

Задержки в очередях СМО с коррелированными потоками заявок

Б.Я. Лихтциндер¹

¹Поволжский государственный университет телекоммуникаций и информатики, Льва Толстого 23, Самара, Россия, 443010

Аннотация. Статья посвящена анализу задержек в очередях систем массового обслуживания (СМО), с коррелированными потоками заявок общего вида. Для трафика пакетов в мультисервисных сетях характерна высокая степень коррелированности. На основании интервальных методов анализа, получены соотношения, обобщающие формулу Хинчина-Поллячека для среднего значения времени ожидания в системах массового обслуживания, с потоками заявок общего вида. Основными анализируемыми параметрами при выводе указанных формул являются временные интервалы между соседними заявками. Показано, что значения временных задержек в очередях, зависят от дисперсии и индекса дисперсии случайной величины, характеризующей степень дообслуживания обрабатываемых заявок.

1. Введение

Во все периоды исследования трафика телекоммуникационных систем анализ производился на основе теории массового обслуживания. Предметами исследования здесь являются потоки заявок, подлежащие обработке некоторым ресурсом, ограниченной производительности.

Ограниченность ресурса и случайный характер поступления заявок приводят к отказам в обработке или образованию очередей. Первые работы, посвященные анализу телетрафика принадлежат А. К. Эрлангу. Потоки заявок представляли потоки запросов на соединения в сетях с коммутацией каналов. Поскольку заявки поступали от большого числа независимых абонентов, потоки таких заявок можно было определить, как стационарные, ординарные без последствия, или как рекуррентные, с экспоненциальным распределением интервалов времени между соседними заявками. Потоки запросов на соединение, поступающие в узел телефонной станции, являются суперпозицией большого числа потоков малой интенсивности, поступающих от независимых абонентов. Поэтому модель стационарного пуассоновского потока достаточно хорошо описывает реальные потоки в телефонных станциях с коммутацией каналов. Появление телекоммуникационных сетей, с пакетной коммутацией, в особенности мультисервисных сетей, показало невозможность применения для их анализа пуассоновских моделей потоков заявок.

Появляется множество различных моделей с не экспоненциальными законами распределения вероятностей интервалов между заявками [1]. Однако, все указанные интервалы считаются взаимно независимыми. Такие модели плохо отражают свойства потоков, поскольку потоки пакетов в реальных сетях сильно коррелированы и носят пачечный характер [1,3]. Описание таких потоков может быть осуществлено с помощью моделей самоподобных

процессов [4, 5]. Главным недостатком моделей самоподобных процессов является их сложность. Подавляющее число исследований в области самоподобных потоков происходит в отрыве от процессов их обработки в системе массового обслуживания. Попытки получения аналитических соотношений, определяющих задержки и размеры очередей в СМО с самоподобными потоками, обычно сопровождаются значительными трудностями и не привели к существенным практическим результатам. Поэтому аналитическое исследование не только самого потока, но и СМО, в которой он обрабатывается, представляется мало реальным [1]. Действительно, очереди в СМО образуются в результате взаимодействия двух потоков, один из которых - поток заявок, поступающих в систему, а другой - поток заявок, покидающих очередь. Указанные потоки сильно коррелированы. В результате, суммарный поток теряет долговременную зависимость, характерную для самоподобных процессов [3].

Недостаточная эффективность применения моделей самоподобных процессов к анализу пакетного телетрафика привели к появлению целого ряда моделей, управляемых цепью Маркова. Частный случай подобных моделей получил название "модели с гиперэкспоненциальными потоками событий" [6]. В дальнейшем потоки получали названия МС- потоков (Markov Chain), МАР- (Markovian Arrival Process) и их обобщение - групповые Марковские входные потоки ВМАР (Batch Markovian Arrival Process). Эволюция подобных моделей хорошо представлена в обзоре [1]. Здесь же отмечается, важность учета корреляционных свойств телетрафика, поскольку наличие положительной корреляции существенно увеличивает размеры очередей и задержки в СМО. Влияние корреляционных зависимостей во входящем потоке телетрафика неоднократно рассматривалось в работах автора этой статьи, где была установлена непосредственная зависимость среднего размера очереди от коэффициентов корреляции входного потока заявок. [2, 3]

2. Время дообслуживания

Вся классическая теория СМО основана на предположении о взаимной независимости поступающих заявок. Даже известная классификация Кендалла [7] не предусматривает возможности учета корреляционных свойств исследуемых потоков, а учитывает лишь законы распределения соответствующих вероятностей. Существует множество работ, рассматривающих различные законы распределения вероятностей интервалов между соседними заявками, не пуассоновские распределения вероятностей чисел заявок на некотором временном интервале, и ряд других. Однако, они предполагают взаимную независимость поступающих заявок.

Одним из перспективных, на наш взгляд, направлений изучения пакетного трафика является Интервальный метод [3], позволяющий заменить анализ интервалов времени между соседними заявками и интервалов времени обработки заявок, анализом одной случайной величины - числом заявок, поступающих в течение последовательных интервалов времени обработки заявок. Нами показано, что дисперсия и корреляционные свойства указанной случайной величины, при заданной загрузке, полностью характеризуют средний размер очереди в СМО [8, 9].

В этой статье, в качестве единой случайной величины, характеризующей входной поток заявок и его обработку в СМО, мы рассмотрим другую случайную величину u_i , представляющую разность между временем обслуживания заявки τ_i и интервалом времени ϑ_i между двумя соседними заявками на i -м интервале $u_i = \tau_i - \vartheta_i$. Если интервал обработки τ_i окажется больше интервала ϑ_i , то обслуживание заявки на данном интервале времени не успеет завершиться. Поэтому величину u_i называют временем дообслуживания. Указанная величина входит как единственная внешняя переменная в уравнение баланса, установленное Линдли [10],

$$\begin{aligned} w_i &= w_{i-1} + u_i \\ u_i &= \tau_i - \vartheta_i \quad \text{если} \quad w_{i-1} + \tau_i - \vartheta_i \geq 0 \end{aligned} \quad (1)$$

$$u_i = -w_{i-1} \quad \text{если} \quad w_{i-1} + \tau_i - \vartheta_i < 0.$$

Во всех случаях, когда выполняется второе неравенство, $w_i = 0$. Очевидно, что $w_{i-1} = \sum_{j=-\infty}^{i-1} u_j$.

Л. Клейнрок в своей работе [11] отмечал определяющее влияние указанной величины на размеры задержек и очередей в СМО. Полагая величины τ_i и ϑ_i не коррелированными и взаимно независимыми, он существенно ограничил, требования, предъявляемые к потокам. Однако, потоки пакетов в мультисервисных телекоммуникационных сетях не удовлетворяют указанным требованиям, поскольку имеют высокую степень корреляции.

На рисунке 1. показаны доли влияния корреляционных свойств потока видеотрафика на средние размеры очередей (верхний график) и доля, полученная без учета влияния корреляции (нижний график) при одинаковых значениях коэффициента загрузки ρ . Разница впечатляет и показывает, что корреляционными зависимостями потоков пакетного трафика пренебрегать нельзя.

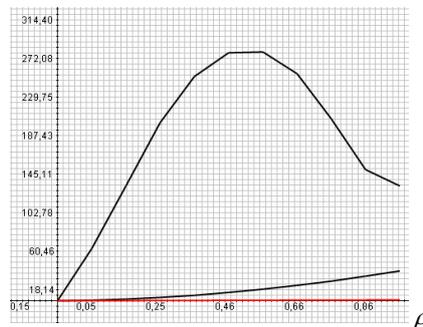


Рисунок 1. Доли влияния корреляционных свойств потока видеотрафика на средние размеры очередей.

Возведем в квадрат обе части уравнения (1). После усреднения на значительном интервале времени, получим: $\overline{u_i \cdot w_{i-1}} + \frac{1}{2} \overline{u_i^2} = 0$. Подставляя значение $w_{i-1} = \sum_{j=-\infty}^{i-1} u_j = w + \sum_{j=1}^{i-1} u_j$ получим:

$$\overline{u_i \cdot w_{i-1}} + \frac{1}{2} \overline{u_i^2} = \overline{w u_i} + u_i \sum_{j=1}^{i-1} \overline{u_j} + \frac{1}{2} \overline{u_i^2} = \overline{w_0 u_i} + u_i (\sum_{j=1}^i \overline{u_j} - u_i) + \frac{1}{2} \overline{u_i^2} = 0, \text{ или}$$

$$u_i \sum_{j=1}^i \overline{u_j} - \frac{1}{2} \overline{u_i^2} = -\overline{w u_i}.$$

Принимая гипотезу о затухании корреляционной связи, можем считать, что $\overline{w_0 u_i} = \overline{w_0} \overline{u_i}$, следовательно,

$$u_i \sum_{j=1}^i \overline{u_j} - \frac{1}{2} \overline{u_i^2} = -\overline{w} \overline{u_i}. \quad (2)$$

Первый член указанной разности представляет среднее значение произведения функции на ее интеграл (сумму), поэтому $u_i \cdot \sum_{j=1}^i \overline{u_j} = \sum_{k=0}^{\infty} R_u(k) = \sum_{k=1}^{\infty} R_u(k) + \overline{u_i^2}$. Здесь $R_u(k)$ - значение корреляционного коэффициента переменной u_i , при сдвиге на k интервалов. Подставляя в (2), получим:

$$\sum_{k=1}^{\infty} R_u(k) + \overline{u_i^2} - \frac{1}{2} \overline{u_i^2} = \sum_{k=1}^{\infty} R_u(k) + \frac{1}{2} \overline{u_i^2} = -\overline{w_0} \overline{u_i}, \text{ или}$$

$$\sum_{k=1}^{\infty} R_u(k) + \frac{1}{2} D_u + \frac{1}{2} \overline{(u_i)^2} = -w \overline{u_i}, \quad \text{Откуда,}$$

$$\overline{w} = \frac{D_u + 2 \sum_{k=1}^{\infty} R_u(k) + \overline{(u_i)^2}}{2(-\overline{u_i})}.$$

Произведем замену переменной $v_i = -u_i$, при этом, значение числителя не изменится, и в

ведем в рассмотрение индекс дисперсии $J_v = 1 + 2 \sum_{k=1}^{\infty} r_v(k)$, где $r_v(k) = \frac{R_v(k)}{D_v}$ - нормированное

значение корреляционного коэффициента. Тогда, получим фундаментальное соотношение, обобщающее известную формулу Хинчина Поллачека и пригодное для анализа коррелированных входных потоков заявок:

$$\overline{w} = \frac{D_v J_v}{2 v_i} + \frac{\overline{v_i}}{2}. \quad (3)$$

3. Алгоритм формирования массива случайной переменной v_i , характеризующей график.

Рассмотрим последовательность обработки исходных данных.

Вводятся массивы длин пакетов l_i [B] и моментов прихода пакетов t_i [мкс]. Устанавливается предельное значение числа пакетов M , порядковый номер пришедшего пакета $i = 0$; , начальный момент отсчета $t_0[c] = 0$ [мкс], начальное значение переменной $v_0 = 0$ [мкс]. Суммарное время обработки пакетов $T_0[c] = 0$ [мкс], Суммарное время ожидания $W_0 = 0$, начальное значение времени ожидания в очереди $w_0 = 0$, значение пропускной способности канала $C_0 [\frac{M \text{ бит}}{c}]$.

1. Переход к очередному пакету $i := i + 1$.
2. Вводится момент $t_{i+1}[c]$ времени поступления последующего пакета.
3. Определяется межпакетный интервал $\mathcal{G}_i = t_{i+1} - t_i$ [мкс].
4. Вычисляется коэффициент $k = \frac{8}{C_0} [\frac{\text{мкс}}{B}]$.
4. Вводится длина очередного пакета в байтах - l_i [B].
5. Определяется время обработки пакета в микросекундах $\tau_i = k l_i [\frac{\text{мкс}}{\text{пак}}]$.
6. Определяется суммарное время обработки пакетов в микросекундах $T_i = T_{i-1} + \tau_i$ [мкс].
7. Определяется время дообслуживания $u_i := \tau_i - \mathcal{G}_i$, $v_i := u_i$ [мкс].
8. Определяется остаточное время $\Delta_i = w_{i-1} + u_i$ [мкс].
9. Если $\Delta_i < 0$, то $v_i := -w_{i-1}$ [мкс].
10. Запись v_i в массив V.
11. Определяется текущее время ожидания в очереди $w_i = w_{i-1} + v_i$ [мкс].
12. Определяется суммарное время обработок $T_i = T_{i-1} + \tau_i$ [мкс].
13. Суммарное время ожидания $W_i = W_{i-1} + w_i$ [мкс].

14. Если $i < M - 1$, переход к п.1.
 15. Среднее значение задержки $\bar{w} = \frac{W_M}{M}$ [мкс].
 16. Коэффициент $\rho_0 = \frac{T_M}{t_M}$. загрузки
 17. Средняя интенсивность $\lambda = \frac{M}{t_M}$.
 18. Среднее время обработки пакета $\bar{\tau} = \frac{T_M}{M}$ [мкс].
 19. Средний интервал между пакетами $\bar{g} = \frac{t_M}{M}$.
 20. Средний размер очереди $\bar{q} = \bar{w} \lambda$ [мкс].
 21. Вывод результатов и массива V [мкс].
 22. Организуется 10 циклов для значений C_x , соответствующих $\rho_x = 0,1; 0,2, \dots, 0,9; 1,0$.
- $C_x = C_0 \frac{\rho_0}{\rho_x}$, и строятся графики для всех средних значений параметров и зависимость C_x от коэффициента ρ_x .

4. Заключение

Таким образом, нами получено фундаментальное соотношение (3), напрямую связывающее дисперсию и индекс дисперсии переменной v_i со средним значением задержки обслуживания в однолинейной СМО.

Математическое ожидание, дисперсия и индекс дисперсии указанной случайной величины полностью определяют средние размеры очередей в СМО, при заданном значении коэффициента ее загрузки, а приведенный алгоритм показывает способ получения значений этой величины.

5. Литература

- [1] Вишневский, В.М. Системы массового обслуживания с коррелированными входными потоками и их применение для моделирования телекоммуникационных сетей / В.М. Вишневский, А.Н. Дудин / Автоматика и телемеханика. – 2017. – Т. 8. – С. 3-59.
- [2] Лихтциндер, Б.Я. Интервальный метод анализа мультисервисного трафика сетей доступа // Электросвязь. – 2015. – № 12. – С. 52-54.
- [3] Лихтциндер, Б.Я. Трафик мультисервисных сетей доступа (интервальный анализ и проектирование) – М.: Горячая линия – Телеком, 2018. – 290 с.
- [4] Leland, W.E. On the Self-Similar Nature of Ethernet Traffic / W.E. Leland, S. Taqqu Murad, W. Willinger, D.V. Wilson // J. IEEE/ACM Transact. Networking. – 1994. – Vol. 2(1). – P. 1.
- [5] Цыбаков, Б.С. Модель телетрафика на основе самоподобного случайного процесса // Радиотехника. – 1999. – № 5. – P. 24-31.
- [6] Кузьмин, И.В. Основы моделирования сложных систем: учебное пособие для студентов вузов – Киев: Вища школа. Головное изд-во, 1981. – 360 с.
- [7] Клейнрок, Л. Вычислительные системы с очередями – М.: Мир, 1979. – Т. 2.
- [8] Лихтциндер, Б.Я. Корреляционные свойства длин очередей в системах массового обслуживания с потоками общего вида // ИКТ. – 2015. – Т. 13, № 3. – С. 276-280.
- [9] Лихтциндер, Б.Я. Корреляционные связи в пачечных потоках систем массового обслуживания // Телекоммуникации. – 2015. – № 9. – С. 8-12.

- [10] Lindley, D.V. The Theory of Queues with a Single Server // Proc. Cambridge philosophical Society. – 1952. – Vol. 48. – P. 277-289.
- [11] Клейнрок, Л. Теория массового обслуживания – М.: Книга по требованию, 2013. – 429 с.

Delays in QS queues with correlated application flows

В.Ya. Likhttcinder¹

¹Povolzhskiy State University of Telecommunications and Informatics, L. Tolstoy street 23, Samara, Russia, 443010

Abstract. The article is devoted to the analysis of delays in queues of queuing systems (QS), with correlated flows of requests of a general form. Packet traffic in multiservice networks is characterized by a high degree of correlation. Based on interval analysis methods, relations are obtained that generalize the Khinchin-Pollyachek formula for the average value of the waiting time in queuing systems with general application flows. The main parameters analyzed when deriving these formulas are time intervals between neighboring applications. It is shown that the values of time delays in queues depend on the variance of the variance index and the variance index of a random variable characterizing the degree of after-service of the processed orders.