# 8. МОДЕЛИРОВАНИЕ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ

**8.1. Понятие СМО**

*Системой массового обслуживания* (СМО) называется любая система предназначенная для обслуживания каких-либо заявок (требований), посту­пающих на нее в случайные моменты времени.

В качестве процесса обслуживания могут быть представлены различные по своей физической природе процессы функционирования экономических, производственных, технических и других систем. Примеры систем массового обслуживания следующие: потоки поставок продукции некоторому предприятию, потоки деталей и комплектующих изделий на сборочном конвейере цеха, заявки на обработку информации ЭВМ от удаленных терминалов и т. д. При этом характерным для работы таких объек­тов является случайное появление заявок (требований) на обслу­живание и завершение обслуживания в случайные моменты време­ни, т. е. стохастический характер процесса их функционирования. Остановимся на основных понятиях массового обслуживания, не­обходимых как при аналитическом, так и при имитационном подходе.

Работа любой системы массового обслуживания состоит в выполнении поступающего на ее вход потока ***заявок***. Заявки поступают в некоторые, в общем случае случайные, моменты времени. Обслуживание заявки продолжается какое-то время, также случайное, после чего канал освобождается для обслуживания следующей заявки. Предмет теории массового обслуживания – установление зависимостей между характером потока заявок, производительностью отдельного канала обслуживания, числом каналов и эффективностью обслуживания.

Различают СМО *с отказами и* *СМО с очередью.* В СМО с отказами заявка, пришедшая в момент, когда все каналы заняты, получает отказ, покидает СМО и в дальнейшем в процессе ее работы не участвует. В СМО с очередью заявка, при­шедшая в момент занятости всех каналов, не покидает СМО, а становится в оче­редь и ждет, пока не освободится какой-нибудь канал. Число мест в очереди m может быть как ограниченным, так и неограниченным. При m = О СМО с оче­редью превращается в СМО с отказами. Очередь может иметь ограничения не только по количеству стоящих в ней заявок (длине очереди), но и по времени ожидания (такие СМО называются «системами с нетерпеливыми клиентами»).

СМО с очередью различаются не только по ограничениям очереди, но и по *дисциплине обслуживания:* обслуживаются ли заявки в порядке поступления, или в случайном порядке, или же некоторые заявки обслуживаются вне оче­реди (так называемые «СМО с приоритетом»). Приоритет может иметь несколько градаций или рангов.

Аналитическое исследование СМО является наиболее простым, если все потоки событий, переводящие ее из состояния в состояние, — простейшие (ста­ционарные пуассоновские). Это значит, что интервалы времени между события­ми в потоках имеют показательное распределение с параметром, равным интен­сивности соответствующего потока. Для СМО это допущение означает, что как поток заявок, так и поток обслуживании — простейшие. Под *потоком обслужи­вании* понимается поток заявок, обслуживаемых одна за другой одним непрерыв­но занятым каналом. Этот поток оказывается простейшим, только если время обслуживания заявки *Тобс* представляет собой случайную величину, имеющую показательное распределение. Параметр этого распределения  есть величина, обратная среднему времени обслуживания. Вместо «поток обслуживании — простейший» часто говорят «время обслужива­ния — показательное». Условимся в дальнейшем для краткости всякую СМО, в которой все потоки простейшие, называть *простейшей* СМО. В этой главе мы будет рассматривать главным образом простейшие СМО.

Если всё потоки событий простейшие, то процесс, протекающий в СМО, представляет собой марковский случайный процесс с дискретными состояниями и непрерывным, временем. При выполнении некоторых условий для этого процес­са существует финальный стационарный режим, при котором как вероятности со­стояний, так и другие характеристики процесса не зависят от времени.

Задачи теории массового обслуживания — нахождение вероятностей раз­личных состояний СМО, а также установление зависимости между заданными параметрами (числом каналов n, интенсивностью потока заявок , распределе­нием времени обслуживания и т, д.) и *характеристиками эффективности* ра­боты СМО. В качестве таких характеристик могут рассматриваться, например, следующие:

* среднее число заявок *А,* обслуживаемое СМО в единицу времени, или *абсолютная пропускная способность* СМО;
* вероятность обслуживания- поступившей заявки Q или *относительная пропускная способность* СМО;*;*
* вероятность отказа Ротк т.е вероятность того, что поступившая заявка не будет обслужена, получит отказ; Ротк = 1 - q;

Рассмотри процессы, протекающие в системе массового обслуживания.

**8.2.Мнемоническое обозначение СМО.**

В теории массового обслуживания приняты очень удобные сокращенные обозначения для различных СМО, позволяющие легко охарактеризовать систему. В основе этих обозначений лежит трехбуквенная комбинация вида А/В/N, где:

А — описывает распределение (или задает характер закона распределения) интервалов поступления заявок;

В — описывает распределение длительностей обслуживания заявок;

N — задает количество обслуживающих приборов в СМО.

Иногда, когда СМО является системой с ограниченной емкостью накопителя (или с ограниченной очередью), приведенное обозначение расширяется до четырех букв А/В/N/К, где последняя буква (на самом деле число, как и N) К задает емкость накопителя (количество мест ожидания).

Приведенные трех или четырех буквенные обозначения называют обозначениями Кендалла. В этих обозначениях А и В могут принимать значения из следующего набора символов {M, D, Ek, Hk, G, U}. При этом:

а) А или В=M, если распределение интервалов поступления или длительностей обслуживания заявок является экспоненциальным (М — от слова Markovian — Марковский);

б) А или В=D, если интервалы поступления или длительности обслуживания являются детерминированными (D — Determinate);

в) А или В=Ek, если соответствующие распределения являются Эрланговскими порядка k (E — Erlang);

г) А или В=Hk, в случае гиперэкспоненциальных распределений порядка k (H — Hyperexponential);

д) А или В= G, в случае распределений общего (произвольного) вида (G — General — общий, общего вида);

е) А или В= U — при равномерных распределениях соответствующих случайных величин (U — Uniform distribution — равномерное распределение).

Так, например, обозначение вида:

М/М/1 означает СМО с простейшим потоком на входе и экспоненциально распределенной длительностью обслуживания заявок в приборе (один)

D/Е2/3/5 — СМО с регулярным потоком на входе, длительностью обслуживания, распределенной по закону Эрланга 2-го порядка, тремя обслуживающими приборами и пятью местами ожидания;

М/G/2 — СМО с простейшим потоком на входе, длительностью обслуживания, распределенная по закону произвольного вида, и двумя обслуживающими приборами.

В случае СМО с неоднородной нагрузкой используются обозначения вида, где символ вектора над буквами А и В указывает на неоднородность нагрузки, а индекс Н задает количество классов заявок. Например, — это обозначение СМО с одним обслуживающим прибором, четырьмя классами заявок, которые образуют на входе системы простейшие потоки и имеют общие законы распределения длительностей обслуживания.

**8.3. Потоки обслуживания**

Характеристики потока событий, поступающих на вход системы были рассмотрены в предыдущей теме.

* 1. **Время обслуживания**

Кроме характеристик входного потока заявок, режим работы СМО зависит также от производительности самой системы: числа каналов n и быстродействия каждого канала.

Одной из важнейших характеристик, связанных с производительностью системы является *время обслуживания одной заявки Tоб*, которая может быть как случайной так и регулярной величиной.

Для практики особый интерес представляет случай, когда закон распределения времени обслуживания является экспоненциальным с параметром μ:

g(t) = μe-μ, (t>0),

где μ – величина, обратная среднему времени обслуживания заявки.

Особая роль, которую играет в теории массового обслуживания показательный закон распределения времени обслуживания связана со следующим свойством: *если в какой-то момент времени t0 происходит обслуживание заявки, то закон распределения оставшегося времени обслуживания не зависит от того, сколько времени обслуживание уже продолжалось.*

Показательный закон не является универсальным законом распределения времени обслуживания. Часто время обслуживания лучше описывается другими законами, например, нормальным или Эрланга. Однако, пропускная способность и другие характеристики СМО сравнительно мало зависят от вида закона распределения времени обслуживания, а зависят, главным образом, от его среднего значения. Поэтому, в теории массового обслуживания чаще всего пользуются допущением, что время обслуживания распределено áî показательному закону.

**8.5. СМО с отказами**

Системы массового обслуживания делятся на системы с отказами и системы с ожиданием.

В системах с отказами заявка, поступившая в момент, когда все каналы обслуживания заняты, немедленно получает отказ, покидает систему и в дальнейшем в процессе обслуживания не участвует.

Пусть имеется n-канальная СМО с отказами. Рассмотрим конечное множество состояний этой системы:

*z0* – свободны все каналы;

*z1* – занят один канал;

. . . . . . . . . . . . . . . . . . .

*zk* – заняты k каналов;

. . . . . . . . . . . . . . . . ..

*zn* – заняты все n каналов.



*Определим вероятности состояния системы pk(t) для любого момента времени в предположении, что поток заявок простейший, с интенсивностью λ, время обслуживания показательное, с параметром μ.*

*Поскольку оба потока заявок в системе (заявок и обслуживания) являются простейшими, то процесс, протекающий в системе будет марковским.*

*Очевидно, что для любого момента времени*

*.*

*Составим дифференциальные уравнения для всех вероятностей состояний системы. Для этого, зафиксируем момент времени t и найдем вероятность pk(t+Δt) того, что в момент (t+Δt) система будет находиться в состоянии zk.*

*Для состояния z0 это может произойти двумя способами:*

*событие A – в момент времени t система находилась в состоянии z0 и осталась в этом состоянии. Вероятность этого события равна вероятности того, что за время Δt на вход системы не пришла ни одной заявки:*

**

*Следовательно, P(A)=p0(t)(1-λ˙Δt).*

*событие B – вероятность того, что система была в состоянии z1 и перешла в состояние z0. Вероятность этого события равна:*

**

*Следовательно, P(B)=p1(t)μ˙Δt.*

*Таким образом:*

*p0(t+Δt)=p0(t)(1-λ˙Δt)+ p1(t)μ˙Δt.*

**

*Аналогично составляются дифференциальные уравнения для других состояний системы. Для состояния zk вероятность pk(t+Δt) определиться как сумма вероятностей трех событий:*

*событие A – в момент времени t система находилась в состоянии zk и осталась в этом состоянии. Вероятность этого события равна вероятности того, что за время Δt на вход системы не пришла ни одной заявки и ни одна из k заявок из системы не ушла (не обслужилась):*

**

*Следовательно, P(A)=pk(t)[1-(λ+kμμ)˙Δt].*

*событие B – вероятность того, что система была в состоянии zk-1 и перешла в состояние zk. (пришла одна заявка). Вероятность этого события равна:*

*P(B)=pk-1(t)λ˙Δt.*

*событие C – вероятность того, что система была в состоянии zk+1 и перешла в состояние zk. (обслужена одна заявка). Вероятность этого события равна:*

*P(С)=pk+1(t)(k+1)μ˙Δt.*

*Таким образом:*

*pk(t+Δt)= pk(t)[1-(λ+kμμ)˙Δt]+ pk-1(t)λ˙Δt+ pk+1(t)(k+1)μ˙Δt.*

**

*Составим уравнение для последней вероятности pn:*

*pn(t+Δt)≈ pn(t)(1-nμ˙Δt)+ pn-1(t)λ˙Δt.*

**

*Таким образом, получена система дифференциальных уравнений для вероятностей состояний системы:*

**

*. . . . . . .*

**

*. . . . . . (0<k<n),*

**

*Эти уравнения называются уравнениями Эрланга.*

Вероятности *pk(t)* характеризуют среднюю загрузку системы и ее изменение с течением времени.

Вероятность *pn(t)=Pотк* есть вероятность того, что заявка, пришедшая в систему в момент времени t получит отказ.

Величина q(t)=1-pn(t) называется пропускной способностью системы.

Введем обозначение α=λ/μ и назовем величину α *приведенной плотностью потока заявок*. Эта величина есть также среднее число заявок, приходящееся на среднее время обслуживания одной заявки: α=λmобсл.

В новых обозначениях вероятности pk принимает вид:



Приведенные выше формулы выражают вероятности pk через p0. Для того, чтобы выразить эти вероятности через характеристики системы α и n, воспользуемся условием нормировки:



откуда



Окончательное выражение для вероятностей состояния системы принимают вид:

 (0≤k≤n).

Вероятность отказа (все каналы заняты):



Для одноканальной системы (n=1):

.

Относительная пропускная способность :



Формулы Эрланга и их следствия были получены в предположении о показательном распределении времени обслуживания заявок. Однако исследования показали, что эти формулы справедливы при любом законе распределения времени обслуживания, лишь бы входной поток был простейшим.

**7.4. СМО с ожиданием**

Система массового обслуживания называется системой с ожиданием, если заявка, заставшая все каналы занятыми, становится в очередь и ждет, пока не освободится какой-нибудь канал.

Если время ожидания заявки в очереди ничем не ограничено, то система называется *чистой системой с ожиданием*. Если оно ограничено некоторыми условиями, то система называется системой смешанного типа. Ограничения, наложенные на ожидание могут быть различного типа, например:

* ограничение на время пребывания заявки в очереди;
* ограничение на длину очереди;
* ограничение на время пребывания заявки в системе.

В системах с ожиданием существенную роль играет так называемая *дисциплина очереди*. Каждый тип системы с ожиданием имеет свои особенности и математическую теорию. Мы остановимся на простейшем случае смешанной системы, являющимся обобщением задачи Эрланга для системы с отказами.

Рассмотрим СМО с n каналами, на вход которой поступает простейший поток с параметром λ. Время обслуживания заявок также имеет показательное распределение с параметром μ. Заявка, заставшая все каналы занятыми, становится в очередь и ожидает обслуживания. Время ожидания заявки в очереди ограничено некоторым сроком Tож. Если до истечения этого срока заявка не будет обслужена, то она покидает систему. Срок ожидания обслуживания будем полагать случайной величиной с показательным распределением и параметром ν. Очевидно, что при ν→∞, система смешанного типа превращается в чистую систему с отказами, а при ν→0, система смешанного типа превращается в чистую систему с ожиданиями.

Отметим, что в предположении о показательном распределении срока ожидания пропускная способность системы не зависит от того, обслуживаются ли заявки в порядке очереди ли в случайно порядке: для каждой заявки закон распределения оставшегося времени ожидания не зависит от того, сколько времени заявка стояла в очереди.

* для любого k≤n 
* для любого s≥1: 

В приведенных выше формулах в качестве сомножителя присутствует вероятность p0. Определим эту вероятность из дополнительного условия:



Введем обозначения:

λ/μ=λmtобсл= α; ν/μ=νmtобсл=β.

Параметра α и β выражают соответственно среднее число заявок и среднее число необслуженных заявок приходящееся на среднее время обслуживания одной заявки.

В новых обозначениях приведенные выше выражения принимают вид:

 (0<k≤n)

; (s≥1).

Зная вероятности состояния системы можно определить и другие интересующие нас характеристики, в частности вероятность того, что заявка покинет систему не обслуженной. Определим эту вероятность из следующих соображений: при установившемся режиме вероятность Pн есть отношение среднего числа заявок, уходящих из очереди в единицу времени. Определим среднее число заявок, находящихся в очереди:



Чтобы найти вероятность Pн, нужно среднее число заявок в очереди умножить на среднюю плотность уходов (определим среднее число заявок, покидающих систему) и умножим на интенсивность входного потока заявок:



Относительная пропускная способность системы: q=1-Pн.

Очевидно, что пропускная система с ожиданиями выше, чем пропускная способность системы с отказами и пропускная способность увеличивается с увеличением среднего времени ожидания mtож=1/ν.

Рассмотрим, во что превратиться система с ожиданиями при изменении параметра β. Очевидно, что при β→∞ система с ожиданиями превращается в чистую систему с отказами, а при β→0 – в чистую систему с ожиданиями. В такой системе вероятность того, что заявка уйдет из системы не обслуженной, равна нулю. Однако, в такой системе не всегда имеется предельный стационарный режим при t→∞. Такой режим существует только при α<n, т.е., когда среднее число заявок, приходящееся на время обслуживания одной заявки не выходит за пределы возможностей n-канальной системы. В противном случае, число заявок в очереди будет неограниченно возрастать.

Полагая, что α<n, найдем предельные вероятности состояния системы (β→0):



Отсюда найдем:

 (0≤k≤n).

 (s≥0).

Среднее число заявок в очереди:



**7.5 Простейшая многофазовая СМО с очередью.**

Анализ многофазовых СМО в общем случае затруднен, тем что входящий поток каждой последующей фазы является выходным потоком предыдущей и в общем случае имеет последействие.Однако *если на вход С МО с неограниченной очередью поступает простейший поток заявок, а время обслуживания показательное, то выходной поток, этой СМО* — *простейший,* с той же интенсивностью *,* что и входящий. Из этого следует, что многофазовую СМО с неограниченной очередью перед каждой фазой, простейшим входящим потоком заявок и показательным временем обслуживания на каждой фазе можно анализировать как простую последовательность простейших СМО.

Если очередь к фазе ограничена, то выходной поток этой фазы перестает быть простейшим и вышеуказанный прием может применяться только в качестве приближенного.