Вестник УГАТУ
Уфа : УГАТУ, 2011 Т. 15, № 5 (45). С. 22–26

УДК 621.396.67

В. Х. Багманов, А. Х. Султанов, А. М. Комиссаров

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ИСПОЛЬЗОВАНИЯ СИСТЕМЫ ОБСЛУЖИВАНИЯ ПОТОКА ЗАЯВОК С ПРИОРИТЕТАМИ НА ОСНОВЕ СТАТИСТИЧЕСКИХ ДАННЫХ

В статье приводятся методы определения оптимальных долей низкоприоритетного потока, обслуживаемого параллельными приборами при наличии только статистических данных о потоках заявок. Приводится сравнение результатов расчетов системы обслуживания с моделью, использующей информацию о загруженности очередей приборов. *Модель системы массового обслуживания с приоритетами*; *распределение потоков*; *цикл обслуживания*

ВВЕДЕНИЕ

С целью повышения эффективности использования систем массового обслуживания (вычислительные комплексы, сети передачи данных, транспортные системы) используют параллельную обработку и в зависимости от свойств заявок – обслуживание по приоритету. Заявки образуют поток: поступают через случайные временные интервалы, длительность облуживания также случайна.

Распределительное устройство решает, в какую из альтернативных цепочек приборов поступит заявка, причем цепочки могут иметь разную производительность, с целью минимизации среднего времени нахождения заявок в системе. В [1, 2] показано, что наилучший результат достигается при учете длины очереди и пропускной способности прибора в каждом пути, такие учитываемые параметры называют метрикой.

Однако бывают ситуации (в сети передачи данных), когда доводить информацию о текущей длине очередей очень накладно, либо информация при достижении распределительного устройства оказывается неактуальной, в этом случае учитываются статистические характеристики потока заявок и производительность приборов. Такая задача называется задачей оптимального распределения потоков, ее решение приводится в [3, 4] для модели потока M/M/1. Если в системе имеются заявки двух приоритетов, то на распределение потока низкоприоритетных заявок будет влиять поведение приоритетных потоков, задача усложняется, хотя теория систем массового обслуживания с приоритетами подробно описана [2, 5]. При множестве обслуживающих приборов, параллельных цепей обслуживания при наличии разных приоритетов заявок задача распределения потоков становится очень трудоемкой, необходимо применять достаточно простой метод расчета долей потока, позволяющий повысить эффективность использования сети.

В статье приводятся методы определения оптимальных долей низкоприоритетного потока, обслуживаемого параллельными приборами при наличии только статистических данных о потоках заявок. Приводится сравнение результатов расчетов системы обслуживания с моделью, использующей информацию о загруженности очередей приборов.

1. ЗАДАЧА РАСПРЕДЕЛЕНИЯ ПОТОКОВ С ПРИОРИТЕТАМИ

В задачах распределения потоков учитываются параметры потока заявок, обычно в качестве модели обслуживания потока заявок в приборах используют модель M/M/1 – интенсивность поступления заявок является пуассоновским процессом, а длительность обслуживания заявки распределено по экспоненциальному закону [2, 4]. В зависимости от топологии сети, пропускных способностей и ограничения на количество цепей обслуживающих приборов определяют, какие доли исходных потоков по каким путям проходят, минимизируя общее среднее время задержки при прохождении через систему обслуживания:

$$T = \frac{1}{\gamma}\sum_{k=1}^{N}\sum_{l=1}^{N}\frac{f_{kl}}{C_{kl} - f_{kl}} \to \min, \qquad (1)$$

где N – число обслуживающих приборов, γ – полный внешний поток, поступающий в систему (заявка/с), C_{kl} – пропускная способность цепи обслуживающих приборов kl (заявка/с), f_{kl} – искомый поток, проходящий по линии kl (заявка/с).

Контактная информация: (347)273-06-89

Сначала находят минимальные пути, при $f_{kl} = 0$, по метрике

$$w_{kl} = \frac{\partial T}{\partial f_{kl}} = \frac{C_{kl}}{(C_{kl} - f_{kl})^2}. \quad (2)$$

Затем, изменяя величину f_{kl}, находят минимум выражения (1), однако распределяющее устройство не идентифицирует потоки, а принимает решение для конкретной заявки на основе текущей информации о параметрах системы и каждая заявка направляется в соответствии с метрикой. В [1, 4] предлагается для каждого пакета выбирать направление по наименьшей метрике:

$$d_{kl} = \frac{\partial T}{\partial f_{kl}} = \frac{n_{kl}+1}{1-\rho_{kl}} \frac{1}{C_{kl}}, \quad (3)$$

где n_{kl} – число заявок, находящихся в очереди в обслуживающем приборе kl, $\rho_{kl} = f_{kl} / C_{kl}$ – коэффициент загрузки прибора kl. Производная задержки по потоку может быть интерпретирована как остаточный период занятости прибора.

Рассмотрим пример на рис. 1, поток λ передается по системе, на участке между распределяющими устройствами A и B имеются два пути, нужно распределить исходный поток по параллельно расположенным приборам с целью минимизации времени задержки. Решение для данного примера получается из равенства метрик вида (2)

$$f_1 = \frac{\sqrt{C_1}\left(\lambda - \left(C_2 - \sqrt{C_1 C_2}\right)\right)}{\sqrt{C_1} + \sqrt{C_2}}, \quad (4)$$

$$f_2 = \frac{\sqrt{C_2}\left(\lambda - \left(C_1 - \sqrt{C_1 C_2}\right)\right)}{\sqrt{C_1} + \sqrt{C_2}}.$$

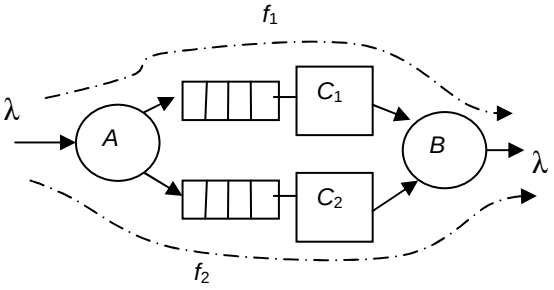

Рис. 1. Распределение потока λ по двум путям

На рис. 2 представлен график распределения оптимальных путевых потоков, при небольшом потоке используется только одна линия с большей пропускной способностью, когда входной поток превышает порог $C_1 - \sqrt{C_1 C_2}$, часть его направляется по второму пути.

Алгоритм распределения может направить заявки по путям, при алгоритме, использующем в качестве метрики (3), учитываются: длина очереди заявок в буфере, интенсивность поступающего потока и пропускная способность прибора. В качестве метрики для алгоритма распределения можно взять величину $(n_{kl} + 1)/C_{kl}$, каждому алгоритму распределения будет соответствовать некоторое распределение потоков внутри сети. Оптимальный алгоритм распределения не обязательно дает решение задачи оптимального распределения потоков [1].

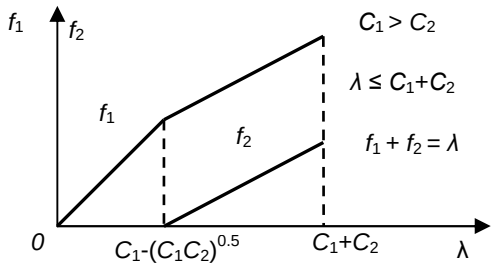

Рис. 2. Оптимальные путевые потоки f_1 и f_2 для рассматриваемого примера

Рассмотрим подобный случай для сети с приоритетами, представленный на рис. 3. Между узлами A и B имеются два параллельно работающих прибора, по ним проходят потоки λ_1 и λ_2, и подается третий поток λ_3, заявки которого имеют меньший приоритет, необходимо распределить последний поток между двумя путями так, чтобы минимизировать время задержки.

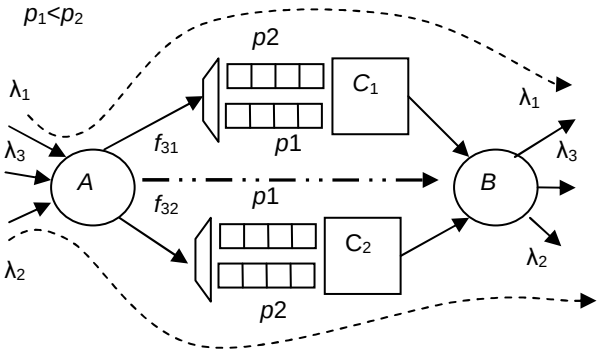

Рис. 3. Распределение низкоприоритетного потока по двум приборам с приоритетами

Среднее время, проводимое заявкой определенного приоритета в очереди, определяется по формуле

$$W_p = \frac{\sum_{i=1}^{P} \rho_i \frac{\overline{x_i^{(2)}}}{\overline{x_i}}}{\left(1 - \sum_{i=p}^{P} \rho_i\right)\left(1 - \sum_{i=p+1}^{P} \rho_i\right)}, \quad p=1, 2,\ldots, P, \quad (5)$$

где $\overline{x_i}$ – первый момент времени обслуживания требования из класса i; $\overline{x_i^2}$ – второй момент времени обслуживания требования из класса i; $\rho_i = \lambda_i \overline{x_i}$ – коэффициент использования потока i-го приоритета (коэффициент загрузки); λ_i – интенсивность поступления заявок i-го приоритета; p – значение приоритета, чем больше значение, тем выше приоритет [2, 4].

В случае с двумя приоритетами, для заявок с низким приоритетом, с учетом того, что время обслуживания для обоих приоритетов распределено одинаково и описывается экспоненциальным распределением, среднее время, проводимое заявкой в буфере и приборе

$$T_j = x + \frac{\lambda_j + f_{3j}}{(C_j - \lambda_j)(C_j - \lambda_j - f_{3j})}, \quad j = 1, 2, \quad (6)$$

где j – номер пути.

По аналогии с выражением (1) для низкоприоритетного трафика γ_{np} с использованием (6) можно получить функцию, минимизирущую общее среднее время задержки для низкоприоритетных заявок:

$$T_{\text{пр}} = \frac{1}{\gamma_{\text{пр}}} \times \times \sum_j \left(\frac{f_{3j}}{C_j} + \frac{f_{3j}(\lambda_j + f_{3j})}{(C_j - \lambda_j)(C_j - \lambda_j - f_{3j})} \right) \to \min. \quad (7)$$

Производная выражения по низкоприоритетному потоку

$$\frac{\partial T_j}{\partial f_{3j}} = \frac{1}{C_j} + \frac{(C_j - \lambda_j)(\lambda_j + 2f_{3j}) - f_{3j}^2}{(C_j - \lambda_j)(C_j - \lambda_j - f_{3j})^2}. \quad (8)$$

2. РЕШЕНИЕ ЗАДАЧИ С ИСПОЛЬЗОВАНИЕМ ЦИКЛА ОБСЛУЖИВАНИЯ

При более сложной сети обслуживания метод, описанный в предыдущем разделе, усложняется. Другой подход к расчету данного случая заключается в том, чтобы перейти от схемы на рис. 3 к схеме на рис. 1, при этом необходимо заменить пропускные способности на эквивалентные величины, зависящие от интенсивности высокоприоритетных потоков, на цикл обслуживания – интервал времени, начинающийся в момент поступления неприоритетного требования на обслуживание в прибор и заканчивающийся в момент, когда прибор вновь готов взять на обслуживание неприоритетное требование [5].

В источнике [5] для такой замены предлагается следующая формула:

$$C_{\text{э}} = C_{\text{пр}}(1 - \rho_{vp.}), \quad (9)$$

где $C_{\text{пр}}$ – интенсивность обслуживания или пропускная способность низкоприоритетного потока, ρ_{vp} – коэффициент загрузки, создаваемый высокоприоритетным потоком.

Рассмотрим схему представленную на рис. 4: поток λ_1 проходит по пути 1, часть этого потока можно пустить по второму пути, при этом заявки первого потока имеют меньший приоритет по сравнению с заявками второго потока, необходимо минимизировать среднее время прохождения заявок потока λ_1. Для решения задачи распределения потоков можно использовать выражения (7) и (8), в этом случае для первого (верхнего) пути высокоприоритетный поток λ_1 равен нулю. Второй подход заключается в использовании выражения (9) и сведения схемы к рис. 1.

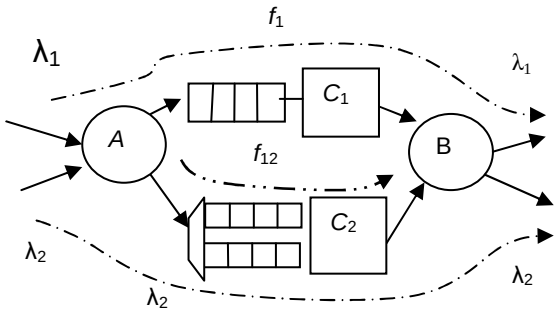

Рис. 4. Распределение потока λ_1 между основным путем и путем, где пакеты этого потока становятся низкоприоритетными

3. МОДЕЛЬ, УЧИТЫВАЮЩАЯ ЗАГРУЗКУ БУФЕРОВ СИСТЕМЫ

Рассмотрим пример $\lambda_1 = 0{,}625$ (заяв/с); $\lambda_2 = 0{,}31$ (заяв./с); $C_1 = 1{,}67$ (заяв./с); $C_1 = 1{,}04$ (заяв./с); коэффициенты загрузок соответственно $\rho_1 = 0{,}375$; $\rho_2 = 0{,}3$, поток без приоритета $\lambda_3 = 0{,}625$, без использования второго пути среднее время задержки неприоритетных пакетов по (6) $T = 3{,}48$ (с). При использовании второго пути и подходе с формулой (9) $f_{31} = 0{,}418$; $f_{32} = 0{,}207$, среднее время задержки $T = 1{,}7$. При подходе с использованием (7) $f_{31} = 0{,}4$; $f_{32} = 0{,}255$, при этом общее время задержки для низкоприоритетного трафика $T = 1{,}393$. Проведем моделирование. Для моделирования подобной схемы подходит Matlab 7 с пакетом SimEvents, на рис. 5 представлена модель схемы рис. 3.

Пакет потока λ_3 выбирает маршрут по наименьшей метрике вида $(n_{vp} + n_{\text{пр}} + 1)/C$, время моделирования 5000 единиц модельного времени, результаты представлены в табл. 1.

Рассмотрим пример для схемы, представленной на рис. 4. $\lambda_1 = 0{,}625$; $\lambda_2 = 0{,}31$; $C_1 = 1{,}67$; $C_1 = 1{,}04$; коэффициенты загрузок соответственно $\rho_1 = 0{,}375$; $\rho_2 = 0{,}3$.

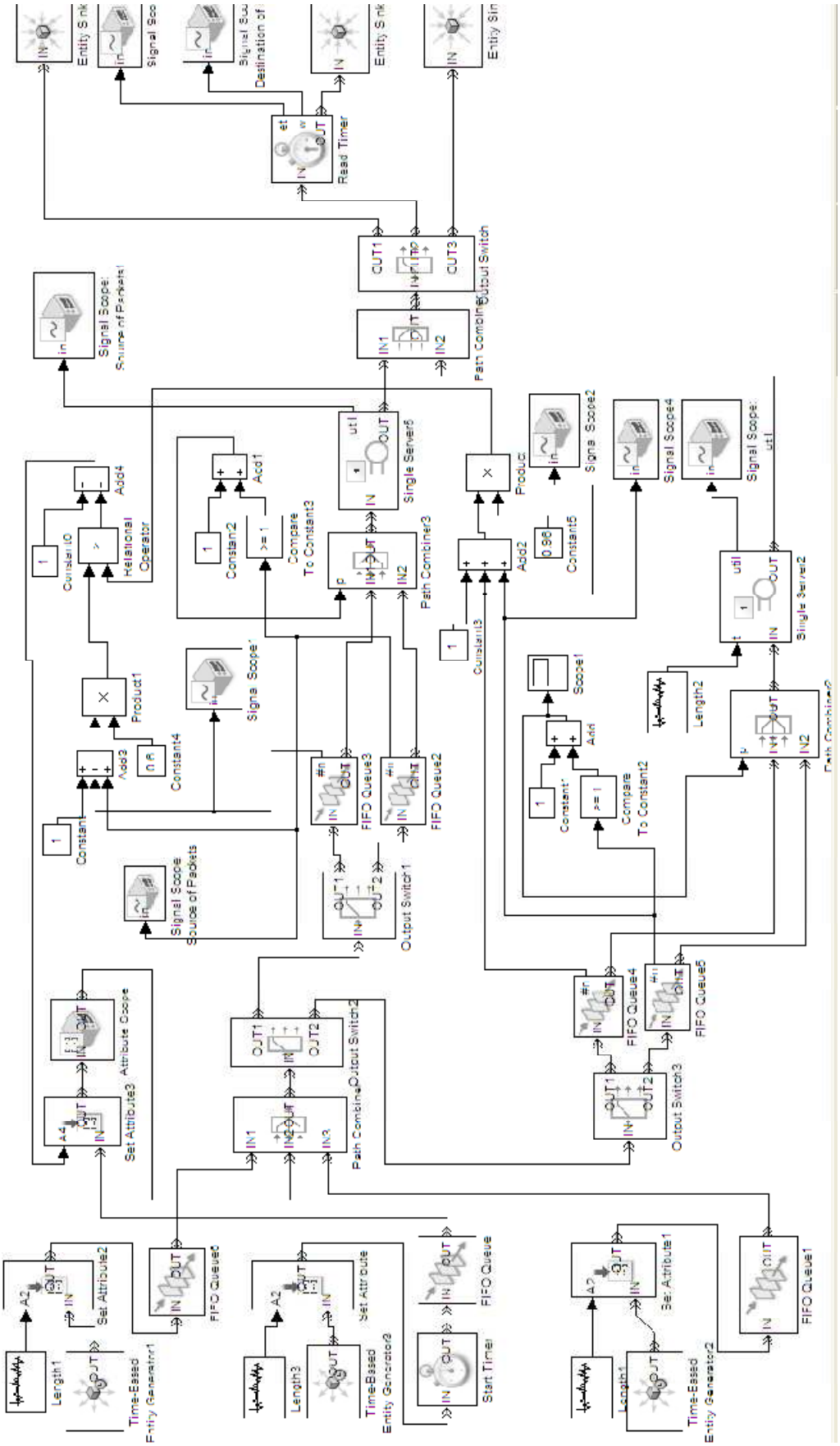

Рис. 5. Модель системы обслуживания, представленной на рис. 3, построенная в пакете Matlab 7

Без ответвления части потока λ_1 средне время задержки для пакетов этого потока $T = 0{,}96$. При подходе с использованием формулы (9) $f_{31} = 0{,}622$; $f_{32} = 0{,}002$ среднее время задержки $T = 1{,}17$. При подходе с использованием (7) $f_1 = 0{,}62$; $f_{32} = 0{,}005$, при этом общее время задержки для потока λ_1 составит $T = 0{,}6$.

Таблица 1

Результаты моделирования для схемы рис. 3

Метрика $(n_{vp}+n_{np}+1)/C$	Среднее время задержки $T \approx 1{,}35$	Число заявок 3098	Поток $\lambda_3 = 0{,}625$
1 путь	$T_1 \approx 1{,}2$	2246	$f_{31} \approx 0{,}434$
2 путь	$T_2 \approx 1{,}7$	852	$f_{31} \approx 0{,}16$
Метрика $(n_{np}+1)/C$	Среднее время задержки $T \approx 1{,}45$	Число заявок 3098	Поток $\lambda_3 = 0{,}625$
1 путь	$T_1 \approx 1{,}35$	2355	$f_{31} \approx 0{,}45$
2 путь	$T_2 \approx 1{,}7$	743	$f_{31} \approx 0{,}14$

Результаты, полученные при использовании модели, представленной на рис. 4, приведены в табл. 2.

Таблица 2

Результаты моделирования для схемы рис. 4

Метрика $(n_{vp}+n_{np}+1)/C$	Среднее время задержки $T \approx 0{,}77$	Число заявок 3098	Поток $\lambda_1 = 0{,}625$
1 путь	$T_1 \approx 0{,}73$	2810	$f_{31} \approx 0{,}55$
2 путь	$T_2 \approx 1{,}1$	288	$f_{31} \approx 0{,}05$

ЗАКЛЮЧЕНИЕ

В статье сравниваются два подхода к определению оптимальных долей низкоприоритетного потока, распределяемых по параллельным цепям обслуживающих приборов, с использованием статистических характеристик потоков заявок и производительности обслуживающих приборов, с целью уменьшить среднее время нахождения заявок в системе обслуживания и соответственно повысить эффективность использования системы обслуживания. Первый подход представляет собой алгоритм решения задачи оптимального распределения потоков, где метрика прибора выводится из параметров потоков обоих приоритетов, при сложной структуре системы подход достаточно сложный. При втором подходе метрика учитывает параметры низко-приоритетного потока и цикл обслуживания прибора, т. е. параметры приоритетного потока закладываются в производительность обслуживающего прибора, метод расчета пи этом упрощается. При сравнении результатов вычисления с моделью, в которой учитываются длина очередей обслуживающих приборов, т. е. среднее время задержки заявок минимально, видно, что второй метод хуже распределяет потоки.

СПИСОК ЛИТЕРАТУРЫ

1. **Башарин Г. П., Бочаров П. П., Коган Я. А.** Анализ очередей в вычислительных сетях. М.: Наука, 1989. 336 с.

2. **Крылов В. В., Самохвалова С. С.** Теория телетрафика и ее приложения. СПб.: БХВ-Петербург, 2005. 288 с.

3. **Вишневский В. М.** Теоретические основы проектирования компьютерных сетей. М.: Техносфера, 2003. 512 с.

4. **Клейнрок Л.** Вычислительные системы с очередями. М.: Мир, 1979. 588 с.

5. **Конвей Р. В.** Теория расписаний. М.: Наука, 1975. 360 с.

ОБ АВТОРАХ

Багманов Валерий Хусаинович, проф. каф. телекоммуникац. систем. Дипл. физик (МГУ, 1975). Д-р техн. наук по системн. анализу, управлению и обработке информации (УГАТУ, 2007). Иссл. в обл. матем. моделирования и обработки сигналов.

Султанов Альберт Ханович, проф., зав. той же каф. Дипл. инженер по многоканальн. электросвязи (Новосибирск. электротехн. ин-т, 1973). Д-р техн. наук по управлению в техн. системах (УГАТУ, 1996). Иссл. в обл. телекоммуникац. систем, аэрокосмическ. систем, оптоэлектроники.

Комиссаров Аркадий Михайлович, ст. преп. той же каф. Дипл. инженер по многоканальн. телекоммуникац. сист. (УГАТУ, 2002). Иссл. в обл. телекоммуникац. систем, систем массового обслуживания.