

УДК 004.415.2.032.24

И.В.Стручков (асп., каф. АиВТ), В.М.Ицыксон, к.т.н., доц., В.Ф.Мелехин, д.т.н., проф.

ДИНАМИЧЕСКАЯ ОПТИМИЗАЦИЯ НАГРУЗКИ В МНОГОПРОЦЕССОРНОЙ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЕ

Задача оптимального распределения нагрузки в настоящее время становится одним из важнейших вопросов организации параллельных и распределенных вычислений. Это связано с двумя наблюдающимися тенденциями в сфере высокопроизводительных вычислений:

- подавляющее большинство вычислительных систем с производительностью 1 ТФлоп и выше строится по принципу систем с распределенной памятью, иначе называемых массивно-параллельными вычислителями (МРР).
- стремительно набирает обороты возникшая в 90-х годах прошлого века технология метакомпьютинга, или GRID.

Основной отличительной чертой этих систем является то, что для них решающую роль в увеличении производительности играет не оптимальная компиляция под заданную архитектуру вычислителя, а оптимальное разбиение программы на подзадачи и распределение их по имеющимся вычислительным ресурсам.

Ограничимся рассмотрением только многопроцессорных вычислительных систем с распределенной памятью. Для указанного класса систем на этапе проектирования параллельной программы выделяются логические программные модули-задачи, каждая из которых может быть размещена в локальной памяти одного процессора.

Целью этапа *статического распределения нагрузки* [1] является оптимизация размещения задач по вычислительным узлам на основе априорных данных до начала выполнения программы. Вопросом оптимизации могут быть следующие характеристики системы задач:

- количество логически различающихся задач и размер одной задачи;
- количество реплик задачи одного типа для распараллеливания однотипных операций;
- размещение задач по процессорам.

Для формального описания параллельной программы на этапе проектирования и статического распределения предполагается использовать граф задач-генераторов и задач-преобразователей данных [2].

Хотя для многих программ статического распределения нагрузки может оказаться достаточно, тем не менее, возникает потребность и *динамического перераспределения нагрузки* в процессе выполнения программы. Целью динамического перераспределения может быть:

1. повышение производительности системы за счет учета реальной загруженности вычислительных узлов и имеющихся ресурсов;
2. повышение надежности системы за счет резервирования избыточных вычислительных узлов и переноса вычислений в случае выхода из строя какого-либо из работающих узлов;
3. поддержка динамически реконфигурируемых или мобильных систем, в которых состав вычислительных узлов и других ресурсов может изменяться в процессе выполнения программы.

При динамическом перераспределении нагрузки возникают задачи сбора информации о загруженности и работоспособности узлов системы, принятия решения об оптимальном реконфигурировании, переноса вычислений между узлами, поддержания адекватности таблиц маршрутизации данных.

Представляется разумным объединение статического распределения нагрузки для начального конфигурирования системы и динамического перераспределения для улучшения характеристик системы в процессе функционирования. Именно такой Подход положен в основу разрабатываемого *адаптивного метода оптимального распределения нагрузки в многопроцессорной вычислительной системе*.

В результате данной работы будут получены:

- необходимые формализмы и математический аппарат для адаптивного метода оптимального распределения нагрузки в многопроцессорной системе;
- требования и спецификации на компоненты программной системы распределения нагрузки на основе адаптивного алгоритма;
- имитационные и аналитические модели компонентов для верификации и сравнения возможных реализаций;
- действующий прототип системы распределения нагрузки на основе адаптивного алгоритма.

ЛИТЕРАТУРА:

1. Hluchy L., Dobrovodsky M., Dobrucky M. Static Mapping Methods for Processor Networks. Institute of Computer Systems, Slovak Academy of Sciences.
2. Стручков И.В., Ицыксон В.М., Мелехин В.Ф. Исследование эффективности коммуникационных систем для многопроцессорного комплекса с распределенной памятью // XXXII Неделя науки СПбГПУ: Материалы межвузовской научной конференции. - СПб.: Изд-во СПбГПУ, 2004.