглощения на границе Γ_1 , где $\omega|_{\Gamma_1} = 0$ и отражения на границе Γ_2 , где $\text{grad}\omega|_{\Gamma_2} = 0$.

В задаче анализа систем с переменными параметрами законов поступления и обслуживания, зависящими от состояния системы (длины очереди к ресурсу), вышеуказанные распределения $\psi_k(y_1)$ и $\varphi_k(y_2)$ определяются отдельно для каждой из подобластей D_k^i ($k=1, \dots, m; i=1, 2, \dots$). Области D_k^i физически соответствует период функционирования узла, на котором в узле находится ровно k заявок (индекс k определяет уровень по оси x_1 , i – уровень по x_2). Этот период характеризуется своими интенсивностями поступления и обслуживания заявок, а также дисперсиями этих времен, зависящими от состояния системы. Такой подход означает решение уравнения Колмогорова с переменными коэффициентами сноса и диффузии, зависящими от длины очереди в узле (рисунок 2б).

Приведены результаты машинных экспериментов по оценке точности метода двумерного диффузионного приближения. Точность методов исследована для диапазонов изменения

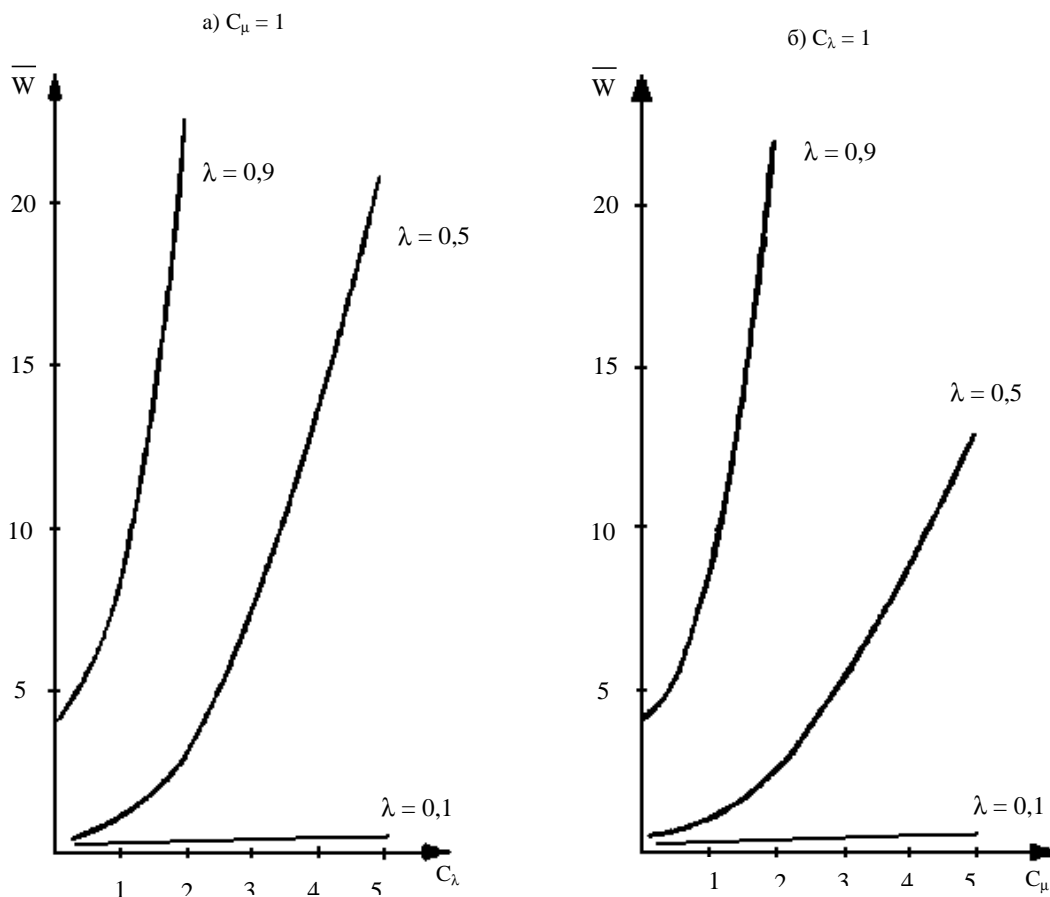


Рисунок 3. Зависимость времени задержки в узле при различных значениях интенсивности λ входного трафика (время обслуживания нормированное): а) – от коэффициента вариации входного потока; б) – от коэффициента вариации времени обслуживания.

коэффициента загрузки от 0,1 до 0,9 и коэффициентов вариаций распределений длин интервалов между заявками во входном потоке и времени обслуживания от 0 до 5. Значения среднего числа заявок в узле сравнивались с результатами имитационного моделирования.

В качестве одного из параметров моделирования задавалось количество циклов занятости, которое в зависимости от загрузки изменялось от 1000 до 20 000. Для отдельных узлов сети результаты двумерного диффузионного приближения совпадают с результатами имитационного моделирования в пределах 2-5% и в некоторых крайних случаях отклоняются до 10%. При расчете сетевых моделей добавляются ошибки декомпозиции, поэтому погрешность получалась несколько выше – до 15%.

Эмпирические распределения и гистограммы реальных потоков, полученные в результате измерений вычислительной нагрузки и потоков, циркулирующих в системе, также свидетельствуют об отличии потоков от пуассоновских, в особенности в системах с коммутацией

пакетов, в которых в режиме насыщения доминируют регулярные потоки. Исследованиями установлена нестационарность характеристик информационных потоков и неприемлемость аппроксимации гистограмм и эмпирических распределений экспоненциальным распределением. На основные показатели производительности систем, такие как время задержки сообщения, вероятность переполнения буферной памяти, время занятости канала и других существенное влияние оказывает вид распределения времени между соседними заявками во входном потоке и времени обслуживания. Эти факты отражены на рисунке 3.

Эксперимент также подтвердил, что моменты более высокого порядка на характеристики системы оказывают несущественное влияние. Сказанное подтверждает правомерность использования методов двумерной диффузионной аппроксимации как методов двухмоментной аппроксимации для анализа сетевых моделей сложных систем управления.

Список использованной литературы:

1. Кругликов В.К., Тарасов В.Н. Анализ и расчет сетей массового обслуживания методом двумерной диффузионной аппроксимации. – Изв. АН СССР – Автоматика и телемеханика, 1983. №8. – 74-84с.
2. Кругликов В.К., Тарасов В.Н. Приближенный метод декомпозиции разомкнутой стохастической сети. – Изв. АН СССР – Техническая кибернетика, 1983. №6. – 142-147 с.