

УДК 621.39

Приближенный анализ вероятностно-временных характеристик многофазной системы массового обслуживания

В. А. Наумов, П. О. Абаев

*Кафедра систем телекоммуникаций
Российский университет дружбы народов
Россия, 117198, Москва, ул. Миклухо-Маклая, 6*

В статье предложен метод расчёта основных вероятностно-временных характеристик системы массового обслуживания, состоящей из нескольких фаз и обслуживающей несколько потоков заявок. Получены приближенные формулы для среднего и квантилей времени пребывания заявки в системе, предложен алгоритм расчёта и проведён предварительный численный анализ.

КЛЮЧЕВЫЕ СЛОВА: многофазная система массового обслуживания, среднее время пребывания, квантиль времени пребывания.

Введение

В последнее время потребитель уже привык к мысли, что современная связь — это совокупность предоставляемых ему разнообразных услуг. Дальнейшее развитие телекоммуникационных сетей основывается на концепции построения сети связи следующего поколения (ССП), суть которой заключается в создании на основе существующих сетей IP универсальной сети, которая позволяла бы передавать любые виды информации, такие как речь, видео, аудио, графику, а, следовательно, предоставлять широкий спектр телекоммуникационных и мультимедийных услуг. При этом для инициации и модификации сеансов связи в сетях СПП международными организациями по стандартизации был принят протокол SIP (Session Initiation Protocol).

Хорошо известно, что качество функционирования сетей телекоммуникаций существенно зависит от своевременной и надёжной доставки сигнального трафика. Отличительной особенностью сигнализации с использованием протокола SIP является отсутствие выделенных канальных ресурсов для передачи сигнального трафика, который обслуживается на базе общих транспортных механизмов, реализованных в сети передачи данных общего пользования оператора связи на базе протокола IP. Однако не совсем ясно как переход от сигнализации по выделенным каналам к сигнализации по сети передачи данных общего пользования сказывается на качестве функционирования сигнализации и, в частности, на задержке сигнальных сообщений. Важными характеристиками качества сигнализации являются как средние задержки сигнальных сообщений, так и квантили функций распределения задержек.

В данной работе исследуется многофазная система массового обслуживания с несколькими входящими потоками. Предложены методы расчёта для среднего значения и квантилей времени пребывания заявки в системе. Проведён численный анализ погрешности получаемых значений формул. Разработанный метод применим для расчёта моделей передачи сигнальных сообщений протокола SIP по сети IP/MPLS [1].

1. Постановка задачи

Рассматривается система массового обслуживания, состоящая из N последовательных однолинейных СМО с накопителями неограниченной ёмкости и постоянным временем обслуживания (рис. 1). На систему поступают два типа заявок: основные и фоновые. Интенсивность потока и время обслуживания основных заявок в i -узле будем обозначать λ_0 и d_i , а интенсивность потока и время обслуживания фоновых заявок в i -узле — λ_i и b_i , $i = \overline{1, N}$. При этом поступающие на каждый узел потоки заявок предполагаются пуассоновскими.

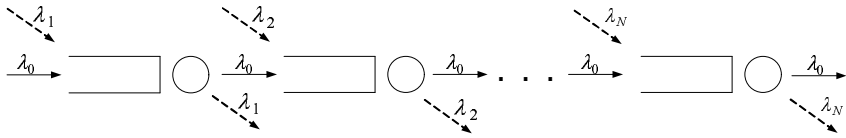


Рис. 1. Обслуживание последовательной цепочкой приборов

Пусть ξ_i — время пребывания заявки основного потока в i -узле, $i = \overline{1, N}$, ξ — время её пребывания в многофазной СМО, η_i — время обслуживания заявки, а ψ_i — время ожидания начала обслуживания заявки в i -узле, $i = \overline{1, N}$. Математические ожидания этих величин будем обозначать $M\psi_i = w_i$, $M\xi_i = D_i$, а функцию распределения (ФР) времени ожидания обозначим $W_i(x) = P\{\psi_i < x\}$. Для введённых случайных величин справедливы соотношения:

$$\begin{aligned} \xi_i &= \eta_i + \psi_i, \\ \xi &= \sum_{i=1}^N \xi_i. \end{aligned} \tag{1}$$

Целью исследований является разработка метода расчёта основных ВВХ СМО — среднего значения $D = M\xi$ и квантилей Q_p времени пребывания заявки основного потока в системе соответствующих уровню p .

2. Метод вычисления вероятностно-временных характеристик

Для расчёта величины среднего значения D времени пребывания заявок основного потока в системе воспользуемся соотношениями (1). Среднюю задержку заявок основного потока в i -узле можно вычислить по формуле $D_i = w_i + d_i$. Средняя задержка заявок основного потока в системе с момента их поступления на первый прибор и до момента завершения обслуживания на последнем приборе находится как сумма средних задержек во всех узлах системы

$$D = \sum_{i=1}^N D_i. \tag{2}$$

Для расчёта k -ого момента $m_i^{(k)} = M\eta_i^k$ времени обслуживания заявок η_i можно воспользоваться формулой

$$m_i^{(k)} = \frac{\lambda_0}{\lambda_0 + \lambda_i} \cdot d_i^k + \frac{\lambda_i}{\lambda_0 + \lambda_i} \cdot b_i^k, \quad k = 1, 2. \tag{3}$$

Отсюда вытекает, что для коэффициента вариации c_i длительности обслуживания в i -узле справедлива следующая формула:

$$c_i^2 = \frac{(\lambda_0 + \lambda_i)(\lambda_0 d_i^2 + \lambda_i b_i^2)}{(\lambda_0 d_i + \lambda_i b_i)^2} - 1. \tag{4}$$

В сделанных выше предположениях среднее время ожидания как основных, так и фоновых заявок в i -узле даётся известной формулой Поллачека–Хинчина [2]

$$w_i = \frac{\rho_i m_i}{1 - \rho_i} \left(\frac{1 + c_i^2}{2} \right), \quad (5)$$

где $\rho_i = \lambda_0 d_i + \lambda_i b_i$ — нагрузка на i -узел.

Для больших значений аргумента функция распределения $W_i(x)$ времени ожидания в i -узле хорошо аппроксимируется [3] экспоненциальной функцией вида

$$W_i(x) \approx 1 - \varphi_i e^{-v_i x}, \quad (6)$$

где v_i — параметр затухания и φ_i — асимптотический коэффициент. Параметр затухания v_i является единственным положительным числом и имеет минимальную действительную часть среди корней уравнения

$$\alpha(v_i) \cdot \beta(-v_i) = 1, \quad (7)$$

где $\alpha_i(s)$ — преобразование Лапласа–Стилтьеса (ПЛС) ФР интервалов между поступлениями заявок в i -узел, $\beta_i(s)$ — ПЛС ФР длительности обслуживания [4] в i -узле. ПЛС ФР интервалов между поступлениями заявок и ПЛС ФР длительности обслуживания в рассматриваемом случае имеют следующий вид

$$\alpha_i(s) = \frac{(\lambda_0 + \lambda_i)}{(\lambda_0 + \lambda_i) + s}, \quad (8)$$

$$\beta_i(s) = \frac{\lambda_0}{(\lambda_0 + \lambda_i)} \cdot e^{-s d_i} + \frac{\lambda_i}{(\lambda_0 + \lambda_i)} \cdot e^{-s b_i}. \quad (9)$$

Таким образом, уравнение (7) можно переписать в виде

$$\frac{\lambda_0}{(\lambda_0 + \lambda_i) + v_i} \cdot e^{d_i v_i} + \frac{\lambda_i}{(\lambda_0 + \lambda_i) + v_i} \cdot e^{b_i v_i} = 1. \quad (10)$$

Корень этого уравнения может быть найден численно с помощью итерационного алгоритма, например, методом бисекции. Для определения асимптотического коэффициента φ_i можно воспользоваться простой приближенной формулой, предложенной в [5]:

$$\varphi_i \approx v_i w_i. \quad (11)$$

Обозначив $T_i = \frac{1}{v_i} \ln(\varphi_i)$, формулу (6) можно записать в виде смещённого экспоненциального распределения

$$W_i(x) \approx 1 - e^{-v_i(x - T_i)}. \quad (12)$$

Нетрудно видеть, что эта аппроксимация эквивалентна представлению времени ожидания заявки в i -узле в виде суммы двух слагаемых. Первое слагаемое является случайной величиной имеющей, экспоненциальное распределение с параметром v_i , а второе слагаемое является постоянной величиной равной параметру сдвига T_i . Аналогично, в виде суммы двух слагаемых можно представить величину задержки заявок основного потока в i -узле. Первое слагаемое, как и при аппроксимации времени ожидания, является случайной величиной, имеющей экспоненциальное распределение с параметром v_i , а второе слагаемое является постоянной величиной, равной $T_i + d_i$. Если допустить, что длительности ожидания заявок в различных узлах являются независимыми величинами, имеющими смещённое экспоненциальное распределение (12), то функцию распределения задержки заявок основного потока в системе можно аппроксимировать смещённым распределением Эрланга порядка N со смещением $\Delta = \sum_{i=1}^N (T_i + d_i)$ и параметрами распределения Эрланга $v_1, v_2, \dots, v_N$.

3. Пример вычислений и анализ точности

Анализ проведём для случая, когда определённые выше величины не зависят от номеров узлов. Формулы для расчёта средних времён ожидания и пребывания заявки основного потока в системе примут вид

$$w = \frac{\rho m}{1 - \rho} \left(\frac{1 + c^2}{2} \right), \quad (13)$$

$$D = N (w + d). \quad (14)$$

Квантиль Q_p может быть вычислена по формуле

$$Q_p = q_p + N (T + d), \quad (15)$$

где q_p — квантиль уровня p распределения Эрланга с N одинаковыми этапами и параметром каждого этапа, равным v . Заметим, что квантиль q_p является единственным положительным корнем уравнения

$$(1 - p) e^{v q_p} = \sum_{k=0}^{N-1} \frac{(v q_p)^k}{k!}. \quad (16)$$

Для оценки точности формул за основу были приняты результаты имитационного моделирования.

Построенная математическая модель иллюстрирует процесс передачи сигнального сообщения SIP по IP/MPLS сети, поэтому расчёты проведём в предположении, что основной поток заявок — это сигнальные сообщения протокола SIP длиной 407 байт, а фоновый поток — это медиаконтент, передаваемый в пакетах протокола IP максимальной длины (1500 байт). В качестве среды передачи рассмотрим сеть с пропускной способностью 10 Мбит/с (10BASE-T), состоящую из а) 4 и б) 8 маршрутизаторов. Считаем, что основная нагрузка равна 0,4 Эрл, а величина фоновой нагрузки варьируется таким образом, чтобы суммарная нагрузка на каждой фазе модели не превышала значения 1. Среднее время обслуживания основных заявок вычисляется по формуле $d = \frac{407 \cdot 8 \cdot 1000}{V_{\text{Eth}}}$, а фонового трафика — $b_i = \frac{1500 \cdot 8 \cdot 1000}{V_{\text{Eth}}}$, $\forall i = \overline{1, N}$, где V_{Eth} — скорость передачи (Мбит/с).

На рисунках 2 и 3 представлены графики зависимости от нагрузки 95%-квантили времени пребывания заявки в системе и относительной погрешности вычислений.

Анализ показывает, что при невысокой загрузке сети IP/MPLS (до 0,7 Эрл) задержка сообщений не превосходит 10 мс во всем диапазоне исследуемых параметров, а квантили задержки не превышают 20 мс, что укладывается в нормы требований МСЭ-Т. Загрузка сети выше значения 0,7 Эрл. приводит к недопустимым задержкам сигнальных сообщений — до 100 мс. Приближенная формула (14) для среднего времени пребывания заявки позволяет получить значения, относительная погрешность которых в среднем не превышает 5%, для квантили этот показатель немного выше, однако не превосходит 10%.

Заключение

В статье изложены результаты исследований, связанных с разработкой методов расчёта основных ВВХ многофазных СМО с поступающими заявками нескольких типов. В работе были получены формулы для расчёта среднего и квантилей времени пребывания заявок в системе. Проведён сравнительный анализ с результатами имитационного моделирования, позволяющий говорить о получении хороших приближений для значений среднего и квантилей времени пребывания.

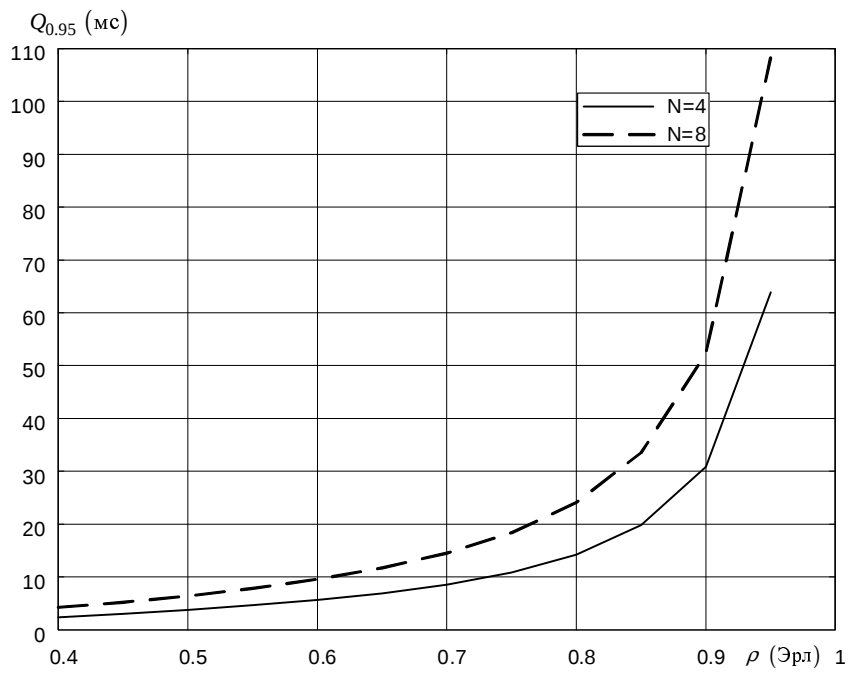


Рис. 2. 95%-квантиль времени пребывания заявки в системе

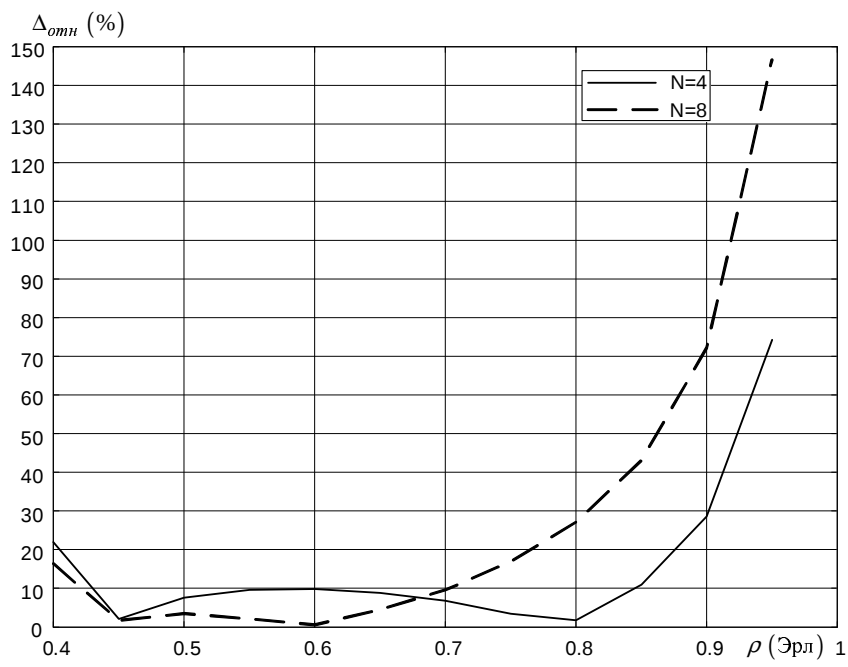


Рис. 3. Относительная погрешность расчётов 95%-квантили времени пребывания в зависимости от нагрузки

Литература

1. *Chakraborty S. S., Prasad R., Fathi H.* Optimization of SIP Session Setup Delay for VoIP in 3G Wireless Networks // IEEE Transactions on Mobile Computing. — Vol. 5, No 9. — 2006. — Pp. 1121–1132.
2. *Гнеденко Б. В., Коваленко И. Н.* Введение в теорию массового обслуживания. — М.: КомКнига, 2005. — 400 с.
3. *Fredericks A. A.* A Class of Approximations for the Waiting Time Distribution in a $GI|G|1$ Queueing System // Bell System Technical Journal. — Vol. 61, No 3. — 1982. — Pp. 295–325.
4. *Наумов В. А.* О максимальном полюсе преобразования Лапласа-Стилтьеса времени ожидания // Сб. «Модели систем распределения информации и их анализ». — М.: Наука, 1982. — С. 77–82.
5. Mean Waiting Time Approximation in the $G|G|1$ Queue / D. L. Jagerman, B. Balcioglu, T. Altioik, B. Melamed // Queueing Systems. — Vol. 46. — 2004. — Pp. 481–506.

UDC 621.39

Approximate Performance Analysis of Queues in Series

V. A. Naumov, P. O. Abaev

*Telecommunications Systems Department
Peoples' Friendship University of Russia
6, Miklukho-Maklaya str., Moscow, 117198, Russia*

Signaling system is the key element of the telecommunication infrastructure. Migration from traditional circuit-based networks to packet-based requires preserving quality and reliability of SS7. Hence, the performance analysis of SIP signaling system will play an essential role in optimizing network QoS provisioning. In this paper we investigate the behavior of SIP protocol over IP/MPLS network. In order to analyze SIP reliability and performance we develop an analytical model, which is multiphase queueing system with two incoming flows. Finally, we present formulas and numerical results to estimate mean and quantiles of sojourn time distribution.