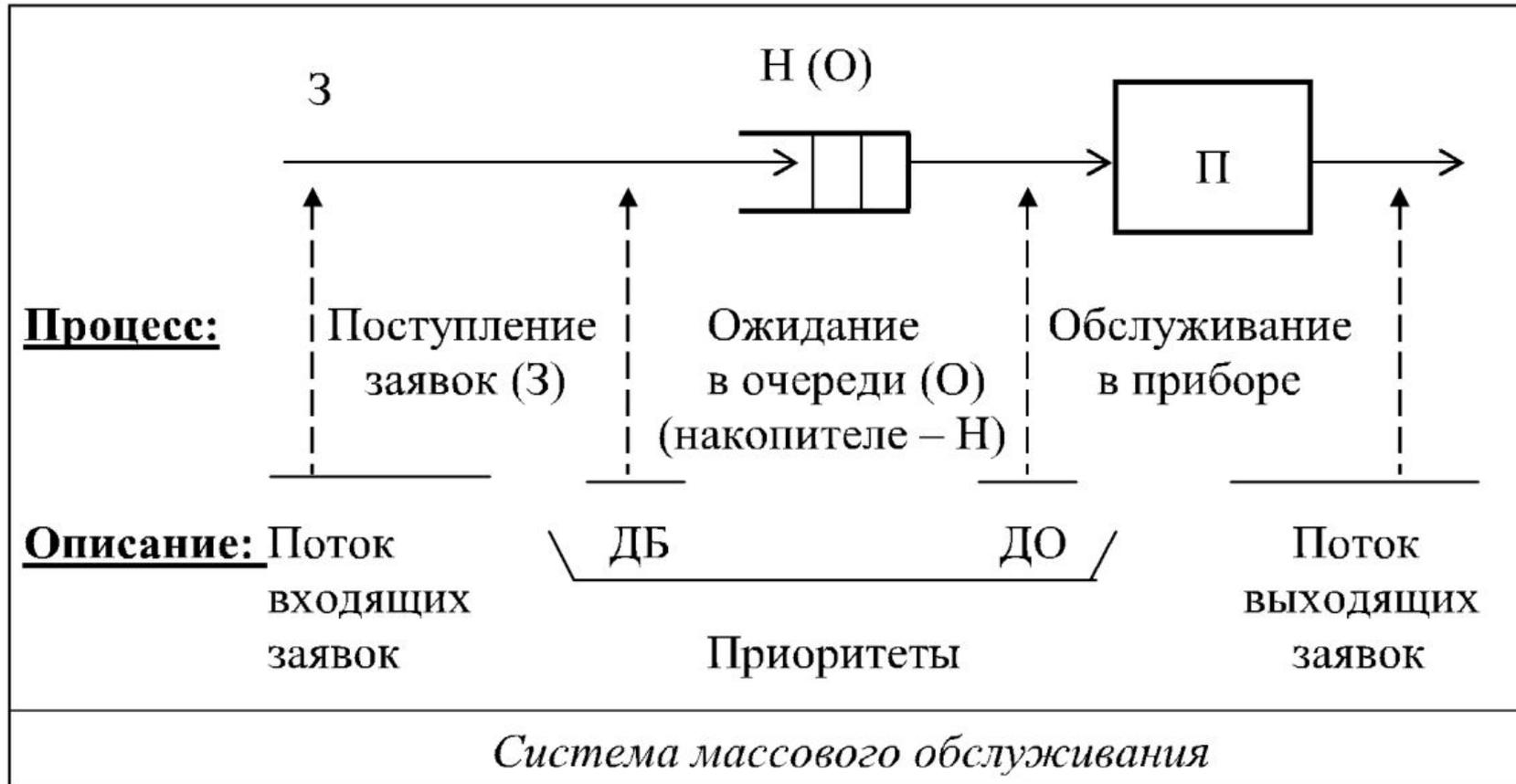


Система массового обслуживания

обслуживания

Система массового обслуживания (СМО) – математический (абстрактный) объект, содержащий один или несколько *приборов П* (каналов), обслуживающих заявки **З**, поступающие в систему, и *накопитель Н*, в котором находятся заявки, образующие очередь **О** и ожидающие обслуживания.



Заявка (требование, запрос, вызов, клиент) – объект, поступающий в СМО и требующий обслуживания в обслуживаемом приборе. Совокупность заявок, распределенных во времени, образуют **поток заявок**.

Обслуживающий прибор или просто **прибор (устройство, канал, линия)** – элемент СМО, функцией которого является обслуживание заявок. В каждый момент времени в приборе на обслуживании может находиться только одна заявка.

Обслуживание – задержка заявки на некоторое время в обслуживаемом приборе.

Длительность обслуживания – время задержки (обслуживания) заявки в приборе.

Накопитель (буфер) – совокупность мест для ожидания заявок перед обслуживающим прибором. Количество мест для ожидания определяет **ёмкость накопителя**. Заявка, поступившая на вход СМО, может находиться в двух состояниях:

- в состоянии **обслуживания** (в приборе);
- в состоянии **ожидания** (в накопителе), если все приборы заняты обслуживанием других заявок.

Заявки, находящиеся в накопителе и ожидающие обслуживания, образуют **очередь** заявок. Количество заявок, ожидающих обслуживания в накопителе, определяет **длину очереди**.

Дисциплина буферизации – правило занесения поступающих заявок в накопитель (буфер).

Дисциплина обслуживания – правило выбора заявок из очереди для обслуживания в приборе.

Приоритет – преимущественное право на занесение (в накопитель) или выбор из очереди (для обслуживания в приборе) заявок одного класса по отношению к заявкам других классов.

Сеть массового обслуживания

Сеть массового обслуживания (СеМО) – совокупность взаимосвязанных СМО, в среде которых циркулируют заявки. Основными элементами СеМО являются узлы (У) и источники заявок (И) (рис.а).

Узел сети представляет собой систему массового обслуживания.

Источник – генератор заявок, поступающих в сеть и требующих определенных этапов обслуживания в узлах сети.

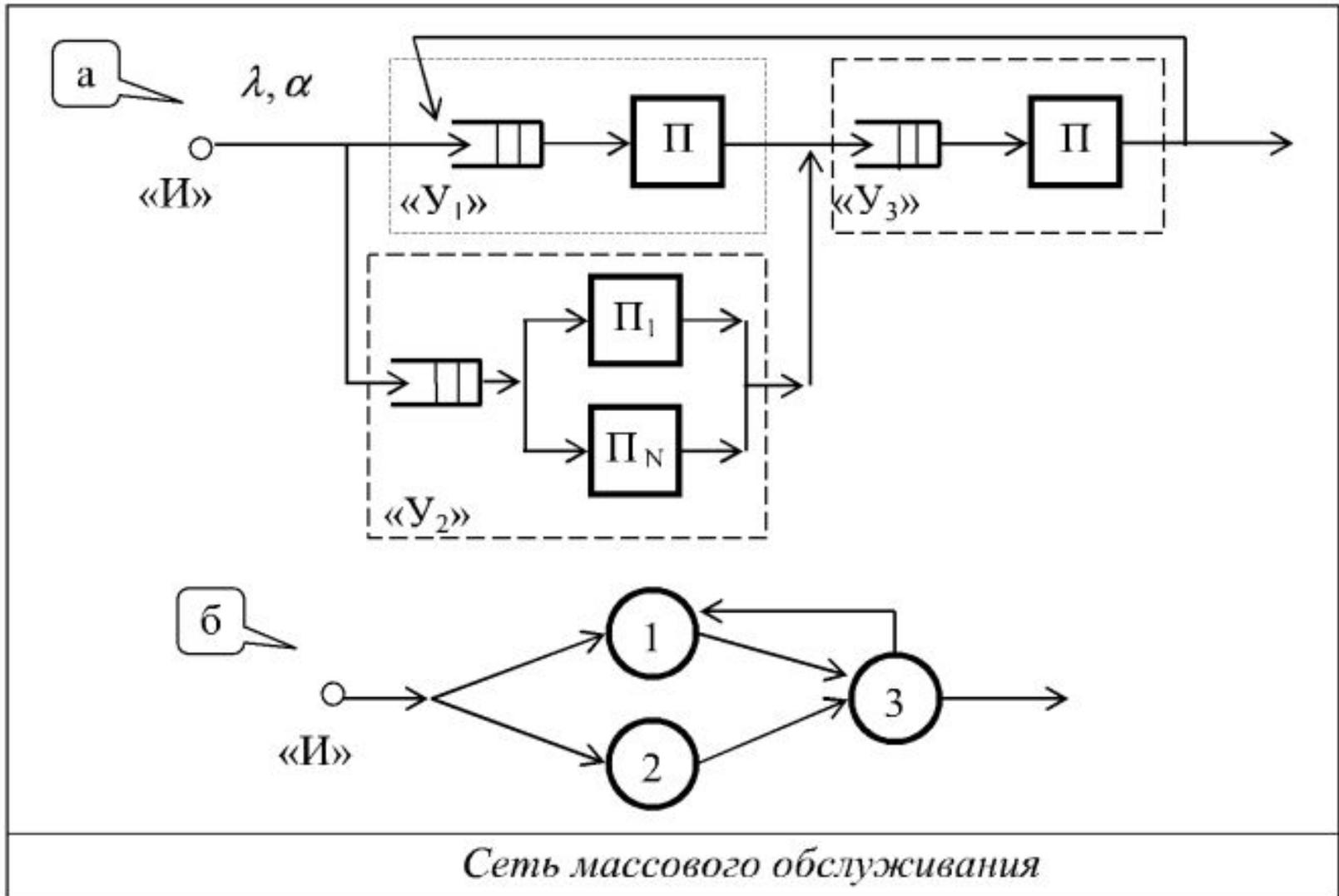
Для упрощенного изображения СеМО используется граф СеМО.

Граф СеМО – ориентированный граф, вершины которого соответствуют узлам СеМО, а дуги отображают переходы заявок между узлами (рис.б).

Переходы заявок между узлами СеМО, в общем случае, могут быть заданы в виде вероятностей передач.

Путь движения заявок в СеМО называется **маршрутом**.

Сеть массового обслуживания



Стратегии управления потоками

заявок

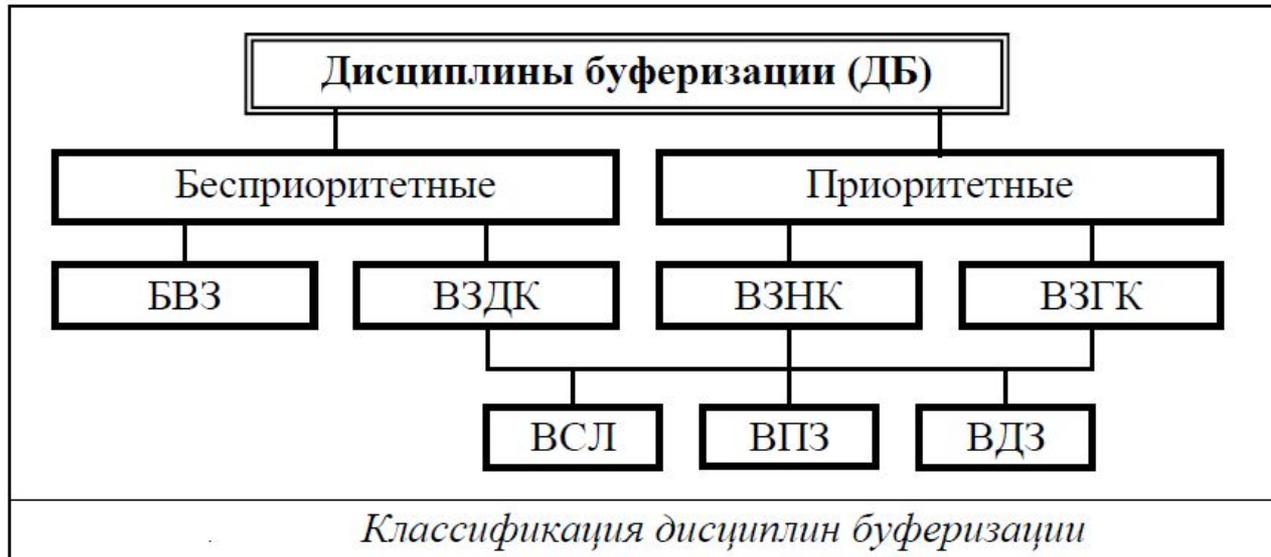
Стратегия управления потоками заявок в моделях массового обслуживания задается в виде:

- дисциплины буферизации (ДБ);
- дисциплины обслуживания (ДО).

ДБ и ДО могут быть классифицированы по следующим признакам:

- наличие приоритетов между заявками разных классов;
- способ (режим) вытеснения заявок из очереди (для ДБ) и назначения заявок на обслуживание (для ДО);
- правило вытеснения или выбора заявок на обслуживание;
- возможность изменения приоритетов.

Одна из возможных **классификаций дисциплин буферизации** в соответствии с перечисленными признаками представлена на рис.:



В зависимости от *наличия или отсутствия приоритетов* между заявками разных классов все ДБ могут быть разбиты на две группы:

- **бесприоритетные;**
- **приоритетные.**

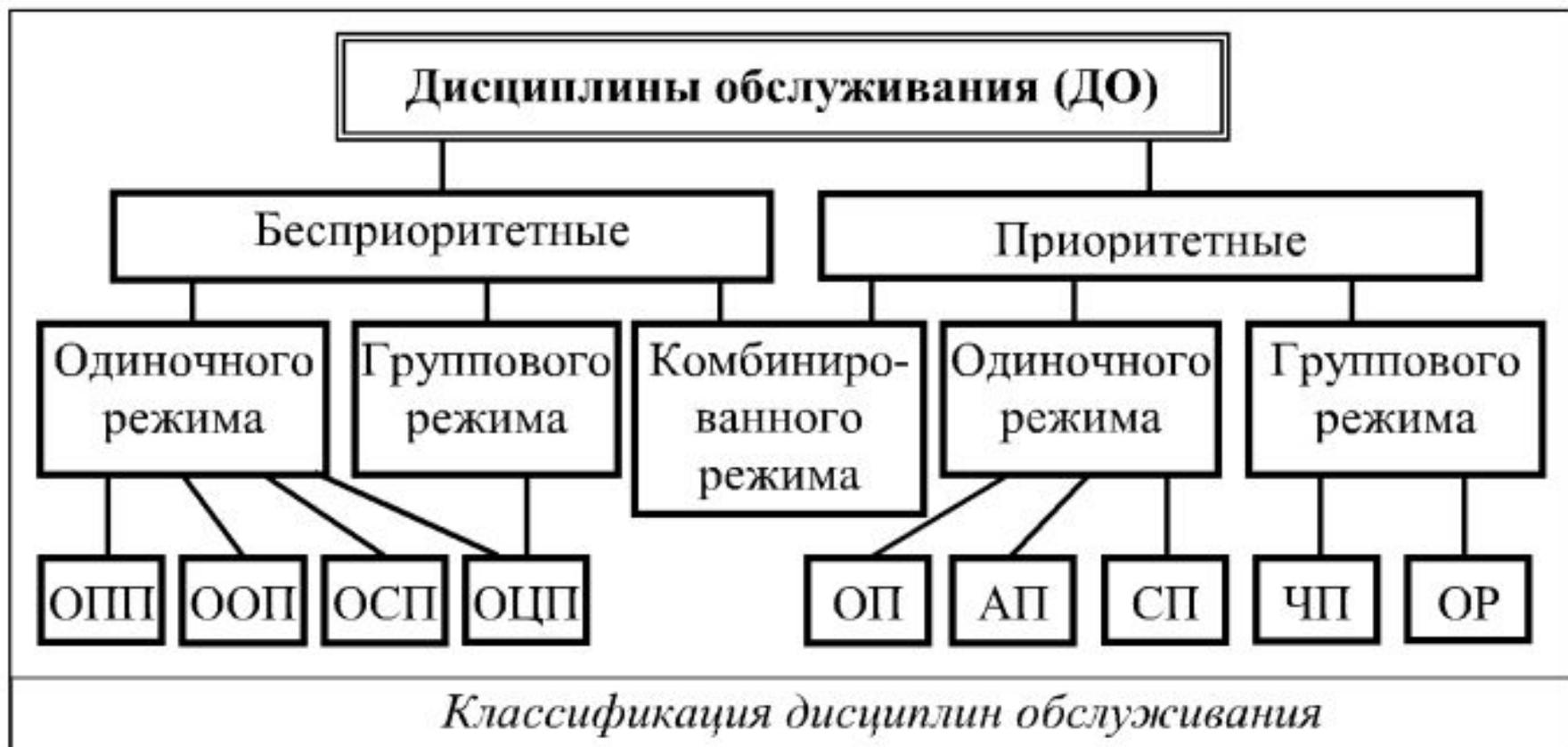
По *способу вытеснения заявок из накопителя* можно выделить следующие классы ДБ:

- **без вытеснения заявок (БВЗ)** – заявки, поступившие в систему и заставшие накопитель заполненным до конца, теряются;
- **с вытеснением заявки данного класса (ВЗДК)**, то есть такого же класса, что и поступившая;
- **с вытеснением заявки самого низкоприоритетного класса (ВЗНК);**
- **с вытеснением заявки, принадлежащей группе низкоприоритетных классов (ВЗГК).**

Два первых класса относятся к **бесприоритетным ДБ**, а остальные – к **приоритетным**.

ДБ могут использовать следующие *правила вытеснения заявок из накопителя*:

- **вытеснение случайное (ВСЛ);**
- **вытеснение последней заявки (ВПЗ)**, то есть поступившей в систему позже всех;
- **вытеснение «долгой» заявки (ВДЗ)**, то есть находящейся в накопителе дольше всех.



В зависимости от *наличия или отсутствия приоритетов* между заявками разных классов все ДО, как и ДБ, могут быть разбиты на две группы:

- **бесприоритетные;**
- **приоритетные.**

По *способу назначения заявок на обслуживание* ДО могут быть разделены на дисциплины:

- **одиночного режима;**
- **группового режима;**
- **комбинированного режима.**

В ДО **одиночного режима** всякий раз на обслуживание *назначается только одна заявка* (просмотр очередей с целью назначения на обслуживание в приборе очередной заявки выполняется после обслуживания каждой заявки).

В ДО **группового режима** всякий раз на обслуживание *назначается группа заявок* одной очереди (просмотр очередей с целью очередного назначения на обслуживание выполняется только после обслуживания всех заявок ранее назначенной группы). В предельном случае назначаемая на обслуживание группа заявок может включать в себя все заявки данной очереди. Заявки назначенной на обслуживание группы *последовательно выбираются из очереди* и обслуживаются прибором, после чего на обслуживание назначается следующая группа заявок другой очереди в соответствии с заданной ДО.

Комбинированный режим – комбинация одиночного и группового режимов, когда часть очередей заявок обрабатывается в одиночном режиме, а другая часть – в групповом.

ДО могут использовать следующие *правила выбора заявок на обслуживание*:

бесприоритетные:

обслуживание в порядке поступления (ОПП или FIFO – First In First Out), когда на обслуживание выбирается заявка, поступившая в систему раньше других;

обслуживание в обратном порядке (ООП или LIFO – Last In First Out) когда на обслуживание выбирается заявка, поступившая в систему позже других;

обслуживание в случайном порядке (ОСП), когда на обслуживание заявка выбирается случайным образом;

обслуживание в циклическом порядке (ОЦП), когда на обслуживание заявки выбираются в процессе циклического опроса накопителей в последовательности $1, 2, \dots, N$ (N – количество накопителей), после чего указанная последовательность повторяется;

приоритетные:

с **относительными приоритетами** (ОП), означающими, что приоритеты учитываются только в моменты завершения обслуживания заявок при выборе новой заявки на обслуживание и не влияют на процесс обслуживания низкоприоритетной заявки в приборе; другими словами, поступление в систему заявки с более высоким приоритетом по сравнению с обслуживаемой в приборе не приводит к прерыванию обслуживаемой заявки;

с **абсолютными приоритетами** (АП), означающими, что, в отличие от ОП, при поступлении высокоприоритетной заявки обслуживание заявки с низким приоритетом прерывается и на обслуживание принимается поступившая высокоприоритетная заявка; при этом прерванная заявка может быть возвращена в накопитель или удалена из системы; если заявка возвращена в накопитель, то её дальнейшее обслуживание может быть продолжено с прерванного места или начато заново, то есть с самого начала;

со **смешанными приоритетами** (СП), представляющими собой любую комбинацию бесприоритетного обслуживания, ОП и АП;

с **чередующимися приоритетами** (ЧП), являющимися аналогом ОП и проявляющимися только в моменты завершения обслуживания группы заявок одной очереди и назначения новой группы;

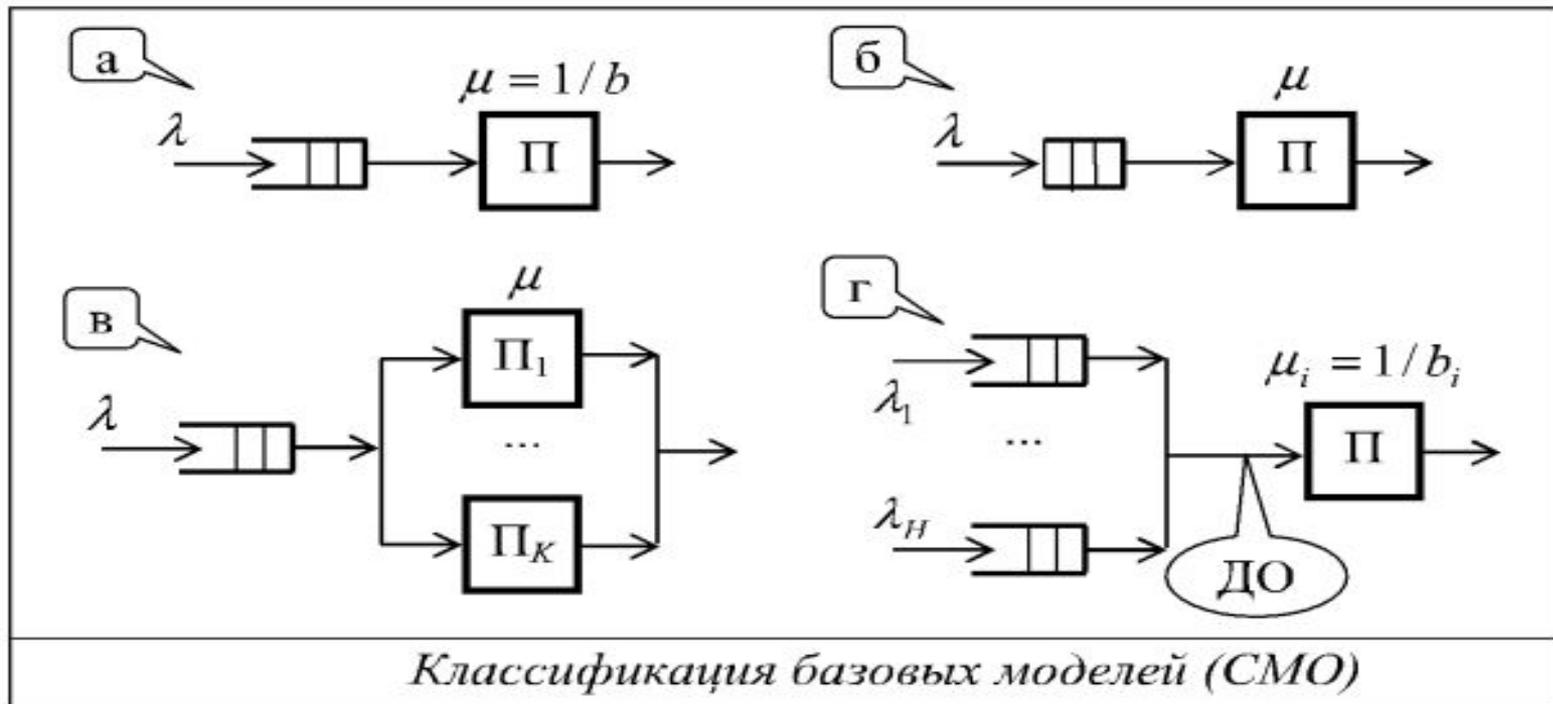
обслуживание по расписанию (ОР), когда заявки разных классов (находящиеся в разных накопителях) выбираются на обслуживание в соответствии с некоторым расписанием (планом), задающим последовательность опроса очередей заявок, например, в случае трех классов заявок (накопителей) расписание может иметь вид: {1, 2, 1, 3, 1, 2}.

Классификация моделей массового обслуживания.

Базовые модели.

При моделировании реальных систем с дискретным характером функционирования широкое применение находят базовые модели в виде СМО, которые могут быть классифицированы (рис.):

- по числу мест в накопителе;
- по числу обслуживающих приборов;
- по количеству классов заявок, поступающих в СМО.



Базовые модели.

1. По числу мест в накопителе СМО делятся на системы:

- **без накопителя**, в которых заявка, поступившая в систему и заставшая все обслуживающие приборы занятыми обслуживанием более высокоприоритетных заявок, получает отказ и теряется; такие системы называются **СМО с отказами**;
- **с накопителем ограниченной ёмкости (СМО с потерями)**, в которых поступившая заявка теряется, если она застаёт накопитель заполненным до конца;
- **системы с накопителем неограниченной ёмкости (СМО без потерь)**, в которых для любой поступившей заявки всегда найдется место в накопителе для ожидания.

2. По количеству обслуживающих приборов СМО делятся на:

- **одноканальные** (рис. а, б, г), содержащие один прибор Π ;
- **многоканальные** (рис.в), содержащие K обслуживающих приборов Π_1, \dots, Π_K ($K > 1$).

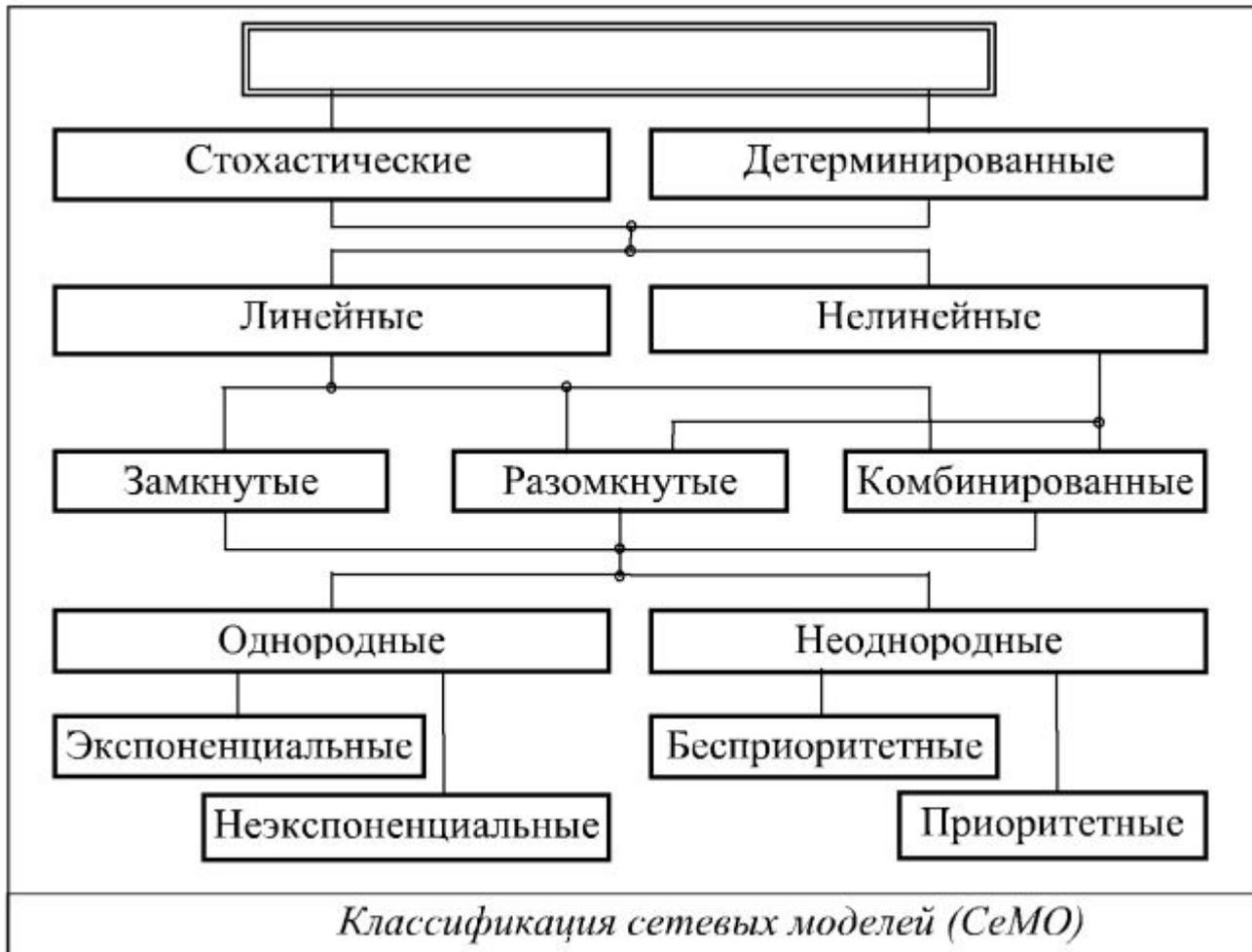
3. По количеству классов (типов) заявок, поступающих в СМО, различают системы:

- **с однородным потоком** заявок (рис. а, б, в);
- **с неоднородным потоком** заявок (рис.г).

В СМО, представляющей собой абстрактную математическую модель, заявки относятся к разным классам в том случае, если они в моделируемой реальной системе различаются хотя бы одним из следующих факторов:

- **длительностью обслуживания**;
- **приоритетами**.

Сетевые модели.



Сетевые модели.

1. В зависимости от *характера процессов поступления и обслуживания заявок* в сети СеМО делятся на:

- **стохастические**, в которых процессы поступления и/или обслуживания заявок носят случайный характер, то есть интервалы времени между поступающими заявками и/или длительности их обслуживания в узлах представляют собой случайные величины, описываемые соответствующими законами распределений;
- **детерминированные**, в которых интервалы времени между поступающими заявками и длительности их обслуживания в узлах являются детерминированными величинами.

2. По *виду зависимостей, связывающих интенсивности потоков заявок в разных узлах*, СеМО делятся на:

- **линейные**, если эти зависимости линейные;
- **нелинейные**, если эти зависимости являются нелинейными.

Сетевые модели.

В *линейных* СеМО, как это следует из определения, интенсивность потока заявок в узел j связана с интенсивностью потока заявок в узел i линейной зависимостью:

$$\lambda_j = \alpha_{ij} \lambda_i,$$

где α_{ij} – коэффициент пропорциональности, показывающий, во сколько раз отличаются интенсивности потоков заявок в узел j и в узел i ($i, j = \overline{1, n}$).

Поскольку указанная зависимость справедлива для любой пары узлов, это выражение можно записать в несколько ином виде и выразить интенсивность поступления заявок во все узлы $j = \overline{1, n}$ через одну и ту же интенсивность, например, через интенсивность λ_0 потока заявок, поступающих в СеМО из источника заявок:

$$\lambda_j = \alpha_j \lambda_0.$$

В последнем выражении коэффициент пропорциональности $\alpha_j \geq 0$ показывает, во сколько раз интенсивность потока заявок в узел j ($i, j = \overline{1, n}$) отличается от интенсивности источника заявок, и называется *коэффициентом передачи*.

Коэффициент передачи можно трактовать как *среднее число попаданий заявки в данный узел за время ее нахождения в сети*.

Сетевые модели.

В *нелинейных* СеМО интенсивности потоков заявок в узлах связаны более сложными нелинейными зависимостями, что значительно усложняет их исследование.

Нелинейность СеМО может быть обусловлена:

- **потерей заявок в сети, например из-за ограниченной емкости накопителей в узлах;**
- **размножением заявок в сети, заключающимся, например, в формировании нескольких новых заявок после завершения обслуживания некоторой заявки в одном из узлов сети.**

Таким образом, СеМО является линейной, если в ней заявки не размножаются и не теряются.

3. По числу циркулирующих в сети заявок различают СеМО:

- **разомкнутые;**
- **замкнутые;**
- **замкнуто-разомкнутые.**

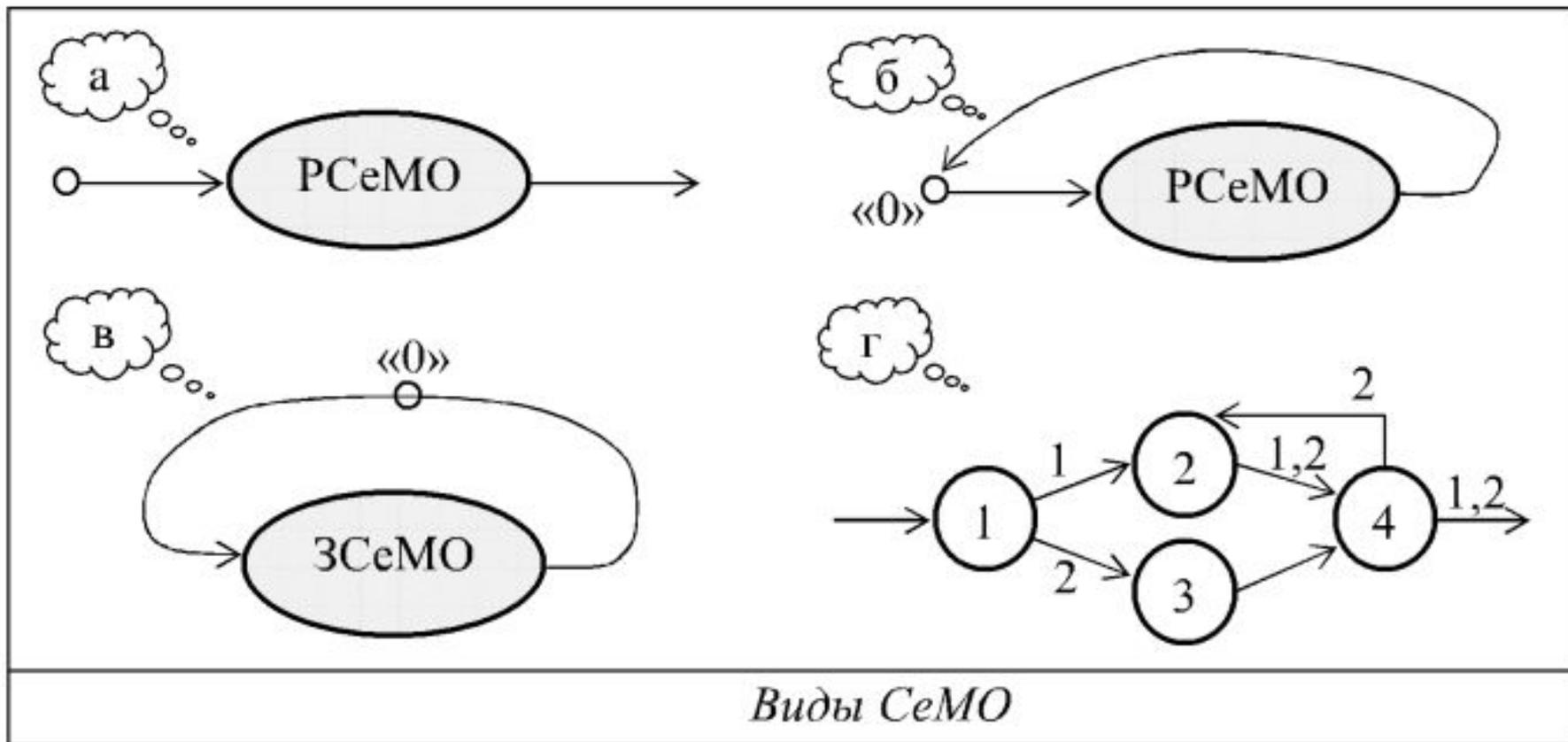
Сетевые модели.

Разомкнутая (открытая) СеМО (РСеМО) содержит один или несколько *внешних независимых источников* заявок, которые генерируют заявки в сеть независимо от числа заявок, находящихся в сети (рис.а). В РСеМО одновременно может находиться *любое число заявок*, в том числе, и сколь угодно большое, то есть от 0 до бесконечности. С РСеМО связана внешняя среда, из которой поступают заявки в сеть и в которую они возвращаются после обслуживания в сети. Внешняя среда в РСеМО обозначается обычно как нулевой узел "0", и РСеМО, в этом случае, изображается в виде рис.б.

Замкнутая (закрытая) СеМО (ЗСеМО) не содержит *независимых внешних источников* заявок и характеризуется тем, что в ней циркулирует *постоянное число заявок M* (рис.в).

Замкнуто-разомкнутая СеМО (комбинированная) представляет собой комбинацию ЗСеМО и РСеМО, в которую, кроме постоянно циркулирующих в сети M^* заявок, из внешнего независимого источника поступают заявки такого же или другого класса, при этом суммарное число заявок в сети M больше либо равно M^* .

Сетевые модели.



Сетевые модели.

4. По *типу циркулирующих заявок* различают СеМО:

- **однородные**, в которых циркулирует один класс заявок (однородный поток заявок);
- **неоднородные**, в которых циркулирует несколько классов заявок (неоднородный поток заявок), различающихся хотя бы одним из следующих факторов:
 - *длительностями обслуживания в узлах;*
 - *приоритетами;*
 - *маршрутами.*

Маршруты заявок разных классов задаются путем указания номеров классов заявок на соответствующих дугах сети (рис. Г).