УДК 681.324

П.Е. ПУСТОВОЙТОВ, канд. техн. наук,
ЭЛЬ САЕД АБДЕЛААЛ ЭЛЬ САЕД МОХАМЕД

МИНИМАКСНОЕ УПРАВЛЕНИЕ ДИСЦИПЛИНОЙ ОБСЛУЖИВАНИЯ В КОМПЬЮТЕРНЫХ СЕТЯХ

Розглянуто задачу управління обслуговуванням сукупності нестаціонарних потоків на вході вузла комп'ютерної мережі. Введено критерії ефективності управління. Отримано співвідношення, що задають параметри раціонального управління.

It was considered the task of transient flow aggregation control on network hub entrance. It was introduced the criteria of control affectivity. Equations that set parameters of rational control were got.

Введение. Анализ литературы. По мере глобального роста и интенсивности использования Internet и корпоративных объединенных компьютерных сетей на передний план выходит ряд требований к качеству функционирования сетей, среди которых одним из самых важных является эффективная борьба с перегрузкой. При этом для удовлетворения этих новых требований недостаточно просто увеличить пропускную способность сети. Необходимы результативные методы управления трафиком и дисциплиной обслуживания потоков в узлах сети. Дело в том, что при возникновении перегрузки реальной становится опасность потери пакетов. Эти потерянные пакеты необходимо передавать повторно, что заметно увеличивает нагрузку на сеть и является причиной значительных задержек [1]. Более серьезный негативный феномен, называемый «глобальной синхронизацией» [2], состоит в следующем. Как правило, эффект переполнения очередей затрагивает несколько очередей одновременно. При этом в результате принятия мер для снятия перегрузки суммарный объем трафика резко падает и в течение определенного времени сеть используется не в полной мере. Это приводит к увеличению объемов трафика и возникает новый цикл «перегрузка-недогрузка».

При анализе возникающих здесь ситуаций для выбора мер противодействия перегрузкам необходимо иметь в виду, что поток сообщений для каждого узла сети представляет собой суперпозицию некоторого числа неоднородных потоков разной интенсивности и содержащих пакеты разной, в среднем длины. Известные методы борьбы с перегрузкой [3 – 5] учитывают эти обстоятельства в недостаточной мере.

Перечисленные обстоятельства определяют **цель работы:** рассмотрение методов борьбы с перегрузкой, основанных на управлении дисциплиной обслуживания с учетом различий в интенсивностях потоков, суммарное воздействие которых формирует перегрузку.

Основные результаты. Понятно, что рассмотрение вопросов управления дисциплиной обслуживания невозможно без предварительного описания потоков сообщений, поступающих на вход узла. Эти потоки являются нестационарными. Непосредственный анализ наблюдений за трафиком позволяет выявить наличие суточных, недельных и сезонных колебаний интенсивности потоков. При этом, практически в каждом потоке, амплитуда сезонных колебаний заметно меньше амплитуды недельных и особенно суточных колебаний, которая является самой высокой. С учетом этого обстоятельства выберем интервал наблюдений меньшим четверти периода сезонных колебаний и займемся отысканием амплитуды суточных и недельных колебаний. Разобьем интервал наблюдений на совокупность подынтервалов длиной Δ (например, $\Delta = \frac{1}{2}$ часа). Зафиксируем количество пакетов, поступающих в пределах каждого подынтервала. Обработаем теперь совместно случайные значения числа пакетов, приходящих в одно и то же время суток и в одноименные дни недели (например, интервал с $10^{\underline{30}}$ до $11^{\underline{00}}$ вторника). Вычислим среднее значение и дисперсию числа пакетов, поступающих в каждый из подынтервалов. При этом получим совокупность средних значений $m_1, m_2, ..., m_n$ случайного числа пакетов и их дисперсий $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$, $n = 24*7 = 168$. Введем теперь модель, описывающую зависимость интенсивности входящего потока от времени

$$\lambda(t) = a_0(a_1 + a_2\cos(\tfrac{2\pi}{24}t))\cos(\tfrac{2\pi}{168}t)\,. \qquad (1)$$

Параметры a_0, a_1, a_2 модели (1) найдем, используя данные о реальных наблюдениях, по методу наименьших квадратов. Введем функционал

$$J = \sum_{k=1}^{n}[m_k - \lambda(t_k)]^2 = \sum_{k=1}^{n}[m_k - a_0 - (a_1 + a_2\cos\tfrac{\pi}{12}t)\cos\tfrac{\pi}{74}t]^2\,. \qquad (2)$$

Минимизируем (2) методом Нелдера-Мида [6]. При этом получим оценки $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$, подстановка которых в (1) дает прогнозируемые моделью значения интенсивности входящего потока в любой момент времени t. Адекватность модели может быть проверена по критерию Фишера [7] с учетом анализа автокорреляции остатков [8]. Полученные описания динамики интенсивности потоков могут быть использованы для отыскания управления обслуживанием потоков на входе узла. В соответствии с принятой в настоящее время схемой BRFG разделения процессора [2] или какой-либо другой схемой, ресурс системы обработки в каждом цикле обслуживания распределяется определенным образом между потоками.

Пусть в систему поступает суперпозиция n разнородных по интенсивности $\lambda_1, \lambda_2, ..., \lambda_n$ и длине пакетов потоков, а соответствующие интенсивности обслуживания образуют набор $\mu_1, \mu_2, ..., \mu_n$. Введем вектор $X = (x_1, x_2, ..., x_n)$, где x_j – доля цикла обслуживания, выделяемая для обработки j-го потока. Тогда $\mu_j x_j$ – эффективная интенсивность обслуживания пакетов j-го потока, $j = 1, 2, ..., n$.

Как известно [9, 10], вероятность того, что в одноканальной системе обслуживания с пуассоновским потоком интенсивности λ на входе длина очереди составляет s заявок, определяется выражением

$$P_{1+s} = (1-\alpha)\alpha^{s+1},$$

где $\alpha = \dfrac{\lambda}{\mu}$ – приведенная интенсивность входящего потока. Тогда с учетом разделения ресурса процессора, в соответствии с компонентами вектора X, вероятность того, что длина очереди j-го потока будет равна s имеет вид

$$P_{1+s,j} = \left(1 - \frac{\lambda_j}{\mu_j x_j}\right)\left(\frac{\lambda_j}{\mu_j x_j}\right)^{s+1} = (1-\alpha_j)\,\alpha_j^{s+1}. \qquad (3)$$

С учетом (3) легко рассчитать вероятность того, что длина очереди будет не меньше какого-то конкретного критического значения q_0:

$$P_j(s \ge q_0) = \sum_{s=q_0}^{\infty} P_{1+s,j} = \sum_{s=q_0}^{\infty} (1-\alpha_j)\alpha_j^{s+1}.$$

При $k = s - q_0$ имеем

$$P_j(s \ge q_0) = (1-\alpha)\sum_{k=0}^{\infty} \alpha_j^{k+q_0+1} = (1-\alpha)\alpha_j^{q_0+1}\sum_{k=0}^{\infty} \alpha_j^k = \alpha_j^{q_0+1}.$$

Введем теперь минимаксное управление распределением ресурса процессора, минимизирующее в каждом цикле распределения максимальную вероятность того, что длина очереди будет не ниже критической. Ясно, что реализация такого управления приведет к тому, что вероятности критической ситуации для всех потоков будут равны между собой:

$$P_1(s \ge q_0) = P_2(s \ge q_0) = ... = P_n(s \ge q_0) = C$$

или

$$(\lambda_j / \mu_j x_j)^{q_0+1} = C, \; j = 1, 2, ..., n. \qquad (4)$$

Неизвестная константа C в (4) отыскивается из условия нормировки.

В результате получим

$$x_j = \frac{\rho_j}{\sum_{j=1}^{n} \rho_j}, \; j = 1, 2, ..., n. \qquad (5)$$

Замечательное свойство полученного управления состоит в том, что оно не зависит от критической длины q_0. Близкий результат получается, если рациональное распределение ресурса искать с использованием аддитивного критерия – среднее число очередей, длина которых не ниже критической.

Формальное описание задачи: найти вектор X, минимизирующий

$$F(x) = \sum_{j=1}^{n} \left(\frac{\rho_j}{x_j}\right)^{q_0+1} \qquad (6)$$

и удовлетворяющий ограничениям

$$\sum_{j=1}^{n} x_j = 1, \; x_j \ge 0, \; j = 1, 2, ..., n. \qquad (7)$$

Задача (6) – (7) легко решается методом неопределенных множителей Лагранжа. Функция Лагранжа при этом имеет вид

$$\Phi(x) = \sum_{j=1}^{n} \left(\frac{\rho_j}{x_j}\right)^{q_0+1} + \lambda\left(\sum_{j=1}^{n} x_j - 1\right).$$

Далее имеем

$$\frac{d\Phi(x)}{dx_j} = -\frac{\rho_j^{q_0+1}(q_0+1)}{x_j^{q_0+2}} + \lambda = 0, \; j = 1, 2, ..., n.$$

Отсюда

$$x_j = \left(\frac{\lambda}{\rho_j^{q_0+1}(q_0+1)} \right)^{\frac{1}{q_0+1}} = \left(\frac{\lambda}{q_0+1} \right)^{\frac{1}{q_0+1}} \rho_j^{\frac{q_0+1}{q_0+2}}, \; j = 1, 2, ..., n. \qquad (8)$$

Подставим (8) в (7):

$$\sum_{j=1}^{n} x_j = \left(\frac{\lambda}{q_0+1} \right)^{-\frac{1}{q_0+2}} \sum_{j=1}^{n} \rho_j^{\frac{q_0+1}{q_0+2}} = 1.$$

Тогда

$$\left(\frac{\lambda}{q_0+1} \right)^{-\frac{1}{q_0+2}} = \frac{1}{\sum_{j=1}^{n} \rho_j^{\frac{q_0+1}{q_0+2}}}. \qquad (9)$$

Наконец, подставляя (9) в (8), окончательно имеем

$$x_j = \frac{\rho_j^{\frac{q_0+1}{q_0+2}}}{\sum_{j=1}^{n} \rho_j^{\frac{q_0+1}{q_0+2}}}, \; j = 1, 2, ..., n. \qquad (10)$$

Ясно, что при практических, достаточно больших, значениях q_0

$$x_j \cong \frac{\rho_j}{\sum_{j=1}^{n} \rho_j}, \; j = 1, 2, ..., n, \qquad (11)$$

что совпадает с (5).

Таким образом, полученные выражения позволяют определить рациональное распределение ресурса процессора, минимизирующее негативные последствия перегрузки. Понятно, что в связи с нестационарностью входящих потоков нестационарным должно быть и управление распределением ресурса. Реализовать на практике такое управление технологически удобно, используя интервальное, кусочно-постоянное управление. При этом цикл обновления управления естественно выбрать так, чтобы максимальное отклонение интенсивностей входящих потоков от соответствующих средних, в пределах цикла, значений не превосходило заданное значение. Соответствующее аналитическое выражение отображается неравенством

$$\max_j \max_{T_0} \max_{0 \le t \le T_0} \left(\lambda_j(t) - \frac{1}{T_0} \int_0^{T_0} \lambda_j(t) dt \right) \le \varepsilon,$$

которому должно удовлетворять искомое значение периода обновления.

Выводы. Введена математическая модель описания нестационарных входящих потоков и процедура оценивания параметров модели. На их основе предложена технология рационального распределения ресурса процессора, обеспечивающая эффективное противодействие перегрузкам в узлах компьютерных сетей. Получено соотношение для расчета периода обновления управления распределением ресурса.

Список литературы: **1.** *Таненбаум Э.* Компьютерные сети. – СПб.: Питер, 2003. – 992 с. **2.** *Столлингс В.* Современные компьютерные сети. – СПб.: Питер, 2003. – 783 с. **3.** *Karr P., Partridge C.* Improving Roud-Trip Estimates in Reliable Transport Protocols // ACM Trans. On Coup. Systems. – 1991. – № 3. – P. 88–97. **4.** *Goyal R.* Improving Performance of TCP over the ATM Service // Proc., ICC. – 1997. – № 6 – P. 34–51. **5.** *Jacobson V.* Congestion Avoidence and Control // Proc. Comp. Communication Review. – 1995. – № 1. – P. 87–104. **6.** *Мину М.* Математическое программирование. – М.: Наука, 1990. – 488 с. **7.** *Крамер Г.* Математические методы статистики. –М.: Мир, 1975. – 563 с. **8.** *Сигел Э.Ф.* Практическая бизнес-статистика. – М.: Вильямс, 2004. – 1051 с. **9.** *Риордан Дж.* Вероятностные системы обслуживания. – М.: Связь, 1996. – 184 с. **10.** *Гнеденко Б.В., Коваленко И.Н.* Введение в теорию массового обслуживания. – М.: Наука, 1987. – 336 с.

Поступила в редакцию 30.03.2006