Лекция № 9. Основные понятия теории массового обслуживания

Цель лекции

Сформировать у обучающихся теоретические знания об основных понятиях, принципах и математических методах теории массового обслуживания. Изучить структуру, классификацию и закономерности функционирования систем массового обслуживания, характеристики входящих потоков заявок, особенности организации очередей и обслуживания. Ознакомить с методами определения вероятностей состояний системы, показателей пропускной способности, загрузки каналов и времени ожидания заявок. Развить навыки анализа эффективности одноканальных и многоканальных систем массового обслуживания и применения математических моделей для решения практических задач.

План лекции

1. Понятие теории массового обслуживания и история ее развития.
2. Основные задачи и области применения теории массового обслуживания.
3. Структура и основные элементы систем массового обслуживания.
4. Классификация систем массового обслуживания.
5. Системы массового обслуживания с отказами и с ожиданием.
6. Одноканальные и многоканальные системы с отказами.
7. Основные показатели эффективности функционирования СМО.
8. Одноканальные системы массового обслуживания с неограниченной очередью.
9. Многоканальные системы массового обслуживания с неограниченной очередью.
10. Понятие случайного процесса и его роль в теории массового обслуживания.
11. Марковские случайные процессы и потоки событий.
12. Простейший поток событий и его основные свойства.
13. Графы состояний систем массового обслуживания.
14. Практическое применение математических моделей СМО для оценки эффективности обслуживания.

1. Понятие теории массового обслуживания и история ее развития

Теория массового обслуживания представляет собой раздел прикладной математики, изучающий закономерности функционирования систем, в которых поступающие заявки обслуживаются одним или несколькими обслуживающими устройствами.

В зарубежной научной литературе теория массового обслуживания известна под названием Queueing Theory - теория очередей.

Основоположником теории массового обслуживания считается датский математик и инженер Ангер Крарруп Эрланг (1878–1929). В 1909 году он опубликовал работу, посвященную применению теории вероятностей к телефонным разговорам.

Исследования Эрланга заложили основу математического анализа потоков вызовов и эффективности телефонных станций. В его честь была названа единица измерения интенсивности телефонной нагрузки - эрланг.

Значительный вклад в развитие теории потоков событий внес советский математик Александр Яковлевич Хинчин (1894–1959). Его исследования в области теории вероятностей и случайных процессов способствовали формированию математических основ теории массового обслуживания.

В настоящее время теория массового обслуживания широко используется при исследовании информационных, производственных, транспортных и организационных систем.

Система массового обслуживания (СМО) - система, предназначенная для многократного выполнения однотипных операций по обслуживанию поступающих заявок.

Основная цель теории массового обслуживания заключается в разработке рекомендаций по рациональной организации работы обслуживающих систем, определению необходимого количества каналов и обеспечению высокой эффективности их функционирования.

2. Основные задачи и области применения теории массового обслуживания

Теория массового обслуживания предназначена для исследования зависимости эффективности работы системы от ее организационных и технических характеристик.

Основными факторами, определяющими эффективность СМО, являются характер поступающего потока заявок, количество обслуживающих каналов, производительность каждого канала и установленные правила обслуживания.

Методы теории массового обслуживания позволяют оценивать вероятность возникновения очередей, время ожидания клиентов, загрузку оборудования, пропускную способность и вероятность отказа в обслуживании.

Системы массового обслуживания встречаются в различных областях деятельности. К ним относятся телефонные станции, компьютерные сети, серверы, маршрутизаторы, автозаправочные станции, кассы, медицинские учреждения, транспортные терминалы, производственные участки и другие объекты.

Например, в компьютерной сети заявки могут представлять собой запросы пользователей, а обслуживающими устройствами выступают серверы.

При исследовании такой системы необходимо определить, сколько запросов может быть обработано за единицу времени, какова средняя продолжительность ожидания и какое количество серверов требуется для обеспечения заданного качества обслуживания.

3. Структура и основные элементы систем массового обслуживания

Любая система массового обслуживания состоит из нескольких взаимосвязанных элементов.

Основными элементами СМО являются входящий поток заявок, очередь, механизм обслуживания и выходящий поток заявок.

Входящий поток заявок

Интенсивность поступления λ

Очередь

Ожидание освобождения обслуживающего канала

Система обслуживания

Канал 1

Канал 2

Канал 3

Интенсивность обслуживания одного канала μ

Выходящий поток

Обслуженные заявки покидают систему

Входящий поток заявок - последовательность требований, поступающих в систему для обслуживания.

Очередь - совокупность заявок, ожидающих начала обслуживания.

Каналы обслуживания - устройства или исполнители, непосредственно выполняющие операции по обработке заявок.

Выходящий поток заявок - последовательность заявок, покидающих систему после обслуживания.

Процесс функционирования системы включает поступление заявки, ожидание при необходимости, обслуживание и последующий выход из системы.

4. Классификация систем массового обслуживания

Системы массового обслуживания классифицируются по количеству обслуживающих каналов, их взаимному расположению, возможности ожидания и дисциплине обслуживания.

По количеству каналов различают одноканальные и многоканальные системы.

Одноканальные СМО имеют одно обслуживающее устройство, тогда как многоканальные системы содержат несколько устройств, способных обслуживать заявки одновременно.

По расположению обслуживающих каналов системы подразделяются на системы с последовательным и параллельным обслуживанием.

В зависимости от возможности формирования очереди выделяют системы с отказами, с неограниченным ожиданием и с ограниченным ожиданием.
По организации обслуживания различают системы с одинаковым приоритетом заявок и системы, в которых отдельным заявкам предоставляется преимущество.

5. Системы массового обслуживания с отказами и с ожиданием

СМО с отказами - система, в которой заявка, поступившая при занятости всех обслуживающих каналов, получает отказ и покидает систему.
Такие системы используются в ситуациях, когда формирование очереди невозможно или не предусмотрено.
СМО с ожиданием - система, в которой поступившая заявка при занятости всех обслуживающих каналов становится в очередь.
В системах с неограниченным ожиданием отсутствуют ограничения на количество заявок в очереди и продолжительность их ожидания.
СМО с ограниченным ожиданием - система, в которой на процесс ожидания устанавливаются определенные ограничения.
Такие ограничения могут касаться максимальной длины очереди, времени ожидания или общего времени нахождения заявки в системе.

6. Одноканальные и многоканальные системы с отказами

В одноканальной системе с отказами рассматриваются два возможных состояния:

- S_0 - обслуживающий канал свободен;
- S_1 - обслуживающий канал занят.

Если интенсивность поступления заявок равна λ, а интенсивность обслуживания - μ, то для простейшей марковской модели установившиеся вероятности состояний определяются формулами:

$$p_0 = \frac{\mu}{\lambda + \mu}, \qquad p_1 = \frac{\lambda}{\lambda + \mu}.$$

Вероятность отказа в одноканальной системе равна вероятности занятости обслуживающего канала:

$$P_{\text{отк}} = p_1.$$

В многоканальной системе с отказами одновременно могут обслуживаться несколько заявок.
Если система имеет n каналов, ее состояния характеризуются количеством занятых обслуживающих устройств.
При пуассоновском поступлении заявок, экспоненциальном времени обслуживания и одинаковых независимых каналах вероятности состояний определяются соотношениями Эрланга:

$$p_k = \frac{a^k}{k!} p_0, \qquad k = 0, \dots, n,$$

где $a = \lambda/\mu$- предложенная нагрузка, а

$$p_0 = \left(\sum_{k=0}^{n} \frac{a^k}{k!} \right)^{-1}.$$

Вероятность отказа определяется вероятностью того, что все обслуживающие каналы заняты:

$$P_{\text{отк}} = p_n.$$

Полученные зависимости позволяют оценивать эффективность многоканальной системы и определять необходимое количество обслуживающих устройств.

7. Основные показатели эффективности функционирования СМО

Для оценки эффективности систем массового обслуживания используются специальные количественные показатели.

Абсолютная пропускная способность A - среднее количество заявок, которое система обслуживает за единицу времени.

Относительная пропускная способность Q - доля поступивших заявок, которые успешно обслуживаются системой:

$$Q = \frac{A}{\lambda}.$$

Вероятность отказа $P_{\text{отк}}$ - вероятность того, что поступившая заявка покинет систему без обслуживания.

Если каждая заявка либо обслуживается, либо получает отказ, то:

$$P_{\text{отк}} = 1 - Q.$$

Дополнительными характеристиками эффективности являются среднее число заявок в системе $L_{\text{сист}}$, средняя длина очереди $L_{\text{оч}}$, среднее время пребывания заявки в системе $T_{\text{сист}}$ и среднее время ожидания $T_{\text{оч}}$.

Важным показателем является также степень загрузки обслуживающих каналов.

Перечисленные характеристики позволяют исследовать производительность системы, определять причины возникновения очередей и оценивать необходимость изменения числа обслуживающих устройств.

8. Одноканальные системы массового обслуживания с неограниченной очередью

В одноканальной системе с неограниченной очередью имеется один обслуживающий канал, а количество ожидающих заявок не ограничивается.

Состояния системы определяются количеством заявок, находящихся на обслуживании или в очереди.

Состояние S_0 соответствует свободному каналу, S_1 - занятому каналу без очереди, S_2 - занятому каналу и одной ожидающей заявке.

В общем случае состояние S_k означает наличие k заявок в системе, из которых одна обслуживается, а остальные ожидают.

В простейшей модели M/M/1 коэффициент загрузки определяется выражением:

$$\rho = \frac{\lambda}{\mu},$$

где λ - интенсивность входящего потока, а μ - интенсивность обслуживания.

Если $\rho < 1$, система может функционировать в устойчивом стационарном режиме.

Если $\rho \geq 1$, стационарное распределение с конечной средней длиной очереди отсутствует.

При условии устойчивости среднее количество заявок в системе определяется формулой:

$$L_{\text{сист}} = \frac{\rho}{1 - \rho}.$$

Среднее количество заявок в очереди:

$$L_{\text{оч}} = \frac{\rho^2}{1-\rho}.$$

Среднее время пребывания заявки в системе:

$$T_{\text{сист}} = \frac{1}{\mu - \lambda}.$$

Среднее время ожидания в очереди:

$$T_{\text{оч}} = \frac{\rho}{\mu - \lambda}.$$

Эти формулы позволяют оценивать эффективность одноканальных обслуживающих систем при выполнении допущений модели M/M/1.
В презентации рассматривается задача о работе одного причала для разгрузки судов.
Интенсивность поступления судов составляет 0,4 судна в сутки, а средняя продолжительность разгрузки - 2 суток.
Следовательно, интенсивность обслуживания составляет 0,5 судна в сутки, а коэффициент загрузки равен 0,8.
При допущениях модели M/M/1 среднее число судов в системе составляет 4, среднее число судов в очереди - 3,2, а среднее время пребывания судна в системе - 10 суток.
Данный пример показывает, что даже при интенсивности поступления заявок ниже интенсивности обслуживания высокая загрузка системы может приводить к значительному увеличению времени ожидания.

9. Многоканальные системы массового обслуживания с неограниченной очередью

Многоканальная СМО содержит несколько обслуживающих устройств, работающих параллельно.
Поступившая заявка направляется на свободный канал. Если все каналы заняты, заявка становится в очередь.
Состояния системы определяются количеством занятых каналов и числом ожидающих заявок.
Если число заявок меньше количества каналов, свободные устройства могут оставаться незагруженными.
Когда число заявок достигает количества каналов, все устройства оказываются занятыми.
Дальнейшее поступление заявок приводит к формированию очереди.
В отличие от одноканальной модели, интенсивность завершения обслуживания зависит от количества занятых каналов.
Если в системе занято k каналов и каждый имеет интенсивность обслуживания μ, общая интенсивность завершения обслуживания составляет:

$$\mu_k = k\mu, \qquad k \le n.$$

При занятости всех n каналов интенсивность обслуживания достигает значения $n\mu$.
Для простой модели M/M/n с неограниченной очередью условием устойчивости является:

$$\lambda < n\mu.$$

При исследовании многоканальных систем определяются средняя длина очереди, время ожидания, загрузка каналов и другие показатели эффективности.
Результаты расчетов используются для выбора оптимального количества обслуживающих устройств с учетом необходимой производительности системы.

10. Понятие случайного процесса и его роль в теории массового обслуживания

Функционирование системы массового обслуживания представляет собой случайный процесс, поскольку поступление заявок и продолжительность обслуживания обычно зависят от случайных факторов.

Случайный процесс - процесс изменения состояния системы во времени в соответствии с определенными вероятностными закономерностями.

В теории массового обслуживания часто рассматриваются случайные процессы с дискретными состояниями и непрерывным временем.

Процесс называется процессом с дискретными состояниями, если множество его возможных состояний можно перечислить, а переход между ними происходит скачкообразно.

Процесс с непрерывным временем характеризуется тем, что моменты переходов между состояниями не фиксируются заранее и могут принимать произвольные значения времени.

Например, состояние системы обслуживания может измениться при поступлении новой заявки или завершении обслуживания ранее поступившей заявки.

Исследование таких процессов позволяет определять вероятности состояний системы и прогнозировать ее поведение при различных условиях нагрузки.

11. Марковские случайные процессы и потоки событий

Особое значение в теории массового обслуживания имеют марковские случайные процессы.

Марковский случайный процесс - случайный процесс, в котором вероятностные характеристики дальнейшего поведения системы определяются ее текущим состоянием и не зависят от предшествующей истории при условии известного настоящего состояния.

Использование марковского подхода существенно упрощает математический анализ сложных систем.

Одним из основных понятий теории массового обслуживания является поток событий.

Поток событий - последовательность однородных событий, происходящих в случайные моменты времени.

Примерами являются поток телефонных вызовов, запросов к серверу, покупателей в магазине и отказов оборудования.

Основной характеристикой потока является его интенсивность:

$$\lambda = \frac{N}{T},$$

где N - количество поступивших событий за интервал наблюдения T. Это выражение дает оценку средней интенсивности потока по наблюдениям.

В зависимости от статистических свойств различают стационарные, регулярные и другие виды потоков событий.

Стационарный поток характеризуется неизменностью вероятностных свойств при сдвиге временного интервала.

Регулярный поток представляет собой последовательность событий, возникающих через одинаковые промежутки времени.

При исследовании реальных систем необходимо учитывать, что поступление заявок может происходить неравномерно, а интенсивность потока способна изменяться в зависимости от времени и внешних условий.

12. Простейший поток событий и его основные свойства

Простейший поток событий - поток, который одновременно обладает свойствами стационарности, ординарности и отсутствия последействия.

Стационарность означает, что вероятностные характеристики потока не зависят от положения рассматриваемого временного интервала.

Ординарность означает, что вероятность появления двух и более событий за достаточно малый промежуток времени пренебрежимо мала по сравнению с вероятностью появления одного события.
Отсутствие последействия означает, что число событий на непересекающихся временных интервалах статистически независимо.
Простейший поток называют также стационарным пуассоновским потоком.
Вероятность появления ровно k событий за интервал времени t определяется законом Пуассона:

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Данный закон широко применяется для математического описания входящих потоков заявок в системах массового обслуживания.
Использование простейшего потока позволяет строить аналитические модели и определять основные вероятностные характеристики исследуемой системы.

13. Графы состояний систем массового обслуживания

При исследовании систем массового обслуживания используются графы состояний.
Граф состояний СМО - графическое представление возможных состояний системы и переходов между ними.
Вершины графа соответствуют состояниям системы, а направленные дуги отражают возможные переходы между состояниями.
Переходы могут происходить вследствие поступления новой заявки или завершения обслуживания.
В марковских моделях дуги часто обозначаются интенсивностями соответствующих переходов.
Для одноканальной СМО с отказами граф содержит два состояния: свободный и занятый канал.
Для многоканальной системы количество состояний увеличивается в зависимости от числа обслуживающих устройств и возможности формирования очереди.
Использование графов состояний позволяет наглядно представить структуру модели и составить математические уравнения для определения вероятностей ее состояний.

14. Практическое применение математических моделей СМО

Методы теории массового обслуживания применяются для анализа и оптимизации различных технических и организационных систем.
В компьютерных сетях они позволяют исследовать задержки передачи данных, загрузку каналов связи и эффективность серверного оборудования.
В транспортных системах - определять время ожидания транспортных средств, пропускную способность терминалов и необходимое количество обслуживающих устройств.
В производственных системах - оценивать загруженность оборудования, продолжительность выполнения операций и влияние технологических ограничений на производительность.
Результаты исследования используются для принятия решений о количестве обслуживающих каналов, изменении дисциплины обслуживания, распределении ресурсов и повышении эффективности функционирования системы.

Заключение

Теория массового обслуживания является важным разделом прикладной математики, позволяющим исследовать закономерности функционирования систем, в которых осуществляется обслуживание поступающих заявок. Ее методы широко применяются в технических, производственных, транспортных, информационных и организационных системах.

Основными элементами систем массового обслуживания являются входящий поток заявок, очередь, каналы обслуживания и выходящий поток. В зависимости от количества обслуживающих каналов, возможности ожидания и установленных правил обработки заявок выделяют различные виды СМО, отличающиеся структурой и характеристиками функционирования.

Особое значение имеет математическое моделирование одноканальных и многоканальных систем с отказами и ожиданием. Использование вероятностных моделей позволяет определять вероятности состояний системы, пропускную способность, среднюю длину очереди, время ожидания и степень загрузки обслуживающих устройств.

Основой математического анализа СМО являются теория вероятностей, случайные процессы, теория потоков событий и марковские модели. Применение этих методов обеспечивает возможность количественной оценки эффективности систем и прогнозирования их поведения при различных условиях нагрузки.

Таким образом, изучение основных понятий теории массового обслуживания формирует теоретическую основу для дальнейшего исследования сложных систем, разработки имитационных моделей и принятия обоснованных решений по оптимизации процессов обслуживания.

Контрольные вопросы

1. Что изучает теория массового обслуживания?
2. Кто является основоположником теории массового обслуживания?
3. Какой вклад в развитие теории массового обслуживания внесли А. К. Эрланг и А. Я. Хинчин?
4. Что понимается под системой массового обслуживания?
5. Какие основные задачи решает теория массового обслуживания?
6. В каких областях применяются системы массового обслуживания?
7. Назовите основные элементы СМО.
8. Что понимается под входящим потоком заявок?
9. Что такое очередь и канал обслуживания?
10. По каким основным признакам классифицируются СМО?
11. Чем одноканальные системы отличаются от многоканальных?
12. В чем заключаются особенности систем массового обслуживания с отказами?
13. Что понимается под СМО с ожиданием?
14. Чем отличаются системы с ограниченным и неограниченным ожиданием?
15. Какие состояния имеет одноканальная система с отказами?
16. Как определяется вероятность отказа в одноканальной системе?
17. Какие состояния рассматриваются в многоканальной системе с отказами?
18. Как определяется вероятность отказа в многоканальной СМО?
19. Что понимается под абсолютной пропускной способностью СМО?
20. Что характеризует относительная пропускная способность?
21. Какие основные показатели эффективности используются при исследовании СМО?
22. Что характеризует коэффициент загрузки обслуживающего канала?
23. Какие состояния имеет одноканальная СМО с неограниченной очередью?
24. При каком условии одноканальная СМО с неограниченной очередью функционирует в устойчивом режиме?

Список литературы

1. Гнеденко Б. В., Коваленко И. Н. Введение в теорию массового обслуживания. - 2-е изд., перераб. и доп. - М.: Наука, 1987. - 336 с.
2. Shortle J. F., Thompson J. M., Gross D., Harris C. M. Fundamentals of Queueing Theory. - 5th ed. - Wiley, 2018. - 576 p.
3. Harchol-Balter M. Performance Modeling and Design of Computer Systems: Queueing Theory in Action. - Cambridge: Cambridge University Press, 2013. - DOI: 10.1017/CBO9781139226424.