Лекция 16.11

Распределение числа полученных требований в потоке Пальма

На предыдущих лекциях мы рассмотрели ряд частных случаев потока Пальма, которые позволяют охватить потоки с произвольным значением коэффициента вариации длины интервала между требованиями γ_τ. В частности, при $\gamma_\tau = 0$ можно использовать модель регулярного потока, при $0 < \gamma_\tau < 1$ – модель потока Эрланга, при $\gamma_\tau = 1$ – модель простейшего потока, а при $\gamma_\tau > 1$ – модель потока Кокса. Сейчас мы разберем другой вопрос: как распределено число требований в потоке Пальма?

Напомним, что мы описывали потоки с помощью случайной функции $N(t)$ – числа требований, которые поступят к моменту t. Это случайная функция, принимающая дискретные целочисленные значения. Распределение $N(t)$ характеризуется вероятностями

$$P_n(t) = \mathcal{P}\{N(t) = n\}, \quad (n = \overline{0, \infty}). \quad (1)$$

Как мы помним, поток Пальма полностью характеризуется законом распределения интервала между требованиями $a(\tau)$. Наша цель – выразить вероятности (1) через функцию $a(\tau)$.

Обозначим моменты прихода последовательных требований в потоке Пальма через $\{Y_k\}_{k=1}^{\infty}$. Плотность вероятностей

Y_k будет обозначаться через $f_k(y)$. Далее, как обычно, функцию распределения интервала между требованиями обозначим через $A(t)$.

Рассмотрим вероятность отсутствия требований в интервале $[0,t]$. Она получается из (1) при $n=0$. Очевидно, отсутствие требований в интервале $[0,t]$ эквивалентно тому, что момент прихода *первого* требования Y_1 оказался больше t

$$P_0(t) = P\{Y_1 > t\} \qquad (2)$$

по определению

с другой стороны

$$P\{Y_1 < t\} = A(t). \qquad (3)$$

значит

$$P_0(t) = 1 - A(t). \qquad (4)$$

Теперь выразим вероятность $P_n(t)$. Понятно, что при произвольном $n>0$

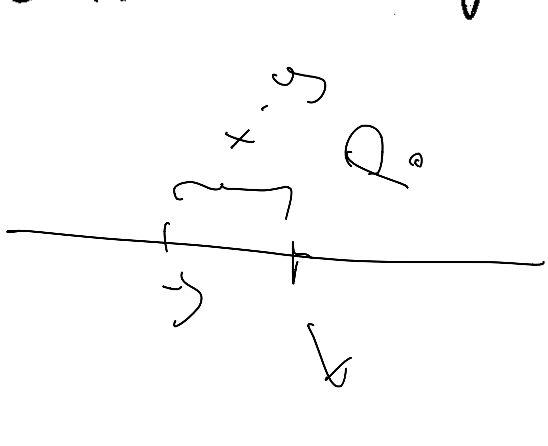
$$P_n(t) = P\{Y_n \le t, Y_{n+1} > t\}. \qquad (5)$$

Эту вероятность можно выразить через закон распределения Y_n следующим образом

$$P_n(t) = \int_0^t f_n(y)[1 - A(t-y)]dy. \qquad (6)$$

В этом равенстве $f_n(y)dy$ дает элементарную вероятность того, что точка Y_n попала *левее* t, а второй множитель в подынтегральном выражении дает вероятность того, что $Y_{n+1} - Y_n$ окажется больше $t - y$. В результате Y_{n+1} будет лежать *правее* точки t.

Осталось получить закон $f_n(y)$.



Для его вычисления заметим, что
$$Y_n = Y_{n-1} + \tau_n, \quad (n \geq 2), \tag{7}$$
причем Y_{n-1} и τ_n независимы. При этом, очевидно,
$$Y_1 = \tau_1. \tag{8}$$
В этих соотношениях τ_n обозначает n-й интервал в потоке Пальма. По определению потока Пальма все τ_n независимы и распределены по закону $a(\tau)$.

Из уравнений (7) и (8) следует, что функции f_n удовлетворяют рекуррентным соотношениям
$$f_n(y) = \int_0^y a(y-x) f_{n-1}(x)dx, \quad (n \geq 2), \tag{9}$$
причем
$$f_1(y) = a(y). \tag{10}$$

Распределение числа требований в простейшем потоке

Рассмотрим в качестве примера случай простейшего потока, когда
$$a(\tau) = \lambda e^{-\lambda\tau}. \tag{11}$$
В этом случае нетрудно показать, что
$$f_n(y) = \lambda e^{-\lambda y} \frac{(\lambda y)^{n-1}}{(n-1)!}, \quad (n \geq 1). \tag{12}$$

Докажем этот факт по индукции. При $n=1$ согласно (10) выражение (12) справедливо, то есть база индукции имеет место.

Теперь допустим, что (12) имеет место при произвольном n. Тогда согласно рекуррентному соотношению (9) будем иметь:

$$f_{n+1}(y) = \int_0^y \lambda e^{-\lambda(y-x)} \lambda e^{-\lambda y} \frac{(\lambda y)^{n-1}}{(n-1)!} dy =$$

$$= \lambda e^{-\lambda y} \int_0^y \lambda \frac{(\lambda y)^{n-1}}{(n-1)!} dy =$$

$$= \lambda e^{-\lambda y} \frac{(\lambda y)^n}{n!}, \qquad (13)$$

то есть получили соотношение вида (12), в котором n заменено на $n+1$. Таким образом, формула (12) доказана.

Теперь обратимся к формуле (6). По определению

$$A(\tau) = \mathcal{P}\{\tau_k < \tau\} = \int_0^\tau \lambda e^{-\lambda s} ds =$$

$$= 1 - e^{-\lambda\tau}, \quad (\tau > 0). \qquad (14)$$

Подставляя выражение (14) в (6), будем иметь

$$P_n(t) = \int_0^t \lambda e^{-\lambda y} \frac{(\lambda y)^{n-1}}{(n-1)!} e^{-\lambda(t-y)} dy =$$

$$= e^{-\lambda t} \int_0^t \lambda \frac{(\lambda y)^{n-1}}{(n-1)!} dy =$$

$$= e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad (n \geq 1). \qquad (15)$$

По формуле (4) с учетом (14) находим

$$P_0(t) = e^{-\lambda t}, \qquad (16)$$

поэтому равенство (15) справедливо при любых $n \geq 0$, то есть

$$P_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad (n \geq 0) \qquad (17).$$

Иначе говоря, в простейшем

потоке число полученных требований распределено по закону Пуассона с параметром λt.

Накопитель очереди и исполнительная часть системы обслуживания

Вернемся к общей схеме СМО, которую мы рассматривали на первых лекциях (см. рис. 1)

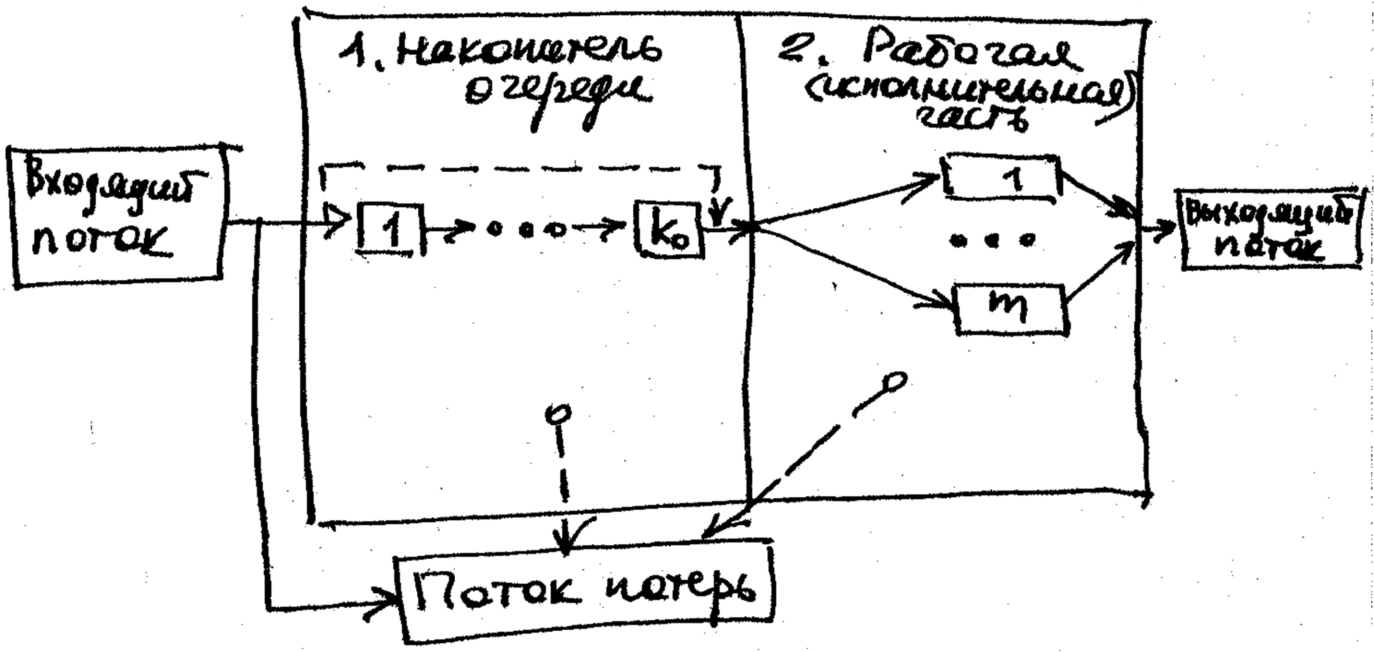


Рис. 1. Общая схема СМО: k – емкость системы, m – число каналов обслуживания, $k_0 = k - m$ – емкость накопителя

До настоящего момента мы занимались построением моделей входящего потока. Теперь хотя бы кратко следует сказать о двух других основных элементах СМО – накопителе и рабочей (или исполнительной) части. Оба этих элемента СМО работают в тесном контакте, и поэтому их удобно рассмотреть совместно.

Если в системе имеются свободные каналы обслуживания, то тогда вновь пришедшее требование из входящего потока сразу передается на один из них, минуя накопитель (пунктирная линия в обход накопителя на рис. 1).

Накопитель включается в работу только при отсутствии свободных каналов. Его назначение – сохранять попавшие в него требования и передавать в надлежащем порядке на обслуживание. Основными характеристиками накопителя являются: число одновременно формируемых очередей, емкость (объем) накопителя, дисциплина очереди и сохранность требований.

В большинстве моделей СМО рассматривается одна общая очередь. Такая организация работы накопителя целесообразна, если все требования идентичны и все каналы обслуживания обладают одинаковыми характеристиками. В противном случае могут возникать несколько очередей. Например, в приоритетных системах могут выделяться очереди привилегированных требований. Иногда рассматривают и СМО с неидентичными каналами.

Емкость накопителя k_0 может быть произвольным целым числом от 0 до бесконечности. Значение $k_0 = 0$ соответствует системе с отказами. Пусть N обозначает число требований в системе, а m – число ее каналов обслуживания. В системе с отказами при $N < m$ вновь поступившее требование сразу идет на обслуживание, а при $N = m$ безвозвратно теряется. Такие необслуженные требования образуют поток потерь. Значения $N > m$ в системе с отказами невозможны.

Пусть теперь $k_0 > 0$ и конечно.

Тогда емкость системы $K = K_0 + m$, при $N < m$ вновь поступившее требование сразу пойдет на обслуживание, при $m \le N < K$ оно встанет в очередь, а при $N = K$ потеряется. Такая система называется системой с ожиданием и отказами. Чем больше K_0, тем вероятность отказа будет меньше.

В общей теории рассматриваются системы, характеризуемые $K_0 = \infty$. Такие системы называются системами с ожиданием. В них потери отсутствуют, и все требования, поступающие на вход СМО, могут встать в очередь.

Дисциплиной очереди называют порядок выбора требований из очереди для передачи на освобождающиеся каналы обслуживания. Наиболее употребительными дисциплинами являются: обслуживание в порядке поступления (ОПП), обслуживание в обратном порядке (ООП) и обслуживание в случайном порядке (ОСП). В зарубежной литературе такие дисциплины фигурируют как FCFS (first come - first serve), LCFS (last come - first serve) и RANDOM.

В предыдущих рассмотрениях считалось, что накопитель обеспечивает полную сохранность попавших в него требований. Применяются и более общие модели, в которых допускаются потери из очереди. Например, время ожидания в очереди может ограничиваться в силу нетерпеливости клиентов или каких-либо технологических ограничений. Необслуженные требования, покинувшие накопитель, присоединяются к потоку потерь.

В накопителе требования не подвергаются никакой обработке, они лишь ждут передачи на канал обслуживания. Все работы по обслуживанию локализованы в исполнительной (рабочей) части. Основными ее характеристиками являются: число каналов, число фаз обслуживания на каждом канале, закон распределения времени обслуживания, очередность загрузки каналов и сохранность требований.

Каналы обслуживания соединены параллельно и не блокируют друг друга. Возможны ситуации, когда часть каналов свободна. Очередность их загрузки вновь поступившими требованиями задается правилом, аналогичным
дисциплине
дисциплине очереди. Если все свободные каналы идентичны, то выбор одного из них под загрузку обычно происходит по случайному закону (аналог дисциплины ОСП). Выбор канала, освободившегося раньше других, называется холодной загрузкой (аналог дисциплины ОПП), а выбор канала, освободившегося последним по счету, называется горячей загрузкой и является аналогом дисциплины ООП.

Фазой называется организационно законченный этап полного обслуживания требования в рабочей части. При наличии нескольких фаз обслуживание называется многофазным.

Для любого канала обслуживания время обслуживания Х, вообще говоря, распределено по случайному закону. Обозначим плотность вероятности X через $b(x)$. Функция $b(x)$ является важнейшей характеристикой СМО. Здесь также, как и при анализе интервалов во входящем потоке, вводят среднее время обслуживания

$$\overline{x} = M[X] = \int_0^{\infty} x\, b(x)dx, \qquad (18)$$

дисперсию времени обслуживания

$$\sigma_x^2 = D[X] = \int_0^{\infty} (x-\overline{x})^2 b(x)dx, \qquad (19)$$

коэффициент вариации

$$\gamma_x = \frac{\sigma_x}{\overline{x}} \qquad (20)$$

и интенсивность обслуживания

$$\mu = \frac{1}{\overline{x}}. \qquad (21)$$

Обычно считается, что функция $b(x)$ одинакова для всех требований. Тогда если канал работает постоянно, без простоев, то поток выходящих с каждого канала будет представлять собой поток Пальма с законом распределения интервала, равным $b(x)$.

Любая СМО состоит из описанных выше стандартных элементов. Задать численную систему — это значит задать численно параметры m и k, а также ℓ — число источников нагрузки, $a(\tau)$ — закон распределения интервалов между требованиями и $b(x)$ — закон распределения времени обслуживания. Отметим, что при моделировании $b(x)$ обычно используют те же самые законы, что и при моделировании $a(\tau)$.

Классификация СМО по Кендаллу

В ТМО принята очень удобная система сокращенных обозначений СМО, которая позволяет с помощью короткой символики полностью охарактеризовать наиболее важные качественные особенности рассматриваемой СМО. Эта система была разработана еще в середине прошлого века английским ученым Д. Кендаллом.

Существует две варианта обозначений по Кендаллу — трехбуквенный и пятибуквенный. Трехбуквенная система применяется для описания систем с ожиданием. Она основана на следующих предположениях:

1) Входящий поток считается потоком Пальма;
2) Число источников нагрузки $\ell = \infty$;
3) Емкость накопителя $K_0 = \infty$;
4) число каналов обслуживания m произвольно;
5) все каналы идентичны, обслуживание однофазное.

При этих предположениях сокращенное обозначение СМО имеет вид $A/B/m$, причем буквы A и B задают соответственно закон распределения интервала между требованиями во входящем потоке и закон распределения времени обслуживания.

Буквы A и B обозначают закон, выбранный из стандартного набора распределений: M (показательный закон), E_r (закон Эрланга порядка r), H_r (гиперэкспоненциальный закон порядка Г), D (детерминированный закон), G (закон общего вида)

Пятибуквенная система описывает системы с ожиданием и потерями. Допустим, что система имеет ограниченную емкость k и l одинаковых источников нагрузки, то есть отличные от предположений 2) и 3) сделанных при введении трехбуквенной системы. Все остальные допущения сохраняются. Тогда обозначение такой системы имеет вид $A/B/m/k/l$, причем здесь буква A характеризует поток исходящий от одного источника.

Простейшим примером системы СМО класса с ожиданием является $M/M/1$, представленная на рис. 2.

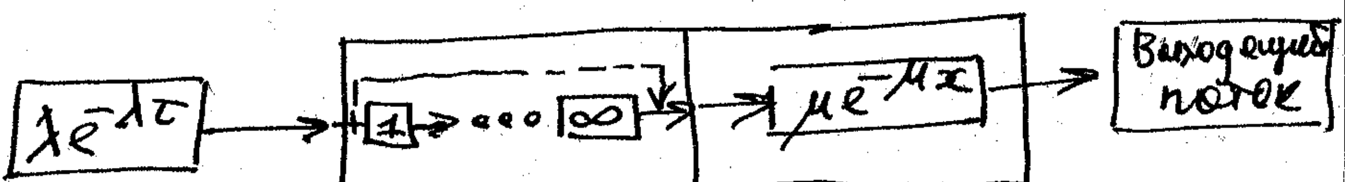


Рис. 2. Система $M/M/1$

В этой системе входящий поток является простейшим ($a(\tau)=\lambda e^{-\lambda\tau}$), закон распределения времени обслуживания является показательным ($b(x)=\mu e^{-\mu x}$), имеется один канал ($m=1$) и бесконечно много мест ожидания в накопителе ($k_0=\infty$).

Примером применения пятибуквенного обозначения может служить система $M/M/m/k/l$, представленная на рис. 3.

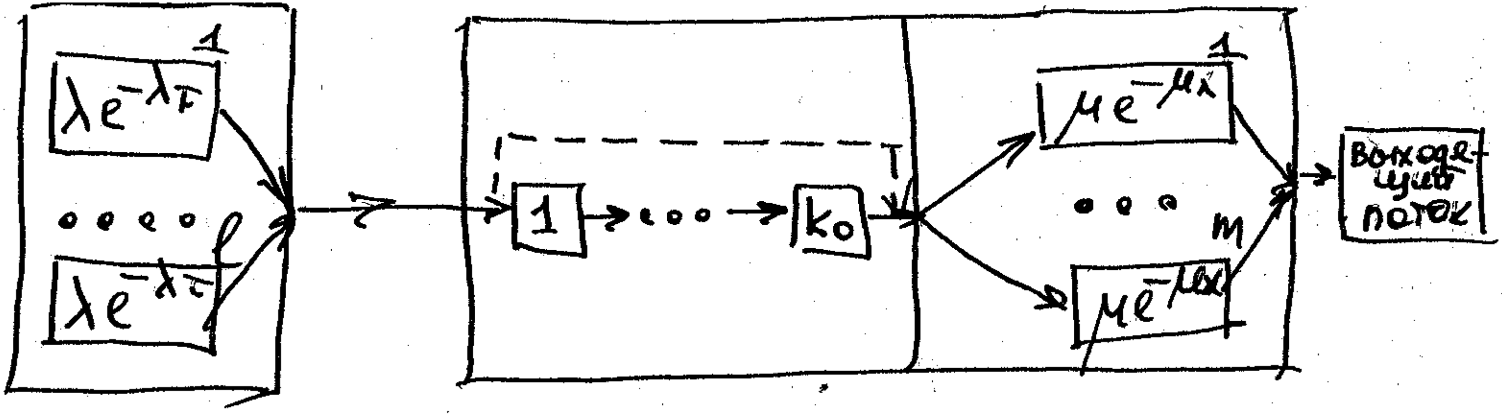


Рис. 3. Система M/M/m/k/l

Все три целевых параметра m, k, l могут принимать произвольные значения, включая и бесконечные