АНАЛИЗ ДЛИНЫ ОЧЕРЕДИ В СИСТЕМАХ МАССОВОГО ОБСЛУЖИВАНИЯ ТИПА G/G/1

Меркулова И.А.

Поволжская Государственная Академия Телекоммуникаций и Информатики

В числе основных требований предъявляемых к сетям связи является гарантия определенного качества обслуживания (Quality of service, QoS) каждого соединения.

Основной задачей Интернет-провайдера при построении сети передачи данных является обеспечение потенциальных пользователей гарантированной скоростью доставки трафика и требуемым качеством обслуживания [3]. Требуемое качество обслуживания в сети не может быть обеспечено только методами построения оптимального маршрута передачи данных. Помимо этого требуются эффективные протоколы и алгоритмы управления классами, очередями и потоками в устройствах коммутации трафика.

Образование очередей происходит в том случае, когда интерфейс занят, например, если интенсивность поступления пакетов становится больше интенсивности их обслуживания. Если же он свободен, то пакеты передаются без всякой дополнительной обработки. Все стандартные очереди работают по принципу FIFO и если очередь заполнена, но приходят новые пакеты, то происходит «отброс хвоста».

Основные параметры очереди характеризуются свойствами входящего потока требований, потока обслуживания и дисциплины очереди.

Для предотвращения потерь пакетов при кратковременном многократном превышении среднего значения интенсивности трафика единственным средством служит буфер большого объема. Чем больше объем памяти буфера, тем менее вероятны потери пакетов при перегрузках.

Ниже произведено исследование поведения длины очереди на базе модели массового обслуживания типа G/G/1.

Допустим, что Интернет-провайдер в своей сети имеет N пользователей, каждому из них он предоставляет соответствующее качество обслуживания c [5] и требуемую полосу пропускания h:

$[0; h_1(c_1)]$ – первый пользователь, $[0; h_2(c_2)]$ – второй пользователь, … $[0; h_N(c_N)]$ – N – ый пользователь.

Введем обозначения:

n – номер пакета;

$X_n = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$ – время ожидания пакета n в очереди; $Y_n = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$ – время обслуживания пакета n;

$Z_n = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix}$ – интервал времени между поступлениями n и n+1 пакета.

Аналогично уравнению Линдли [4], запишем уравнение для скоростей:

$$\frac{1}{X_{n+1}} = \frac{1}{X_n} + \frac{1}{Y_n} - \frac{1}{Z_n} \geq 0, \tag{1}$$

где $\frac{1}{X_{n+1}}$ и $\frac{1}{X_n}$ - скорость передачи $n+1$ и n пакета по каналу провайдера соответственно,

$\lambda = \frac{1}{Z_n}$ – интенсивность поступления n пакета, $\mu = \frac{1}{Y_n}$ – интенсивность обслуживания n пакета.

Из (1) следует:

$$P\left[\frac{1}{X_{n+1}} \leq \frac{1}{X}\right] = P\left[\frac{1}{X_n} + \mu - \lambda \leq \frac{1}{X}\right]. \tag{2}$$

Введем величину $U = \mu - \lambda$.

Обозначим через $K(U)$ – функцию распределения величины U, следовательно:

$$K(U) = P[\mu - \lambda \le U]. \tag{3}$$

При $U \succ 0$ буфер не переполнен и устройство обслуживания готово обработать следующий поступающий пакет. Если $U \prec 0$, то буфер переполнен и следующий пакет, который поступает на обрабатывающее устройство, становится в очередь на обслуживание.

Из (2) следует, что:

$$F_{n+1}\left(\frac{1}{X}\right) = \int_0^\infty P\left[\mu - \lambda \le \frac{1}{X} - \frac{1}{X_n}\right] dF_n\left(\frac{1}{X_n}\right), \tag{4}$$

где $F_n\left(\frac{1}{X_n}\right)$ - интегральная функция распределения скорости передачи, относящаяся к n пакету.

Преобразуя (4) и перейдя к пределу при $n \to \infty$, получим, что $F\left(\frac{1}{X}\right)$, функция распределения скорости передачи, удовлетворяет уравнению [2]:

$$F\left(\frac{1}{X}\right) = \int_0^\infty F\left(\frac{1}{Y}\right) dK\left(\frac{1}{X} - \frac{1}{Y}\right) \text{ при } X \ge 0. \tag{5}$$

Уравнение (5) можно переписать по-другому с учетом замены переменных в виде $U = \frac{1}{X} - \frac{1}{Y}$:

$$F\left(\frac{1}{X}\right) = \int_0^{\frac{1}{X}} F\left(\frac{1}{X} - U\right) dK(U) \text{ при } X \ge 0, \tag{6}$$

где $K(U)$ определяется из (3), при известной функции распределения входного потока [1]. Функция распределения интенсивности обслуживания задается произвольно в соответствии с режимом работы коммутатора.

Решение уравнения (6) согласно методике, приведенной в [2], позволяет аналитически определить длину очереди в системе G/G/1.

Произведем анализ длины очереди для случая, когда в сети провайдера два пользователя $N = 2$. Плотность распределения интенсивности поступления пакетов определена в [1] и равна:

$$\omega_\lambda(x) = \frac{1}{2(1-m_1)(1-m_2)}\left[(2-x)\cdot sign(2-x) - (m_1 - x + 1)\cdot sign(m_1 - x + 1) - (m_2 - x + 1)\times\right.$$
$$\left.\times\, sign(m_2 - x + 1) + (m_1 + m_2 - x)\cdot sign(m_1 + m_2 - x)\right],$$

где m_i - параметр со значением из интервала [0;1].

Плотность распределения обслуживания пакетов задаем произвольно, например, равномерное распределение:

$$\omega_\mu(x) = \frac{1}{2 - m_1 - m_2}.$$

Определим преобразование Лапласа для $\omega_\lambda(x)$ и $\omega_\mu(x)$, соответственно:

$$A(s) = \frac{1}{2\cdot s^2\cdot(1-m_1)(1-m_2)}\cdot e^{-2s}\left[1 + e^{-s(m_1+m_2-2)}\right],$$

$$B(s) = \frac{1}{2 - m_1 - m_2}\cdot\frac{1}{s}\cdot e^{-2s}\left[e^{-s(m_1+m_2-2)} - 1\right].$$

Согласно [2], найдем преобразование Лапласа для плотности распределения величины U $k(U)$:

$$K(s) = \frac{e^{s(m_1+m_2-2)} - e^{-s(m_1+m_2-2)}}{2\cdot s^2(1-m_1)(1-m_2)(2-m_1-m_2)\cdot(-s)}.$$

Далее, после некоторых преобразований находим:

$$\psi_+(s) = \frac{e^{s(m_1+m_2-2)} - e^{-s(m_1+m_2-2)} + 2\cdot s^3(1-m_1)(1-m_2)(2-m_1-m_2)}{2\cdot s^2(1-m_1)(1-m_2)(2-m_1-m_2)},$$

$$\psi_-(s) = -s .$$

Наконец, определяем вероятность того, что поступающий пакет застает систему свободной, т.е. не ожидает обработки:

$$R = \lim_{s\to 0} \frac{\psi_+(s)}{s} = \frac{-m_1^2 + 4m_1m_2 - 2m_1 + 2 - m_2^2 - 2m_2}{6 - 6m_1 - 6m_2 + 6m_1m_2}$$

Если задать $m_1 = 0{,}2$ и $m_2 = 0{,}4$, то $R = 0{,}319$.

Интересным представляется анализ величины R для случаев $N = 3, 4, 5...$ при разнообразных режимах работы коммутатора.

Литература

1. Меркулова И.А.. Доступ в Интернет с гарантированной скоростью доставки трафика//Инфокоммуникационные технологии, 2004, Том 2, № 2.- с.27-30.
2. Клейнрок Л. Теория массового обслуживания. – М.: Машиностроение, 1979,432 с.
3. Львов С.П. Интернет-доступ с гарантированной скоростью доставки трафика//ЭС, 2001, № 11.- с.30-33.
4. Д. Кокс, У. Смит. Теория очередей. – М.: Мир, 1966,220 с.
5. Лагутин В.С., Степанов С.Н. Телетрафик мультисервисных сетей связи. - М.: Радио и связь, 2000, 320 с.