

Т. И. АЛИЕВ, Л. А. МУРАВЬЕВА-ВИТКОВСКАЯ

## ПРИОРИТЕТНЫЕ СТРАТЕГИИ УПРАВЛЕНИЯ ТРАФИКОМ В МУЛЬТИСЕРВИСНЫХ КОМПЬЮТЕРНЫХ СЕТЯХ

Рассматриваются модели и методы исследования характеристик функционирования мультисервисных компьютерных сетей, в которых применяются приоритетные стратегии управления трафиком. Показано, что за счет изменения стратегий можно обеспечить выполнение требований к качеству обслуживания сетей.

*Ключевые слова:* мультисервисные компьютерные сети, приоритетные стратегии управления, качество обслуживания, неоднородный трафик, модель функционирования.

**Введение.** Мультисервисные компьютерные сети (КС) предназначены для передачи различных типов данных, образующих неоднородный трафик. Функционирование сетей подчинено требованиям, предъявляемым к качеству обработки и передачи данных каждого типа [1]. Одна из актуальных задач при построении мультисервисных КС — необходимость обоснованного выбора технических и программных средств, обеспечивающих заданное качество функционирования сетей при минимальных затратах. Для этого при разработке, модернизации и эксплуатации сетей необходимо располагать сведениями о влиянии различных способов их структурно-функциональной организации на такие характеристики функционирования КС, как время доставки пакетов и вероятность их потерь, загрузка узлов и каналов связи, производительность и т.п. Такие сведения могут быть получены лишь в том случае, если использовать системный подход и рассматривать технические и программные средства как единую систему. Методы анализа мультисервисных КС на основе вероятностного подхода к описанию протекающих в сетях процессов, ориентированные на применение сравнительно простых моделей их функционирования, являются весьма результативными, позволяя в компактной аналитической форме определить все важнейшие характеристики сети. Решение поставленных задач направлено на создание новых методов исследования мультисервисных КС, базирующихся на математическом аппарате теории вероятностей, случайных процессов и теории массового обслуживания.

**Требования к характеристикам мультисервисных КС.** В мультисервисных сетях осуществляется передача данных разных типов: видео- и аудиоданных, речи, компьютерных данных, к которым предъявляются различные требования по качеству обслуживания. Обслуживание данных заключается в обработке передаваемых пакетов в узлах сети и непосредственно в передаче пакетов по каналам связи. При этом задержка пакетов определяется как время прохождения пакетов между конечными узлами.

В рекомендациях [1] определены пять классов QoS и требования к характеристикам обслуживания пакетов каждого класса, выполнение которых гарантирует качественную передачу

соответствующего типа трафика. В качестве основных характеристик сети рассматриваются среднее время задержки пакета, вариация (джиттер) задержки и вероятности потерь пакетов и передачи ошибочных пакетов. Для классов 0 и 1 установлены следующие предельные (верхние) значения среднего времени (IPTD)  $u$  и вариации (IPDV)  $\sigma$  задержки пакета:  $u_0^* = 100$  мс;  $\sigma_0^* = 50$  мс;  $u_1^* = 400$  мс;  $\sigma_1^* = 50$  мс. Для классов 2, 3 и 4 установлены ограничения только на среднее время:  $u_2^* = 100$  мс;  $u_3^* = 400$  мс;  $u_4^* = 1$  с, а ограничения на вариацию задержки отсутствуют. Допустимые вероятности потерь и передачи ошибочных пакетов для всех классов одинаковы и составляют соответственно  $1 \cdot 10^{-3}$  и  $1 \cdot 10^{-4}$ .

Обеспечение требуемого качества функционирования мультисервисных сетей может быть достигнуто за счет применения приоритетных стратегий управления неоднородным трафиком.

**Модели функционирования мультисервисных КС.** В качестве моделей функционирования мультисервисных КС применяются сети массового обслуживания разных классов [2], представляющие собой совокупность взаимосвязанных систем массового обслуживания (СМО), рассматриваемых как базовые модели [3]. Неоднородностью нагрузки в мультисервисных сетях обуславливается необходимость использования в качестве базовых моделей СМО с неоднородным потоком заявок, обслуживание которых реализуется на основе приоритетных стратегий [4].

В качестве базовой модели мультисервисной КС будем использовать одноканальную СМО с накопителем неограниченной емкости и неоднородным потоком [5], в которую поступают  $H$  классов пакетов (заявок), образующие простейшие потоки с интенсивностями  $\lambda_1, \dots, \lambda_H$ . Длительность  $\tau_{b_k}$  обслуживания заявок (обработки или передачи пакета) класса  $k$  распределена по произвольному закону с функцией распределения  $B_k(\tau)$  и средним значением  $b_k^{(1)}$ . Заявки одного класса выбираются для обслуживания в порядке поступления. Обслуживание прерванной заявки продолжается с прерванного места.

**Методы расчета характеристик мультисервисных КС.** Математическое моделирование сетей предполагает применение аналитических и статистических (имитационных) методов. Выбор метода зависит от целей моделирования, сложности исследуемой системы и ее математической модели, требований к точности и общности результатов и т.д.

Применение аналитических методов моделирования обусловлено такими факторами, как возможность проведения всестороннего исследования свойств системы в широком диапазоне изменения параметров; малые затраты времени на получение результатов, что особенно важно на этапе предварительного проектирования, в процессе которого требуется получать оценки характеристик системы в сжатые сроки; возможность в приемлемые сроки рассмотреть несколько различных вариантов организации проектируемой системы в целях их сравнительного анализа.

Основное достоинство имитационных методов моделирования — возможность исследования систем любой сложности с любой степенью детализации.

Наиболее результативным на практике является комбинированный подход, использование которого позволяет разрабатывать одновременно одну или несколько аналитических и имитационных моделей. При этом имитационные модели применяются как для оценки погрешностей приближенных аналитических моделей, так и для детального анализа оптимальной сети, синтезированной на основе приближенных аналитических моделей.

Для рассматриваемой базовой модели были получены математические зависимости для расчета характеристик качества обслуживания в мультисервисной КС.

Выбор заявок из очереди на обслуживание осуществляется в соответствии с приоритетной стратегией управления трафиком, основанной на смешанных приоритетах (СП), задаваемых в виде матрицы приоритетов  $Q = [q_{ij}, i, j = 1, \dots, H]$ , элемент  $q_{ij}$  которой определяет приоритет заявок класса  $i$  по отношению к заявкам класса  $j$  и может принимать следующие значения: 0 — нет приоритета, 1 — приоритет относительный, 2 — приоритет абсолютный [3].

Среднее значение  $u_k^{(1)}$  и вариация  $\sigma_{u_k}$  задержки заявок (пакетов) класса  $k = 1, \dots, H$  определяются по формулам [3]

$$u_k^{(1)} = w_k^{(1)} + v_k^{(1)}; \quad \sigma_{u_k} = \sqrt{w_k^{(2)} + v_k^{(2)} - (w_k^{(1)} + v_k^{(1)})^2}, \quad (1)$$

где  $w_k^{(1)}$ ,  $w_k^{(2)}$  и  $v_k^{(1)}$ ,  $v_k^{(2)}$  — средние значения и вторые начальные моменты времени ожидания начала обслуживания и времени нахождения заявки класса  $k$  на обработке соответственно; время обработки включает в себя время обслуживания заявки в приборе и время ожидания в прерванном состоянии.

Выражения для  $w_k^{(1)}$ ,  $w_k^{(2)}$  и  $v_k^{(1)}$ ,  $v_k^{(2)}$  определяются на основе преобразований Лапласа соответствующих плотностей распределений, полученных в работе [3], и имеют следующий вид:

$$w_k^{(1)} = \frac{\sum_{i=1}^H r_6(i, k) \lambda_i b_i^{(2)}}{2(1 - R_k^{(4)})(1 - R_k^{(5)})}; \quad v_k^{(1)} = \frac{b_k^{(1)}}{(1 - R_k^{(3)})}; \quad (2)$$

$$w_k^{(2)} = \frac{\sum_{i=1}^H r_6(i, k) \lambda_i b_i^{(3)}}{3(1 - R_k^{(4)})^2 (1 - R_k^{(5)})} + \frac{\sum_{i=1}^H r_5(i, k) \lambda_i b_i^{(2)} \sum_{i=1}^H r_6(i, k) \lambda_i b_i^{(2)}}{2(1 - R_k^{(4)})^2 (1 - R_k^{(5)})^2} +$$

$$+ \frac{\sum_{i=1}^H r_4(i, k) \lambda_i b_i^{(2)} \sum_{i=1}^H r_6(i, k) \lambda_i b_i^{(2)}}{2(1 - R_k^{(4)})^3 (1 - R_k^{(5)})}; \quad (3)$$

$$v_k^{(2)} = \frac{b_k^{(2)}}{(1 - R_k^{(3)})^2} + \frac{b_k \sum_{i=1}^H r_3(i, k) \lambda_i b_i^{(2)}}{(1 - R_k^{(3)})^3}, \quad (4)$$

где  $b_i^{(2)}$  и  $b_i^{(3)}$  — начальные моменты длительности обслуживания заявок класса  $i$ :

$$b_i^{(l)} = \int_0^{\infty} \tau^l dB_i(\tau), \quad i, k = 1, \dots, H; \quad l = 1, 2, \dots$$

В выражениях (1)–(4) использованы следующие обозначения:  $r_g(i, k)$  — коэффициенты, принимающие значения 0 и 1 в зависимости от значений элементов  $q_{ik}$  матрицы приоритетов и позволяющие выделить заявки классов  $i$  и  $k$ , между которыми установлен тот или иной вид приоритета (относительный приоритет — ОП, абсолютный приоритет — АП, без приоритета — БП или любое их сочетание):  $r_1(i, k) = 0, 5(1 - q_{ik} - q_{ki})(2 - q_{ik} - q_{ki})$ ;  $r_2(i, k) = q_{ik}(2 - q_{ik})$ ;  $r_3(i, k) =$

$$= 0,5q_{ik}(q_{ik} - 1); \quad r_4(i, k) = r_2(i, k) + r_3(i, k); \quad r_5(i, k) = r_1(i, k) + r_2(i, k) + r_3(i, k); \quad r_6(i, k) = r_1(i, k) + r_2(i, k) + r_2(k, i) + r_3(i, k);$$

$$\Lambda_k^{(g)} = \sum_{i=1}^H r_g(i, k)\lambda_i, \quad R_k^{(g)} = \sum_{i=1}^H r_g(i, k)\rho_i \quad \text{— частичные суммарные}$$

интенсивности потоков заявок и суммарные загрузки соответственно, где  $\rho_i = \lambda_i b_i^{(1)}$  — загрузка, создаваемая заявками  $i$ -го класса,  $i, k = 1, \dots, H; g = 1, \dots, 6$ .

Выражения (1)—(4) позволяют рассчитать среднее значение и вариацию задержки пакетов при передаче в канале связи или в узле сети.

В таблице приведены средние значения  $u_k^{(1)}$  и вариации  $\sigma_k$  задержек пакетов при детерминированном обслуживании и использовании трех стратегий управления трафиком: 1) беспriorитетной; 2) с относительными приоритетами, назначенными в последовательности  $0 \rightarrow 2 \rightarrow 1 \rightarrow 3 \rightarrow 4$  (класс 0 имеет наивысший приоритет, а класс 4 — низший); 3) со смешанными приоритетами, где в отличие от предыдущей стратегии отсутствует приоритет между классами 0, 1 и 2:  $(0, 1, 2) \rightarrow (3 \rightarrow 4)$ .

| Класс | $\lambda_k, c^{-1}$ | $b_k^{(1)}, \text{мс}$ | $b_k^{(2)}, \text{мс}^2$ | Стратегия управления                        |   |   | $u_k^* / \sigma_k^*, \text{мс}$ |
|-------|---------------------|------------------------|--------------------------|---|---|---|---------------------------------|
|       |                     |                        |                          | $u_k^{(1)\text{БП}} / \sigma_k^{\text{БП}}$ | $u_k^{(1)\text{ОП}} / \sigma_k^{\text{ОП}}$ | $u_k^{(1)\text{СП}} / \sigma_k^{\text{СП}}$ |                                 |
| 0     | 5                   | 25                     | 625                      | <b>185/181</b>                              | 55/25                                       | 74/45                                       | 100/50                          |
| 1     | 3                   | 75                     | 5625                     | 235/181                                     | 139/66                                      | 124/45                                      | 400/50                          |
| 2     | 4                   | 30                     | 900                      | <b>190/181</b>                              | 69/33                                       | 79/45                                       | 100/—                           |
| 3     | 4                   | 75                     | 5625                     | 235/181                                     | 285/267                                     | 285/267                                     | 400/—                           |
| 4     | 1                   | 70                     | 4900                     | 230/181                                     | 763/1105                                    | 763/1105                                    | 1000/—                          |

Как следует из таблицы, только стратегия, использующая смешанные приоритеты, обеспечивает выполнение заданных ограничений  $u_k^*$  и  $\sigma_k^*$ ,  $k = 0, 4$ , для всех классов неоднородного трафика.

Для оценки вероятности потери пакетов из-за переполнения буферной памяти в узлах сети можно воспользоваться зависимостью, связывающей производящую функцию  $M_k^*(z)$  числа заявок класса  $k$  в системе с преобразованием Лапласа  $U_k^*(s)$  функции распределения  $U_k(\tau)$  времени пребывания заявок в системе [3]:

$$M_k^*(z) = U_k^*(\lambda_k - \lambda_k z), \quad k = 1, \dots, H. \quad (5)$$

Продифференцировав выражение (5) по  $z$  в точке  $z = 1$ , получим зависимости, связывающие соответствующие начальные моменты числа заявок и времени их пребывания в системе. В частности, для двух первых моментов

$$m_k^{(1)} = \lambda_k u_k^{(1)}; \quad m_k^{(2)} = \lambda_k^2 u_k^{(2)} + m_k^{(1)}, \quad k = 1, \dots, H. \quad (6)$$

Для заданного значения емкости  $E_k$  буферной памяти, отводимой для пакетов класса  $k$ , используя неравенство Чебышева, можно получить оценку вероятности потери  $\varepsilon_k$  как вероятность превышения числа пакетов ( $X_k$ ) класса  $k$  над значением  $E_k$  в узле сети:

$$\varepsilon_k = P(X_k > E_k) < \frac{\lambda_k (\lambda_k \sigma_k^2 + u_k^{(1)})}{(E_k - \lambda_k u_k^{(1)})^2}.$$

Таким образом, выполнение заданных требований к качеству обслуживания в мультисервисной компьютерной сети может быть обеспечено за счет применения стратегий

управления трафиком с динамически изменяющимися приоритетами в зависимости от разных факторов, в частности от времени нахождения пакетов в узле сети. В работе [6] получены рекуррентные формулы для расчета среднего значения времени задержки пакетов разных классов и показано, что введение динамических приоритетов позволяет уменьшить разброс средних значений времени по сравнению со статическими приоритетами. Кроме того, при использовании динамических приоритетов переход от одной дисциплины обслуживания со смешанными приоритетами к другой обеспечивает плавное (непрерывное) изменение характеристик пакетов разных классов, тогда как в классе дисциплин обслуживания со статическими приоритетами это изменение происходит скачкообразно.

**Заключение.** Рассмотренная стратегия управления трафиком на основе смешанных приоритетов позволяет обеспечить требуемые характеристики обслуживания неоднородного трафика в мультисервисных КС за счет эффективного распределения приоритетов между потоками данных разных типов. Разработанные методы расчета позволяют оценить средние значения и вариацию задержки приоритетных потоков пакетов. Моделирование мультисервисных КС с использованием сетевых моделей с большим количеством узлов и структурно-функциональных параметров делает задачу аналитического исследования малоэффективной из-за громоздкости математических выкладок и невозможности получить точные аналитические результаты в явном виде. В этом случае предпочтительным становится использование имитационных методов моделирования, что позволяет получать результаты с высокой точностью и степенью достоверности.

#### СПИСОК ЛИТЕРАТУРЫ

1. ITU-T Recommendation Y.1541 (02/2006) – Network performance objectives for IP-based services // Int'l Telecommunication Union. 2006. Febr.
2. Алиев Т. И., Никульский И. Е., Пяттаев В. О. Моделирование ядра мультисервисной сети с относительной приоритезацией неоднородного трафика // Науч.-техн. вестн. СПбГУ ИТМО. 2009. Вып. 04(62). С. 88—96.
3. Алиев Т. И. Характеристики дисциплин обслуживания заявок с несколькими классами приоритетов // Изв. АН СССР. Сер. Техн. кибернетика. 1987. № 6. С. 188—191.
4. Алиев Т. И., Новиков Г. И. Метрическая теория и мониторинг компьютерных систем: состояние и проблемы // Изв. вузов. Приборостроение. 2000. Т. 43, № 3. С. 40—44.
5. Алиев Т. И. Основы моделирования дискретных систем: Учеб. пособие. СПб: СПбГУ ИТМО, 2009. 363 с.
6. Алиев Т. И., Муравьева Л. А. Система с динамически изменяющимися смешанными приоритетами и ненадежным прибором // Автоматика и телемеханика. 1988. № 7. С. 99—106.

#### Сведения об авторах

**Тауфик Измайлович Алиев**

— д-р техн. наук, профессор; Санкт-Петербургский государственный университет информационных технологий, механики и оптики, кафедра вычислительной техники; E-mail: aliev@d1.ifmo.ru

**Людмила Александровна Муравьева-Витковская**

— канд. техн. наук, доцент; Санкт-Петербургский государственный университет информационных технологий, механики и оптики, кафедра вычислительной техники; E-mail: mur-lada@yandex.ru

Рекомендована кафедрой  
вычислительной техники

Поступила в редакцию  
18.01.11 г.