**Теория массового обслуживания**

**1.1. История теории массового обслуживания**

В последние годы одним из ведущих направлений развития сетевой промышленности делаются сети беспроводной передачи информации. Для оценки данных этих сетей пользуют всевозможные имитационные и стохастические модели. Одним из раскладов к исследованию сетей считается ТМО.

ТМО появилась в начале X века, основателем теории является А.К. Эрланг, который трудился в телефонной компании и занимался изучением телефонных сетей. В последующем данная теория получила активное развитие и использование в различных областях науки, экономики, производства. Дело в том, что данная теория изучает довольно нередко встречающуюся ситуацию, когда имеется некоторый прибор (или ресурс) и большое количество заказов на его использование (в ТМО говорят «обслуживание на приборе»). Естественно, что в такой ситуации может появиться очередь и задержки в обслуживании. Математическую природу данных явлений и исследует ТМО.

Формальное описание СМО состоит из надлежащих объектов:

* поступающий поток требований;
* обслуживающий прибор или приборы (разного типа), еще их именуют каналом или каналами обслуживания;
* емкость накопителя для ожидания обслуживания (может также отсутствовать или же быть бесконечной);
* продолжительность обслуживания требования на приборе;
* дисциплина обслуживания (в каком порядке требования приходят на прибор и обслуживаются).

**1.2. Основные понятия. Классификация СМО**

При проведении исследования операций часто попадаются структуры, предназначенные для многократного использования при решении аналогичных задач. Процессы, возникающие в таком процессе, стали известны как сервисные процессы, а системы - как СМО. Образцами таких структур можно назвать телефонные структуры, ремонтные мастерские, компьютерные средства, магазины, парикмахерские и так далее.

Любая СМО состоит из фиксированного количества сервисных единиц (устройств, точек, станций), которые мы назовем сервисными каналами. Каналы могут быть линиями связи, рабочими точками, компьютерами, продавцами и так далее. По количеству каналов СМО делится на одноканальные и многоканальные.

Заказы обычно приходят не регулярно в СМО, а спонтанным образом, образуя так называемый спонтанный поток заказов (требований). Сервисные приложения, вообще говоря, также работают в течение некоторого спонтанного времени. Спонтанный характер потока запросов и времени обслуживаний приводит к тому, что СМО загружается неравномерно: в некоторые периоды времени накапливается очень большое количество запросов (они либо ставятся в очередь, либо оставляют СМО не обслуживаемыми), а другие СМО работают с недогрузкой или холостым ходом.

Предметом ТМО есть построения математических моделей, которые связывают заданные требования работы СМО (число каналов, их продуктивность, характер потока запросов и так далее). С показателями продуктивности СМО, описывающими его умениями управляться с потоком приложений.

В качестве уровня эффективности СМО используют следующие данные: среднее количество заказов, поданных за единицу времени; среднее количество заказов в очереди; средняя продолжительность ожидания обслуживания; шанс неудачи в обслуживании без ожидания; шанс того, что количество заказов в очереди превзойдет определенное значение и так далее.

СМО делится на два основных типа (классы): СМО с неудачами и СМО с ожиданием (очередь). В СМО с неудачами приложение, полученное в то время, когда все каналы не свободны, получает отказ, оставляет СМО и не участвует в дальнейшем процессе обслуживания (к примеру, приложение для телефонного разговора в то время, когда все каналы не свободны, получает отказ и оставляет СМО без присмотра). В СМО с ожиданием заказ, которая приходит в то время, когда все каналы не свободны, не уходит, а вместо этого оказывается в очереди на обслуживание.

СМО с ожиданием делятся на различные типы в зависимости от того, как организована очередь: с конечной или бесконечной длиной очереди, с ограниченной продолжительностью ожидания и т. д.

Для классификации СМО важна служебная дисциплина, которая определяет порядок выбора заказов по количеству полученных и порядок их распределения по свободным каналам. Согласно этому критерию обслуживание приложения может быть организовано на основе «первым поступил, первым обслужен», «последним поступил, первым обслужен» (эту процедуру можно использовать, к примеру, при вывозе товаров со склада для обслуживания, потому что последние часто более доступны) или приоритетное обслуживание (когда самые важные запросы обслуживаются первыми). Приоритет может быть либо абсолютным, когда более важный запрос «выдавливает» обычный запрос из службы (к примеру, в случае аварийной ситуации, запланированная работа ремонтных бригад прерывается до ликвидации аварии), и относительным когда более важный запрос получает только «наилучшее» место в очереди.

**1.3. Понятие марковского случайного процесса**

Процесс СМО является спонтанным процессом.

Под случайным процессом подразумевается процесс преобразования во времени состояния структуры в соответствии с вероятностными законами.

Процесс носит название процесс с дискретными состояниями, если его всевозможные состояния $S\_{1}, S\_{2}, S\_{3}, …$ могут быть перечислены заранее, и переход структуры из состояния в состояние протекает мгновенно (внезапно). Процесс именуется процессом с беспрерывным временем, в случае если моменты вероятных переходов структуры из состояния в состояние не фиксированы заблаговременно, а спонтанны.

Процесс работы с СМО - это спонтанный процесс с дискретными состояниями и беспрерывным временем. Это значит, что состояние СМО резко изменяется в спонтанные моменты появления некоторых событий (к примеру, прибытие нового приложения, окончание обслуживания и так далее).

Математический анализ работы СМО значительно упрощается, если в процессе этой работы используется Марков. Спонтанный процесс носит название марковский или спонтанный процесс без последствий, если для любого времени $t\_{0}$ вероятностные свойства процесса в будущем зависит только от его состояния в момент $t\_{0}$ и не зависят от того, когда и как структура перешла в это состояние.

Образец марковского процесса: структура S - счетчик в такси. Состояние структуры в момент времени t характеризуется количеством километров (десятых километров), пройденных транспортным средством до этой точки. Предположим, что в момент времени $t\_{0}$ счетчик показывает $S\_{0}$. Шанс того, что в момент $t>t\_{0}$ счетчик будет показывать то или иное количество километров (точнее, соответствующее количество тенге) $S\_{1}$, зависит от $S\_{0}$, но не зависит от того, в какой момент времени показания счетчика изменялись до момента $t\_{0}$.

Многие процессы можно приблизительно считать марковскими. К примеру, процесс игры в шахматы; структура S - группа шахматных фигурок. Состояние структуры характеризуется количеством вражеских фигурок, сохраненных на доске в момент времени $t\_{0}$. Шанс того, что в момент $t>t\_{0}$ материальное преимущество окажется на любой из сторон противников, зависит, прежде всего, от состояния, в котором в данный момент находится $t\_{0}$, а не от того, когда и в какой последовательности фигурки исчезли с доски до $t\_{0}$.

В некоторых случаях фон процессов, которые рассматриваются в данный момент, возможно просто пренебречь, и марковские модели использовались для их изучения.

При анализе спонтанных процессов с дискретными состояниями удобно использовать геометрическую схему - так называемый граф состояний. Как правило, структурные состояния представлены прямоугольниками (кружками), а всевозможные переходы из состояния в состояние обозначены стрелками, соединяющими состояния.

Пример 1. Спроектировать граф состояний спонтанного процесса: устройство S состоит из 2-ух узлов, каждый из которых может выйти из строя в любой момент времени, после чего начинается моментальное восстановление узла, которое продолжается с заранее неизвестным спонтанным временем.

◄Решение. Всевозможные состояния структуры: $S\_{0}$ — два узла исправны; $S\_{1}$ — 1-ый узел находится в ремонте, 2-ой исправен; $S\_{2}$ — 2-ой узел ремонтируется, 1-ый исправен; $S\_{3}$ — два узла находятся в ремонтных работах. Граф структуры представлен на рисунке 1



Рисунок 1

Стрелка, направленная, к примеру, из $S\_{0}$ в $S\_{1}$, означает переход структуры в момент неудачи 1-го узла, из $S\_{1}$ в $S\_{0}$ — переход в момент окончания ремонтных работ этого узла.

На графе не присутствуют стрелки из $S\_{0}$ в $S\_{3}$ и из $S\_{1}$ в $S\_{2}$. Это объясняется тем, что отказы узлов представляются независимыми друг от друга и, к примеру, можно пренебречь шансом одновременного отказа обоих узлов (переход из $S\_{0}$ в $S\_{3}$) или одновременного завершения ремонтных работ обоих узлов (переход из $S\_{3}$ в $S\_{0}$) возможно пренебречь.►

Для описания марковского спонтанного процесса с дискретными состояниями и беспрерывным течением временем в СМО. Одно из важных понятий теории вероятностей — понятие потока событий.

**1.4. Потоки событий**

Под потоком событий подразумевается последовательность однородных событий, которые идут одно за другим в некоторые спонтанные моменты времени (к примеру, поток вызовов на телефон, поток отказов компьютера, поток клиентов и т.п.).

Поток событий носит название регулярный, в случае если события идут одно за другим через равные промежутки времени. К примеру, поток товаров на конвеире сборочного цеха является регулярным.

Поток событий носит название стационарный, в случае если его вероятностные свойства не зависит от времени. В частности, интенсивности стационарного потока является постоянной величиной: $λ\left(t\right)=λ$. К примеру, поток автомобилей на проспекте города не является постоянным в течение дня, но этот поток можно считать стационарным в течение дня, скажем, в часы пик. Обращаем внимание на то, что в последнем случае фактическое количество проезжающих машин за единицу времени может значительно отличаться друг от друга, но их среднее количество будет постоянным и не будет зависить от времени.

Поток событий носит название обычный, если шанс попадания двух или более событий на небольшой (элементарный) временной интервал $∆t$ пренебрежимо мала по сравнению с шансом попадания в одно событие. Другими словами, поток событий является обычным, если события появляются в нем отдельно, а не в группах. К примеру, поток поездов, приближающихся к станции, является обычным, а поток автомобилей - не обычным.

Поток событий носит название простейший (или стационарным пуассоновским), если он одновременно является стационарным, обычным и не имеет последействие. Название «простейший» связано с тем, что СМО с простейшими потоками имеет простейшее математическое описание. Отметим, что обычный поток не является «самым простым», поскольку он имеет следствие: моменты возникновения событий в таком потоке жестко фиксированы.

Простой поток в качестве предельного получается в теории спонтанных процессов так же естественно, как и в теории вероятностей нормальные распределения получаются в качестве предельного для суммы спонтанных величин: когда суперпозиция достаточно большого количества $n$ независимых, стационарные и обычные потоки (сравнимые по интенсивности $λ\_{1} (i=1, 2, …, n)$) получается поток, ближайший к простейшему с интенсивностью $λ$, равная сумме интенсивностей поступающих потоков, то есть

$$λ=\sum\_{i=1}^{n}λ\_{i}.$$

Рассмотрим на оси времени $Ot$ (рисунок 2) простой поток событий как безграничную последовательность спонтанных точек.



Рисунок 2

Можно показать, что для простейшего потока количество m событий (точек), попадающих на любой участок времени τ, распределен по закону Пуассона

$$P\_{m}\left(τ\right)=\frac{\left(λτ\right)^{m}}{m!}e^{-λτ}, (1)$$

для которого математическое ожидание случайной величины такое же, как ее дисперсии: $a=σ^{2}=λτ$.

В частности, шанс того, что за время т не случится ни единого события (m=0), равна

$$P\_{0}\left(τ\right)=e^{-λτ}. (2)$$

Найдем распределение временного интервала $T$ между спонтанными двумя событиями по соседству простейшего потока.

В соответствии с (2) шанс того, что на участке времени длиной $t$ не будет замечено ни единого из последующих событий, равна

$$P\left(T\geq t\right)=e^{-λt}, (3)$$

а шанс противоположного события, то есть функция распределения случайной величины $T$, есть

$$F\left(t\right)=P\left(T<t\right)=1-e^{-λt}. (4)$$

Плотность вероятностей случайной величины есть производная ее функции распределения (рисунок 3), то есть

$$φ\left(t\right)=F^{'}\left(t\right)=λe^{-λt}. (5)$$



Рисунок 3

Распределение, задаваемое плотностью вероятности (5) или функциеи распределения (4), носит название показательный. Таким образом, временной интервал между двумя спонтанными соседствующими событиями имеет показательное распределение, для которого математичиское ожидание равно среднему квадратичискому отклонению случайной величины

$$a=σ=\frac{1}{λ} (6)$$

и обратно по величине интенсивности потока λ.

Для простого потока с интенсивностью $λ$ шанс попадания хотя бы одного события потока в элементарный (малый) временной интервал $∆t$ равна согласно (4)

$$P\_{∆t}=P\left(T<∆t\right)=1-e^{-λ∆t}≈λ∆t. (7)$$

**1.5. Процесс погибели и размножения**

В ТМО обширное распространение имеет особый класс спонтанных процессов – так называемый процесс погибели и размножения. Именование этого процесса тесно связано с рядом биологических задач, где он считается математической моделью преобразования количества биологических популяций.

Граф состояний процесса погибели и размножения имеет вид, представленный на рисунке 4.



Рисунок 4

Рассмотрим упорядоченное множество состояний структуры $S\_{0}, S\_{1}, S\_{2}, …, S\_{k}$. Переходы могут осуществлять из всякого состояния только в состояния с рядом стоящими номерами, то есть из состояния $S\_{k}$ возможны переходы только в состояние $S\_{k-1}$, или в состояние $S\_{k+1}$.

Допустим, все потоки событий, переводящие структуру по стрелкам графа, наипростейшие с надлежащими интенсивностями $λ\_{k,k+1}$ либо $λ\_{k+1,k}$.

По графу, изображенному на рисунке 4, составим и решим алгебраические уравнения для предельной вероятности состояний (их существование вытекает из возможности перехода из всякого состояния в любое другое и конечности количества состояний).

В соответствии с правилом составления таких уравнений (смотреть формулу 9) получим: для состояния $S\_{0}$

$$λ\_{01}p\_{0}=λ\_{10}p\_{1}, (8)$$

для состояния $S\_{1}- \left(λ\_{12}+λ\_{10}\right)p\_{1}=λ\_{01}p\_{0}+λ\_{21}p\_{2}$, которое с учетом (8) приводится к виду

$$λ\_{12}p\_{1}=λ\_{21}p\_{2}. (9)$$

Аналогично, делая запись уравнения для предельных вероятностей других состояний, можно получить следующую систему уравнений:

$$\left\{\begin{array}{c}\begin{array}{c}λ\_{01}p\_{0}=λ\_{10}p\_{1},\\λ\_{12}p\_{1}=λ\_{21}p\_{2}\\………………………….\end{array}\\λ\_{k-1,k}p\_{k-1}=λ\_{k,k+1}p\_{k}\\………………………….\\λ\_{n-1,n}p\_{n-1}=λ\_{n,n+1}p\_{n}\end{array}\right. (10)$$

к которой добавится нормировочное условие

$$p\_{0}+p\_{1}+p\_{2}+…+p\_{n}=1. (11)$$

Решая систему (10), (11), можно получить

$$p\_{0}=\left(1+\frac{λ\_{01}}{λ\_{10}}+\frac{λ\_{12}λ\_{01}}{λ\_{21}λ\_{10}}+…+\frac{λ\_{n-1,n}…λ\_{12}λ\_{01}}{λ\_{n,n-1}…λ\_{21}λ\_{10}}\right)^{-1}, (12)$$

$$p\_{1}=\frac{λ\_{01}}{λ\_{10}}p\_{0}, p\_{2}=\frac{λ\_{12}λ\_{01}}{λ\_{21}λ\_{10}}p\_{0}, …, p\_{n}=\frac{λ\_{n-1,n}…λ\_{12}λ\_{01}}{λ\_{n,n-1}…λ\_{21}λ\_{10}}p\_{0}. (13)$$

Легко заметить, что в формулах (13) для $p\_{1},p\_{2},…,p\_{n}$ коэфициенты при $p\_{0}$ есть слагаемые, стоящее после единицы в формуле (12). Числители данных коэффициентов представляют произведение всех интенсивностей, стоящих у стрелок, ведущих слева направо до данного состояния $S\_{k}$ $(k=1,2,…,n)$, а знаменатели — произведение всех интенсивностей, стоящих у стрелок, ведущих справа налево до состояния $S\_{k}$.

Пример 2. Процесс погибели и размножения представляется графом (рисунок 5). Определить предельные объективные возможности состояний.



Рисунок 5

◄Решение. По формуле (12) найдем

$$p\_{0}=\left(1+\frac{1}{4}+\frac{2∙1}{3∙4}\right)^{-1}=0,706 ,$$

по (13) $- p\_{1}=\frac{1}{4}∙0,706=0,176, p\_{2}=\frac{2∙1}{3∙4}∙0,706=0,118,$ то есть в установившемся, стационарном режиме структура будет находиться в среднем $70,6\%$ времени в состоянии $S\_{0}$, $17,6\%-$ в состоянии $S\_{1}$ и $11,8\%-$ в состоянии $S\_{2}$.►