# ВВЕДЕНИЕ

На современном этапе развития информационных технологий часто появляется необходимость обработки больших массивов однотипных данных за ограниченное время. Одним из способов решения данной проблемы является применение систем конвейерного типа. В этих системах важной составляющей является поступление требований на обработку.

В результате выполнения данной дипломной работы было разработано Windows-приложение, которое строит расписания обработки данных многих типов, с учетом времени перенастройки оборудования с обработки требования одного типа на другой. Учитывая то, что задача составления расписаний является NP-полной, то для её решения использовался метод градиентного поиска.

Разработанная система производит оптимизацию расписания по нескольким критериям эффективности, программный комплекс условно разделён на три уровня. Каждый из них представляет собой поиск оптимума по одному или нескольким критериям.

Производительность обработки данных при выполнении программ можно повысить путём конвейеризации [1]. Под конвейеризацией выполнения программы подразумевается разделение её на фрагменты, каждый из которых закреплён за соответствующим сегментом конвейера. Выполнение программ реализуется в многостадийной последовательной обрабатывающей системе с заданным порядком сегментов конвейера. Сегменты конвейера следуют строго друг за другом, местоположение сегмента в последовательности определяется его номером. Обозначим количество сегментов в конвейере через L, тогда порядковый номер сегмента, входящего в состав конвейера - . Если в состав конвейера входит L сегментов, то все выполняемые в системе программы должны быть разделены на L фрагментов, каждый из которых закреплён для выполнения за соответствующим сегментов конвейера. Тогда выполнение конвейеризированной программы в составе конвейера предполагает строгое закрепление её l-того фрагмента для выполнения на l-том сегменте конвейера.

В систему обработки данных могут поступать различные типы данных, обозначим их количество через n. Через i обозначим номер множества однотипных данных, которые должны быть обработаны в системе, тогда . Количество элементов в множестве однотипных данных, характеризуемых индексом i, обозначим через . Все данные i-того типа (в количестве ) обрабатываются соответствующей им программой. То есть существует  типов программ, выполняемых в системе, обрабатывающих данные i типа. Таким образом, в системе задаётся строгое соответствие между типом данных и обрабатывающей эти данные программой. Однократное выполнение программы i-го типа позволяет обработать один элемент множества данных i-того типа. Так как множество i-того типа содержит  элементов, то обрабатывающая программа i-того типа должна быть выполнена  число раз. Цель функционирования системы в этом случае состоит в обработке всех поступающих на её вход данных. В конвейеризированной системе, которая состоит из L сегментов, реализуется обработка n типов данных n типами программ. Для повышения эффективности использования ресурсов системы рекомендуется обрабатывать данные партиями [2, 3]. В этом случае «партия» - это совокупность наборов данных одного (i-того, ) типа, которые обрабатываются в программе непосредственно друг за другом. Группа партий – это совокупность партий, обрабатываемых в течение одного из заданных интервалов времени . При этом состав групп партий для каждого временного интервала  определяется в соответствии с вводимыми критериями эффективности.

Выполнение программ в конвейеризированной вычислительной системе требует использования её (системы) ресурсов. В первую очередь такими ресурсами являются процессорное время каждого сегмента конвейера и оперативная память, в которой хранится выполняемая программа и обрабатываемые данные. Чтобы наиболее эффективно с точки зрения использования сегментов конвейера использовать вычислительную систему необходимо планировать запуск программ различных типов, то есть составлять расписания обработки данных разных типов. Так как поступление данных i-того типа инициирует в конвейеризованной системе выполнение программы i-того типа, то под «расписанием» понимается порядок поступления данных разных типов на вход системы или, другими словами, порядок запуска программ разных типов на выполнение.

Из-за того что обработка на всех устройствах происходит последовательно, и обработка каждой партии не может быть разбита на несколько частей, необходимо сформировать порядок поступления партий данных только на первое устройство.

Так как рассматриваемая задача по определению количества и составов партий данных i-ых типов () связана с формированием комплектов, то эффективность решений по составам партий, генерируемым на первом уровне иерархии подсистем, определяется в соответствии количеством комплектов, образовываемых из обработанных в группах партий данных. Тогда наряду с реализацией метода определения количества и составов партий обрабатываемых данных необходима реализация метода формирования комплектов из данных, обрабатываемых в группах партий. Полученное решение по составам комплектов данных позволяет определить эффективность решения по количеству и составам партий. Разработанный метод формирования комплектов также, как и метод определения составов партий, реализует "жадную" стратегию по определению решений (реализует "жадный" подход к определению приближенно эффективных решений по составам комплектов).

# АНАЛИЗ ЛИТЕРАТУРЫ

# МОДЕЛЬ СИСТЕМЫ ПОСТРОЕНИЯ РАСПИСАНИЯ ПАРТИЙ ДАННЫХ

Для обоснования методов оптимизации составов партий, составов групп партий и расписаний обработки партий в рассмотрение введены обозначения и выполнены рассуждения, приведенные ниже ([1]). Через i обозначен идентификатор типа обрабатываемых в системе данных, n – количество типов обрабатываемых в системе данных (),  – количество данных i-го типа, которые должны быть обработаны в системе. В постановке задачи являются заданным, что  (), тогда обрабатывающая программа должна быть выполнена в конвейерной системе  число раз, а для обработки однотипных данных формируются партии. Партия – это совокупность однотипных данных, обработка которых выполняется без переналадки сегментов конвейера. Партия является фиксированной, если в нее входят все данные i-го типа, либо партия может содержать не все  данных i-го типа, в этом случае в системе обрабатывается ни одна, а несколько партий данных этого типа. Через *l* обозначен индекс сегмента конвейерной системы, осуществляющего выполнение *l-*й части программы в системе (). Дисциплина обслуживания выполняемых в системе программ предполагает прохождение данными, которые они обраба­тывают, всех сегментов конвейера, при этом если *l*-й прибор приступил к обработке данных *i*-го типа, обработка не может быть прервана. Все обрабатывающие приборы конвейерной системы характе­ризуются равными и неизменными во времени значениями производительно­сти их работы. Выполнение на каждом *l*-м сегменте назначенной ему части *i*‑ой программы характеризуется длительностью обработки данных.

Обработка партий данных разных типов выполняется в течение интервалов времени задаваемой длительности, для которых введены обозначения: – интервал времени, в течение которого может быть реализована обработка партий данных. Т.к. обработка партий данных i-ых типов () реализуется в течение заданных интервалов времени функционирования системы  (), тогда на основе решения по количеству и составам партий данных различных типов формируются группы партий, каждая группа партий обрабатывается в течение одного из интервалов  (формируется Z групп партий). Группа партий– это совокупность партий, обрабатываемых в течение одного интервала времени  функционирования системы. Т.к. заданными являются ограничения на интервалы времени функционирования системы  (), тогда не все сформированные партий могут быть включены в состав групп партий. Состав групп партий для интервалов  () определяется таким образом, чтобы обеспечить возможность формирования максимального количества комплектов каждого типа.

Через d обозначен идентификатор типа комплекта, который формируется из результатов обработки данных в конвейерной системе, через – общее количество типов формируемых комплектов. Комплект данных (результатов обработки данных) представляет собой совокупность данных, при этом количество данных каждого типа, входящих в комплект одного (d-го) типа, является заданным. Данные всех n типов входят в определенном количестве в состав комплекта каждого типа. Через  обозначено количество данных i-го типа, которые должны входить в один комплект d-го типа. В рассмотрение введена матрица (W), значения элементов  которой равны числу данных i-го типа, которые должны входить в комплект d-го типа. Таким образом, рассматриваются комплекты разных типов и задаются соответствующие им составы. В выполненной постановке задачи введены ограничения на длительность интервалов времени обработки партий, входящих в группы, тогда не все сформированные партии могут быть распределены по группам, не вошедшие в группы партии являются не обработанными. Комплекты разных типов могут быть сформированы только из данных, входящих в партии, включенные в соответствующие группы для обработки.

В тоже время количество обработанных в течение интервалов  данных зависит от количества и составов сформированных партий (т.е. составы групп партий (как следствие, количество сформированных комплектов) является зависящим от количества и составов партий). Решение по порядку обработки партий в каждой из групп является зависящим от составов партий данных в этой группе. Таким образом, определение количества и составов партий данных различных типов, распределение партий по группам, построение расписаний обработки партий групп с учетом ограничений на длительность интервалов  должно обеспечить формирование максимального количества комплектов типов.

Выполнена декомпозиция обобщенной цели функционирования системы на совокупность иерархически упорядоченных целей подсистем (функций, решаемых на иерархически упорядоченных уровнях системы построения расписаний обработки партий данных). Формирование решений осуществляется на уровнях системы следующим образом [1]: первый – решения по составам партий данных, второй – решения по составам групп партий; третий – решения по порядку обработки на сегментах конвейера партий, входящих в группы.

Для формирования решений по составам партий данных i-ых типов () в рассмотрение введены обозначения: – количество партий данных i-го типа (), формируемых на первом уровне принятия решений, элементы  образуют вектор (М); (А)– матрица, элемент  которой – это количество данных i-го типа в h-ой партии (). Решение, формируемое на первом уровне системы имеет вид: [(М), (А)], где (М)– вектор количества партий данных i-ых типов, (А)–матрица количества данных в партиях.

Для формирования решений по составам групп партий введены обозначения:  () – группа партий, обрабатываемых в течение одного из интервалов (); – количество партий данных i-го типа в группе партий ;  – вектор количества данных i-го типа в  партиях в группе . Тогда партии данных i-го типа, входящие в группу партий , определены с использованием набора параметров вида: , а группа партий – совокупность таких наборов вида: , где – количество типов данных, партии которых входят в . Решение, формируемое на втором уровне системы – совокупность групп партий, имеет вид: {()}.

В соответствии с полученным решением по составу групп партий требуется определить порядок обработки партий каждой из групп на приборах системы (должна быть определена последовательность запуска партий каждой из групп на обработку на сегментах конвейера), т.е. расписание обработки партий соответствующей группы. Расписание обработки партий z-ой группы обозначено как , оно представляет собой совокупность последовательностей  запуска партий на обработку на l-ых сегментах конвейера (). Расписание  имеет вид: . Расписание обработки партий  для группы  формируется в предположении, что порядок обработки партий данных является одинаковым на всех L сегментах конвейера. Для формализации вида последовательностей  расписания  в рассмотрении введена матрица порядка обработки партий . Элемент , если партия данных i-го типа занимает в последовательности  j-ю позицию,  в случае, если партия данных i-го типа не занимает в последовательности  j-ю позицию, размер матрицы , где  – количество типов данных в партиях в группе ,  – число партий в последовательностях для группы . Порядок обработки партий группы на всех сегментах одинаков, поэтому достаточно определения одной матрицы порядка . В рассмотрение введена матрица – матрица количества данных i-го типа в партиях, занимающих в последовательности  j-е позиции (элемент  равен количеству данных i-го типа в партии, занимающей j-ю позицию в , размер матрицы ). Решение, формируемое на нижнем (третьем) уровне иерархии подсистем в системе, имеет вид: {}.

В силу наличия ограничений на время функционирования системы не все сформированные партии данных могут быть размещены в группах партий (могут быть обработаны в течение заданных интервалов ()), а цель функционирования системы состоит в обработке данных таким образом, чтобы было сформировано максимальное количество комплектов, тогда эффективное решение {| ()} обеспечивает реализацию этого требования для текущих составов партий. Т.о. для формирования решения по составам комплектов данных на основе решения по группам партий {| ()} в рассмотрение введена матрица (), значения элементов d-го столбца которой соответствуют количеству данных i-ых типов, входящих в комплекты d-го типа (элемент  соответствует количеству данных i-го типа, которые будут входить в состав комплектов d-го типа). Для идентификации количества комплектов, сформированных из данных, обработанных в группах партий (), в рассмотрение введен вектор , значение d-го элемента которого соответствует количеству комплектов d-го типа (элемент  – это количество комплектов d-го типа, сформированных из данных, обработанных в группах партий ()). Для реализации алгоритма определения составов комплектов из обработанных данных в рассмотрение введен вектор , i-ый элемент которого  – это количество данных i-го типа (), обрабатываемых в группах партий в соответствии с текущим решением (), которые могут быть распределены по комплектам  типов. Перед началом реализации алгоритма распределения данных по комплектам элемент  вектора  определяется следующим образом: ().

Для определения вида модели вычислительного процесса обработки партий данных, включенных в группы () введены следующие обозначения: – время обработки данных i-го типа на l-ом сегменте конвейера (); – время переналадки l-го сегмента с обработки данных i-го типа на обработку данных k-го типа; – время первоначальной наладки l-го сегмента на обработку данных i-го типа;  – время начала обработки партии данных i-го типа, занимающей в последовательности  на l-ом сегменте j-ю позицию; – матрица моментов времени начала обработки партий данных i-ых типов, занимающих в  j-е позиции (для группы партий ); – матрица моментов времени начала обработки q-ых данных в партии, занимающей в  j-ю позицию (q – порядковый номер данных в партии в j-ой позиции в  (, где – количество данных в партии, входящей в группу , занимающей j-ю позицию в ). Элементы матрицы  определяются следующим образом: , где , – момент времени начала обработки первых данных в партии, занимающей j-ю позицию в , – количество типов данных, партии которых входят в одну рассматриваемую группу партий , – количество партий разных типов в группе . В рассмотрение введена матрица переналадок (), элементы  которой соответствуют длительностям переналадки сегментов с обработки данных i-го типа на обработку данных k-го типа, элементы – время первоначальной наладки сегментов на обработку данных i-го типа. Определение значений  и  (; ; ) выполняется в соответствии с изложенным ниже подходом для партий данных, входящих в одну группу  (индекс z группы партий  для временных параметров модели опущен при изложении).

Для позиции (q=1) данных i-го типа в (j=1)-ой партии в  выражение для определения  формируется при учете параметра  –времени наладки сегмента на обработку данных i-го типа: . Для  (,) имеем [1]: . В этом выражении первое слагаемое определяет длительность наладки сегмента конвейера на обработку данных в первой позиции в , второе слагаемое–длительность обработки данных в первой партии, предшествующих данным в q-ой позиции. Для первой позиции данных (q=1) (j=2)-ой партии в для  имеем: если  – время переналадки первого сегмента конвейера с обработки данных i-го типа (первая позиция партии в ) на обработку данных другого типа (вторая позиция партии в ), а – количество данных i-го типа в партии в первой (j=1) позиции в , тогда , где значение  определяется следующим образом:

, где .

Первой слагаемое в выражении для  и  определяет длительность наладки сегмента на обработку данных в партии, занимающей первую позицию в , второе слагаемое – длительность обработки данных в первой партии в , третье слагаемое – длительность переналадки сегмента с обработки данных в первой партии в на обработку данных во второй партии в . Для позиций  () значения  определяются выражением вида:

, .

По аналогии обобщенное выражение для определения  при  для последовательности  имеет вид: , моменты времени начала обработки данных в *q*-ой позиции в партии с индексом : . Выражения для  и  сформированы в общем виде следующим образом [1]:

;

;

.

Использование выражений для ,  и ,  () позволяет определить временные характеристики вычислительного процесса выполнения конвейеризированных программ обработки данных на сегментах конвейера. Метод построения расписаний обработки партий предполагает добавление текущей рассматриваемой партии данных i-го типа в конец последовательностей  () и определение эффективного местоположения партии в этих последовательностях. Тогда местоположение рассматриваемой партии в  может быть охарактеризовано текущими (для данного количества партий в  ()) простоями сегментов конвейера при обработке партий, нахо­дящихся в последовательностях .

В работе [1] выполнено формирование вида критерия эффективности принятия решений при построении расписаний обработки партий данных на нижнем уровне иерархии подсистем. Полученный критерий эффективности определения решения по порядку обработки партий в системе имеет вид [1]:

. (2.1)

Первое слагаемое в (2.1) позволяет определить суммарный простой всех L сегментов конвейера перед началом обработки данных в первой позиции (q=1) в первой партии (j=1) в  (). Второе слагаемое определяет суммарный простой сегментов конвейера при переходе от обработки партии данных одного типа (в (j-1)-ой позиции в ) к обработке партии данных другого типа (в j-ой позиции в ). Третье слагаемое позволяет определить суммарный простой сегментов, вызванный ожиданием готовности данных при обработке их внутри партий.

При формировании комплектов после окончания обработки могут быть использованы только данные, партии которых включены в группы (), тогда на втором уровне иерархии решение по составам групп  должно анализироваться с точки зрения количества комплектов, которые могут быть сформировано из данных, входящих в партии, включенные в группы (). Таким образом, на втором уровне иерархии системы на основе решения по составам групп партий  необходимо определить решение по составам комплектов данных, формируемых из результатов их обработки в группах партий. Элемент  () вектора соответствует количеству комплектов d-го типа, сформированных из данных, партии которых включены в группы (), тогда общее количество комплектов, сформированных из данных, партии которых включены в группы, определяется выражением: . Т.к. цель функционирования системы в течение заданных временных интервалов предполагает обработку данных таким образом, чтобы обеспечить формирование максимального количества комплектов, тогда полученное выражение использовано в качестве критерия эффективности решений  по составам групп партий на втором уровне иерархии системы ([1]).

На первом уровне иерархии критерий эффективности решений должен учитывать решение вида , формируемое на этом уровне, и эффективное решение по составам групп партий, формируемое на втором уровне в виде . Решение  характеризуется количеством сформированных комплектов разных типов, следствием из решения по составам групп партий является решение по количеству комплектов каждого типа – вектор  (решение  обеспечивает максимальное количество комплектов). При определении значения критерия , характеризующего решение на первом уровне, используется само это решение и решение по количеству комплектов каждого типа – вектор , характеризующий эффективное решение по составам групп партий . В качестве входных параметров при определении составов партий задается количество  () данных каждого типа, которые входят в партии для обработки, из результатов обработки которых формируются комплекты. Предполагается, что все данные, подаваемые на вход системы для обработки, должны быть использованы при формировании комплектов. Тогда решение  может быть охарактеризовано количеством данных, которые должны быть обработаны и включены в комплекты, но в соответствии с решением в комплекты не вошли. В соответствии с решением количество данных i-го типа, которые должны быть обработаны в системе, определяется выражением, где – элемент матрицы (A), значение которого равно количеству данных i-го типа в h-ой партии (), – количество партий данных i-го типа – i-ый элемент вектора (M). Общее количество данных различных типов, которые должны быть обработаны в системе, в соответствии с решением  определяется выражением . Количество данных различных типов, используемых при формировании комплектов  типов, определяется выражением вида . Тогда количество данных, которые должны быть обработаны в соответствии с входными параметрами решаемой задачи, но не использованы при формировании комплектов, определяется выражением .Полученное выражение используется в качестве критерия эффективности на первом уровне принятия решений.

Значение  – это количество данных в партии, занимающей j-ю позицию в последовательностях  (). Эти рассуждения распространены на последнюю партию в последовательности . Для определения последней партии, а также данных в последней позиции в этой партии использованы обозначения: –количество партий, входящих в группу  (индекс последней партии в ); – количество данных, входящих в последнюю в  партию (индекс последних данных в этой партии). Тогда – момент времени начала обработки последних данных в партии с индексом  на L-ом сегменте конвейера, а время окончания обработки этой партии на этом сегменте определяется выражением вида: . Ограничение на время обработки партий данных, входящих в группы  () , имеет вид:при . Тогда модель многоуровневого программирования для определения эффективных составов партий, составов групп партий и расписаний обработки партий имеет вид [1]:

1) первый (верхней) уровень: , где ;

2) второй уровень: , где ;

3) третий уровень:  (), где

(2.2)



4) ограничения на длительность реализации обработки партий групп  ():

 при .

С учётом модели (2.2) процедура определений эффективных составов партий данных, составов групп партий, обрабатываемых в течение задаваемых интервалов времени функционирования системы (), расписаний обработки партий каждой из групп в конвейерной системе предполагает: 1) определение количества партий данных i-ых типов () и составов этих партий (распределение данных i-ых типов по партиям); 2) определение составов групп партий данных, обрабатываемых в течение задаваемых интервалов времени (); 3) формирование порядка обработки партий (расписания их обработки). Иерархичность функций в системе определяет порядок принятия решений: 1) решение по составам партий данных в виде [(M), (A)], где (М) – вектор количества партий данных различных типов, (А) – матрица количества данных в партиях; 2) решение по составам групп партий в виде {()}; 3) решение по порядкам обработки партий данных, входящих группы () в виде:  (). Т.к. эффективность формирования групп партий (эффективность составов партий) оценивается на основе количества комплектов разных типов, формируемых из данных, обрабатываемых в группах партий, тогда дополнительным решением на втором уровне иерархии системы является решение по составам комплектов, определяемое в виде матрицы составов комплектов , значение элемента  которой соответствует количеству данных i-го типа, включаемых в комплекты d-го типа, и вектора  количества комплектов каждого типа.