- **Система массового обслуживания (СМО)** – это любая система, предназначенная для обслуживания каких-либо заявок (требований, клиентов), поступающих на нее в случайные моменты времени. Состоит из потока (требований) и одного, либо нескольких потоков обслуживания (каналов, линий обслуживания).
- **Примерами** СМО могут служить:
- расчетно-кассовые узлы в банках, на предприятиях;

  персональные компьютеры (сервера в сети), обслуживающие поступающие заявки или требования на решение тех или иных задач;
- станции технического обслуживания автомобилей; АЗС;
- аудиторские фирмы;
- отделы налоговых инспекций, занимающиеся приёмкой и проверкой текущей отчетности предприятий;
- телефонные станции и т. д.
-

  Открытая СМО – когда входящий поток требований не зависит от состояния системы, поступая в неё извне.

- *Замкнутая СМО* – когда поток требований находится внутри системы, зависит от состояния системы. Пример:
  Компьютерный класс, или сложная вычислительная система в которой узлы периодически требует ремонта, перезагрузки или переустановки программного обеспечения (т.е. создают заявки на обслуживание).

  Различают СМО с отказами (потерями) 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 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 может быть как ограниченным, так и неограниченным. 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)
-

Дисциплина очереди определяет принцип, в соответствии с которым поступающие на вход обслуживающей системы требования подключаются из очереди к процедуре обслуживания. Чаще всего используются дисциплины очереди, определяемые следующими правилами:

- ***Упорядоченный тип*: первым пришел – первый обслуживаешься**; first in first out (**FIFO, FF**) - самый распространенный тип очереди.
- ***С приоритетами***: тот же принцип, но требования в потоке делятся на приоритетные и не приоритетные, и образует две очереди (вторая обслуживается только во время, когда нету первой) **PR** (Priority)
- ***Стековый или магазинный:*** *пришел последним — обслуживаешься первым* (**LIFO, LF**) (обойма для патронов, тупик на железнодорожной станции, поездка на лифте).
- ***Равновероятный*** выбор заявки, **SP** (Same Probability)

Бывают и другие дисциплины о.)

Механизм обслуживания *(Дисциплина обслуживания*) - определяется характеристиками самой процедуры обслуживания и структурой обслуживающей системы, в том числе:

- количество каналов обслуживания ;
- продолжительность процедуры обслуживания (вероятностное распределение времени обслуживания требований)–задаёт **интенсивность обслуживания канала** (т.е. интенсивность потока на канале обслуживания при идеальной, максимальной загруженности – когда канал всегда занят);

одинаковость\различие интенсивностей среди каналов обслуживания и возможность задания предпочтений выбора между ними в последнем случае;

- количество требований, удовлетворяемых в результате процедуры;
- вероятность выхода из строя обслуживающего канала;
- возможность\невозможность участия в обслуживании заявки сразу нескольких каналов;
- все каналы могут обслуживать любые заявки – тогда СМО называют ***полнодоступной*** , либо некоторые обслуживают лишь определённые категории заявок (***неполнодоступные СМО***, в таких моделях встречаются случаи с несколькими потоками заявок)

- Для краткой записи СМО приняты т.н. **символы (обозначения) Кендалла**:

Обычно используют **4** символа: первый характеризует входной поток требований, второй — распределение длительностей обслуживания и третий — число приборов в обслуживающей системе. **а 4й** символ - число мест в очереди (0, ∞ или конечн. число).

Приведем перечень общепринятых символов, характеризующих распределения вероятностей, которые ставятся в соответствие моделям массового обслуживания:

M — экспоненциальное распределение продолжительностей интервалов между поступлениями требований или длительностей обслуживания (от определяющего слова «марковский»);

D — детерминированное (или регулярное) распределение длительностей интервалов между поступлениями требований или длительностей обслуживания;

E_n — n-фазное распределение Эрланга

(возникает в потоке Эрланга)

(т.е. пальмовский)

(GU) GI — рекуррентный характер входного потока [1]) без каких-либо специальных предположений относительно функции распределения;

G — общий вид распределения длительностей обслуживания (т. е. не делается никаких конкретизирующих предположений относительно функции распределения).

- Все символы записываются подряд через черту. Например, простейшая СМО с одним каналом и отказами (потерями, нет очереди) :
  M / M / 1 / 0
  Остальные символы (начиная с пятого – не обязательны, характеризуют специфические, «неклассические», особенности устройства СМО).
- Например, далее могут добавить (уже не обязательно) ещё символ, обозначающий ограничение на источники нагрузки (т.е. если в СМО поступает конечное число требований вместо потока требований):
- M / M / 1 / 0 / K (K – целое положит. число)

- И после этого ещё символы для характеризации дисциплины очереди и особого устройства механизма обслуживания (если отличны от FF| FM), например:

  D / M / 1 / 0 / LF / FM

  Также в последнем разряде можно было бы увидеть:

  FM (Full Matrix)– полнодоступная система,
  G – неполнодоступная система (Grading)
  и другое (например, разновидности неполнодоступной системы)

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

среднее число заявок A, обслуживаемое СМО в единицу времени, или *абсолютная пропускная способность* СМО;

вероятность обслуживания поступившей заявки Q или *относительная пропускная способность* СМО; $Q = A / \lambda$;

вероятность отказа $P_{\text{отк}}$, т.е. вероятность того, что поступившая заявка не будет обслужена, получит отказ; $P_{\text{отк}} = 1 - Q$;

среднее число заявок в СМО (обслуживаемых или ожидающих в очереди) $\bar{z}$;

среднее число заявок в очереди $\bar{r}$; (также распространено обозн. $L_{\text{оч.}}$)

среднее время пребывания заявки в СМО (в очереди или под обслуживанием) $\bar{t}_{\text{сист}}$;

среднее время пребыван заявки в очереди $\bar{t}_{\text{оч}}$; среднее число занят каналов $\bar{k}$.

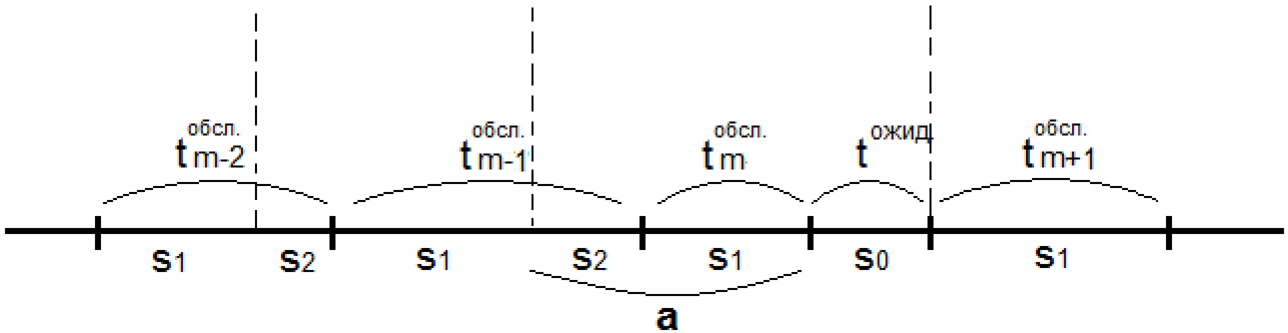
В общем случае все эти характеристики зависят от времени. Но многие СМО работают в неизменных условиях достаточно долгое время, и поэтому для них успевает установитьс режим, близкий к стационарному. (Тогда соотв. процесс Маркова эргодичен) повсюду, не оговаривая этого каждый раз специально, будем вычислять финальные вероятности состояний и финальные характеристики эффективности СМО, относящиеся к предельному, стационарному режиму ее работы.

- **ОПР.** СМО, в которой поток требований простейший, а время обслуживания на всех каналах распределено по показательному закону будем называть *простейшей СМО*
(т.е. из времён обслуживания на отдельной линии обслуживания также складывается простейший поток)

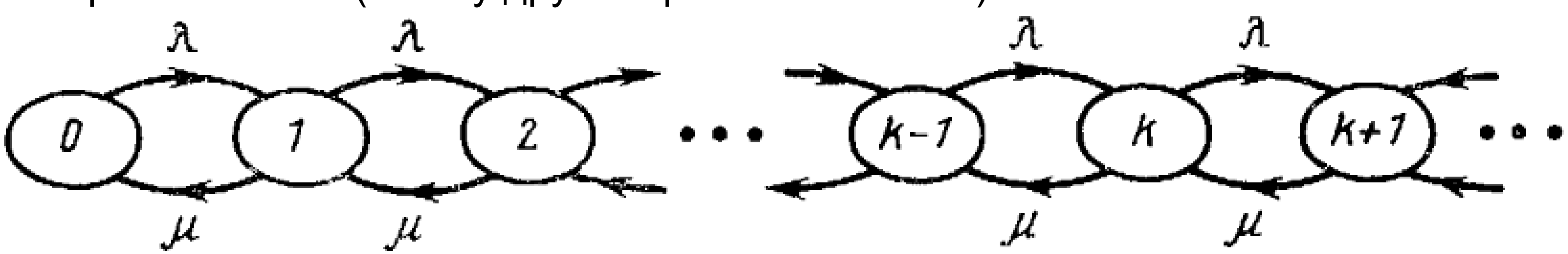
- **ТЕОРЕМА 1.**
Любая простейшая СМО моделируется процессом Маркова.

- **ТЕОРЕМА 2.**
- Любая СМО стандартного механизма обслуживания (полнодоступная, с одинаковыми интенсивностями каналов обслуживания) с Пуассоновским потоком требований и определённым распределением времени обслуживания (временами обслуживания, если выделяют приоритетную категорию заявок и она имеет другое время обслуживания) моделируется процессом Маркова. Причём, в случае стационарности потока требований и экспоненциального распределения времени обслуживания имеем однородный процесс Маркова.

- **Д-ВО** (идея)Теорему можно доказать, начав с рассмотрения схемы работы (поток событий) канала для одноканальной СМО с неограниченной очередью:



- в которой пунктиром обозначены моменты поступления новых заявок и подписаны периоды нахождения системы в состояниях
- S_2 – канал работает, одна заявка в очереди
- S_1 – канал работает, нет заявок в очереди
- S_0 – канал свободен (простаивает)
- Схема моделирующего процесса будет частным случаем схемы гибели и размножения (как и у других простейших СМО):



- Покажем, что все интенсивности переходов именно такие: λ и μ (и что эти величины представляют в самой СМО). Далее можно увидеть, что в других простейших СМО схема составляется по аналогии.