Основы теории вычислительных систем

ПРЕДМЕТ И ЗАДАЧИ

Теория вычислительных систем – инженерная дисциплина, объединяющая методы решения задач проектирования и эксплуатации ЭВМ, вычислительных комплексов, систем и сетей.

Предметом исследования в теории вычислительных систем являются вычислительные системы в аспектах их производительности, надежности и стоимости. В системе выделяются следующие составляющие:

1) технические средства, определяемые конфигурацией системы – составом устройств и структурой связей между ними;
2) режим обработки, определяющий порядок функционирования системы;
3) рабочая нагрузка, характеризующая класс обрабатываемых задач и порядок их поступления в систему.

ПРЕДМЕТ И ЗАДАЧИ

Задачи анализа.

Анализ вычислительных систем – определение свойств, присущих системе или классу систем. Типичная задача анализа – оценка производительности и надежности систем с заданной конфигурацией, режимом функционирования и рабочей нагрузкой. Другие примеры задач: определение длительности занятости (загрузки) процессора, загрузки канала ввода – вывода, определение (оценка) вероятности конфликта при доступе к общей шине.

ПРЕДМЕТ И ЗАДАЧИ

Задачи синтеза.

Синтез – процесс создания вычислительной системы, наилучшим образом соответствующей своему назначению.

Исходными в задаче синтеза являются следующие сведения, характеризующие назначение системы:

1) ***функция системы*** **(класс решаемых задач);**

2) ***ограничения на характеристики системы*****, например на производительность, время ответа, надежность и др.;**

3) ***критерий эффективности*****, устанавливающий способ оценки качества системы в целом. Необходимо выбрать конфигурацию системы и режим обработки данных, удовлетворяющие заданным ограничениям и оптимальные по критерию эффективности.**

ПРЕДМЕТ И ЗАДАЧИ

Задачи идентификации.

При эксплуатации вычислительных систем возникает необходимость в повышении их эффективности путем подбора конфигурации и режима функционирования, соответствующих классу решаемых задач и требованиям к качеству обслуживания пользователей. В связи с ростом нагрузки на систему и переходом на новую технологию обработки данных может потребоваться изменение конфигурации системы, использование более совершенных операционных систем и реализуемых ими режимов обработки. В этих случаях следует оценить возможный эффект, для чего необходимы модели производительности и надежности системы.

МОДЕЛИ И МЕТОДЫ

Модель – физическая или абстрактная система, адекватно представляющая объект исследования.

В теории вычислительных систем используются преимущественно абстрактные модели – описания объекта исследования на некотором языке.

Абстрактность модели проявляется в том, что компонентами модели являются не физические элементы, а понятия, в качестве которых наиболее широко используются математические. Абстрактная модель, представленная на языке математических отношений, называется математической моделью.

МОДЕЛИ И МЕТОДЫ

При оценке производительности первостепенное значение имеет продолжительность вычислительных процессов. При оценке надежности исследуется продолжительность пребывания системы в различных состояниях, которые меняются из-за отказов оборудования и последующего восстановления работоспособности.

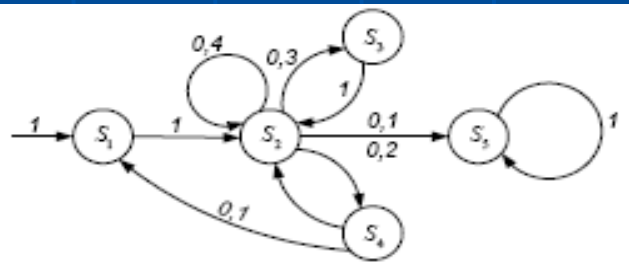
Вероятностный подход к описанию функционирования вычислительных систем приводит к использованию аппарата теории вероятностей и математической статистики в качестве математической базы методов исследования.

Марковские модели

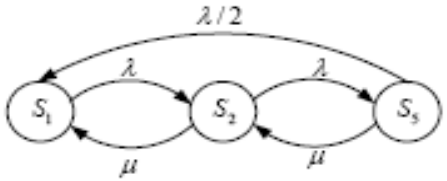
Марковским называется случайный процесс, состояние которого в очередной момент времени $t+\delta$ зависит только от текущего состояния в момент времени t. Это означает, что поведение марковского процесса в будущем определяется текущим состоянием процесса и не зависит от предыстории процесса – состояний, в которых пребывал процесс до момента t.

Марковские модели

Марковская цепь изображается в виде графа, вершины которого соответствуют состояниям цепи и дуги – переходам между состояниями. Дуги (i, j), связывающие вершины si и sj, отличаются вероятностями переходов Pij. На рисунке представлен граф марковской цепи с множеством состояний S={S1,...,S5} , матрицей вероятностей переходов и вектором начальных вероятностей П0= {1,0,0,0,0}



Граф марковской цепи



Граф непрерывной марковской цепи

$$P = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} \begin{array}{c} \begin{array}{ccccc} s_1 & s_2 & s_3 & s_4 & s_5 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0,4 & 0,3 & 0,2 & 0,1 \\ 0 & 1 & 0 & 0 & 0 \\ 0,1 & 0,9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Марковские модели

Марковские цепи классифицируются в зависимости от возможности перехода из одних состояний в другие.

Основными являются два класса:

- поглощающие и
- эргодические цепи.

Поглощающая марковская цепь - содержит поглощающее состояние, достигнув которого, процесс уже никогда его не покидает, т. е. по сути прекращается.

Марковские модели

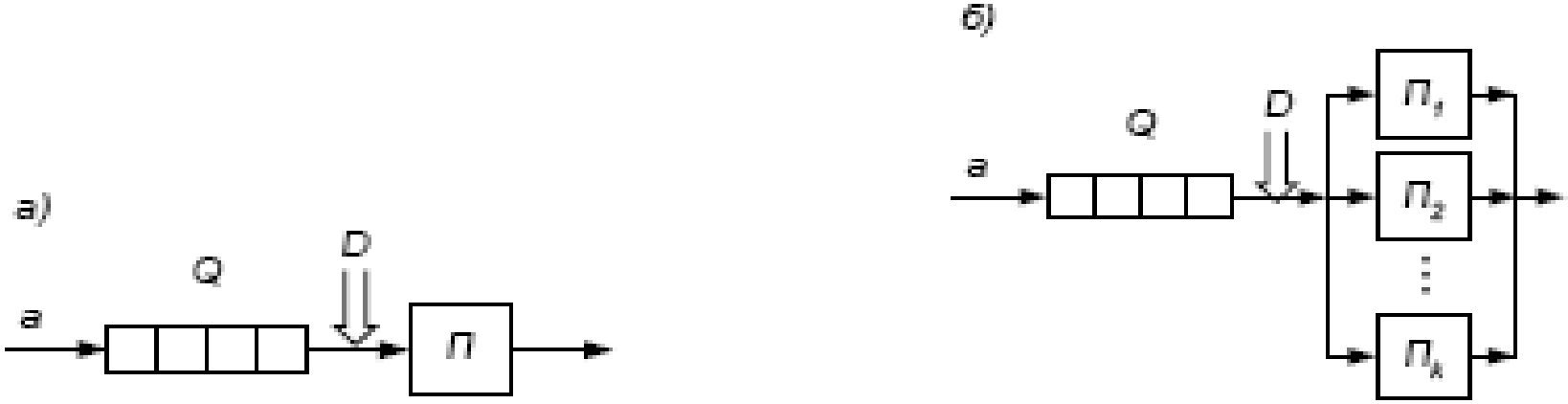
Эргодическая марковская цепь **представляет собой множество состояний, связанных матрицей переходных вероятностей таким образом, что из какого бы состояния процесс ни исходил, после некоторого числа шагов он может оказаться в любом состоянии. Это означает, что в любое состояние эргодической цепи можно перейти из любого другого состояния за сколько-то шагов. По этой причине состояния эргодической цепи называются эргодическими (возвратными).**

Эргодические цепи **широко используются в качестве моделей надежности систем. При этом состояния системы, различающиеся составом исправного и отказавшего оборудования, трактуются как состояния эргодической цепи, переходы между которыми связаны с отказами и восстановлением устройств и реконфигурацией связей между ними, проводимой для сохранения работоспособности системы. Оценки характеристик эргодической цепи дают представление о надежности поведения системы в целом.**

Модели массового обслуживания

Система массового обслуживания **состоит из входящего потока заявок *a*, очереди Q, дисциплины обслуживания D, определяющей порядок выбора заявок из очереди, и обслуживающего прибора П или k одинаковых обслуживающих приборов (каналов) П1…Пк.**

Система, содержащая только один прибор (канал), называется одноканальной, а несколько приборов, – многоканальной.



Одноканальная *(а)* и многоканальная *(б)* система массового обслуживания

Модели массового обслуживания

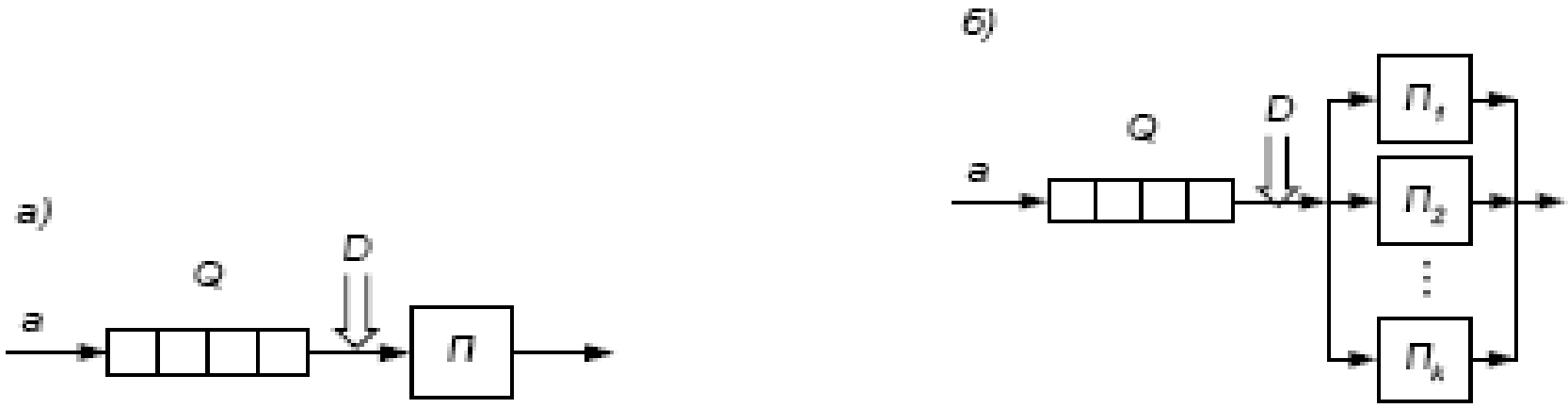
Таким образом, система массового обслуживания характеризуется следующим набором параметров:

1) распределением длительности интервалов между заявками входящего потока p(a);

2) дисциплиной обслуживания заявок D;

3) числом обслуживающих приборов (каналов) K;

4) распределением длительности обслуживания заявок приборами (каналами) p(b).



Одноканальная *(а)* и многоканальная *(б)* система массового обслуживания

Модели массового обслуживания

Процесс функционирования количественно оценивается следующим набором основных характеристик:

1) длиной очереди – числом заявок, ожидающих обслуживания;

2) числом заявок, находящихся в системе (в очереди и на обслуживании приборами);

3) временем ожидания заявки – от момента поступления заявки в систему до начала обслуживания;

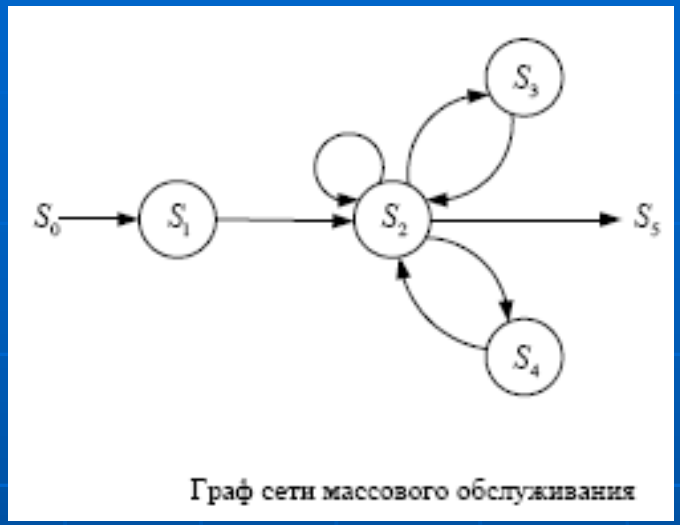
4) временем пребывания заявки в системе – от момента поступления заявки до окончания ее обслуживания, т. е. до выхода из системы.

***) загрузкой системы – средним по времени числом приборов (каналов), занятых обслуживанием (для одноканальной системы загрузка определяет долю времени, в течение которой прибор занят обслуживанием, т. е. не простаивает);**

СМО и стохастические сети

Наряду с основными характеристиками для оценки функционирования системы используются дополнительные характеристики:

- длительность простоя,
- непрерывной занятости приборов
- и др.



Граф сети массового обслуживания

В теории массового обслуживания изучаются и более сложные объекты – сети.

Сеть массового обслуживания – совокупность взаимосвязанных систем массового обслуживания.

Здесь S0 – узел – источник заявок, S1,..S4 – системы массового обслуживания, и S5– узел, представляющий выход из сети. Дуги показывают направления движения заявок по сети.

Стохастические сети

Сетевые характеристики оценивают функционирование сети в целом и включают в себя:

1) число заявок, ожидающих обслуживания в сети;

2) число заявок, находящихся в сети (в состоянии, ожидания и обслуживания);

3) суммарное время ожидания заявки в сети;

4) суммарное время пребывания заявки в сети;

***) загрузку сети – среднее по времени число заявок, обслуживаемых сетью, и одновременно среднее число приборов (каналов), занятых обслуживанием.**

Статистические модели

В тех случаях, когда причинно-следственные отношения в исследуемом объекте трудно охарактеризовать из-за их многообразия, сложности и невыясненной природы процессов или когда эти отношения несущественны, а желательно представить свойства объекта в достаточно компактной форме, используются статистические методы для математического выражения зависимостей между характеристиками и параметрами объекта.

Статистические методы – совокупность способов сбора, анализа и интерпретации данных о некотором объекте или совокупности объектов с целью получения теоретических или практических выводов.

Аналитические методы

Аналитические методы исследования вычислительных систем сводятся к построению математических моделей, которые представляют физические свойства как математические объекты и отношения между ними, выражаемые посредством математических операций.

При построении аналитических моделей свойства объектов описываются исходя из свойств составляющих – физических элементов или элементарных процессов. Для этого используется подходящий математический аналог и с помощью соответствующего математического аппарата строятся выражения, которые связывают показатели, характеризующие элементы.

Имитационные методы

Имитационные методы основаны на представлении порядка функционирования системы в виде алгоритма, который называется имитационной (алгоритмической) моделью.

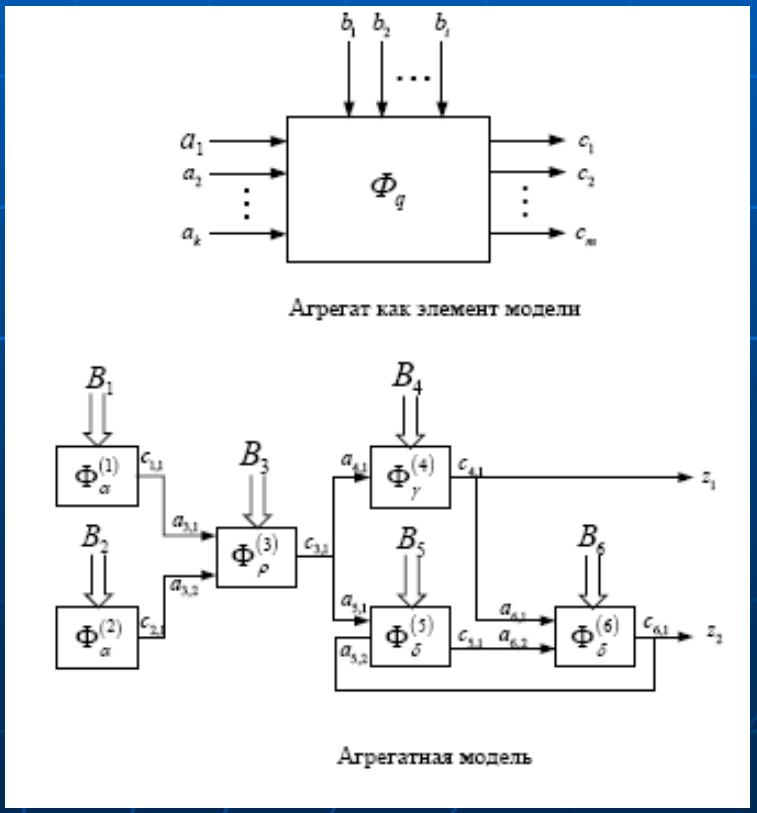
Программа содержит процедуры, регистрирующие состояния имитационной модели и обрабатывающие зарегистрированные данные для оценки требуемых характеристик процессов и моделируемой системы.

Имитационные методы

При построении имитационных моделей широко используется агрегатный подход. Для моделирования заданного класса систем создается набор агрегатов – Ф0,…,Фq - элементов модели. Агрегаты могут соответствовать элементам систем, например процессорам, оперативным запоминающим устройствам, каналам ввода–вывода, каналам передачи данных и другим, воспроизводя определенные аспекты их функционирования.

В качестве агрегатов могут выступать математические объекты, с помощью которых генерируются и преобразуются необходимые процессы.

Так, для моделирования систем на основе сетей массового обслуживания в качестве агрегатов представляются источники потоков заявок, системы массового обслуживания, узлы, управляющие распределением заявок по нескольким направлениям, и т.д.



Экспериментальные методы

Экспериментальные методы **основываются на получении данных о функционировании вычислительных систем в реальных или специально созданных условиях с целью оценки качества функционирования и выявления зависимостей, характеризующих свойства систем и их составляющих. Типичные задачи, решаемые экспериментальными методами,– оценка производительности и надежности системы, определение состава и количественных показателей системной нагрузки в зависимости от прикладной нагрузки и т. д.**

Экспериментальные исследования **выполняются в следующем порядке:**

1. Формулируется цель исследования.

2. Выбирается или разрабатывается методика исследования, которая устанавливает модель исследуемого объекта; способ и средства измерения; способ и средства обработки измерительных данных, а также интерпретация результатов измерений и обработок.

3. Проводятся измерения процесса функционирования объекта в реальных или специально создаваемых условиях.

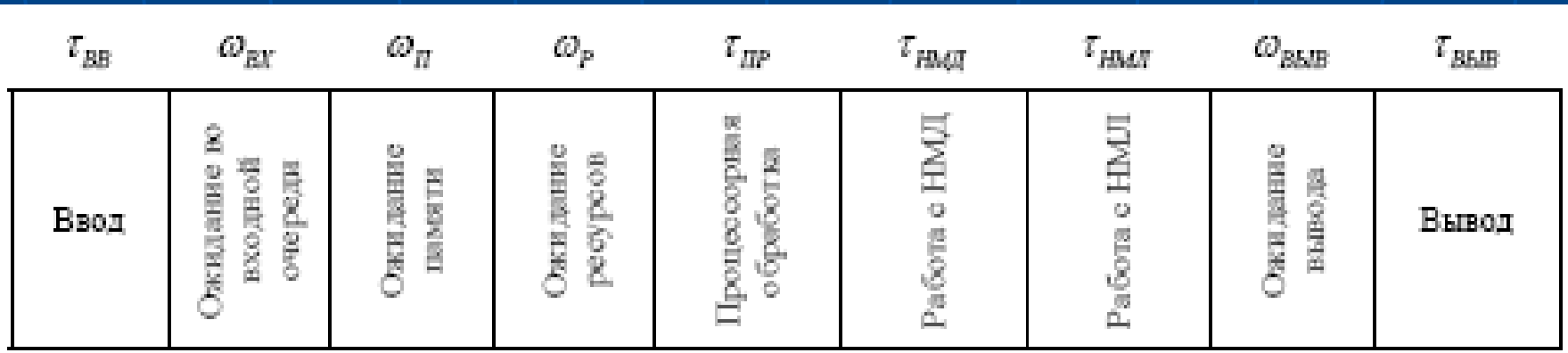
4. Измерительные данные обрабатываются и соответствующим образом интерпретируются.

ПРИНЦИПЫ АНАЛИЗА ПРОИЗВОДИТЕЛЬНОСТИ

Производительность вычислительной системы связана с продолжительностью процессов обработки задач, которая зависит от трех факторов: 1) рабочей нагрузки; 2) конфигурации системы; 3) режима обработки задач.

Применительно к задачам анализа производительности функционирование вычислительной системы рассматривается как $R=\{R1,\ldots,R_{N+P}\}$ - совокупность процессов, связанных с использованием ресурсов системы.

Реализация процесса представляется в виде последовательности фаз, продолжительность пребывания в которых характеризуется значениями *ti* (время использования устройств) и *tw* (время ожидания). Сумма этих значений составляет время пребывания задания в системе.



Профиль вычислительного процесса

Способы описания загрузки ресурсов

Производительность системы непосредственно связана с загрузкой устройств. Загрузка устройства – время, в течение которого устройство занято работой, т. е. не простаивает.

Если t1,...,tk – длительность рабочих интервалов и T – время работы системы, то загрузка устройств на отрезке времени T:

$$\rho = \frac{1}{T}\sum_{k=1}^{K} \tau_k \leq 1$$

Сумма P - называется загрузкой ВС.

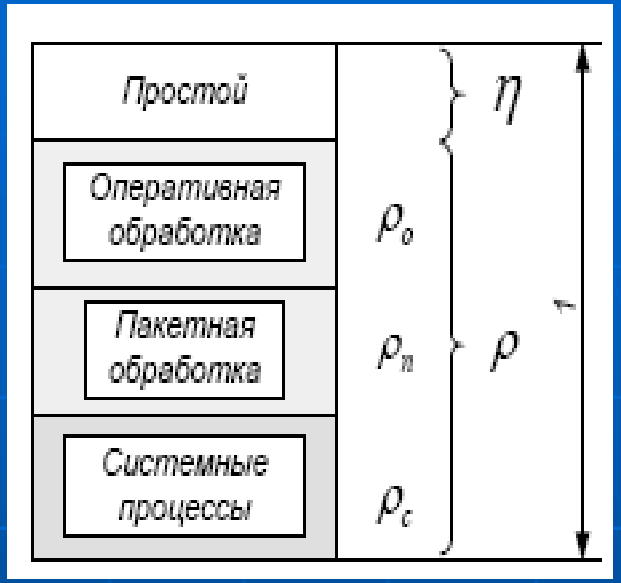
$$P = \sum_{n=1}^{N} \rho_n$$

Способы описания загрузки ресурсов

При анализе производительности большую роль играет не только значение, но и структура загрузки – составляющие, из которых складывается значение р.

Типичная структура представлена на рисунке. В данном случае выделено три класса процессов (видов нагрузки):

- системные процессы,
- пакетная и
- оперативная обработка.



Структура загрузки устройства

Если система функционирует в однопрограммном режиме, причем не простаивает из-за отсутствия нагрузки, P = 1. Если P>1, то производительность системы в этом режиме в P раз выше, чем в однопрограммном.

Таким образом, загрузка системы характеризует производительность системы по отношению к производительности однопрограммного режима.

Модели производительности

Имитационные модели производительности систем общего назначения состоят из трех основных блоков: рабочей нагрузки, планирования работ и выполнения задач. Модель рабочей нагрузки создает потоки заданий, формируемых пользователями на входе системы, и определяет параметры заданий.

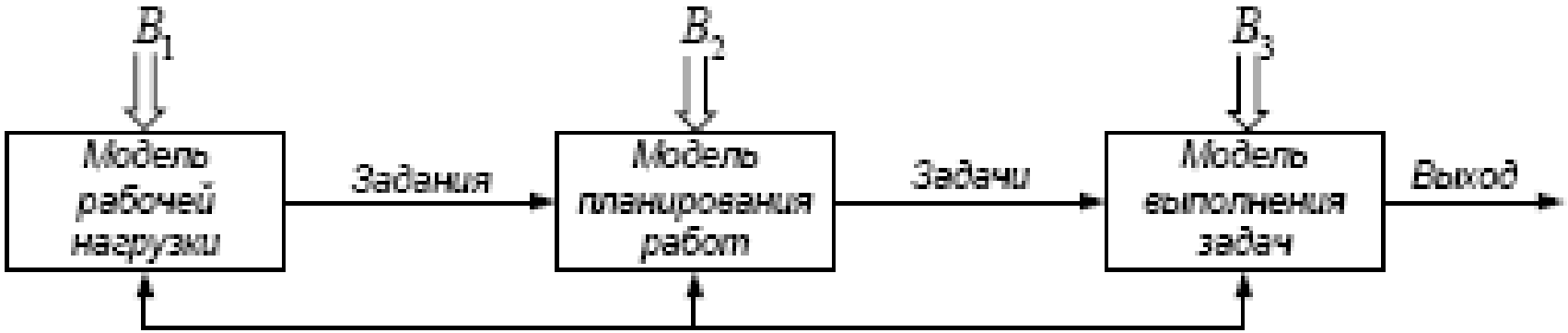
Модель настраивается на конкретный тип нагрузки набором параметров B1.

Модель планирования работ воспроизводит обеспечение заданий ресурсами. Модель настраивается на конкретный режим обработки набором параметров B2 (число разделов или инициаторов, распределение классов задач между инициаторами и т. д.).

Модели производительности

Задания, обеспеченные на фазе планирования ресурсами, образуют задачи, обработка которых воспроизводится моделью выполнения задач. Набор параметров Bi характеризует структуру системы и быстродействие устройств, влияющие на продолжительность выполнения задач.

Состояние процессов в общем случае влияет на состояние процессов планирования и порядок поступления задач в систему: данные о состоянии последующих фаз обработки передаются в предыдущие фазы.



Состав модели производительности

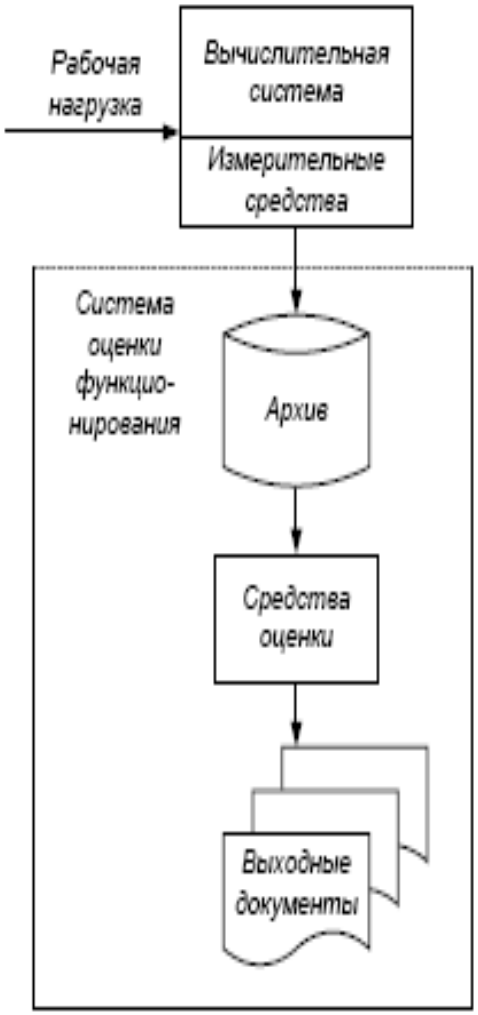
МЕТОДЫ И СРЕДСТВА ИЗМЕРЕНИЙ И ОЦЕНКИ ФУНКЦИОНИРОВАНИЯ

Измерения могут быть направлены на исследование как системы в целом, так и отдельных подсистем.

Объектом измерений является вычислительная система, функционирующая, как правило, в рабочем режиме.

К системе подключаются измерительные средства – мониторы, реагирующие на изменение состояний системы и измеряющие параметры состояний (моменты изменения состояний, продолжительность пребывания в них и др.).

Измерительные данные поступают от мониторов в архив на протяжении заданного промежутка времени, накапливаются и затем обрабатываются.

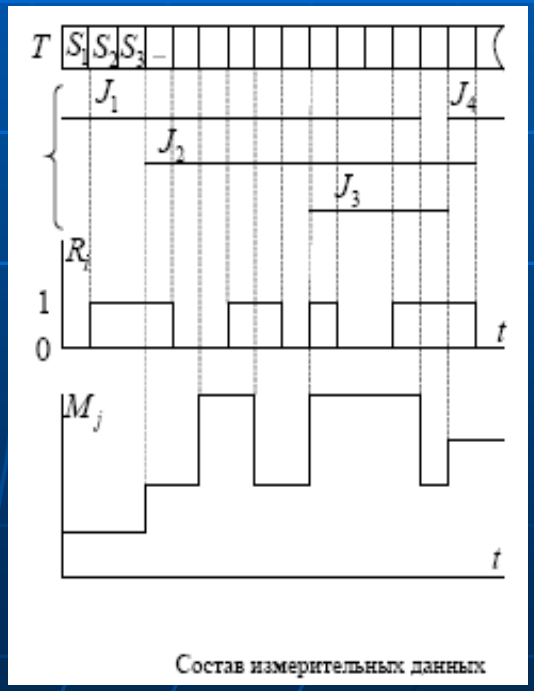


Организация измерений и оценки функционирования

МЕТОДЫ И СРЕДСТВА ИЗМЕРЕНИЙ И ОЦЕНКИ ФУНКЦИОНИРОВАНИЯ

Трассировочный метод измерений основан на регистрации событий, соответствующих моментам изменения состояний вычислительной системы. К таким событиям, в частности, относятся начало и конец ввода задания, шага задания, этапа процессорной обработки, обращения к внешней памяти и т. д. События регистрируются монитором в виде событийного набора данных T, состоящего из последовательности записей s1, s2, …, соответствующих последовательности событий.

Выборочный метод измерений основан на регистрации состояний вычислительной системы в заданные моменты времени, как правило, через промежутки длительностью δ. В моменты t=nδ n=0,1,2…, выборочный монитор регистрирует состояние системы, фиксируя в соответствующих записях данные из управляющих таблиц, или значения электрических сигналов, характеризующих состояния устройств системы.



Состав измерительных данных

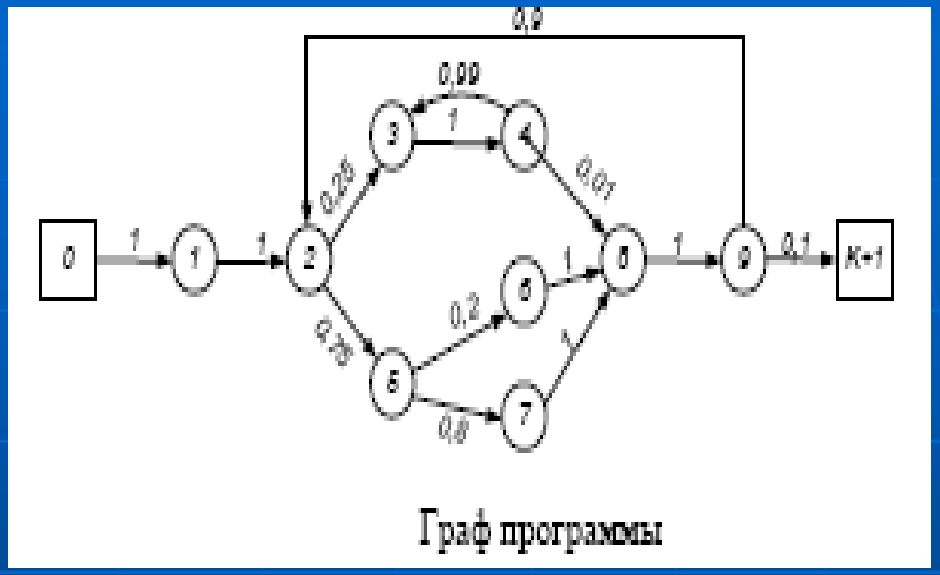
МЕТОДЫ И СРЕДСТВА ИЗМЕРЕНИЙ И ОЦЕНКИ ФУНКЦИОНИРОВАНИЯ

Трассировочные мониторы измеряют отдельные процессы, например обработку одного задания, более точно, чем выборочные. Однако, если функционирование системы оценивается статистическими методами, выборочный монитор обеспечивает такую же точность, как и трассировочный, правда при большей продолжительности измерений.

Основное достоинство выборочных мониторов – возможность измерений сколь угодно быстрых процессов при ограниченном быстродействии.

МОДЕЛИ РАБОЧЕЙ И СИСТЕМНОЙ НАГРУЗКИ

Для определения нагрузки, создаваемой программой в отношении устройств системы, используется марковская модель программы. Как правило, модель представляет собой граф, в вершинах которого, соответствующих операторам программы, отмечены объемы ресурсов, используемых при выполнении оператора, а на дугах – вероятностных переходов к следующим операторам.



Граф программы

Рабочая нагрузка вычислительных систем общего назначения оценивается на основе измерений процесса функционирования. Для оценки рабочей нагрузки выбирается учетная единица работ:

при пакетной обработке – задание, а

при оперативной – взаимодействие пользователя с системой, называемое транзакцией.

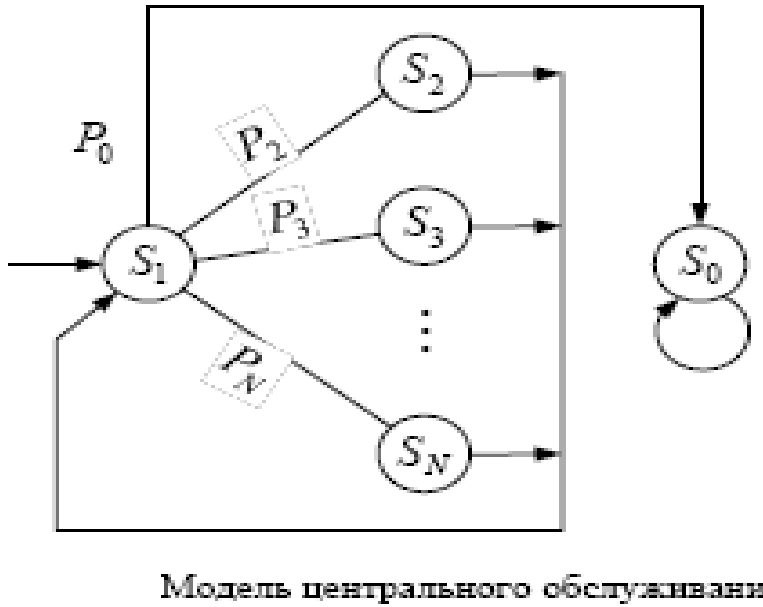
МОДЕЛИ РАБОЧЕЙ И СИСТЕМНОЙ НАГРУЗКИ

В модели центрального обслуживания процесс выполнения программы представляется поглощающей марковской цепью с множеством состояний $s0,s1,\ldots,s_N$, где

$s0$ – поглощающее состояние, а

$s1,\ldots,s_N$ - невозвратные состояния, соответствующие этапам выполнения процесса на устройствах $R1,..,R_N$ (процессор и периферийные),

причем состояние $s1$ отождествляется с этапом процессорной обработки.



Модель центрального обслуживания

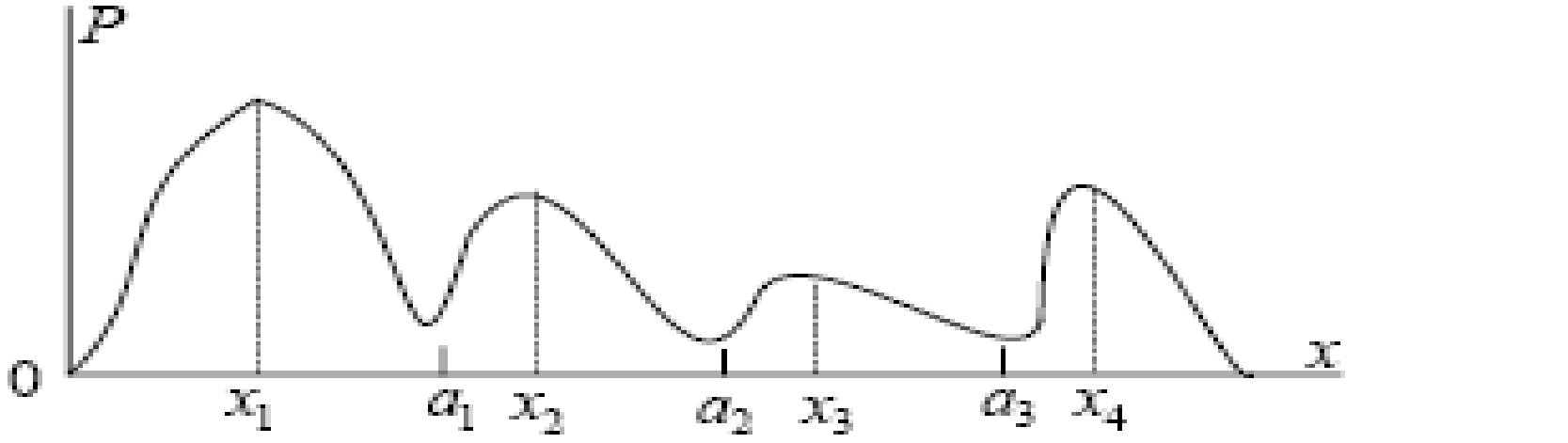
Однородное и неоднородное представление рабочей нагрузки

Рабочую нагрузку, зафиксированную при измерении процесса функционирования системы в достаточном интервале времени, можно представить среднестатистическим заданием, параметры которого – среднее число обращений $n2,..,n_N$ к периферийным устройствам $R2,..., R_N$ и длительностью процессорной обработки и ввода-вывода $V1,..., V_N$ - определяются как статистические средние на множестве выполненных заданий. Представление рабочей нагрузки заданием одного типа со среднестатистическими параметрами называется однородным.

В подавляющем большинстве случаев рабочая нагрузка состоит из неоднородных заданий, существенно различающихся по объему используемых ресурсов – в десятки и даже сотни раз. Различия в ресурсоемкости учитываются при обработке данных путем разбиения заданий на классы, каждый из которых объединяет задания с примерно одинаковыми свойствами, но существенно отличными от свойств заданий других классов.

Однородное и неоднородное представление рабочей нагрузки

Классификация заданий используется для создания мультипрограммных смесей, позволяющих равномерно загружать ресурсы и за счет этого повышать производительность системы, а также при назначении заданиям приоритетов, с помощью которых обеспечивается необходимое время ответа, например малое время для коротких заданий.



. Распределение параметра заданий для неоднородной нагрузки