ВОПРОСЫ ТЕОРИИ И ПОСТРОЕНИЯ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ
(Вывчислительные системы)
1977 год
Выпуск 70

ЮДА 681.142.39:681.323

ОБ ЭФФЕКТИВНОСТИ ИСПОЛЬЗОВАНИЯ
БУФЕРА КОМАНД ПРИ РАБОТЕ ОДНОРОДНОЙ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЫ
В СИНХРОННО-ПРОГРАММНОМ РЕЖИМЕ

Д.К. Дмитриев

Одним из основных свойств однородных вычислительных систем (ОВС) является их гибкость, проявляющаяся, в частности, в возможности реализации в них режимов, связанных с потоками команд и данных [1]. Как правило, в ОВС реализуется режим множественного потока команд с множественным потоком данных, называемый здесь А-режим (автономным режимом). На ОВС возможно пространство аппаратурная реализация другого режима: одиночного потока команд с множественным потоком данных (C-режим). Этот режим используется, например, в системе НИЛС-IV.

При работе в C-режиме специальным выделением элементарные машины (C-ЭМ) системы выполняют над своими операндами одну и ту же программу, которая выдается командой из одной C-ЭМ. C-ЭМ, которую выдает, видит и выполняет программу, называется ведущей (В-ЭМ), а C-ЭМ, которые только исполняют программу, называются ведомыми (В-ЭМ).

Основные достоинства и недостатки C-режима описаны в [2,3]. В [2] показано, что одним из основных факторов, влияющих на эффективность использования этого режима, является продолжительность времени выполнения команд от значений операндов. Установлено, что величина потеря времени зависит от длины программы и числа выполняемых в продольную машину. Показано, что для любой программы существует такое число машин, при котором математическое ожидание потерь времени меньше любо, например заданной, сколько угодно малой величины E.

Величина потеря времени из-за резких затрат времени на выполнение одноразовых команд в разных машинах системы зависит от ко-

пользованного метода синхронизации. В [2] подробно исследован метод покомандной синхронизации C-ЭМ при котором машины, ведущие программу, а машины, выполняющие ее, работают покомандно и поочередно, оставаясь после завершения своей части текущего взаимо-

действия. Ведущая очередная машина выполняется лишь после выполнения текущей. Анализ покомандной синхронизации дает оценку потерь времени для C-режима сверху.

В данной работе будет использован метод синхронизации C-ЭМ с помощью сверхоперативной памяти, входящей в каждую текущую машину.


С точки зрения загрузки устройства вычислительного модуля, для которых буфер команд является источником информации, емкость буфера должна быть такой, чтобы в нем было хотя бы одна команда при заданных колебаниях значений скорости загрузки fC и скорости разгрузки fO. При этом буфер команд должен проектироваться как асинхронный блок, загруженный с некоторой скоростью, если буфер команд неполон, и разгруженный — если буфер команд не пуст. В соответствии со сказанным, сигнал "запрос команды", который может интерпретироваться как "готовность буфера к приему", является функцией от загрузки последнего. Этот сигнал существует до тех пор, пока имеется хотя бы одна загружаемая нежели буфера. Предполагается, что запрос OP на выборку команд схемы отличается от запроса OP на выборку операндов и т.д. Это связано с особенностями обработки кодов команд и вызывающей от нее загрузки процессоров.

Буфер команд тем лучше справляется с задачей стабилизации колебаний fC и fO, и с задачей сохранения числа обращений к OP, чем больше его емкость Е. Однако увеличение емкости буфера увеличи-
вает стоимость вычислительного модуля. В связи с этим задача определения емкости буфера команд становится как задача расчета числа ячеек, которое обеспечивает при заданных колебаниях $f_T$ и $f_0$ установленное значение загрузки процессора. Такая задача решена для случая одной изолированно рабочей машины. Целью нашей работы будет определение влияния на загрузку процессоров машины ОВС количества машин, выполняющих одну и ту же программу. Оценка будет дана для программы, не содержащей ветвлений.

На рис. 1,6 (СВ-ЭМ — схема виртуальной машины, ММС — машина ММС, ВЭМ — ведущая ЭМ, БЭМ — ведомая ЭМ) приведена блок-схема ОВС, в которой используется буфер команд при работе в С-режиме. Как и на рис. 1, а, подчеркнуто, что команды в ОВС не имеют тактов. Иными словами, появление на рис. 1,6 пунктиров, определяет задержки распространения сигналов ни в одной из машин. В данном случае задержки считаются пренебрежимо малыми. Схема виртуальной машины определяет зависимость выдачя заявки в ОВС от очередной команды от виртуальной структурированного сигнала в каждой С-машине. Таким образом, схема виртуальной машины в логическом отношении представляет собой многоходовую схему.

Рассмотрим на примере использование буфера команд заданной емкости $E$ при выполнении программы в C-режиме. Значение $E$ найдено из условия установки загрузки ЭМ в A-режиме. Положим, что при выполнении программы на ЭМ в C-режиме, когда количество машин, выполняющих команды, равное 2, ЭМ, имеющий максимальное время выполнения своей программы в А-режиме, может выполнять кодовую машину. Время выполнения программы определяется после полного заполнения буфера команд. Для упрощения считаем, что команды в ОВС по очереди команды полностью совмещаются в ВЭМ с выполнением текущей, что не нарушает общности и вместе с тем упрощает качественное рассмотрение процесса.

Рассмотрение времени выполнения команд рассматриваемой программы дано на рис. 2, а). На рис. 2, б построена зависимость времени $T_{EM}$ работы ЭМ от числа $x$ выполненных команд: $T_{EM} = \sum_{i=1}^{x} t_i$, где $x = 1, 2, \ldots, \tau$, и $t_i$ — время выполнения $i$-й команды. Определяемая зависимость на этом графике соответствует.

Я) На рис. 2, а представлены значения времени выполнения команд для однопроцессорных ЭМ без совмещения операций. В общем случае эти значения есть промежутки времени между последовательными обращениями к буферу и находятся в интервале $[\frac{1}{f_{min}}, \frac{1}{f_{max}}]$. 

132
наибольшее значение ордината при $x = \tau$. Горизонталь с номером $x$ на рис.2,6 отмечает момент поступления кода $x$-й команды из В-ЭМ на входы буферов команд ЭМ ОВС в С-режиме. Номера команд, находящихся в буфере команд, для каждой ЭМ указаны над соответствующими участками графика.

Как можно видеть, различие скоростей разгрузки буферов команд в С-машине может привести к увеличению времени решения задачи в С-режиме сравнительно с временем её решения в А-режиме. После выполнения третьей команды ЭМ-1 пространства в ожидании поступления кода четвертой команды, вследствие чего график для неё смещается в положение, указанное пунктиром. Аналогичным образом ЭМ-2 пространства в ожидании кода одиннадцатой команды.

Проведенное рассмотрение показывает, что время выполнения программы в С-режиме при синхронизации с помощью буфера команд по-прежнему зависит от емкости этого буфера. Видно, что величина потерь С-режима зависит не только от емкости буфера команд, но и от числа взаимодействующих машин. Легко видеть также, что время выполнения программы в А-режиме $\ell$ машинами также зависит от $\ell$:

$T_{\ell} = \max \{ T_j \}$, где $T_j$ - время выполнения программы в $j$-й ЭМ.

На эффективность использования С-режима, кроме емкости буфера команд, влияет также время обработки ОП за командой в ведущей машине. Эффективность работы ОВС в С-режиме может быть повышена, если запросы на ОП за командами в С-режиме будут выделены собственным, более высоким, чем для А-режима, приоритетом. Изучим указанные факторы методом статистического моделирования.

2. Модель системы. ЭМ системы описывается как двухуровневая двухступенчатая система массового обслуживания без потерь. Узел генерации заявок (оперативная память) выдает заявки в случайные моменты времени, подчиняющиеся закону распределения,

$F_t(\tau) = \begin{cases} 0, & \text{если } \tau < E, \\ P(\tau \geq t), & \text{если } \tau \geq E. \end{cases}$

Здесь $e$ - текущее значение числа заявок в накопителе заявок (буфер команд), $E$ - заданная емкость накопителя заявок, $P(\tau > t)$ - вероятность того, что если заявка поступит в систему в данный момент, то очередная заявка не поступит за время $t$. Узел обслуживания (процессор) имеет закон распределения времени обслуживания.
Изучение моделей \( M_\lambda \) и \( M_\mu \) проводилось методом статистического моделирования для случая простейших потоков \( F_1(t) \) и \( F_2(t) \) с интенсивностями \( \lambda \) и \( \mu \) соответственно. Правдоподобие таких результатов соответствует отношениям \( \lambda/\mu = 10:1 \), что обеспечивает приемлемые для практики значения коэффициентов загрузки процессоров \( \varepsilon \) в пределах 0,65-0,96. Отражение на машинное время обчисления диапазон изменений независимых параметров при моделировании в следующих пределах: \( E \geq 20 \), \( \ell \geq 20 \), \( R \leq 4 \). Значение \( \xi < 0,1 \). Программы моделирования написаны на языке ФОРТРАН-ЦЕНБР для ЭВМ БЭСМ-6 и имеют объем 500 операторов.

3. Обсуждение результатов. Оценка функционирования машины вне системной работы осуществляется по значению среднего коэффициента загрузки процессора

\[
A_\infty = \frac{1}{\lambda \ell} \sum_{j=1}^{\ell} \frac{1}{j} \sum_{i=1}^{\lambda_j} t_{nj}.
\]

где \( \lambda \) - число проведенных испытаний системы, а \( A_\infty \) - средний коэффициент загрузки процессора \( j \)-й ЭМ в \( \lambda \)-м испытании. Значение \( A_\infty \) определяется соотношением

\[
P_{\lambda}\infty = \frac{1}{\ell \lambda} \sum_{i=1}^{\lambda} \frac{1}{j} \sum_{i=1}^{\lambda} t_{nj}.
\]

Здесь \( t_{nj} \) - время выполнения \( i \)-й по счету команды процессора \( j \)-й ЭМ; \( t_{nji} \) - полное время выполнения \( i \)-й по счету команды \( j \)-м процессором: \( t_{nj} = t_{nji} + t_{nj} \); \( t_{nji} \) - время обращения к оперативной памяти за \( i \)-й по счету командой \( j \)-й машины.

Оценка влияния системы машины на значение коэффициента загрузки процессора требует учета времени \( T_A \) выполнения заданной программы всеми машинами. Она производится по формуле

\[
A_\infty = \frac{1}{\lambda \ell} \sum_{j=1}^{\ell} \frac{1}{j} \sum_{i=1}^{\lambda_j} t_{nji}.
\]

где

\[
A_\infty = \frac{1}{\lambda \ell} \sum_{j=1}^{\ell} \frac{1}{j} \sum_{i=1}^{\lambda_j} t_{nji}.
\]

\[
T_A = \max \{ \frac{1}{\lambda \ell} \sum_{j=1}^{\ell} \frac{1}{j} \sum_{i=1}^{\lambda_j} t_{nji} \}.
\]
Оценива функционирования системы в С-режиме производится по значениям средних коэффициентов \( \bar{A}_C \) и \( \bar{A}_S \), определяемых аналогично \( \bar{A}_{ACP} \) и \( \bar{A}_R \). Время выполнения программы в С-режиме для \( j \)-ой ЭМ (\( \bar{\tau}_C \)) отличается от общего времени \( \bar{\tau}_C \) выполнения этой программы системой из \( \ell \) машин не более чем на величину полной разгрузки некоторой задачи емкостью \( E \). При достаточном большом \( \ell \) можно пренебречь этой разностью, вследствие чего для оценки С-режима определилось только значение \( \bar{A}_R \).

Характер изменения указанных величин показан на графиках рис. 3–6.

На рис. 3 показан характер изменения коэффициентов загрузки процессора в А-режиме для разных значений \( E, \ell \) и \( R \). Большому значению \( R \) соответствует меньший приоритетный уровень запроса на команду. Из данных графиков можно видеть, что коэффициент загрузки процессора \( \bar{A}_{ACP} \), служащий при достаточно высоких значениях \( R \) и \( E \) надежным показателем эффективности использования процессора одинарной машины, утрачивает свое значение уже при \( \ell = 3 \). Вместе с тем видно, что уже при \( \ell = 10 \) коэффициент \( \bar{A}_R \) достигает постоянного значения (предельная \( \bar{A}_{ACP} \), пунктирная - \( \bar{A}_{ACP} \)).

Характер зависимости \( \bar{A}_C \) от значений \( E, \ell \) и \( R \) аналитичен полученному для А-режима, хотя и дает несколько более низкие значения коэффициентов загрузки. На рис. 4 приведены графики, иллюстрирующие отношение наполнения коэффициента загрузки процессора при переходе от А-к С-режиму. Из этих графиков следует, что при емкости буфера команд порядка 10 ячеек отношение наполнения коэффициента загрузки при переходе на С-режим составляет около 10% общего времени выполнения программы. Кризис для \( E = 0 \) соответствует С-режиму с полной циклической синхронизацией.

На рис. 5 показано, какие приращения емкости буфера команд необходимо для того чтобы обеспечить коэффициент загрузки процессора на уровне, достигнутом для А-режима. В пределах \( \bar{A}_R = 0,65-0,7 \) такие приращения составляют около 4 ячеек для \( \ell = 20 \).

Данные рис. 6 показывают, что при определенных условиях компенсации падение коэффициента загрузки процессора при переходе в С-режим может быть обеспечено присоединением команд обращения к памяти в С-режиме более высокого приоритета, чем в А-режиме. Это свидетельствует о возможности эффективной организации С-режима при сохранении емкости буфера команд, используемой для А-режима.
Результаты статистического моделирования подтверждают эффективность предложенных архитектурных мер повышения скорости работы системы в С-режиме. Рассмотрены два архитектурных приема: увеличение емкости схемы в оперативной памяти команд и изменение приоритета обращения за командной в системе запросов к оперативной памяти. Первый прием применим в системах с программируемой микроструктурой [4]. Эти системы позволяют программно изменять емкость буфера команд в соответствии с числом машин, занятых в A- и C-режимах. Второй прием позволяет использовать работу в C-режиме при традиционной структуре ЭВМ третьего и четвертого поколений. Он обеспечивает постоянное значение коэффициента загрузки процессора независимо от числа занятых в работе машин.

В моделировании и обсуждении результатов принимала участие Д.К.Ингатенко, за что автор выражает ей благодарность.

Литература


3. ДИМИТРИЕВ Д.К. О способе реализации вставляемых программ при работе однородной вычислительной системы в синхронно-программном режиме. - В кн.: Вопросы теории и построения вычислительных систем. (Вычислительные системы. Вып. 80.) Новосибирск, 1974, с. 103-114.


Поступила в ред.-изд. отд. 29 декабря 1976 года