

## ДИНАМИЧЕСКОЕ ИЗМЕНЕНИЕ ПРИОРИТЕТОВ ДЛЯ ОБЕСПЕЧЕНИЯ ТРЕБУЕМОГО УРОВНЯ КАЧЕСТВА ОБСЛУЖИВАНИЯ В КОМПЬЮТЕРНЫХ СЕТЯХ

В. В. Соснин (Санкт-Петербург)

### Введение

При построении и эксплуатации современных компьютерных сетей с мультимедийным неоднородным трафиком одной из важных задач является обеспечение заданного качества обслуживания (QoS) для различных типов трафика. Для оценки уровня QoS мультимедийного трафика в качестве основных характеристик используются среднее значение задержки  $U$  пакета в сети и вариация  $J$  этой задержки [1]. Выполнение требований QoS может быть реализовано за счет применения различных приоритетных дисциплин обслуживания (ДО) пакетов, в том числе с приоритетами, изменяющимися динамически [2]. Изменение параметров ДО с динамическими приоритетами (ДП) приводит к перераспределению значений  $U$  и  $J$  между различными типами трафика, что позволяет обеспечить требуемый уровень QoS. Для расчетов значений  $U$  каждого из типов трафика можно воспользоваться известными законами сохранения среднего значения задержки [3], которые не зависят от применяемой ДО. В то же время подобные законы сохранения для вариации задержки  $J$  автору неизвестны. В статье делается попытка вывести такой закон для ДО с ДП.

Отличительной особенностью данной работы является рассмотрение неэкспоненциальных законов распределений параметров исследуемой системы. Часто утверждается, что в компьютерных сетях потоки пакетов не являются простейшими, что делает невозможным применение аналитических методов теории массового обслуживания. Наиболее эффективным в этом случае является применение имитационного моделирования, используемого в данной работе.

В рамках работы поставлены и решены следующие задачи: 1) создание имитационной модели промежуточного узла компьютерной сети, использующего ДО с ДП; 2) проведение экспериментов с моделью и анализ свойств ДО с ДП; 3) формулирование рекомендаций по применению ДО с ДП для обеспечения требуемого уровня характеристик QoS.

### Постановка задачи

Исследования проводились при следующих предположениях и допущениях.

1. Не учитывается влияние сложности технической реализации ДО (в реальных условиях полученные с помощью ДО с ДП преимущества могут нивелироваться в результате её значительных требований к вычислительным ресурсам промежуточного узла сети).

2. Промежуточные узлы сети рассматриваются как одноканальные системы обслуживания (реальная техническая реализация маршрутизаторов и коммутаторов может потребовать применения других моделей).

3. Варьирование нагрузки на промежуточный узел проводится за счет изменения интенсивности потока пакетов (результаты могут оказаться иными, если варьировать загрузку, изменяя скорость обработки пакетов в узлах, однако эта ситуация менее интересна в практическом смысле, т.к. в реальных системах изменить производительность работающего устройства невозможно, а изменение загрузки варьированием размера пакета, как показывает анализ статистической информации в [4], оказывает очень незначительное влияние на характер задержек).

4. Исследуются характеристики QoS при прохождении пакетов только через один промежуточный узел сети (методика расчета общесетевых характеристик на основе сведений об узловых характеристиках приведена в [1]).

5. Предполагается, что в промежуточных узлах отсутствуют потери пакетов в результате переполнения буферов.

6. Динамический приоритет рассматривается в относительном смысле, т.е. отсутствуют прерывания в обслуживании текущего пакета и уровень приоритета учитывается только в момент выбора очередного пакета из входного буфера.

7. В качестве закона распределения параметров используется гамма-распределение, позволяющее изменять коэффициент вариации в диапазоне от 0 до  $+\infty$ .

При ДО с ДП приоритет пакетов изменяется в соответствии с функцией приоритетности  $\varphi(t) = k_i(t - t_j)$ , где  $t_j$  – момент поступления  $j$ -го пакета в систему,  $t$  – текущий момент времени,  $k_i \geq 0$  – коэффициент приоритетности, определяющий скорость изменения функции приоритетности пакета  $i$ -го класса ( $i = \overline{1, H}$ ),  $H$  – число классов пакетов. После завершения обслуживания текущего пакета, из очереди выбирается пакет, имеющий наибольшее значение приоритета.

Промежуточный узел сети рассматривается как одноканальная система массового обслуживания (СМО) с бесконечным накопителем. Поступающие в узел пакеты разных типов образуют неоднородный поток заявок. Ниже используются следующие обозначения:  $G_N/G_N/1$  – класс СМО, задающий в символике Кендалла [5] СМО  $G/G/1$  с неоднородным потоком из  $N$  классов заявок;  $b_i$  – среднее время обслуживания заявок  $i$ -го класса,  $a_i$  – средний интервал времени между поступлением заявок  $i$ -го класса,

$\rho_i = \frac{b_i}{a_i}$  – нагрузка, создаваемая заявками  $i$ -го класса,  $V_{bi}$  и  $V_{ai}$  – коэффициенты вариации интервалов между прибытием заявок и времени обслуживания  $i$ -го класса. Для простоты изложения в статье приводятся результаты, полученные для СМО  $G_2/G_2/1$ , которые, как показали результаты экспериментов, легко обобщаются на весь класс систем  $G_N/G_N/1$ .

Для удобства графического представления характеристик СМО  $G_2/G_2/1$  используется параметр  $K = \frac{k_1 - k_2}{k_1 + k_2}$ , причем  $K = -1$  означает, что 2-й класс имеет ОП по отношению к 1-му классу,  $K = 1$  – 1-й класс имеет ОП по отношению к 2-му классу,  $K = 0$  – соответствует бесприоритетной (БП) ДО. При прочих промежуточных значениях  $K$  реализуется ДО ДП. Варьируя  $K$ , можно осуществить плавный переход от ДО БП к ДО ОП. При  $-1 \leq K < 0$  заявки первого класса являются заявками низкоприоритетного класса (ЗНК), а заявки второго класса – заявками высокоприоритетного класса (ЗВК). При  $0 < K \leq 1$  ситуация обратная. Через  $U_i$  обозначим среднее значение, а через  $J_i$  – вариацию задержки заявок  $i$ -го класса. В приводимых результатах имитационного моделирования 95%-й доверительный интервал составляет не более 2% от указываемых значений, поэтому он не показывается на графиках. Расчет вариации задержки выполняется по формуле [1]:  $J = M + \sigma \cdot (3.090 + 1.425 \cdot \frac{\beta_3}{\sigma^3})$ , где  $M$  и  $\sigma$  – соответственно математическое ожидание и среднеквадратическое отклонение задержки. В последнем выражении  $\beta_3 = \alpha_3 + 2M^3 - 3M \cdot \alpha_2$ , где  $\alpha_n$  –  $n$ -й начальный момент задержки.

Имитационная модель описанной СМО была создана в системе GPSS World, в которой отсутствуют штатные средства для реализации ДП [2, 7–9]. Поэтому ниже представлен центральный фрагмент разработанной GPSS-модели, поясняющий способ реализации ДО с ДП.

```

GENERATE v$timeA1          ; создать поток с 1-м классом пакетов
ASSIGN class,1             ; задать уникальный номер класса
ASSIGN koef,0.1           ; назначить коэффициент приоритетности равным 0.1
ASSIGN servicetime,v$timeB1 ; установить предстоящее время обработки в узле
TRANSFER ,beginning       ; перейти к узлу сети

GENERATE v$timeA2          ; создать поток со 2-м классом пакетов
ASSIGN class,2             ; задать уникальный номер класса
ASSIGN koef,1.5           ; назначить коэффициент приоритетности равным 1.5
ASSIGN servicetime,v$timeB2 ; установить предстоящее время обработки в узле
TRANSFER ,beginning       ; перейти к узлу сети

beginning  QUEUE p$class          ; отметить время начала для сбора статистики
           ASSIGN prior,0         ; установить начальное значение приоритета
           MARK                   ; запомнить время попадания пакета в узел
entrance   LINK chain,p$prior,getfacil ; поставить пакет в общую очередь
getfacil   SEIZE p$ribor;         ; занять пакетом прибор (узел сети)
           DEPART p$class         ; отметить время конца для сбора статистики
           ADVANCE p$servicetime ; выполнить обработку пакета
           RELEASE p$ribor        ; освободить прибор
           UNLINK chain,recalc,ALL ; послать все пакеты из общей очереди на _
                                   ; _пересчет приоритета по метке "recalc"
           PRIORITY 0,BU          ; задержать обработанный пакет, пока не _
                                   ; _пересчитается приоритет ожидавших
           UNLINK chain,getfacil,1,BACK ; обработанный пакет выпускает из общей _
                                   ; _очереди пакет с максимальным приоритетом
quit       TERMINATE 1           ; уничтожить обработанный пакет

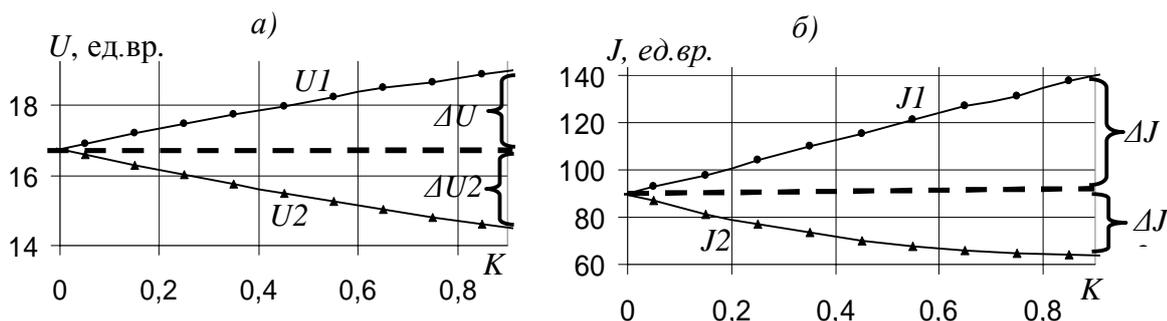
recalc     ASSIGN prior,(M1#p$koef) ; пересчитать приоритет пакета
           TRANSFER ,entrance      ; отправить пакет обратно в общую очередь

```

Важное преимущество описанного способа реализации ДО с ДП – его хорошая масштабируемость: для увеличения числа классов потоков достаточно вставить в начало модели копию блоков с 1-го (generate) по 5-й (transfer), задав нужные параметры новому классу. Модель также позволяет очень просто изменить вид функции приоритетности: для этого достаточно изменить строку с меткой recalc, в которой выражение в скобках задает формулу расчета  $\varphi(t)$ , а M1 есть не что иное, как разность  $(t - t_j)$ . Отметим также возможность раздельного сбора статистики для каждого из классов пакетов с помощью блока QUEUE (в приведенном примере осуществляется сбор статистики времени ожидания в буферной памяти независимо для каждого класса). Верификация модели проводилась путем сравнения результатов имитационного моделирования с результатами аналитического моделирования, полученными в [3, 5, 6] при использовании простейших потоков и экспоненциальных законов распределения нагрузочных параметров. Сравнение результатов оказалось возможным только для значений  $U$ , т.к. для  $J$  аналитических результатов получено не было даже для случая  $M_H/M_H/1$ .

## Обсуждение результатов

На рисунке представлены результаты имитационного моделирования СМО  $G_2/G_2/1$ , в которой нагрузочные параметры классов симметричны:  $v_{bi} = v_{ai} = 0.7$ ,  $a_1 = a_2 = 3$  с,  $b_1 = b_2 = 10$  с,  $\rho_1 = \rho_2 = 0.3$ . Рис. а) иллюстрирует закон сохранения  $U$  [3]: уменьшение средней задержки ЗВК  $U_1$  достигается за счет эквивалентного увеличения средней задержки ЗНК  $U_2$  (т.е.  $\Delta U_1 = \Delta U_2$ ). Однако вариация задержки ЗВК  $J_1$ , как показано на рис. б), изменяется быстрее, чем вариация задержки ЗНК  $J_2$  ( $\Delta J_1 > \Delta J_2$ ). Отсюда следует, что закон сохранения для  $J$  не может формулироваться так же, как и для  $U$ .



Результаты моделирования

а – среднее значение задержки; б – вариация задержки

Ввиду того, что существующие аналитические методы [3, 5, 6] не позволяют дать численную оценку  $J$  для каждого из классов, автором ставилась задача на основе анализа большого числа имитационных экспериментов выявить некоторую закономерность, которая может быть сформулирована как закон сохранения для вариации задержки. В ходе работы были проанализированы значения  $J_1$  и  $J_2$  для более чем 600 различных сочетаний нагрузочных параметров СМО: значения  $v_{ai}$  и  $v_{bi}$  варьировались от 0.1 до 2.1 с шагом 0.2 при разных соотношениях загрузок  $\rho_1/\rho_2$ : 0.1/0.5, 0.4/0.4, 0.2/0.6, 0.2/0.2. Для каждого из таких сочетаний проводилось изменение  $K$  от -0.9 до 0.9 с шагом 0.1. В результате была выявлена закономерность, которую можно рассматривать как закон сохранения вариации задержек при использовании ДО с ДП:  $\rho_1 \cdot \log(J_1) + \rho_2 \cdot \log(J_2) \approx \text{const}$  (при любом  $K$ ). В общем случае для систем с  $N$  классами заявок имеем:  $\sum_{i=1}^N (\rho_i \cdot \log(J_i)) \approx \text{const}$ . Поскольку в современных компьютерных сетях обычно используется не более четырех типов мультимедийного трафика (речь, видеоконференция, обычное видео, видео высокой четкости), то проверка выполнения полученного закона сохранения проводилась для случаев  $N \leq 4$ .

Следующим этапом работы стало получение строгой оценки справедливости предложенного закона сохранения (т.е. определения пределов варьирования константы в правой части уравнения). Результаты удалось обобщить формулой

$$\sum_{i=1}^n (\rho_i \cdot \log(J_i)) = C \cdot (1 \pm 0.02), \quad (1)$$

где  $C$  – константа, меняющаяся не более чем на 2% ( $\Delta = 0.02$ ). Основание логарифма в данном случае не имеет значения, т.к. может быть приведено к любому другому.

Формула (1) позволяет оценить эффект от введения ДО с ДП. Рассмотрим пример, иллюстрирующий ее применение. Пусть в исследуемый узел сети поступает два класса потоков пакетов и используется ДО с ОП, а измерения показали, что  $J_1 = 100$  мс,  $J_2 = 25$  мс,  $\rho_1 = 0.7$ ,  $\rho_2 = 0.2$ . Пусть второй поток – это речевой трафик приложения IP-телефонии, для которого в [1] рекомендуется значение  $J \leq 50$  мс. Очевидно, что значение  $J_2$  обеспечивается с некоторым запасом. Рассчитаем, на сколько можно уменьшить  $J_1$  при увеличении  $J_2$  до рекомендованного значения 50 мс при помощи использования ДО с ДП. В соответствии с формулой (1) имеем:  $C = 0.7 \cdot \ln(100) + 0.2 \cdot \ln(25) \approx 3.867$ . Вследствие погрешности  $\Delta$  реальное значение  $C^*$  может находиться в интервале от  $C \cdot \left(1 - \frac{2\Delta}{1+\Delta}\right)$  до  $C \cdot \left(1 + \frac{2\Delta}{1-\Delta}\right)$ , или округленно:  $C^* \approx C \cdot (1 \pm 2\Delta)$ . Тогда при  $J_2 = 50$  мс значение вариации задержки  $J_1$  будет равно  $J_1 = \exp\left(\frac{C(1 \pm 2\Delta) - \rho_2 \cdot \ln(J_2)}{\rho_1}\right) = \exp\left(\frac{3.867(1 \pm 0.04) - 0.2 \cdot \ln(50)}{0.7}\right) \approx 83 \pm 15$  мс.

### Заключение

Предложен хорошо масштабируемый способ реализации нестандартной для GPSS World имитационной модели узла телекоммуникационной сети с динамическими приоритетами. На основе анализа результатов многочисленных имитационных экспериментов сформулирован закон сохранения вариации задержки, который позволяет оценивать эффект от применения ДО с ДП при различных законах распределений нагрузочных параметров трафика сети.

### Литература

1. ITU-T recommendation Y.1541. Amendment 1: An example showing how to calculate IPDV across multiple sections. Geneva, 2006.
2. **Рыжиков Ю. И.** Имитационное моделирование. Теория и технологии. СПб.: КОРОНА принт; М.: Альтекс-А, 2004. 384 с.
3. **Клейнрок Л.** Вычислительные системы с очередями. М.: Мир, 1979. 600 с.
4. The PingER project: <http://www-iepm.slac.stanford.edu/pinger/> – ссылка актуальна на 10 сентября 2009 г.
5. **Шнепс М. А.** Системы распределения информации. Методы расчета. М.: Связь, 1979. 344 с.
6. Приоритетные системы обслуживания / **Б. В. Гнеденко, Э. А. Даниелян, Б. Н. Дмитриев, Г. П. Климов., В. Ф. Матвеев.** М.: Изд-во Московского ун-та, 1973. 448 с.
7. **Бражник А. И.** Имитационное моделирование: возможности GPSS WORLD. СПб.: Реноме, 2006. 439 с.
8. **Томашевский В., Жданова Е.** Имитационное моделирование в среде GPSS. М.: Бестселлер, 2003. 416 с.
9. GPSS World Reference Manual. Holly Springs, NC, USA: Minuteman Software, 2009. – актуально на 10 сентября 2009 года по Интернет-ссылке [http://www.minutemansoftware.com/reference/reference\\_manual.htm](http://www.minutemansoftware.com/reference/reference_manual.htm)