НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ

ВМИ

БЛАГОТВОРИТЕЛЬНЫЙ
ФОНД В. ПОТАНИНА

Дисциплина «Вероятностные модели»

Тема «Специализированные стационарные СМО и их функциональные характеристики**»**

Разработчики:
А.Т. Латипова, к.ф.м.н., доцент
Е.Л. Шпигель, студент группы ВМИ-532

Обобщенная модель СМО

В данном разделе рассматриваются ***обобщенные (общие) системы массового обслуживания***, в которых есть как входной поток клиентов, так и выходной поток обслуженных клиентов. Время между последовательными поступлениями клиентов и время обслуживания являются экспоненциально распределенными случайными величинами. Эта модель служит основой при рассмотрении специализированных моделей Пуассона.

При рассмотрении общих систем массового обслуживания предполагается, что система функционирует в течение достаточно большого интервала времени, по истечении которого в ее работе наступает стационарный режим. Этот режим функционирования обслуживающей системы противопоставляется переходному (или неустановившемуся) режиму, который превалирует в самый начальный период функционирования системы. Здесь *не рассматриваются переходные режимы работы* систем массового обслуживания, поскольку, во-первых, это связано с серьезными математическими трудностями, а, во-вторых, на практике данные системы обычно предназначаются для работы в течение весьма длительного времени.

В рассматриваемой в этом разделе общей модели системы массового обслуживания предполагается, что и интенсивность поступления клиентов, и интенсивность выходного потока *зависят от состояния* системы, что означает их зависимость от числа клиентов в системе обслуживания. Например, сборщик платы за проезд по автомагистрали в часы интенсивного движения стремится ускорить сбор пошлины. Или в мастерской с фиксированным количеством станков интенсивность их поломки убывает по мере возрастания числа аварийных станков, ибо лишь работающие станки могут выходить из строя.

Введем следующие обозначения.

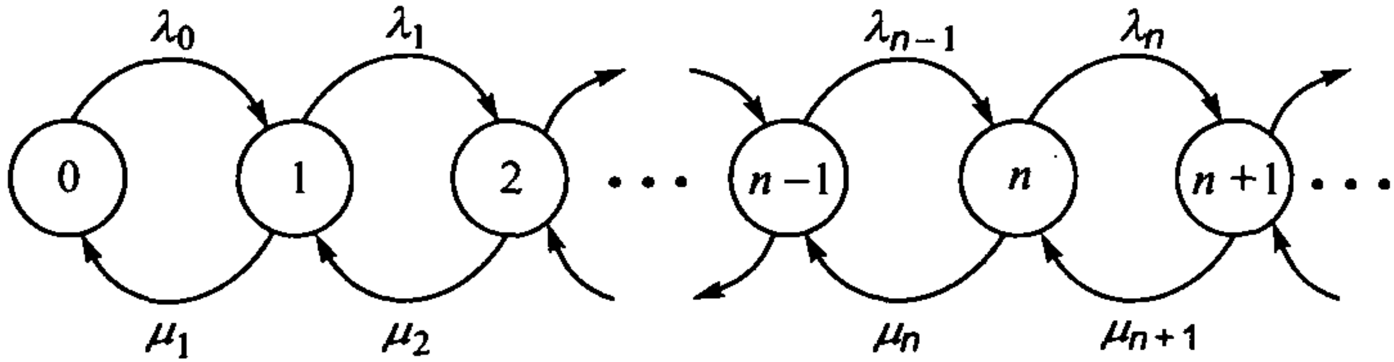
n – число клиентов в системе обслуживания (в очереди и на обслуживании), λ_n – интенсивность поступления в систему клиентов при условии, что в системе уже находится n клиентов,

μ_n – интенсивность выходного потока обслуженных клиентов при условии, что в системе находится n клиентов,

p_n – вероятность того, что в системе находится n клиентов (**финальная**, конечная **вероятность**).

Вероятности p_n определяются из **диаграммы интенсивностей переходов**, представленной на рисунке. Обслуживающая система находится в состоянии n, если в ней имеется n клиентов. Как показано ранее, вероятность появления более одного нового клиента на протяжении малого промежутка времени h стремится к нулю при $h \to 0$. Это означает, что при $h \to 0$ состояние n может быть изменено в двух возможных направлениях: $n-1$, когда с интенсивностью μ_n обслуженный клиент выбывает из системы, и $n+1$, когда клиенты поступают с интенсивностью λ_n. Состояние 0 может измениться лишь к состоянию 1, когда имеет место поступление клиента с интенсивностью λ_0. Заметим, что μ_0 не определено, так как клиенты не могут выбывать из пустой системы обслуживания.

Диаграмма переходов для обобщенной модели



При выполнении условий стационарности *ожидаемые* интенсивности входного и выходного потоков в состоянии n ($n > 0$) должны быть равны. Так как состояние n может изменяться лишь к состояниям $n - 1$ и $n + 1$, отсюда следует, что

$$\lambda_{n-1} p_{n-1} + \mu_{n+1} p_{n+1} = \lambda_n p_n + \mu_n p_n .$$

Для нулевого состояния имеем

$$\mu_1 p_1 = \lambda_0 p_0,$$

откуда

$$p_1 = \frac{\lambda_0}{\mu_1} p_0.$$

Для первого состояния имеем

$$\lambda_0 p_0 + \mu_2 p_2 = \lambda_1 p_1 + \mu_1 p_1,$$

откуда

$$\lambda_0 p_0 + \mu_2 p_2 = \lambda_1 \frac{\lambda_0}{\mu_1} p_0 + \mu_1 \frac{\lambda_0}{\mu_1} p_0.$$

Тогда

$$\mu_2 p_2 = p_0 \lambda_1 \lambda_0 / \mu_1 \text{ или } p_2 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} p_0.$$

Можно доказать, что из формулы $\lambda_{n-1}p_{n-1}+\mu_{n+1}p_{n+1}=\lambda_n p_n+\mu_n p_n$ следует, что для всех $n=1,2,\ldots,N$ выполняется (N – емкость системы)

$$p_n=\frac{\lambda_{n-1}}{\mu_n}p_{n-1}.$$

Тогда получаем **общую формулу для расчета финальных вероятностей** через p_0.

$$p_n=\frac{\prod_{i=0}^{n-1}\lambda_i}{\prod_{j=1}^{n}\mu_j}p_0.$$

Чтобы найти p_0 используется равенство

$$\sum_{n=0}^{N}p_n=1.$$

Задача 1. Бакалейный магазин работает с тремя кассами (обобщенная модель СМО). Вывеска возле касс извещает покупателей, что в любой момент будет открыта дополнительная касса, как только число покупателей в любой очереди превысит 3. Это означает, что если число покупателей меньше четырех, то работать будет лишь одна касса (одна касса работает всегда). Если число покупателей от четырех до шести, то будет работать две кассы. Если имеется больше шести покупателей, будут открыты все три кассы. Покупатели подходят к кассам в соответствии с распределением Пуассона с математическим ожиданием 10 человек в час. Время обслуживания одного покупателя в кассе распределено по экспоненциальному закону со средним 12 минут. Определим в установившемся режиме вероятность p_n, что n покупателей стоят в очереди в кассу.

$$\lambda = 10 чел./час, \mu_1 = \frac{60}{12} = 5 чел./час,$$

$$\mu = \begin{cases} 5, n \le 3 \\ 10, 4 \le n \le 6 \\ 15, n > 6 \end{cases}$$

Рассчитаем вероятности для одной открытой 1 кассы:

$$p_1 = \frac{\lambda_0}{\mu_1} \cdot p_0 = \frac{10}{5} \cdot p_0 = 2 \cdot p_0,$$

$$p_2 = \frac{\lambda_1}{\mu_2} \cdot p_1 = \frac{10}{5} \cdot p_1 = 4 \cdot p_0, p_3 = \frac{\lambda_2}{\mu_3} \cdot p_2 = \frac{10}{5} \cdot p_2 = 8 \cdot p_0,$$

Вероятности для 2 открытых касс:

$$p_4 = \frac{\lambda_3}{\mu_4} \cdot p_3 = \frac{10}{10} \cdot p_3 = 8 \cdot p_0,$$

$$p_5 = \frac{\lambda_4}{\mu_5} \cdot p_4 = \frac{10}{10} \cdot p_4 = 8 \cdot p_0,\ p_6 = \frac{\lambda_5}{\mu_6} \cdot p_5 = \frac{10}{10} \cdot p_5 = 8 \cdot p_0,$$

Для трех открытых касс получаем бесконечную геометрическую прогрессию с постоянным множителем 2/3:

$$p_7 = \frac{\lambda_6}{\mu_7} \cdot p_6 = \frac{10}{15} \cdot p_6 = 8 \cdot \frac{2}{3} \cdot p_0,$$

$$p_8 = \frac{\lambda_7}{\mu_8} \cdot p_7 = \frac{10}{15} \cdot p_7 = 8 \cdot \left(\frac{2}{3}\right)^2 \cdot p_0,$$

...

$$\sum_{n=0}^{\infty} p = 1,$$

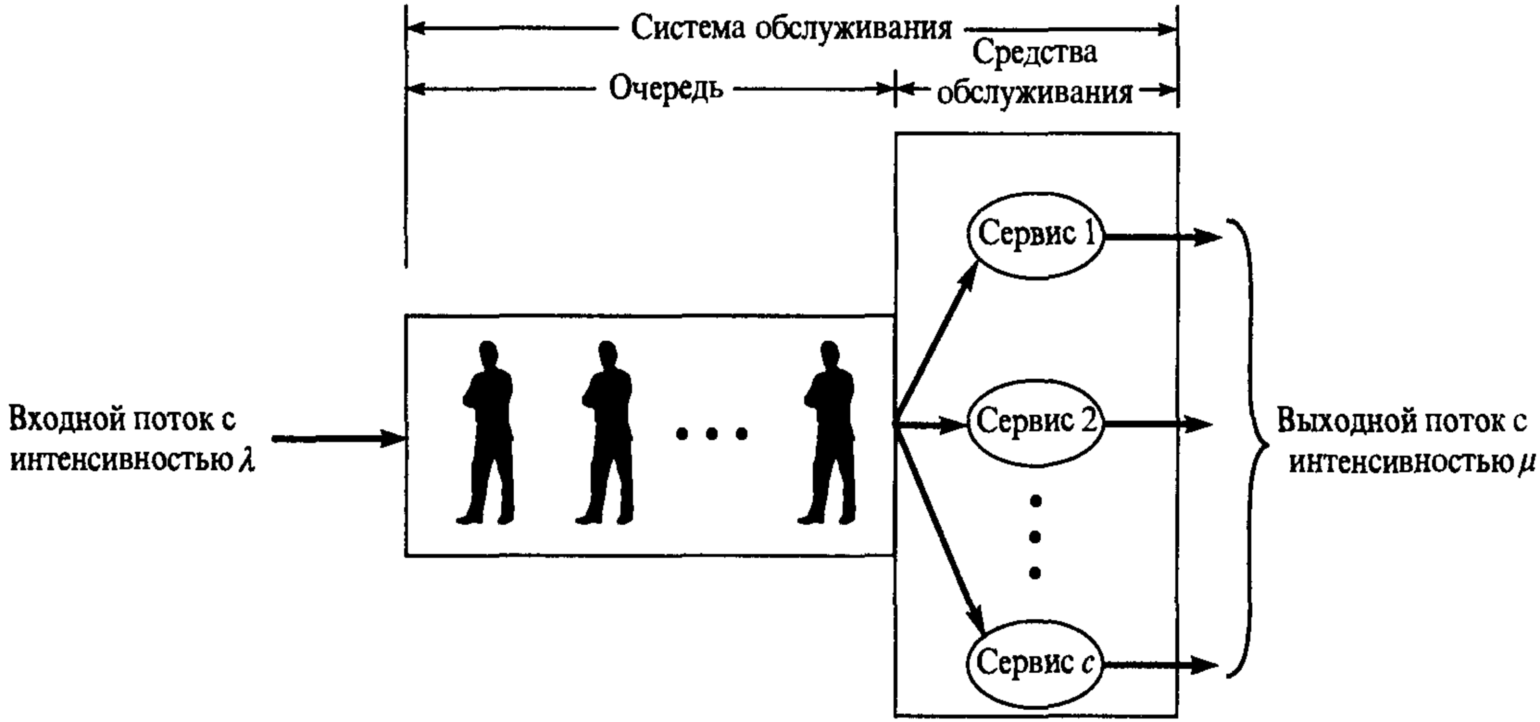
$$p_0 + p_1 + p_2 + \ldots = p_0 \cdot \left(1 + 2 + 4 + 8 + 8 + 8 + \frac{8}{1 - \frac{2}{3}} \right)$$

$$p_0 = \frac{1}{55}.$$

Рассчитав значение p_0, можно найти все остальные финальные вероятности

Специализированные СМО

Структура специализированной системы



На рисунке схематически представлена специализированная система обслуживания пуассоновского типа, в которой параллельно функционируют c идентичных сервисов (средств обслуживания). Ожидающий клиент выбирается из очереди для обслуживания на первом свободном сервисе. Интенсивность поступления клиентов в систему равна λ клиентов в единицу времени. Все параллельные сервисы являются идентичными; это означает, что интенсивность обслуживания каждого сервиса равна μ клиентов в единицу времени. Число клиентов, находящихся в *системе обслуживания,* включает тех, кто уже *обслуживается,* и тех, кто находится в *очереди.*

Обозначения, наиболее подходящие для характеристик системы обслуживания, имеют *следующую структуру*:

(a/b/c):(d/e/f),

где

a — тип распределения моментов времени поступления клиентов в систему,

b — тип распределения времени между появлением элементов выходного потока (времени обслуживания),

c — количество параллельно работающих сервисов (=1,2, ..., ∞),

d — дисциплина очереди,

e — максимальная емкость (конечная или бесконечная) системы (количество клиентов в очереди плюс число клиентов, принятых на обслуживание),

f — емкость (конечная или бесконечная) источника, генерирующего клиентов.

E_k — распределение Эрланга, или гамма-распределение интервалов времени (или, что то же самое, распределение суммы независимых случайных величин, имеющих экспоненциальное распределение),
GD — произвольный (общий) тип распределения моментов поступления клиентов на обслуживание,
G — произвольный (общий) тип распределения продолжительности обслуживания клиентов.

Для дисциплины очереди (символ d) используются следующие обозначения.
FCFS — первым пришел — первым обслуживаешься,
LCFS — последним пришел — первым обслуживаешься,
SIRO — случайный отбор клиентов,
GD — произвольный (общий) тип дисциплины.

В качестве исторической справки заметим, что первые три элемента **(*a*/*b*/*c*)** рассмотренного обозначения были введены Кендаллом (D. G. Kendall) в 1953 году, и в литературе по теории массового обслуживания они фигурируют как **обозначения Кендалла**. Позднее в 1966 году **Ли** (A. M. Lee) добавил к ним символы ***d*** и ***e***. **Тахой** 1968 году был введен последний символ принятых обозначений — *f*.

Задача 2. Описать систему $(M/D/10)$: $(GD/N/\infty)$.

В соответствии с принятыми обозначениями здесь речь идет о системе массового обслуживания с пуассоновским входным потоком (или экспоненциальным распределением интервалов времени между моментами последовательных поступлений клиентов), фиксированным временем обслуживания и десятью параллельно функционирующими сервисами. При этом дисциплина очереди не регламентирована, и максимальное количество допускаемых в систему клиентов равно N. Наконец, источник, "порождающий клиентов", имеет неограниченную емкость.

Функциональные характеристики стационарных систем обслуживания

Основными функциональными характеристиками систем массового обслуживания являются следующие.

L_s — среднее число находящихся в *системе* клиентов,

L_q — среднее число клиентов в *очереди,*

W_s — средняя продолжительность пребывания клиента в *системе,*

W_q — средняя продолжительность пребывания клиента в *очереди,*

$\bar{c}$ — среднее количество занятых средств обслуживания (сервисов).

Покажем, как перечисленные функциональные характеристики получаются (прямо или косвенно) из вероятностей p_n — вероятностей того, что в системе находится n клиентов. В частности, имеем следующее.

$$L_s = \sum_{n=1}^{N} n p_n \ ,$$

$$L_q = \sum_{n=c+1}^{N} (n-c) p_n \ .$$

Зависимость между L_s и W_s (а также между L_q и W_q), известная в литературе по теории массового обслуживания как **формула Литтла**, имеет вид

$$L_s = \lambda_{эфф} * W_s;$$

$$L_q = \lambda_{эфф} * W_q.$$

Эти соотношения справедливы при достаточно общих условиях. Параметр $\lambda_{эфф}$ представляет собой *эффективную* интенсивность поступления клиентов в систему обслуживания.

Параметр $\lambda_{эфф}$ равен (исходной) интенсивности поступления клиентов λ, когда все прибывающие клиенты имеют возможность попасть в обслуживающую систему. Если же некоторые клиенты не имеют такой возможности по той причине, что она заполнена (например, заполненная автостоянка), то $\lambda_{эфф} < \lambda$.

$$\lambda_{эфф} = \lambda(1 - p_{отказа}); \; p_N = p_{отказа};$$

$$\lambda_{потери} = \lambda - \lambda_{эфф} = \lambda * p_{отказа}$$

Существует также прямая зависимость между величинами W_s и W_q. По определению: средняя продолжительность пребывания в системе равна среднему времени пребывания в очереди плюс среднее время обслуживания.

$$W_s = W_q + 1/\mu .$$

Теперь можно получить формулу, связывающую L_S и L_q, умножая обе части последнего соотношения на $\lambda_{эфф}$ и используя формулу Литтла. В результате получаем

$$L_S = L_q + \lambda_{эфф} / \mu .$$

По определению разность между средним числом находящихся в системе клиентов L_s и средним числом клиентов в очереди L_q равна сред*нему количеству занятых узлов* обслуживания $\bar{c}$. Следовательно, имеем

$$\bar{c} = L_s - L_q = \sum_{n=1}^{c} np_n + \sum_{n=c+1}^{N} cp_n .$$

Поэтому *коэффициент использования* узлов обслуживания вычисляется как отношение

$\bar{c} / c$.

Задача 3. Автостоянка для посетителей колледжа имеет всего пять мест. Автомобили прибывают на стоянку в соответствии с распределением Пуассона с интенсивностью шесть автомобилей в час. Время пребывания автомобилей на стоянке является экспоненциально распределенной случайной величиной со средним 30 мин. Посетители, которые не могут найти свободного места на стоянке непосредственно по прибытии, могут временно ожидать освобождения места на территории стоянки. Таких мест для ожидания на стоянке имеется три. Если и стоянка, и все места для ожидания заполнены, то прибывшие автомобили вынуждены искать другую автостоянку. Требуется определить следующее:

1) вероятность p_n того, что в системе находится n автомобилей;
2) эффективную интенсивность поступления автомобилей на стоянку;
3) среднее количество автомобилей на стоянке;
4) среднее время нахождения автомобиля в очереди на территории стоянки;
5) среднее количество занятых мест на автостоянке.

Решение

Прежде всего, заметим, что место для стоянки в рассматриваемой ситуации выступает в роли сервиса, так что система имеет всего $c = 5$ средств обслуживания. Максимальная вместимость системы равна 5 + 3 = 8 автомобилей.

1) Вероятность p_n может быть определена как частный случай из обобщенной модели. Интенсивности поступления и обслуживания по исходным данным равны

$$\lambda_n = 6 \text{ машин/час}$$

$$\mu_n = \begin{cases} 60/30 * n = 2n, & n \le 5; \\ 5*(60/30) = 10, & n = 6,7,8. \end{cases}$$

Найдем финальные вероятности.

n	1	2	3	4	5	6	7	8
p_n	0,14436	0,21654	0,21654	0,16240	0,09744	0,05847	0,03508	0,02105

2) Вероятность отказа в обслуживании равна p_8. Отсюда

$$\lambda_{эфф} = (1 - p_8)\lambda = 5{,}8737 \text{ машин/час.}$$

3) $L_s = 3{,}1286$ машин.

4) $W_s = 0{,}53265$ часа.

5) $\overline{c} = 2{,}9368$ мест