

Содержание

Введение	4
1. Распределение и учет использования машинного времени	6
1.1. Распределение машинного времени	7
1.2. Учет расхода машинного времени	8
1.3. Обеспечение прохождения задач в соответствии с распределением машинного времени	9
2. Принципы включения задач в решение	9
2.1. Обеспечение скорейшего прохождения пакета задач	10
2.2. Удовлетворение требований к порядку прохождения задач....	11
3. Принципы обслуживания задач, находящихся в решении	13
4. Средства, позволяющие пользователю влиять на порядок прохождения задач	14
4.1. Ограничение предлагаемых цен решения задач.....	14
4.2. Повышение предлагаемых цен решения задач.....	14
4.3. Изменение скорости решения задачи, находящейся в рабочей смеси.....	15
5. Обеспечение режимов разделения времени и реального времени. 16	
5.1. Обеспечение режима разделения времени	17
5.2. Обеспечение режима реального времени	18
6. Распределение стоимости машинного времени между решаемыми задачами и источниками недогрузки.....	18
Заключение.....	21
Литература	23

Введение

Постоянный рост быстродействия ЭВМ и увеличение количества и разнообразия их ресурсов приводит к значительному усложнению распределения и учета машинного времени и управления прохождением задач. К сожалению, при разработке операционных систем в настоящее время не уделяется должного внимания автоматизации управления прохождением задач. Хотя операционные системы и имеют в своем составе планировщики, подсистемы бюджетирования, сбора и распечатки статистики, средства наблюдения за ходом решения задач, однако, все эти механизмы не связаны в единый комплекс, объединенный общей целью и единым подходом. Недостатки, вызванные отсутствием единого системного подхода, становятся особенно заметны, когда ЭВМ используется для решения большого количества разнообразных задач разной важности. Средства дифференциации задач по важности в операционных системах развиты крайне слабо. Так, известные механизмы приоритетов, как правило, приводят к потерям машинного времени, которые не поддаются учету и планированию и не стимулируют разумное назначение приоритетов задач и оптимальное использование ЭВМ. Кроме того, администрация вычислительного центра не имеет информации, отражающей дефицитность различных ресурсов ЭВМ, и поэтому очень слабо представляет себе, какой эффект будет получен в результате приобретения какого-то нового оборудования (например, дополнительных объемов оперативной или внешней памяти). Пользователи, в свою очередь, слабо заинтересованы в экономии дефицитных ресурсов.

Ниже предлагаются принципы функционирования автоматизированной системы управления прохождением задач на ЭВМ, реализация которой должна способствовать повышению общей производительности ЭВМ благодаря:

- целенаправленному распределению вычислительных мощностей;
- увеличению пропускной способности системы;
- эффективному совмещению различных режимов использования ЭВМ (пакетного, разделения времени, реального времени).

Основной отличительной особенностью системы является введение условно-денежного исчисления для распределения и учета

машинного времени и выражения в нем стоимости использования задач различных ресурсов ЭВМ. В результате многие алгоритмы операционной системы обретают логическую стройность, поскольку все они должны преследовать общую цель - достижение максимальной окупаемости машины.

В работе над проектом системы принимали участие Н.Е.Балакирев, М.Г.Тонконогов, А.Е.Фирсов. Основные положения проекта обсуждались со многими ведущими специалистами в области системного программирования. Авторы выражают особую признательность М.Р.Шура-Бура, Э.З.Любимскому, А.Н.Мямлину, В.С.Штаркману, И.Б.Задыхайло, А.И.Илюшину, В.Я.Карпову, В.П.Иванникову, И.Н.Силину, Ю.И.Еремину, Б.Ф.Синенкину, принимавшим участие в обсуждении первых вариантов проекта и сделавшим ряд ценных замечаний.

1. Распределение и учет использования машинного времени

При организации прохождения задач в вычислительном центре прежде всего требуется решить, каким образом будет осуществляться определение и учет использования машинного времени. Решение этого вопроса зависит от режима использования ЭВМ.

Если ЭВМ используется в пультовом режиме, то достаточно составить расписание работы пользователей на ЭВМ. Если на ЭВМ организована пакетная обработка задач в однопрограммном режиме, то затраты на решение задачи достаточно хорошо выражаются астрономическим временем ее решения. В этом случае распределение и учет целесообразно осуществлять в терминах астрономического времени работы ЭВМ.

Гораздо сложнее решить вопрос распределения и учета в случае мультипрограммного режима использования ЭВМ. Каждая задача требует для своего решения множество разнообразных ресурсов вычислительной системы (процессор, оперативную память, устройства ввода-вывода, различные таблицы операционной системы и т.п.). Если среди этих ресурсов имеется один самый главный, по отношению к которому остальные ресурсы не являются дефицитными, то распределение и учет можно осуществлять в терминах этого самого главного ресурса. На ранних этапах использования ЭВМ таким самым главным ресурсом был, конечно, процессор. По этой причине и в настоящее время распределение и учет машинного времени при мультипрограммном использовании ЭВМ ведутся, как правило, в терминах процессорного времени. Однако такой подход нельзя считать удовлетворительным, поскольку дефицитность ресурсов зависит от комплектации ЭВМ и характера решаемых задач. Можно с уверенностью утверждать, что для большинства ВЦ нельзя вообще считать главным какой-либо ресурс. Например, трудно определить, какие ресурсы ЭВМ БЭСМ-6 являются в настоящее время самыми дефицитными - процессор, оперативная память, магнитные барабаны, магнитные диски или терминалы.

Распределять и учитывать машинное время в терминах нескольких важных ресурсов практически невозможно, поскольку использование этих ресурсов очень взаимосвязано. Поэтому предлагается ввести единый эквивалент для различных ресурсов - своего рода условные деньги.

Установим стоимость единицы машинного времени в этом условно-денежном исчислении - цену машинного времени. Поскольку в течение недели имеются периоды времени более удобные и менее удобные для работы, то цены машинного времени могут также различаться.

Для выражения затрат машинного времени на решение задачи требуется вычислить цены использования различных ресурсов вычислительной системы. При этом предлагается руководствоваться следующими двумя принципами:

- цены ресурсов должны отражать их дефицитность;
- в случае нормальной (например, среднестатистической для данного момента времени) загрузки вычислительной системы и отсутствия каких-либо источников недогрузки суммарная стоимость использования всех ресурсов должна равняться стоимости машинного времени.

Стоимость решения задачи складывается из двух составляющих - стоимости использования ресурсов и стоимости недогрузки ЭВМ, вызванной прохождением данной задачи (например, задача, которая захватила все имеющиеся в ЭВМ магнитные барабаны и тем самым не позволяет решать другие задачи, должна оплатить все машинное время).

1.1. Распределение машинного времени

Распределение машинного времени на какой-то интервал планирования (например, месяц) предлагается осуществлять следующим образом.

Из общего бюджета организации (стоимости машинного времени за интервал планирования) выделяются общие бюджеты отделов (бюджеты тем), из которых, в свою очередь, выделяются общие бюджеты групп. Из общих бюджетов групп выделяются общие бюджеты отдельных пользователей. При выделении бюджетов любого уровня (отделов, групп или пользователей) вовсе не обязательно полностью распределять исходный бюджет. Остаток этого бюджета можно разрешить совместно использовать некоторым владельцам бюджетов следующего уровня. Например, если руководитель группы не желает делить машинное время между своими сотрудниками, то он может выделить им нулевые бюджеты и разрешить всем пользоваться бюджетом группы. Можно было бы ограничиться таким простым распределением машинного времени, однако для повышения эффективности использования ЭВМ и удобства работы пользователей желательно более детально

планировать работу пользователей на ЭВМ. Для этого нужно вместо общих бюджетов на постоянный интервал планирования задать переменные бюджеты на период времени от начала интервала планирования до момента времени t (момент времени t принадлежит интервалу планирования). Осуществление такого детального планирования предполагается возложить на специальную систему, которая будет работать по следующему алгоритму. На первом этапе все переменные бюджеты задаются в виде линейных функций от времени t , изменяющихся на интервале планирования от нулевого значения до значения общего бюджета (т.е. выделенное каждому пользователю машинное время равномерно распределяется по интервалу планирования). На втором этапе переменные бюджеты корректируются согласно указаниям администрации. На третьем этапе производится коррекция тех переменных бюджетов, владельцы которых сделали соответствующие заявки. В заявках содержится информация об удобных и неудобных для них периодах времени, а также может указываться характер их предполагаемой работы. Следует отметить, что описанные коррекции могут производиться каждый раз, когда возникает в этом потребность.

1.2. Учет расхода машинного времени

Кроме бюджета каждый пользователь имеет расходный счет, на который начисляется плата за решение его задач. Стоимость каждой единицы машинного времени распределяется между решаемыми задачами и источниками недогрузки ЭВМ.

Как уже говорилось, задача оплачивает стоимость использования ресурсов и ту недогрузку ЭВМ, которая вызвана ее прохождением.

На расчетный счет инженерной службы начисляется плата за недогрузку ЭВМ, вызванную сбоями оборудования. Если сбой оборудования привел к аварийному завершению какой-либо задачи, то стоимость ее решения после контрольной точки списывается на инженерную службу (но не более какой-то нормативной суммы, которую позволено пользователю расходовать от одной контрольной точки до другой).

Плата за недогрузку, вызванную сверхнормативными ожиданиями реакций операторов на запросы системы, начисляется на расчетный счет операторской службы.

Если из-за отсутствия подходящих задач нельзя обеспечить нормальную загрузку вычислительной системы, то плата за такую недогрузку распределяется между пользователями, которые еще не

использовали выделенное им машинное время (пропорционально неизрасходованным остаткам выделенных им переменных бюджетов). Таким образом, пользователь будет оплачивать выделенное ему и неизрасходованное время, если оно не было использовано другими пользователями. Об отказе от времени пользователю желательно как можно раньше сообщить системе, занимающейся коррекцией переменных бюджетов (самый надежный вариант - сообщить системе, кто готов использовать это время).

1.3. Обеспечение прохождения задач в соответствии с распределением машинного времени

Для обеспечения заданного распределения машинного времени система стремится пропускать задачи таким образом, чтобы пользователи расходовали машинное время справедливо, т.е. пропорционально выделенным бюджетам.

Определим степень использования бюджета как отношение текущего значения расчетного счета к текущему значению переменного бюджета.

При пропуске задач предпочтение получают те задачи, у которых ниже степень использования бюджета. Если несколько владельцев нулевых бюджетов пользуется совместно бюджетом более высокого уровня, то наряду с собственными расчетными счетами они имеют один суммарный. Степень использования бюджета для них вычисляется как отношение текущего значения суммарного расчетного счета к текущему значению их совместного бюджета.

2. Принципы включения задач в решение

При включении в решение пакетных задач (режимы реального времени и разделения времени будут рассмотрены позднее) система должна руководствоваться следующими двумя принципами:

- стремиться обеспечить скорейшее прохождение имеющегося пакета задач;
- стремиться удовлетворить требованиям, предъявляемым к порядку прохождения задач.

2.1. Обеспечение скорейшего прохождения пакета задач

Чтобы обеспечить скорейшее прохождение пакета задач, нужно максимально загружать ресурсы вычислительной системы в соответствии с их дефицитностью. Для этого необходимо знать потребности задач в различных ресурсах.

Каждая задача требует для своего решения в течение различных интервалов времени работы различных устройств ЭВМ (например, процессоров, каналов, устройств ввода-вывода и т.д.). В силу заложенного в современных ЭВМ принципа параллелизма в некоторые интервалы могут работать несколько устройств одновременно. Сумма таких интервалов времени при мультипрограммном режиме может колебаться. Эта сумма будет минимальной при решении задачи в однопрограммном режиме (она будет равна времени нахождения задачи в решении). Назовем такую минимальную сумму интервалов временем решения задачи.

Для определения времени решения задачи вовсе не обязательно пропускать ее в однопрограммном режиме, т.к. и в мультипрограммном режиме операционная система может точно вычислять, насколько время нахождения задачи в решении превосходит минимально необходимое. Например, операционная система ДИСПАК для ЭВМ БЭСМ-6 вычисляет время решения каждой задачи (там оно называется коммерческим временем).

Для количественного выражения загрузки ресурсов введем понятие расхода ресурса. Расход ресурса - произведение объема ресурса на приращение времени решения задачи за период владения этим ресурсом, т.е. произведение объема ресурса на время владения им при однопрограммном режиме работы. Таким образом, время пассивности задачи (когда ее не обслуживают) не учитывается при определении расходов ресурсов и задачей не оплачивается. Будем предполагать, что известна следующая информация о каждой задаче:

- требуемое процессорное время,
- требуемые расходы ресурсов (включая и расход процессорного времени), заданные в процентном отношении к требуемому времени решения задачи.

Если пользователь не задает требуемых расходов ресурсов, то они берутся из хранящейся в системе информации о результатах предыдущего прохождения данного варианта задачи, либо устанавливаются равными некоторым стандартным значениям. В этом случае, а также при неверном указании пользователем

требуемых расходов ресурсов может возникнуть недогрузка ЭВМ, стоимость которой войдет в стоимость решения данной задачи.

Следует отметить, что для определения степени загрузки ресурсов находящимися в решении задачами требуется найти оптимальный режим их решения, т.е. установить скорости решения для каждой задачи. Скорость решения задачи - это приращение времени решения задачи за единицу машинного времени. Назовем ценой решения задачи стоимость единицы ее времени решения. Поскольку предполагаемые расходы ресурсов в единицу времени решения для каждой задачи заданы, то, зная предполагаемые цены ресурсов, можно высчитать для каждой задачи и предполагаемую номинальную цену ее решения.

Поскольку цены ресурсов отражают их дефицитность, то стремиться максимально загрузить ресурсы вычислительной системы - это значит подбирать такую смесь и такие скорости решения задач, чтобы максимизировать сумму предполагаемых номинальных стоимостей решения задач за единицу машинного времени, равную

$$\sum_{i=1}^n ПЦ_i * СК_i \quad ,$$

где

ПЦ_i - предполагаемая номинальная цена i-ой задачи,

СК_i - скорость решения i-ой задачи,

n - количество задач, находящихся в смеси.

2.2. Удовлетворение требований к порядку прохождения задач

Второй принцип включения задач в решение - удовлетворение требований, предъявляемых к порядку прохождения задач - вообще говоря, противоречит первому принципу. Нарушение "оптимального" порядка прохождения задач может привести к снижению загрузки ЭВМ. Если внеочередное прохождение какой-либо задачи вызвало недогрузку ЭВМ, то плата за эту недогрузку должна войти в стоимость решения данной задачи. Для изменения "оптимального" порядка прохождения задач предлагается использовать механизм коррекции предполагаемых цен задач.

К порядку прохождения задач предъявляются следующие требования:

- расходование машинного времени должно производиться в соответствии с его распределением,

- при прохождении задач предпочтение должно отдаваться коротким задачам (с малым временем решения) и тем задачам, которые долго ожидают включения в решение.

Для удовлетворения этих требований вместо предполагаемых номинальных цен задач можно использовать так называемые предлагаемые цены, которые вычисляются по следующей формуле:

$$\text{ПРЕДЛЦ} = \begin{cases} \text{ПЦ} * \text{K1}, & \text{если } \text{ВОТ} = -1 \\ \text{ПЦ} * (\text{K2} + \text{K3} * \text{ВОТ}), & \text{если } -1 < \text{ВОТ} < 0 \\ \text{ПЦ} * (1 + \text{K4} * \text{ВОТ} * \text{ВОЖ} / \text{ВРШ}), & \text{если } \text{ВОТ} \geq 0 \end{cases}$$

где:

ПРЕДЛЦ - предлагаемая цена задачи (ее повышение делает задачу более "выгодной" для включения в решение, а понижение - менее "выгодной");

ПЦ - предполагаемая номинальная цена задачи;

ВОТ - величина отставания степени использования бюджета задачи от степени использования бюджета организации (разница между степенью использования бюджета организации и степенью использования бюджета задачи). Если общий бюджет задачи уже исчерпан, то величина отставания ВОТ имеет значение -1.

ВОЖ - время ожидания включения задачи в решение

ВРШ - требуемое задаче время решения.

Отношение ВОЖ/ВРШ можно считать временем ожидания, при вычислении которого в качестве единицы измерения использовано требуемое задаче время решения. Это время ожидания умножается на величину ВОТ для того, чтобы "сила подталкивания" задач с низкой степенью использования бюджета возрастала постепенно, поскольку поспешное включение в решение только что введенной "отстающей" задачи может вызвать значительную недогрузку ЭВМ (и следовательно, значительную переплату за решение этой задачи);

K1, K2, K3 - параметры, позволяющие администрации вычислительного центра изменять "силу торможения" пользователей с высокой степенью использования бюджета. Поскольку понижение предполагаемой цены задачи приводит к соответствующему понижению платы за ее решение, то остаток фактической стоимости решения такой задачи будет оплачен "отстающими" пользователями, из-за отсутствия задач которых стало выгодным решать со скидкой задачи "опережающих" пользователей;

К4 - параметр, позволяющий изменять "силу подталкивания" пользователей с низкой степенью использования бюджета.

Следует отметить, что увеличение предполагаемой цены задачи не влечет за собой автоматического увеличения платы за ее решение. Плата за решение "отстающей" задачи увеличится только на стоимость фактической недогрузки ЭВМ, вызванной ее внеочередным прохождением.

3. Принципы обслуживания задач, находящихся в решении

Обслуживание находящихся в решении задач должно производиться в соответствии с принципами включения задач в решение, т.е. требуется обеспечивать заданные при включении скорости решения задач. Поскольку требуемые задачам расходы ресурсов известны, то можно считать, что для каждого ресурса заданы пропорции, в которых требуется делить этот ресурс между задачами. Такое деление можно осуществлять, например, отдавая каждый раз предпочтение той задаче, которая "отстает" по использованию данного ресурса. Подсчет полученных задачами ресурсов целесообразно вести в течение небольших интервалов астрономического времени. В этом случае будет обеспечен быстрый отклик терминальных задач на короткие запросы пользователей. Кроме того, такой алгоритм обслуживания позволит чутко реагировать на изменение характера решения задачи. Например, в тот период, когда фоновая задача будет писать промежуточные результаты на внешнюю память, ее скорость решения возрастет за счет того, что этой задаче будет отдаваться предпочтение при выделении процессорного времени.

Однако следует отметить, что описанный простейший алгоритм далеко не решает очень сложную проблему эффективного планирования в условиях неравномерного использования задачей ресурсов. Пути решения этой проблемы следует, по-видимому, искать в двух направлениях. Во-первых, вычисление оптимальных скоростей решения задач целесообразно осуществлять не только при изменении состава рабочей смеси, но и при обнаружении изменения характера решения каких-либо задач в смеси. Во-вторых, можно попытаться разработать приемлемые способы описания неравномерного характера задачи и его учета на этапе включения задач в решение.

4. Средства, позволяющие пользователю влиять на порядок прохождения задач

Пользователь может влиять на порядок прохождения задач посредством изменения предлагаемых цен решения своих задач. Ему предоставляются две возможности - задать потолок для предлагаемых цен и (или) задать добавок к предлагаемым ценам.

4.1. Ограничение предлагаемых цен решения задач

Поскольку цены решения задач зависят от цен ресурсов, а те, в свою очередь, от цены машинного времени, то они могут заметно меняться. Например, цена решения задачи днем может намного превышать цену ее решения ночью. Если пользователь не желает решать задачу по дневным ценам, то он может ограничить ее предлагаемую цену таким образом, что она войдет в решение только ночью. При этом пользователю предоставляется возможность задать ограничение в виде функции от времени окончания задачи (либо астрономического, либо относительного от момента ввода задачи). Практически, достаточно иметь возможность задавать простую функцию - ломаную из 3-4 отрезков. Возможность предлагать решить свою задачу по низкой цене позволяет пользователю вводить в пакет "резервные" задачи, за решение которых он будет платить лишь часть фактической стоимости. Оставшаяся часть стоимости решения таких задач будет оплачиваться виновниками недогрузки ЭВМ, из-за которых стало выгодно решать резервные задачи со скидкой.

4.2. Повышение предлагаемых цен решения задач

Обеспечить срочное прохождение своих задач пользователь может посредством повышения их предлагаемых цен решения.

Поскольку срочное прохождение задач, как правило, вызывает недогрузку ЭВМ, то необходимо уметь контролировать этот процесс. Для этого предлагается внутри каждого бюджета выделить специальный "срочный" бюджет, который бы показывал, какую сумму позволено его владельцу расходовать на оплату недогрузок ЭВМ, вызванных срочным прохождением его задачи. Этот срочный бюджет распределяется между пользователями по такому же алгоритму, как и обычный. Таким образом, для каждого пользователя кроме описанных ранее переменного бюджета и расчетного счета

имеются еще переменный срочный бюджет и срочный расчетный счет. Плата за недогрузку ЭВМ, вызванную повышением пользователем предлагаемых цен решения своих задач, начисляется и на обычный расчетный счет и на срочный.

Возможности пользователя повышать предлагаемые цены определяются величиной неизрасходованной части срочного бюджета.

Предлагаемая добавочная стоимость решения задачи (произведение времени решения на заданную пользователем добавочную цену) не может превышать величину неизрасходованной части срочного бюджета. Добавочную цену пользователь задает в виде функции от времени окончания задачи (заданного абсолютно либо относительно момента ввода задачи). Например, это может выглядеть следующим образом:

ДОПЛАТА 25(+15), 15(17.15), 0(17.30)

что означает:

- если задача будет решена не позже, чем через 15 минут после ввода, то пользователь готов переплатить 25 условно-денежных единиц за каждую единицу времени решения задачи;
- если задача будет решена до 17 ч. 15 м., то он согласен на увеличение цены в пределах от 25 единиц до 15 (в соответствии со временем окончания);
- если решение задачи будет завершено до 17 ч. 30 м., то пользователь согласен на повышение цены в пределах от 15 единиц до нуля;
- если задача не успеет посчитаться до 17 ч. 30 м., то пользователь будет оплачивать ее решение как решение обычной несрочной задачи.

Следует подчеркнуть, что увеличение предлагаемой цены решения задачи увеличивает предельную величину той возможной недогрузки ЭВМ, которую можно допустить ради внеочередного прохождения данной задачи. Это вовсе не означает, что такая недогрузка будет действительно иметь место и задача должна ее оплатить. Предлагаемая цена служит для "торговли" за право войти в решение, но в отличие от аукциона задача оплачивает только действительную стоимость "товара" - стоимость своего решения.

4.3. Изменение скорости решения задачи, находящейся в рабочей смеси

С помощью изменения предлагаемой цены можно влиять на скорости решения задач, находящихся в рабочей смеси. Это можно осуществлять путем обращения с любого терминала к специальной системе - системе оперативного управления и наблюдения. При этом нужно указать требуемую скорость решения задачи и добавочную стоимость, которую пользователь готов уплатить за такое повышение скорости. Если система не согласна на такую "сделку", то она сообщает пользователю величину добавочной стоимости, требуемую для обеспечения указанной скорости, а также сообщает скорость решения, которую она готова обеспечить за предложенную сумму.

Кроме того, данная система позволит изменять указания о срочности тех задач пользователя, которые еще не включены в решение. Система оперативного управления обеспечит также выдачу пользователям всевозможных справок, например: Когда войдет в решение такая-то задача? Когда завершится решение такой-то задачи? Сколько нужно переплатить за решение задачи к указанному времени? Каковы цены на ресурсы?

5. Обеспечение режимов разделения времени и реального времени

Для обеспечения быстрого вхождения в решение задач, работающих в режиме разделения времени или реального времени, предлагается заранее бронировать для них нужные ресурсы.

Будем различать ресурсы разделяемые и закрепляемые. Разделяемыми будут называться такие ресурсы, которые могут легко разделяться во времени между задачами, находящимися в рабочей смеси.

Разделяемыми ресурсами являются процессоры, каналы ввода-вывода. Те ресурсы, которые не могут разделяться во времени, закрепляются за задачами на какие-то периоды их решения. Такие ресурсы будем называть закрепляемыми. Закрепляемыми ресурсами могут быть устройства ввода-вывода, оперативная и внешняя память, программные каналы операционной системы. Чаще всего именно ограниченность закрепляемых ресурсов не позволяет включать в рабочую смесь нужные задачи. По этой причине дефицитные закрепляемые ресурсы стараются смоделировать на разделяемых ресурсах и менее дефицитных закрепляемых ресурсах (организация виртуальной памяти, многоуровневая внешняя память с автоматическим перемещением информации между уровнями, виртуальные устройства ввода-вывода).

Для закрепляемых ресурсов введем понятие операции Т-резервирования. Операция Т-резервирования требует задания четырех параметров - тип ресурса, количество ресурса, время ожидания освобождения ресурса, описание подмножества задач. В результате выполнения такой операции для задач указанного подмножества будет обеспечено получение указанных ресурсов в течение указанного времени после запроса. Если задача не входит в указанное подмножество, то она может получить зарезервированный ресурс только в том случае, если предполагаемое время ее нахождения в ЭВМ не превосходит указанного времени ожидания.

В отличие от обычного бронирования, когда предназначенные кому-то ресурсы никому другому не выдаются, при операции Т-резервирования они могут выдаваться коротким задачам во временное пользование. Тем самым сокращается время простоев забронированных ресурсов. Если модифицировать операцию Т-резервирования таким образом, чтобы вместо времени ожидания указывать астрономическое время освобождения ресурсов, то с помощью этой операции можно будет обеспечивать завершение всех задач к определенному моменту времени (например, к моменту утренней профилактики ЭВМ). Для этого достаточно зарезервировать все программные каналы, указав интересующее нас астрономическое время и пустое подмножество задач. В решение будут входить только те задачи, которые должны завершиться к указанному сроку. Следует отметить, что при задании операций резервирования должен быть указан расчетный счет, на который будет начисляться плата за вызванную этой операцией недогрузку ЭВМ.

5.1. Обеспечение режима разделения времени

Итак, быстрое вхождение задач, работающих в режиме разделения времени, можно обеспечить с помощью операции Т-резервирования. Быстрый отклик на запросы терминалов обеспечивается описанным в гл. 3 алгоритмом обслуживания находящихся в решении задач. Пользователь за терминалом может управлять скоростью решения своих задач с помощью описанных возможностей системы оперативного управления и наблюдения. Диалоговым подсистемам, обслуживающим несколько терминалов, будут предоставлены возможности для распределения стоимости своего решения между работающими пользователями.

Система, обеспечивающая коррекцию переменных бюджетов по заявкам пользователей, будет также уметь составлять расписание

работы пользователей за терминалами. При этом она будет учитывать указания администрации, пожелания пользователей, характер их предполагаемой работы за терминалом, состояния их бюджетов и расчетных счетов.

5.2. Обеспечение режима реального времени

Режим реального времени с точки зрения системы отличается от пакетного тремя особенностями:

- с помощью операций T-резервирования обеспечено быстрое вхождение задач реального времени;
- при задании добавочных цен можно указывать не время окончания задачи, а желательную скорость решения;
- прохождение задач реального времени не зависит от значения переменных бюджетов (бюджеты задач реального времени считаются неисчерпаемыми).

Плата за решение задач реального времени начисляется на расчетные счета точно так же, как и для всех остальных задач.

6. Распределение стоимости машинного времени между решаемыми задачами и источниками недогрузки

Стоимость машинного времени делится между его потребителями - решаемыми задачами и источниками недогрузки - следующим образом. Каждый потребитель имеет одну или несколько расходных статей, на которые начисляются соответствующие суммы. Скорость начисления на расходные статьи (сумма, начисляемая за единицу астрономического времени) меняется в зависимости от изменения различных условий - скоростей решения задач, цен на ресурсы, влияния различных источников недогрузки и т.п.

Тем не менее, можно считать, что все астрономическое время разбито на интервалы, на которых сохраняются постоянные скорости начисления. В конце каждого такого интервала стоимость машинного времени делится между расходными статьями пропорционально их скоростям начисления.

Различаются следующие потребители машинного времени:

- решаемые задачи;
- пользователи, из-за отсутствия задач которых невозможно нормально загрузить ЭВМ;
- пользователи, которые зарезервировали для своих задач закрепляемые ресурсы;

- инженерная служба;
- операторская служба.

У каждой задачи, находящейся в решении, имеются следующие расходные статьи:

а) номинальная стоимость - стоимость израсходованных ресурсов с учетом указания пользователя об ограничениях на цену решения задачи;

б) штраф за сверхнормативные излишки закрепляемых ресурсов.

Закрепляемые ресурсы отличаются от разделяемых тем, что они не могут перераспределяться. Если задача полностью захватит какой-либо ресурс, то другие задачи решаться не смогут. Стоимость захваченного ресурса может быть гораздо меньше недогрузки ЭВМ, вызванной такой монополизацией закрепляемого ресурса. Эта разница должна быть уплачена задачей в качестве штрафа за излишек закрепляемого ресурса. Нормативным количеством ресурса будем считать такое количество, доля стоимости которого в общей стоимости требуемых задачи ресурсов равна доле стоимости этого ресурса в стоимости машинного времени.

в) штраф за "подталкивание" задачи пользователей с низкой степенью использования бюджетов.

Этот штраф берется с задач "отстающих" пользователей, которые были включены в решение вне очереди посредством повышения их предлагаемых цен. Пользователь может стать "отстающим" либо в результате того, что поздно представил свои задачи, либо из-за того, что его задачи "неудобны" для решения. В обоих случаях за "подталкивание" его задач естественно брать дополнительную плату в соответствии с вызванной недогрузкой.

г) штраф за внеочередное включение срочных задач;

Этот штраф берется с тех задач, за срочное решение которых пользователь готов переплатить определенную сумму из своего "срочного" бюджета.

д) штраф за увеличение скорости решения срочных задач.

Увеличение скорости решения срочной задачи может приводить к недогрузке, которую эта задача должна оплатить. Как уже говорилось в гл.4, пользователь имеет возможность влиять на скорость решения своей задачи, находящейся в рабочей смеси. Как и предыдущий, этот штраф (точнее - соответствующая ему доля стоимости задачи) также берется из "срочного" бюджета пользователя.

е) штраф за неравномерность решения задачи.

Штраф берется с задач, которые расходуют свои ресурсы неравномерно. Эта неравномерность приводит к недогрузке ЭВМ. Если пользователь неверно задал требуемые задаче расходы ресурсов, и это привело к недогрузке ЭВМ, то на задачу будет также начислен штраф за неравномерность (поскольку нет необходимости различать эти ситуации).

Пользователи, из-за отсутствия задач которых невозможно нормально загрузить ЭВМ, имеют одну общую расходную статью - штраф за непоставку задач. Начисленная сумма делится между теми пользователями, у которых степень использования бюджетов ниже средней (степени использования бюджета организации). При этом деление производится пропорционально неиспользованным остаткам их текущих бюджетов.

Пользователи, зарезервировавшие для своих задач закрепляемые ресурсы, имеют столько расходных статей, сколько было выдано различных операций резервирования. На эти статьи начисляются штрафы за резервирование ресурсов.

Инженерная служба имеет три расходные статьи.

а) Штраф за отключение неисправных ресурсов.

Этот штраф берется в том случае, когда исправных устройств ЭВМ оказалось меньше, чем предусмотрено в штатной конфигурации.

б) Штраф за задержки в решении задач из-за неисправности оборудования.

Этот штраф налагается в том случае, когда из-за неисправности устройств ЭВМ или носителей информации произошли задержки в решении задач.

в) Штраф за лишние обращения к устройствам.

Этот штраф берется тогда, когда явных задержек в решении задач не зафиксировано, но известно, что были случаи повторения операций обмена.

Кроме того, на расчетный счет инженерной службы перечисляются и потери, вызванные аварийным завершением задач из-за неисправности оборудования. Организация такого перечисления с учетом нормативов на стоимость решения задачи от одной контрольной точки до другой принципиальных затруднений не вызывает.

Операторская служба имеет одну расходную статью - штраф за сверхнормативные ожидания ответов оператора на запросы операционной системы.

В данной работе не представляется целесообразным описывать алгоритмы начисления на каждую расходную статью. В основу этих алгоритмов положены два следующих принципа.

а) Назначение цен ресурсов производится путем вычисления их дефицитности в предшествующие периоды работы ЭВМ и экстраполяции на текущий момент. Исходя из этих назначенных цен и объемов расходуемых задачей ресурсов вычисляется ее номинальная стоимость. При этом учитываются указания пользователей об ограничениях на цену решения задачи.

б) Для нахождения величины каждого штрафа определяется влияние соответствующего источника недогрузки на суммарную номинальную стоимость всех решаемых в ЭВМ задач.

Следует отметить важное значение того факта, что стоимость машинного времени распределяется не только между задачами и основными источниками недогрузки ЭВМ, но также подробно расписывается по перечисленным выше расходным статьям. Информация о таком распределении позволяет каждому пользователю ясно видеть пути снижения стоимости решения своих задач. Эта информация поможет администрации организовать более эффективное использование имеющихся вычислительных мощностей.

Заключение

Система управления прохождением задач неразрывно связана с операционной системой. В ней можно выделить следующие функционально различающиеся компоненты:

- планировщик включения задач в решение;
- планировщик обслуживания находящихся в решении задач;
- система учета расходов машинного времени;
- система оперативного управления и наблюдения;
- система распределения машинного времени (с ее помощью выделяются и корректируются бюджеты, составляется расписание работы пользователей за терминалами);
- система обеспечения графика работы ЭВМ. Эта система будет следить за соблюдением графика работы ЭВМ (обеспечивать завершение задач к моменту сброса операционной системы, обеспечивать переходы на особые режимы работы и т.п.).

Реализация этих компонент для конкретной ЭВМ требует аппроксимации описанных в данной работе "идеальных" алгоритмов. При этом приходится искать компромисс между желанием иметь более точные алгоритмы и желанием минимизировать связанные с ними накладные расходы. В поиске такого компромисса существенную помощь может оказать моделирование.

Описанные выше принципы положены в основу системы управления прохождением задач СУПРОЗ, которая будет функционировать в рамках операционной системы ДИСПАК для ЭВМ БЗСМ-6.

Литература

1. L.J. Mc Kell, J.V. Haneen, L.E. Heitger. Charging for Computing Resources. Computing Surveys. Vol 11, Ho 2, June 1979.
2. C.R. Symons. A Cost Accounting Formula for Multiprogramming Computers. The Computer Journal. Vol 14, Ko 1.