УДК 519.872

Н.Н. Максимова, О.И. Сергамасова

ТЕОРИЯ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ И ЕЕ ПРИЛОЖЕНИЯ

В статье рассматриваются понятие теории массового обслуживания и ее приложения. Представлены составление уравнений Колмогорова и нахождение финальных вероятностей. Рассмотрены также примеры. Статья носит обзорный характер.

The paper is a review, in which the concept of queuing theory and its applications is considered. The paper presents the Kolmogorov equations set-up and finding of final probabilities. Some examples are given.

Введение

При исследовании операций часто приходится сталкиваться с работой своеобразных систем, называемых системами массового обслуживания (СМО). Примерами таковых могут служить телефонные станции, билетные кассы, магазины и т. п.

Каждая СМО состоит из какого-то числа обслуживающих единиц, которые будем называть «каналы обслуживания». Каналами могут быть линии связи, кассиры, продавцы и др. СМО могут быть одноканальными и многоканальными.

Всякая СМО предназначена для обслуживания какого-то потока заявок, поступающих в какие-то случайные моменты времени. Обслуживание заявки продолжается какое-то случайное время $T_{об}$, после чего канал освобождается и готов к приему следующей заявки [1]. Случайный характер потока заявок и времен обслуживания приводит к тому, что в какие-то периоды времени на входе СМО скапливается излишне большое число заявок (они либо становятся в очередь, либо покидают СМО необслуженными); в другие же периоды СМО будет работать с недогрузкой или вообще простаивать.

Работа СМО представляет собой случайный процесс с дискретными состояниями и непрерывным временем.

Предмет теории массового обслуживания – построение математических моделей, связывающих заданные условия работы СМО (число каналов, их производительность, правила работы, характер потока заявок) с интересующими нас характеристиками – показателями эффективности СМО, описывающими ее способность справляться с потоком заявок. В качестве таких показателей могут применяться разные величины, – например: среднее число заявок, обслуживаемых СМО в единицу времени; среднее число занятых каналов; среднее число заявок в очереди и среднее время ожидания обслуживания и т.д.

Системы массового обслуживания делятся на типы и классы по ряду признаков. Первое деление: СМО с отказами и СМО с очередью. В СМО с отказами заявка, поступившая в момент, когда все каналы заняты, получает отказ, покидает СМО и в дальнейшем процессе обслуживания не участвует. В СМО с очередью заявка, пришедшая в момент, когда все каналы заняты, не уходит, а становится в очередь и ожидает возможности быть обслуженной. На практике чаще встречаются СМО с очередью.

СМО с очередью подразделяются на разные виды в зависимости от того, как организована очередь – ограничена она или не ограничена. Ограничения могут касаться как длины очереди, так и времени ожидания. При анализе СМО должна учитываться также и «дисциплина обслуживания» – заявки могут обслуживаться либо в порядке поступления (раньше пришла, раньше обслуживается), либо в случайном порядке. Нередко встречается так называемое обслуживание с приоритетом – некоторые заявки обслуживаются вне очереди.

Существуют СМО с многофазовым обслуживанием, состоящим из нескольких последовательных этапов, или «фаз».

Кроме этих признаков, СМО делятся на два класса: «открытые» и «замкнутые». В открытой СМО характеристики потока заявок не зависят от того, в каком состоянии сама СМО (сколько каналов занято), в замкнутой – зависят [1].

Оптимизация работы СМО может производиться под разными углами зрения: с точки зрения организаторов СМО или с точки зрения обслуживаемых клиентов. С первой точки зрения желательно «выжать все, что возможно» из СМО и добиться того, чтобы ее каналы были предельно загружены. С точки зрения клиентов желательно уменьшение очередей, которые зачастую приводят к бессмысленной трате сил и времени, а в конечном итоге – к понижению производительности труда. При решении задач оптимизации в теории массового обслуживания существенно необходимы «системный подход», полное и комплексное рассмотрение всех последствий каждого решения.

Основы марковских процессов. Уравнения Колмогорова

Состояние СМО определяется количеством занятых каналов обслуживания и числом мест в очереди. Естественно, что эти параметры являются целочисленными и меняются скачкообразно в случайные моменты времени, определяемые появлением заявок во входном потоке. Исследование такой системы существенно упрощается, если переход СМО из одного состояния в другое может быть описан марковским процессом.

Пусть имеется система S, состояние которой изменяется во времени случайным образом и представляет собой случайный процесс. Процесс называется марковским, если для любого момента времени t_0 вероятностные характеристики процесса в будущем зависят только от его состояния в данный момент t_0 и не зависят от того, когда и как система пришла в это состояние.

Если некоторая система меняет свое состояние скачкообразно, причем переходы из одного состояния в другое обладают марковским свойством, то случайный процесс в такой системе – марковская цепь.

Марковская цепь называется дискретной, если переход из одного состояния в другое происходит в строго фиксированные промежутки времени, отделенные друг от друга равными интервалами. Если же эти переходы возможны в любой момент времени t, то соответствующая марковская цепь называется непрерывной. Переход из одного состояния в другое может быть отображен графом состояний, в котором вершины представляют собой возможные состояния системы, а дуги графа отражают переходы из одного состояния в другое. Если две вершины i и j соединяются дугой (i,j), то это означает, что возможен непосредственный переход из состояния i в состояние j.

Поскольку переход из одного состояния в другое для СМО возможен в любой момент времени, определяемый появлением заявки во входном потоке, то для изучения СМО применяются непрерывные марковские цепи.

Одна из важнейших задач теории марковских процессов вообще и ТМО в частности заключается в нахождении вероятностей состояний цепи. Эти вероятности для непрерывных марковских цепей определяются с помощью дифференциальных уравнений Колмогорова.

Рассмотрим некоторую произвольную систему. Система имеет n состояний $S_1, S_2, ..., S_n$. Процесс перехода из одного состояния в другое описывается непрерывной цепью Маркова. Пусть $p_i(t)$ – вероятность того, что в момент времени t система будет находиться в состоянии S_i $(i = 1,2,..., n)$. Поскольку система не может находиться одновременно в двух состояниях, то события, заключающиеся в нахождении системы в состояниях $S_1, S_2, ..., S_n$, несовместны и образуют полную систему событий. Отсюда следует:

$$\sum_{i=1}^{n} p_i(t) = 1.$$

Это соотношение называется условием нормировки. Задача заключается в определении вероятности любого состояния $p_i(t)$ в любой момент времени t.

Введем понятие вероятности перехода системы из состояния *i*, где она находилась в момент времени t, в состояние *j* за время Δt $p_{ij}(t, \Delta t)$. Очевидно, что $p_{ij}(t, \Delta t)$ представляет собой условную вероятность того, что в момент времени $t + \Delta t$ система окажется в состоянии S_j при условии, что в момент времени t она находилась в состоянии S_i: $p_{ij}(t, \Delta t) = p(S_j(t + \Delta t) / S_i(t))$.

Предел отношения вероятности перехода $p_{ij}(t, \Delta t)$ к длине интервала времени Δt назовем плотностью вероятности перехода

$$\lambda_{ij}(t) = \lim_{\Delta t \to 0} \frac{p_{ij}(t, \Delta t)}{\Delta t}.$$

Плотность вероятности перехода определим только для случаев $i \neq j$.

Если $\lambda_{ij}(t) = const$, то марковская цепь называется однородной. В противном случае, когда $\lambda_{ij}(t)$ являются функциями времени, цепь называется неоднородной. При расчетах вероятностей состояний марковской цепи предполагается, что все эти плотности вероятностей переходов λ_{ij} известны. Если у каждой дуги графа состояний системы проставить плотность вероятности перехода по данной дуге, то полученный граф назовем размеченным графом состояний (см. рис. 1).

Сформулируем теперь общее правило составления уравнений Колмогорова. В левой части каждого из них стоит производная вероятности какого-то *i*-го состояния, в правой – сумма произведений вероятностей всех состояний, из которых идут стрелки в данное состояние, на интенсивности соответствующих потоков событий, минус суммарная интенсивность всех потоков, выводящих систему из данного состояния, умноженная на вероятность данного *i*-го состояния.

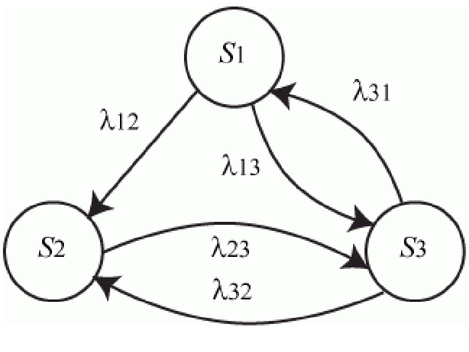

Рис. 1. Размеченный граф состояний.

Начальные условия для системы уравнений Колмогорова определяются начальным состоянием системы. Например, если начальное состояние было S_2, то начальные условия имеют вид: $p_1(0) = 0$; $p_2(0) = 1$; $p_3(0) = 0$;…; $p_n(0) = 0$.

Рассмотрим систему с размеченным графом состояний, изображенным на рис. 1. Система уравнений Колмогорова для этого случая в соответствии с правилом будет иметь вид:

$$\frac{dp_1}{dt} = -(\lambda_{12} + \lambda_{13})p_1 + \lambda_{31}p_3,$$

$$\frac{dp_2}{dt} = -\lambda_{23}p_2 + \lambda_{12}p_1 + \lambda_{32}p_3,$$

$$\frac{dp_3}{dt} = -(\lambda_{31} + \lambda_{32})p_3 + \lambda_{13}p_1 + \lambda_{23}p_2.$$

Чтобы система уравнений Колмогорова имела единственное решение при заданных начальных условиях, необходимо исключить любое из уравнений системы и заменить его условием нормировки.

Итак, решение системы без одного уравнения с условием нормировки определяет в любой момент времени поведение вероятностей состояний марковской цепи при заданных начальных условиях.

Получить это решение можно с помощью любых численных методов (например, Рунге-Кутта, Эйлера-Коши и т.д.), реализуемых на ЭВМ. Только в самых простых случаях система уравнений Колмогорова может быть проинтегрирована в квадратурах.

Теперь поставим вопрос: что будет происходить с вероятностями состояний при $t \to \infty$? Будут ли $p_1(t), p_2(t),...$ стремиться к каким-то пределам? Если эти пределы существуют и не зависят от начального состояния системы, то они называются финальными вероятностями состояний. В теории случайных процессов доказывается, что если число n состояний системы конечно и из каждого из них можно (за конечное число шагов) перейти в любое другое, то финальные вероятности существуют [1].

Финальные вероятности будем обозначать теми же буквами $p_1, p_2,...$, что и сами вероятности состояний, но под ними уже не переменные величины (функции времени), а постоянные числа. Очевидно, они тоже образуют в сумме единицу.

Финальную вероятность состояния S_i можно истолковать как среднее относительное время пребывания системы в этом состоянии.

Если вероятности $p_1, p_2,...$ постоянны, то их производные равны нулю. Значит, чтобы найти финальные вероятности, нужно все левые части в уравнениях Колмогорова положить равными нулю и решить полученную систему уже не дифференциальных, а линейных алгебраических уравнений. Можно и не писать уравнений Колмогорова, а прямо по графу состояний написать систему линейных алгебраических уравнений. Если перенести отрицательный член каждого уравнения из правой части в левую, то получим сразу систему уравнений, где слева стоит финальная вероятность данного состояния p_i умноженная на суммарную интенсивность всех потоков, ведущих из данного состояния, а справа – сумма произведений интенсивностей всех потоков, входящих в i-е состояние, на вероятности тех состояний, из которых эти потоки исходят.

$$(\lambda_{12} + \lambda_{13})p_1 = \lambda_{31}p_3,$$

$$\lambda_{23}p_2 = \lambda_{12}p_1 + \lambda_{32}p_3,$$

$$(\lambda_{31} + \lambda_{32})p_3 = \lambda_{13}p_1 + \lambda_{23}p_2.$$

Для решения данной системы необходимо исключить одно из уравнений и добавить условие нормировки:

$$p_1 + p_2 + p_3 = 1.$$

Приложения теории систем массового обслуживания

Схема гибели и размножения

Часто в системах самого различного назначения протекают процессы, которые можно представить в виде модели «гибели и размножения».

Граф состояний такого процесса показан на рис. 2.

Рис. 2. Схема «гибели и размножения».

Особенностью модели является наличие прямой и обратной связей с каждым соседним состоянием для всех средних состояний; первое и последнее состояния связаны только с одним «соседом».

Название модели – «гибель и размножение» – связано с представлением, что стрелки вправо означают переход к состояниям, связанным с ростом номера состояния («рождение»), а стрелки влево – с убыванием номера состояний («гибель»).

Составлять уравнения Колмогорова нет необходимости, так как структура регулярна, необходимые формулы приводятся в справочниках.

Для приведенных на рис. 2 обозначений формулы имеют вид:

$$p_0 = \left(1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} + \ldots + \frac{\lambda_{n-1,n}\ldots\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\lambda_{21}\lambda_{10}}\right)^{-1}.$$

$$p_1 = \frac{\lambda_{01}}{\lambda_{10}} p_0, \, p_2 = \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} p_0, \ldots, p_n = \frac{\lambda_{n-1,n\ldots}\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\lambda_{10}} p_0.$$

Пример 1

Процесс «гибели и размножения» представлен графом на рис. 3. Найти предельные вероятности состояний.

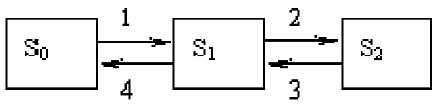

Рис. 3. Схема «гибели и размножения».

Решение

По формулам, указанными выше, найдем:

$$p_0 = \left(1 + \frac{1}{4} + \frac{2\cdot 1}{3\cdot 4}\right)^{-1} = 0.706, \; p_1 = \frac{1}{4}\cdot 0.706 = 0.176,$$

$$p_2 = \frac{2\cdot 1}{3\cdot 4}\cdot 0.706 = 0.118,$$

т.е. в установившемся, стационарном режиме в среднем 70,6% времени система будет находиться в состоянии S_0, в 17,6% – в состоянии S_1 и в 11,8% – в состоянии S_2.

Пример 2

Имеется система из двух одинаковых и работающих параллельно компьютеров. Требуется определить надежностные характеристики этой системы.

Решение

В этой системе возможны три состояния: S_1 – оба компьютера исправны; S_2 – один компьютер исправен, другой ремонтируется; S_3 – оба компьютера неисправны и ремонтируются. Будем полагать, что процессы отказов и восстановлений – однородные марковские, одновременный выход из строя обоих компьютеров, как и одновременное восстановление двух отказавших компьютеров, практически невозможны.

Поскольку компьютеры одинаковые, то с точки зрения надежности, неважно, какой именно компьютер неисправен в состоянии S_2, важно, что один.

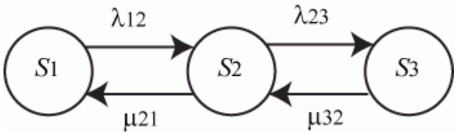

Рис. 4. Схема «гибели и размножения».

С учетом сказанного, ситуация моделируется схемой «гибели и размножения» (рис. 4).

λ_{12}, λ_{23} – интенсивности потоков отказов;

μ_{21}, μ_{32} – интенсивности потоков восстановлений.

Пусть среднее время безотказной работы каждого компьютера t =10 сут., а среднее время восстановления одного компьютера $t_в$ =0.1 сут.

Тогда интенсивность отказов одного компьютера будет равна:

$$\lambda = \frac{1}{t} = \frac{1}{10} = 0.1 \text{ (1/сут)},$$

а интенсивность восстановления одного компьютера:

$$\mu = \frac{1}{t_в} = \frac{1}{0.1} = 10 \text{ (1/сут)}.$$

В состоянии S_1 работают оба компьютера, следовательно:

$$\lambda_{12} = 2 \cdot \lambda = 2 \cdot 0.1 = 0.2 \text{ (1/сут)}.$$

В состоянии S_2 работает один компьютер, значит:

$$\lambda_{23} = \lambda = 0.1 = 0.1 \text{ (1/сут)}.$$

В состоянии S_2 восстанавливается один компьютер, тогда:

$$\mu_{21} = \mu = 10 \text{ (1/сут)}.$$

В состоянии S_3 восстанавливаются оба компьютера:

$$\mu_{32} = 2 \cdot \mu = 20 \text{ (1/сут)}.$$

Найдем вероятность состояния, когда обе машины исправны:

$$p_1 = \frac{1}{1 + \frac{\lambda_{12}}{\mu_{21}} + \frac{\lambda_{12}\lambda_{23}}{\mu_{21}\mu_{32}}} = \frac{1}{1 + \frac{0.2}{10} + \frac{0.2 \cdot 0.1}{10 \cdot 20}} = \frac{1}{1 + 0.02 + 0.0004} = 0.98.$$

Найдем вероятность второго состояния S_2 (работает один компьютер):

$$p_2 = \frac{\lambda_{12}}{\mu_{21}} \cdot p_1 = 0.02 \cdot 0.98 = 0.0196.$$

Аналогично вычисляется и p_3. Хотя найти p_3 можно и так:

$$p_3 = 1 - (p_1 + p_2) = 1 - (0.98 + 0.0196) = 1 - 0.9996 = 0.0004$$

Пример 3

Работа системы обслуживания клиентов, состоящей из двух работающих касс, задается графом состояний, изображенным на рис. 5.

Здесь состояние S_1 – обе кассы работают; S_2 – первая касса не обслуживает, а вторая – обслуживает; S_3 – обслуживает только первая касса; S_4 – обе кассы не обслуживают. Интенсивности переходов задаются величинами: λ_1= 2, μ_1= 1, λ_2= 3, μ_2= 2, λ_3= 3, μ_3= 3. Составить систему уравнений Колмогорова при заданных параметрах.

Решение

Система уравнений Колмогорова в соответствии с вышеизложенным правилом будет иметь вид:

$$\frac{dp_1}{dt} = -(\mu_1 + \mu_2)p_1 + \lambda_1 p_2 + \lambda_2 p_3,$$
$$\frac{dp_2}{dt} = -(\lambda_1 + \lambda_2 + \lambda_3)p_2 + \mu_1 p_1 + \mu_3 p_3 + \mu_2 p_4,$$
$$\frac{dp_3}{dt} = -(\lambda_2 + \mu_3 + \lambda_1)p_3 + \mu_2 p_1 + \lambda_3 p_2 + \mu_1 p_4,$$
$$\frac{dp_4}{dt} = -(\mu_1 + \mu_2)p_4 + \lambda_1 p_3 + \lambda_2 p_2.$$

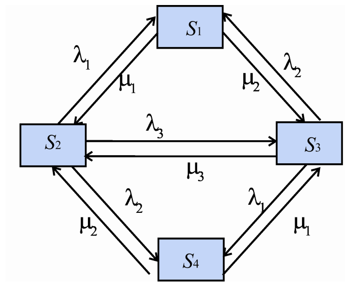

Рис. 5. Граф состояний.

Учтем, что $p_1 + p_2 + p_3 + p_4 = 1$, выразим отсюда p_4 и подставим в систему, при этом можно не учитывать последнее уравнение. Получим систему из трех уравнений с тремя неизвестными.

$$\frac{dp_1}{dt} = -(\mu_1 + \mu_2)p_1 + \lambda_1 p_2 + \lambda_2 p_3,$$
$$\frac{dp_2}{dt} = -(\lambda_1 + \lambda_2 + \lambda_3)p_2 + \mu_1 p_1 + \mu_3 p_3 + \mu_2(1 - p_1 - p_2 - p_3),$$
$$\frac{dp_3}{dt} = -(\lambda_2 + \mu_3 + \lambda_1)p_3 + \mu_2 p_1 + \lambda_3 p_2 + \mu_1(1 - p_1 - p_2 - p_3).$$

При заданных интенсивностях система примет вид:

$$\frac{dp_1}{dt} = -3p_1 + 2p_2 + 3p_3,$$
$$\frac{dp_2}{dt} = -2p_1 - 10p_2 + p_3 + 2,$$
$$\frac{dp_3}{dt} = p_1 + 2p_2 - 9p_3 + 1.$$

Пример 4

Работа системы, состоящей из двух независимо работающих приборов, задается графом состояний, изображенным на рис. 6. Здесь состояние S_1 – оба приборы исправны и работают; S_2 – первый прибор вышел из строя, а второй – исправно работает; S_3 – второй прибор вышел из строя, а первый – исправно работает; S_4 – оба прибора вышли из строя. Интенсивности переходов задаются величинами: $\lambda_1 = 2$, $\mu_1 = 1$, $\lambda_2 = 3$, $\mu_2 = 2$. Найти финальные вероятности для системы S.

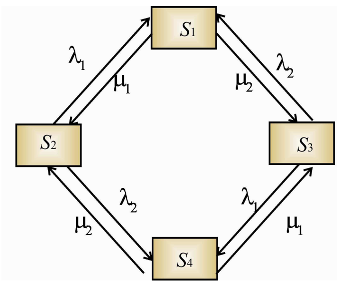

Рис. 6. Граф состояний.

Решение

Система алгебрических уравнений, описывающих стационарный режим данной системы, имеет вид:

$$(\mu_1 + \mu_2)p_1 = \lambda_1 p_2 + \lambda_2 p_3,$$
$$(\lambda_1 + \lambda_2)p_2 = \mu_1 p_1 + \mu_2 p_4,$$
$$(\lambda_1 + \lambda_2)p_3 = \mu_1 p_4 + \mu_2 p_1,$$
$$(\mu_1 + \mu_2)p_4 = \lambda_1 p_3 + \lambda_2 p_2.$$

Учтем, что $p_1 + p_2 + p_3 + p_4 = 1$. Выражаем p_4, подставляем в систему, при этом можно не учитывать последнее уравнение. Получим следующую систему:

$$(\mu_1 + \mu_2)p_1 = \lambda_1 p_2 + \lambda_2 p_3,$$
$$(\lambda_1 + \lambda_2)p_2 = \mu_1 p_1 + \mu_2(1 - p_1 - p_2 - p_3),$$
$$(\lambda_1 + \lambda_2)p_3 = \mu_1(1 - p_1 - p_2 - p_3) + \mu_2 p_1.$$

При заданных интенсивностях система примет вид:

$3p_1 = 2p_2 + 3p_3,$
$7p_2 = -p_1 - 2p_3 + 2,$
$6p_3 = p_1 - p_2 + 1.$

Решив систему уравнений, получим $p_1 = 0.31$, $p_2 = 0.19$, $p_3 = 0.19$, $p_4 = 0.31$. То есть в предельном стационарном режиме система S в среднем 31% времени будет находиться в состоянии S_1 (оба приборы исправны и работают), 19% – в состоянии S_2 (первый прибор вышел из строя, а второй – исправно работает), 19% – в состоянии S_3 (второй прибор вышел из строя, а первый исправно работает), 31% – в состоянии S_4 (оба прибора вышли из строя).

Заключение

За последние десятилетия в самых разных областях народного хозяйства возникла необходимость в решении вероятностных задач, связанных с работой систем массового обслуживания.

Математическая дисциплина, изучающая модели реальных систем массового обслуживания, получила название теории массового обслуживания. Задача теории массового обслуживания – установить зависимость результирующих показателей работы системы массового обслуживания (вероятности того, что требование будет обслужено; математического ожидания числа обслуженных требований и т.д.) от входных показателей (количество приборов в системе, параметров входящего потока требований и т.д.). Установить такие зависимости в формульном виде можно только для простых систем массового обслуживания. Изучение же реальных систем проводится путем имитации, или моделирования их работы на ЭВМ с привлечением метода статистических испытаний [2–6].

Анализ СМО упрощается, если в системе протекает марковский процесс, тогда систему можно описать обыкновенными дифференциальными уравнениями, а предельные вероятности – линейными алгебраическими уравнениями.

В дальнейшем планируется не только составить уравнения Колмогорова, но и решить их с помощью ППП Matlab для задачи моделирования транспортного потока на нерегулируемом пересечении. А также вычислить основные характеристики: средние длины очередей, вероятность не застать мест для ожидания в очереди, среднее время ожидания в очереди. И провести анализ характеристик при различных значениях интенсивности.

1. Вентцель, Е.С. Исследование операций. – М.: Наука, 1988. – 208 С.
2. Ивченко, Г.И. Теория массового обслуживания / Г.И. Ивченко, В.А. Каштанов, И.Н. Коваленко – М.: Высшая школа, 1982. – 310 с.
3. Овчаров, Л.А. Прикладные задачи теории массового обслуживания. – М.: Машиностроение, 1969. – 544 с.
4. Хинчин, А.Я. Работы по математической теории массового обслуживания. – М.: Физматгиз, 1963. – 264 с.
5. Кофман, А. Массовое обслуживание. Теория и приложения / А. Кофман, Р. Крюон. – М.: Мир, 1965. – 272 с.
6. Тараканов, К.В. Аналитические методы исследования систем / К.В. Тараканов, Л.А. Овчаров, А.Н. Тырышкин. – М.: Сов. радио, 1974. – 386 с.
7. Клейнрок, Л. Вычислительные системы с очередями / Л. Клейнрок. – М.: Мир, 1979. – 254 с.
8. Романцев, В.В. Моделирование систем массового обслуживания: Учеб. пособие / В.В. Романцев, С.А. Яковлев. – СПб.: СПбГЭТУ, 1995. – 104 с.