Лекция 07.09

Понятие обслуживания и системы обслуживания

Мы будем заниматься теорией массового обслуживания (ТМО). Методы этой теории являются одними из наиболее употребительных в прикладной математике! Для подтверждения этого утверждения можно привести следующий наглядный пример.

На рубеже нового тысячелетия в США проводилось множество всевозможных опросов и рейтингов. Было принято решение опросить крупнейшие американские компании, какие математические методы они систематически используют в своей работе. Этот вопрос задавался компаниям, входящим в индекс S@P 500 (Standard and Poor's), считающийся важнейшим показателем деловой активности в США, в индекс входят 500 американских компаний с самой высокой капитализацией. Результаты опроса были таковы: более 90% опрошенных используют методы математической статистики, почти 80% — методы оптимизации, на третьем месте примерно с 60% оказались методы теории массового обслуживания. Таким образом, ТМО оказалась одним из наиболее употребительных и практически значимых разделов прикладной математики.

ТМО возникла более ста лет тому назад, официальной датой ее появления считается 1909 год. Не так давно этот юбилей отмечался специалистами во всем мире. На начальное развитие теории большое влияние оказали работы датского ученого А.К. Эрланга (1878–1929). Первоначально задачи ТМО касались расчета числа линий связи на телефонных станциях, расчета запаса товаров на складах, расчета числа продавцов и кассиров для появившихся в начале прошлого века универмагов.

В последующие десятилетия круг задач ТМО существенно расширился, усложнились и сами постановки задач. ТМО является обязательным элементом образования прикладных математиков, ориентированных на приложения в экономической сфере.

Понятие обслуживания и системы обслуживания

Обслуживанием называют многократное выполнение однотипных разовых работ по заявкам (запросам), поступающим в произвольные моменты времени. При этом длительность каждой из работ также может быть произвольной.

В этом определении следует обратить внимание на ряд моментов. Во-первых, речь идет именно о многократном обслуживании, поступлении заявок на обслуживание, то есть о целом потоке заявок. Если заявки поступают редко и в небольшом количестве то каких-либо серьезных математических проблем при анализе таких систем не возникает.

Во-вторых, важно, что заявки могут поступать именно в произвольные моменты времени и каждая такая заявка выполняется также в течение заранее неизвестного времени. Иначе говоря, отсутствует жесткое расписание в получении и выполнении заявок. В результате возможны такие эффекты, как возникновение очереди заявок на обслуживание, простои системы и т.п.

В-третьих, совершенно не играет роли физическое содержание процесса обслуживания. Совершенно не важно, какие именно рабочие операции производятся при выполнении поступивших заявок.

В абстрактных математических моделях процесса обслуживания основной его характеристикой является затраченное время. Обслужить данный объект — это значит определенным образом его видоизменить в соответствии с полученной заявкой, переработать, причем контролируя при этом затраченное время.

Такая общая трактовка термина „обслуживание" позволяет применять математические модели ТМО при решении самых разнообразных задач прикладного содержания: естественнонаучных, военных, социологических, технических и, конечно, экономических.

Процессы обслуживания чрезвычайно часто встречаются на практике во всех сферах экономики: производстве, торговле, бизнесе, бытовом обслуживании, финансовом образовании, логистике, информационных технологиях. Например, модели обслуживания могут описывать обработку деталей рабочим, выполнение заказов фирмой, перевозку пассажиров на транспорте, погрузку и разгрузку кораблей в порту или вагонов на железной дороге, выдачу книг читателям в библиотеке, передачу информации через компьютерные сети, работу различного рода предприятий бытового обслуживания, мастерских, поликлиник, агентств, бюро и т.п.

Классические модели систем обслуживания не затрагивают экономическую деятельность этих систем. Отличительная особенность нашего курса состоит в том,

что мы будем рассматривать также и
экономические показатели работы систем
обслуживания, такие как затраты, выручка,
прибыль рентабельность.

С самого начала изучения моделей
обслуживания столкнулись с целым рядом
трудностей. Главная из них состоит в том,
что, как говорилось выше, моменты посту-
пления заявок являются произвольными,
то есть с математической точки зрения
должны трактоваться, как случайные
величины. Первые задачи ТМО были
связаны с анализом работы телефонных
станций. Попытки каким-то образом
описать поступление заявок на разговоры
по телефону с помощью детерминирован-
ных моделей не дали никакого результата.
Тщательное хронометрирование потоков вы-
зовов показало, что случайные колебания
моментов звонков по самой своей природе
является основной чертой рассматриваемого
класса задач. Похожая ситуация имеет место
и в других прикладных задачах обслужива-
ния. Например, введение номерков в поли-
клиниках или графика подачи деталей на
обработку в цехе способны лишь частично
сгладить случайные колебания моментов
прихода заявок в систему.

Следует отметить, что и время обслужи-
вания также, вообще говоря, на практике
является случайным. Здесь прежде всего
сказывается субъективный фактор, то есть
присутствие человека в процессе выполнения
заявок. Кроме того сами заявки, будучи
однотипными, все-таки несколько отличаются
по своему содержанию. Например, жалобы
больных при визите к врачу сугубо индиви-
дуальны, длительность поездки в такси у каждого
пассажира своя, заказанная читателем в

библиотеке книга может находиться в произвольной точке книгохранилища и т.д. Все это диктует необходимость использования в ТМО <u>вероятностных моделей</u>.

Совокупность сил и средств (включая и человека), организационно объединенная в единое целое и предназначенная для целей обслуживания, называется <u>системой обслуживания</u>.

Системы обслуживания являются основным предметом изучения в ТМО. С учетом сказанного выше о случайных факторах при обслуживании ТМО представляет собой некоторый специфический раздел <u>теории вероятностей</u>. Вероятностный подход здесь является в настоящее время общепринятым. Он позволяет адекватно описать и строго количественно изучить реальные системы обслуживания.

При применении методов ТМО реальные системы обслуживания заменяются некоторыми математическими моделями. Здесь необходимо ввести еще одно важное понятие.

<u>Системой массового обслуживания</u> (или сокращенно СМО) называют составленную по особым правилам математическую модель реальной системы обслуживания.

СМО формируется из стандартных элементов, конкретный вид, число и способ соединения которых зависит от содержания решаемой задачи. В настоящее время разработан и детально изучен широкий класс самых разнообразных СМО. Исходя из потребностей практики постоянно появляются новые модели СМО, продолжается изучение ранее предложенных моделей. Все это позволяет решать большое число различных прикладных задач, связанных с обслуживанием.

Рассмотрим типовую схему простейшей СМО, представленную на рис. 1.

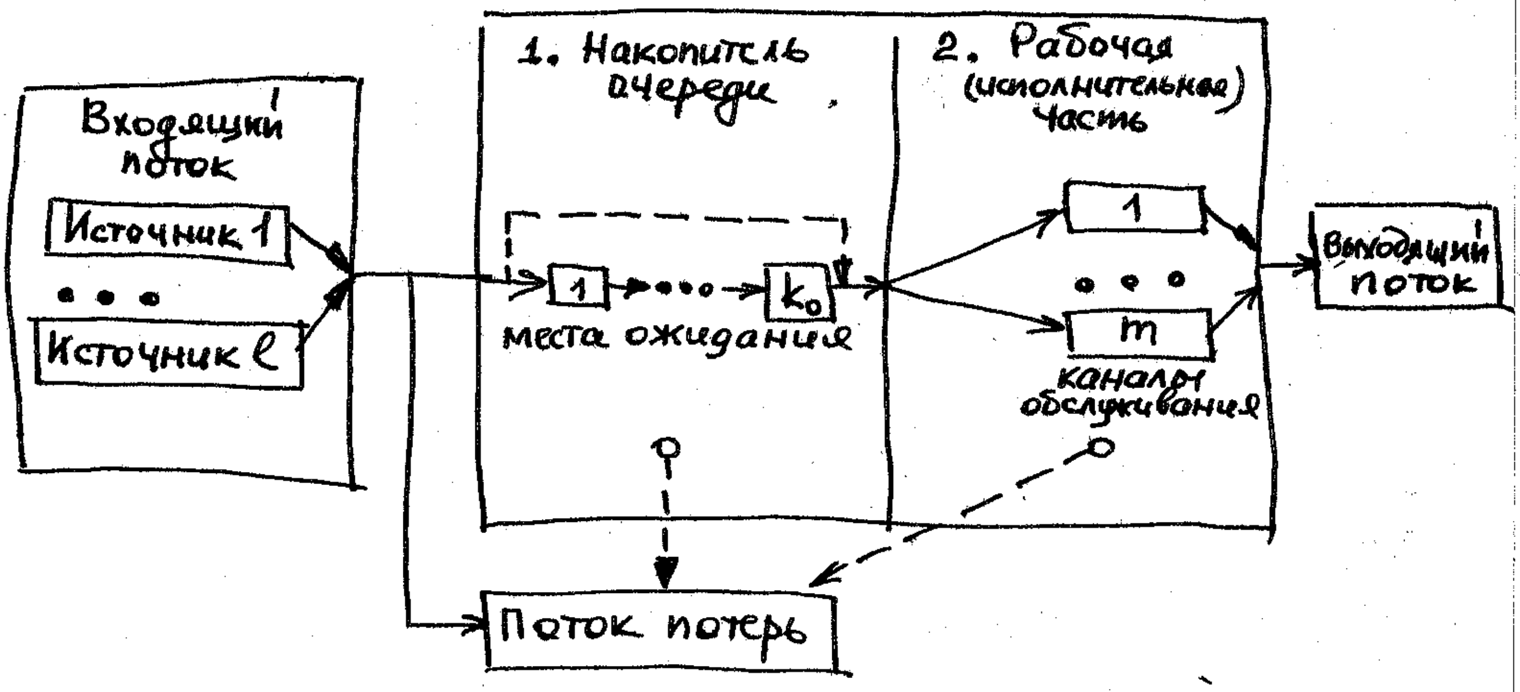

Рис. 1. Общая схема СМО: k – емкость системы, m – число каналов обслуживания, ℓ – число источников нагрузки, $k_0 = k - m$ – емкость накопителя

Эта схема включает следующие элементы: входящий поток требований, накопитель очереди, рабочую (исполнительную) часть, выходящий поток полностью обслуженных требований, поток потерь, составленный из требований, не попавших в систему, а также требований, потерянных из накопителя и рабочей части.

Входящий поток представляет собой случайный процесс, моделирующий процесс поступления требований в систему. Он является суммой элементарных потоков заявок от всех потенциальных клиентов, которые в этой теории называются источниками нагрузки. Их общее число ℓ называют числом источников нагрузки. Во многих задачах это число достаточно велико. Поэтому часто рассматривают асимптотический случай $\ell = \infty$. Существуют задачи, це принципиально важно учитывать, что l ограничено (например, задачи об обслуживании станков). Вместе с тем, как показывают расчеты, уже при сравнительно небольшом l (порядка 40–50 источников нагрузки) число все результаты, отвечающие таким значениям l практически совпадают с результатами, отвечающими $l = \infty$. Поэтому на практике при больших, но конечных l используют аппроксимацию $l = \infty$.

В любой СМО обязательно присутствует исполнительная (рабочая) часть системы, непосредственно выполняющая заявку на обслуживание.

Совокупность элементов рабочей части, необходимых для полного обслуживания одного требования называется <u>каналом обслуживания</u> (иногда используют термин „линия обслуживания").

Канал обслуживания включает все технологическое оборудование и всех работников, которое привлекается для выполнения отдельно взятой разовой работы.

Число каналов обслуживания m называют максимально возможное число одновременно обслуживаемых требований.

В зависимости от имеющегося числа каналов обслуживания различают одноканальные ($m = 1$), многоканальные ($1 < m < \infty$) системы, а также системы немедленного обслуживания ($m = \infty$). Режим работы СМО, отвечающий $m = \infty$, может быть реализован на практике, если постоянно держать в резерве достаточное число временных работников, готовых в любой момент включиться в работу, а также соответствующее оборудование.

Входящий поток и рабочая часть являются обязательными элементами СМО, без них система просто не может существовать. Во многих системах к ним добавляется еще один третий основной элемент — накопитель очереди. Его функции — сохранять требования, ожидающие обслуживания, и передавать их в надлежащем порядке в рабочую часть для обработки. Конкретная физическая реализация накопителя может быть различной в зависимости от профиля системы и ее специфики: „живая" очередь, стол заказов, бюро регистрации, склад и т.п.

Максимально возможное число требований k, которые одновременно могут находиться в системе, называется емкостью системы.

Очевидно, при максимальном заполнении системы будут заняты и все каналы обслуживания. Следовательно, емкость накопителя $k_0 = k - m$. На практике значение k_0 всегда ограничено, но в общей теории рассматривается и предельный случай $k_0 = \infty$. Он применяется для тех СМО, в которых вероятность переполнения накопителя весьма мала.

Характеристики состояния СМО

Работа системы, представленной на рис. 1, осуществляется следующим образом. Если вновь пришедшее требование застает хотя бы один свободный канал обслуживания, то оно сразу без ожидания поступает на обслуживание. Если же все каналы заняты, но есть свободное место в накопителе, то требование встает в очередь. Наконец, если заняты все места ожидания и все каналы, то требование безвозвратно теряется.

При окончании обслуживания какого-либо требования это полностью обслуженное требование поступает в <u>выходящий поток</u>. На его место из накопителя приходит в порядке очередности требование из очереди. <u>Поток потерь</u> формируется требованиями, которые не смогли попасть в систему, требованиями, которые покинули накопитель, не дождавшись начала обслуживания, а также требованиями, покинувшими рабочую часть до полного завершения их обслуживания.

Состояние СМО на рис.1 можно характеризовать числом требований в системе $N(t)$, числом занятых каналов обслуживания $N_{зк}(t)$, а также длиной очереди в накопителе $N_{оч}(t)$. К основным показателям работы СМО относятся также число требований поступивших на вход системы $N_{вх}(t)$, число выполненных требований $N_{вып}(t)$, число потерянных требований $N_{пот}(t)$, а также число требований, прошедших через накопитель $N_{нак}(t)$. Все эти характеристики будут зависеть от выбранного момента времени t и определяют экономические показатели работы системы.

Особо следует остановиться на расчете <u>длины очереди</u>. Проблема расчета длины очереди является одной из тех центральных проблем, которые стимулировали развитие всей теории в целом. Именно тот факт, что А.К. Эрланг дал строгое и корректное решение этой задачи, было расценено современниками как триумф теории массового обслуживания. Достаточно сказать, что сам термин „теория массового обслуживание" (введенный на русском языке А.Я. Хинчиным) в буквальном переводе с английского (the queueing theory) и немецкого (die Warteschlangentheorie) языков означает много иное, как „теория очередей".

Типы систем обслуживания

В зависимости от задания структурных параметров k и m выделяют несколько типов СМО. Пусть вначале $m < \infty$, то есть рассматривается одноканальная или многоканальная система.

Если в системе отсутствует накопитель $(k = m, k_0 = 0)$, то тогда требование, подошедшее в момент, когда все каналы обслуживания заняты, получит отказ в обслуживании и безвозвратно потеряется. Такую систему называют „системой с потерями" или „системой с отказами".

Другой крайний случай соответствует $k_0 = \infty$, когда СМО снабжена неограниченным накопителем. В этой системе все вновь подошедшие требования смогут встать в очередь и дожидаться обслуживания. Такая система называется „системой с ожиданием".

Промежуточный вариант $0 < k_0 < \infty$ относится к случаю „системы с ожиданием и потерями".

Наконец, в случае $m = \infty$ имеем „систему немедленного обслуживания", в которой нет ни ожидания, ни потерь, а все требования, поступившие на вход системы, сразу направляются на один из каналов обслуживания, среди которых всегда есть свободные.