Многопроцессорные вычислительные системы
2006. №1(11)

УДК 519.683

© 2006 г. **В.А. Анненков,**
Я.В. Катуева, канд. техн. наук
(Институт автоматики и процессов управления ДВО РАН, Владивосток)

ИЗМЕРЕНИЕ ОСНОВНЫХ ХАРАКТЕРИСТИК ПРОГРАММНО-АППАРАТНОЙ СРЕДЫ LINUX-КЛАСТЕРА ALEPH-ЦЕНТРА КОЛЛЕКТИВНОГО ПОЛЬЗОВАНИЯ ИАПУ ДВО РАН[1]

В работе описывается инструментальное средство Onto Dev, методы разработки интерфейса с его помощью, способы расширения доступного при разработке инструментария, а также опыт использования данного инструментального средства.

Введение

Эффективное решение крупномасштабных задач численного моделирования и проведение вычислительного эксперимента в фундаментальных и прикладных научных исследованиях возможно только с использованием высокопроизводительных вычислительных ресурсов. Рост вычислительных возможностей во многом определяется интенсивным развитием средств параллельной работы в аппаратуре ЭВМ. В качестве параллельных вычислителей широко используются открытые системы массового параллелизма, состоящие из стандартных компонентов, в том числе массовых серийных микропроцессоров. Для создания подобных кластерных компьютеров сформировался как рынок аппаратных средств, так и требуемый для распределенной обработки набор программных компонентов, состоящий из некоммерческого свободно распространяемого программного обеспечения.

Такие мультипроцессорные системы (вычислительные кластеры) представляют собой набор процессорных узлов, соединенных между собой в единую систему с помощью высокоскоростной шины или коммутатора [1 – 3]. Вычислительными модулями подобных систем являются серийные процессоры с локальной памятью. Преимущества вычислителей данного класса очевидны: они имеют низкую стоимость, надежные и живучие в

[1] Работа выполнена в рамках гранта ДВО РАН 06-1-П14-052 по программе №14 фундаментальных исследований ОЭММПУ РАН и гранта ДВО РАН 06-П15-056 по программе №15 фундаментальных исследований ОЭММПУ РАН.

том плане, что при отказе одного из процессоров оставшиеся процессоры потенциально работоспособны и система может продолжать выполнять свою задачу, правда, с несколько меньшей производительностью, и, наоборот, можно наращивать вычислительную мощность за счет добавления новых вычислительных модулей.

Любая параллельная вычислительная среда обладает своими характерными особенностями, поэтому для работы с ней необходимо знать не только предметную область прикладных исследований, но и технологии параллельного программирования, архитектуру и характеристики параллельного компьютера. Естественной проблемой является то, как наилучшим способом использовать имеющиеся аппаратные средства для конкретной вычислительной задачи.

Основной для параллельного выполнения программы на кластере является модель передачи сообщений. Параллельная программа в данной модели представляет собой систему независимо функционирующих процессов, взаимодействующих друг с другом посредством передачи сообщений. Взаимодействие процессоров через коммуникационную среду требует дополнительных затрат времени. Степень совмещения межпроцессорных обменов с вычислениями и время, необходимое для выполнения межпроцессорных обменов, являются одними из факторов, определяющих эффективность выполнения параллельных программ на многопроцессорных ЭВМ с распределенной памятью [4]. Поэтому знание и учет характеристик коммуникационной среды необходимы на этапе декомпозиции последовательного алгоритма.

В данной работе исследуются характеристики программно-аппаратной среды Linux – кластера ALEPH центра коллективного пользования ИАПУ ДВО РАН.

Постановка задачи

Linux – кластер ALEPH, установленный в Институте автоматики и процессов управления ДВО РАН (http:/www.dvo.ru/bbc), представляет собой многомашинный параллельный вычислитель, состоящий из 5 однородных вычислительных узлов [1].

Вычислительные узлы Linux-кластера ALEPH имеют следующую конфигурацию (табл. 1).

Таблица 1

Процессоры	Pentium IV HyperThreading 3.0GHz
Кэш-память второго уровня	1 Mb
Оперативная память	2Gb на вычислительный узел (DDR 400)

Кластер состоит из 10 (5x2) процессоров. Операционная система – Gentoo Linux с ядром 2.6.15 SMP (Gentoo Base System version 1.6.14). В качестве коммуникационной среды на кластере работает Gigabit Ethernet с коммутатором D-Link.

Программное обеспечение реализует поддержку параллельных вычислений, используя стандарт MPI, реализация LAM-7.1.1. Кроме этого, машины кластера можно использовать для запуска последовательных приложений.

Таким образом, по своей архитектуре кластер ALEPH представляет собой SMP-серверы, объединенные локальной сетью. Структура кластера представлена на рис. 1.

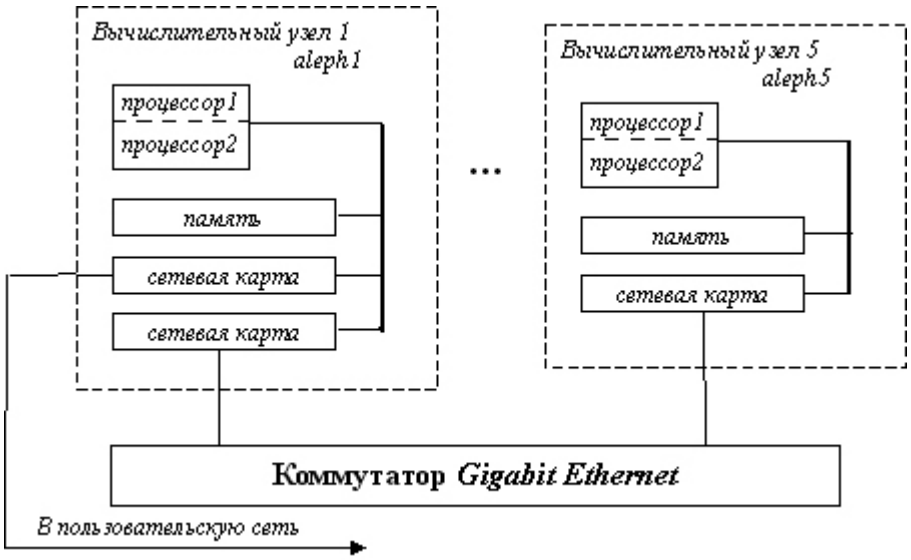

Рис. 1. Структура Linux – кластера ALEPH (пользовательский уровень).

Но поскольку в реализации системы межпроцессорных пересылок MPI и классической постановке для системы UNIX предполагается использовать однопроцессорную архитектуру, то, с точки зрения параллельного приложения, архитектура кластера будет выглядеть следующим образом (рис. 2).

Процессоры, входящие в состав одного вычислительного узла, делят между собой общую память, пересылки между этими парами процессоров происходят через общую память без задействования общего сетевого интерфейса. В связи с этим необходимо изучать количественные характеристики межпроцессорного взаимодействия на уровне SMP-узла.

При такой структуре параллельного вычислителя необходимо исследовать следующие коммуникационные характеристики:

латентность при передаче данных между двумя процессорами одного вычислительного узла;

скорость передачи данных от одного процессора вычислительного узла другому процессору этого же узла;

латентность при передаче данных между двумя вычислительными

узлами;

скорость передачи данных между двумя вычислительными узлами.

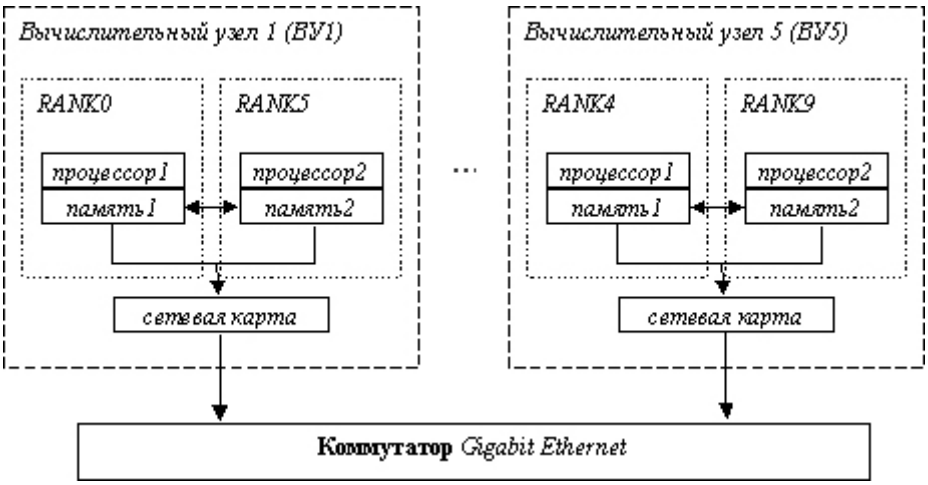

Рис. 2. Параллельная архитектура Linux-кластера ALEPH (уровень MPI-приложения).

Методика измерения основных характеристик программно-аппаратной среды

Обмены между процессорами происходят путем передачи по сети сообщений, содержащих численную или иную информацию. Для оценки мощности вычислителя важна такая характеристика как скорость передачи информации по сети. Основных характеристик производительности коммуникационных сетей в кластерных системах две: латентность – время начальной задержки при посылке сообщений – и пропускная способность сети, определяющая скорость передачи информации по каналам связи [3]. При этом важны не столько пиковые характеристики, заявляемые производителем, сколько реальные, достигаемые на уровне пользовательских приложений, – например, на уровне MPI-приложения. В частности, после вызова пользователем функции посылки сообщений, сообщение последовательно пройдет через целый набор слоев, определяемых особенностями организации ПО и аппаратуры, прежде чем покинуть процессор, что и порождает латентность. Поэтому максимальная скорость передачи по сети достигается на больших сообщениях, когда латентность, возникающая лишь в начале, не столь заметна на фоне непосредственно передачи данных.

Для определения количественных характеристик межпроцессорного обмена применялась стандартная методика (http://parallel.ru/testmpi/), разработанная в НИВЦ МГУ А.Н. Андреевым и Вл.В. Воеводиным [6].

Количественные характеристики межпроцессорного взаимодействия в кластере ALEPH

Количественные характеристики межпроцессорного обмена определялись при применении средств MPI/LAM.

Пропускная способность сети представляет собой количество информации, передаваемой между узлами в единицу времени (байт в секунду).

Латентностью называется время, затрачиваемое программным обеспечением и устройствами, по которым осуществляется обмен информацией, на подготовку к передаче сообщений по данному каналу. Латентность (задержка) — один из основных показателей эффективности коммуникационной инфраструктуры кластера. Эта задержка включает в себя программный компонент (задержка, определяемая стеком протоколов TCP/IP, а также возможные накладные расходы, связанные с копированием передаваемых данных из буферов пользователя в буферы ядра операционной системы) и аппаратный компонент. В последний, кроме задержек на портах коммутаторов, входят и задержки на сетевых платах. Латентность измеряется как время, необходимое на передачу сигнала или сообщения нулевой длины. При измерении латентности для снижения влияния погрешности и низкого разрешения системного таймера важно повторить операцию посылки сигнала и получения ответа большое число раз.

После проведения тестов были получены следующие показатели латентности для Linux-кластера ALEPH (табл. 2).

Таблица 2

	Латентность (10^{-6}с)
Межпроцессорный обмен внутри SMP-узла	1
Межпроцессорный обмен между SMP-узлами	60

Согласно методике измерения пропускной способности сети [6] проводились следующие измерения:

пропускная способность однонаправленных пересылок внутри узла;

пропускная способность однонаправленных пересылок между узлами.

Результаты измерения пропускной способности представлены на графиках рис. 3 – 6, где по горизонтали показана длина сообщения в байтах (Bytes), по вертикали – скорость передачи сообщения в мегабайтах в секунду (MB/sec).

Показатели производительности межпроцессорных обменов между SMP-узлами, выполненных на блокирующих операциях ***MPI_Send*** и ***MPI_Recv***, представлены на рис. 3 (для длинных сообщений) и рис. 4 (для коротких).

Показатели производительности межпроцессорных обменов внутри SMP-узла, выполненных на блокирующих операциях ***MPI_Send*** и ***MPI_Recv***, представлены на рис. 5 (для длинных сообщений) и рис. 6 (для коротких сообщений).

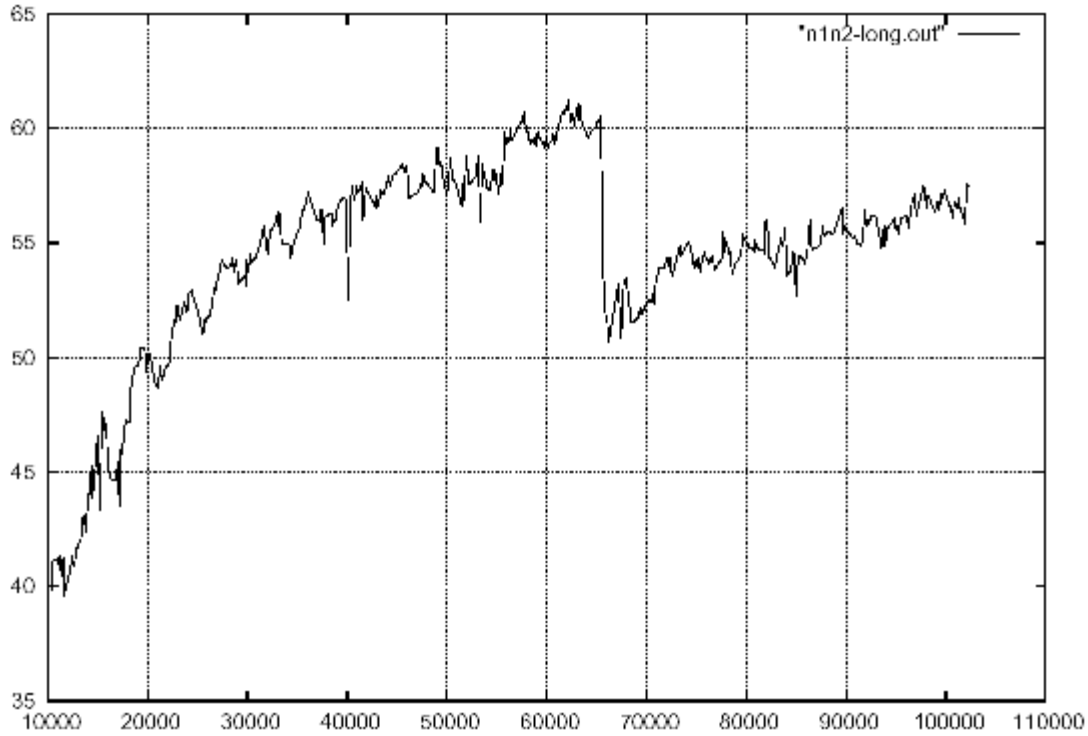

Рис.3. Производительность межпроцессорного обмена между SMP-узлами с применением блокирующих операций для длинных сообщений.

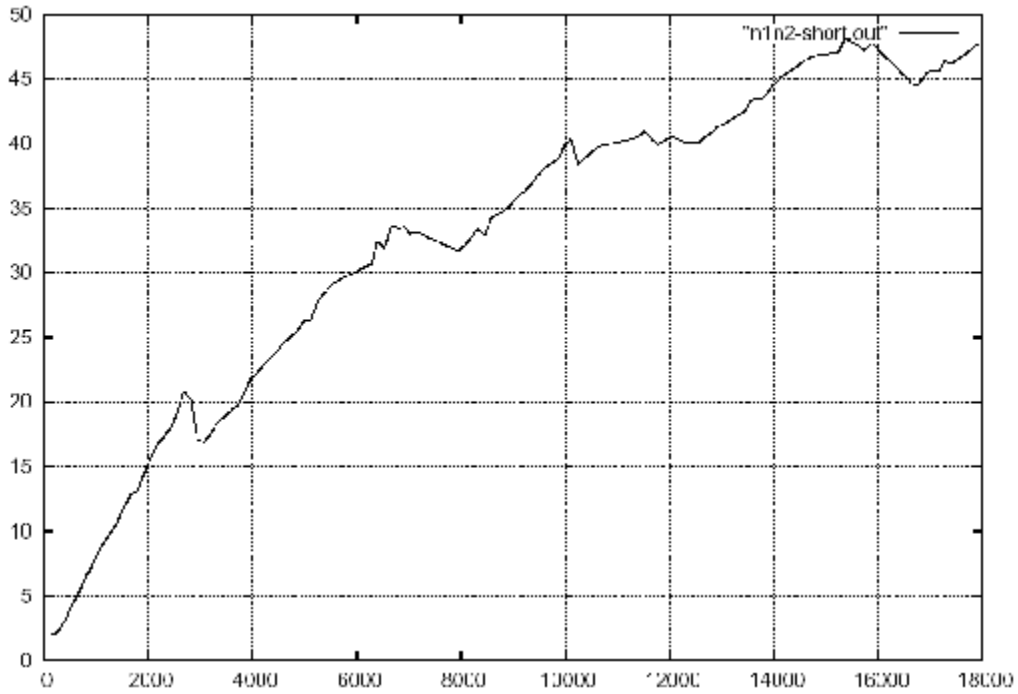

Рис. 4. Производительность межпроцессорного обмена между SMP-узлами с применением блокирующих операций для коротких сообщений.

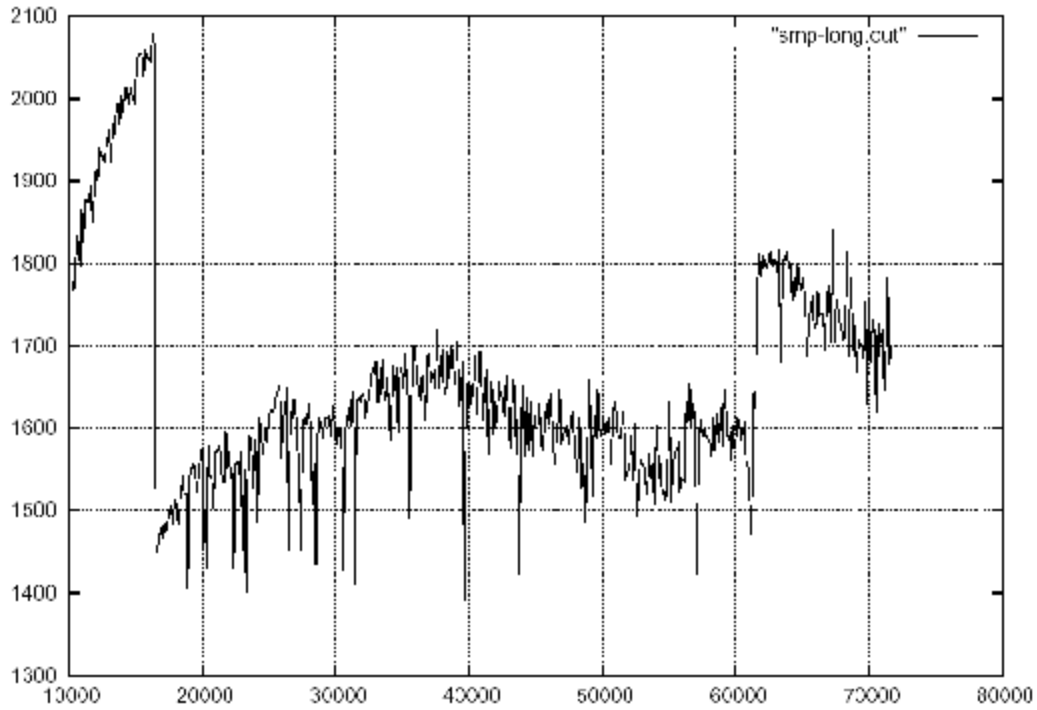

Рис.5. Производительность межпроцессорного обмена внутри SMP-узла с применением блокирующих операций.

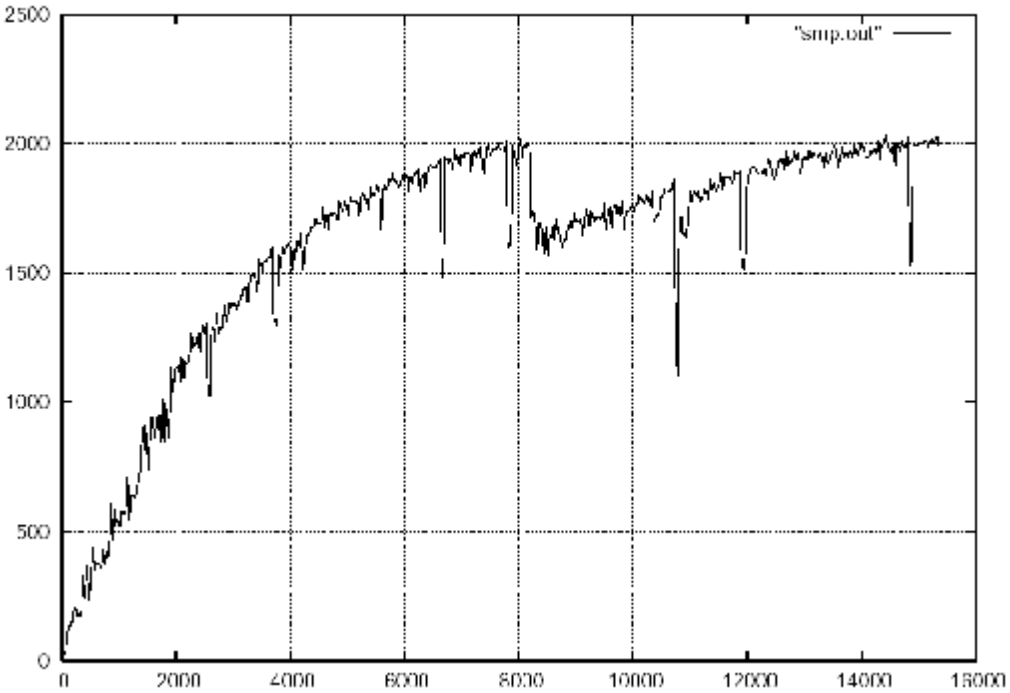

Рис. 6. Производительность межпроцессорного обмена внутри SMP-узла с применением блокирующих операций для коротких пересылок.

Резкое уменьшение скорости обмена между процессорами внутри узла на рис.5 обусловлено внутрипроцессорным кешированием. В дальнейшем наблюдаются рост и стабилизация скорости передачи данных.

Заключение

В работе исследовались характеристики коммуникационной среды вычислительного кластера. Значения данных показателей необходимы при создании параллельных программ и анализе их эффективности. Латентность и пропускная способность являются ограничивающими факторами вычислителей подобного класса, что необходимо учитывать при декомпозиции параллельного алгоритма. Соотношение между скоростью процессоров и скоростью обмена данными в коммуникационной среде является определяющим при построении эффективных параллельных приложений.

ЛИТЕРАТУРА

1. *Крюков В.А*. Разработка параллельных программ для вычислительных кластеров и сетей // Информационные технологии и вычислительные системы. – 2003. – № 1,2.
2. *Воеводин. В.В., Воеводин. Вл.В*. Параллельные вычисления. СПб.: БХВ-Петербург, 2002.
3. *Андреев А.Н., Воеводин Вл.В., Жуматий С.А*. Кластеры и суперкомпьютеры – близнецы или братья? // Открытые системы. – 2000. – № 5, 6. – С. 9-14.
4. *Андреев А.Н., Антонов А.С., Воеводин В.В., Воеводин Вл.В., Жуматий С.А*. Комплексный подход к анализу эффективности программ для параллельных вычислительных систем // тр. Всеросс. науч. конф. "Высокопроизводительные вычисления и их приложения". – Черноголовка, 2000. – С. 18-19.
5. *Антонов А.С., Крысанов Б.Ю*. Вычислительный полигон как средство исследования программно-аппаратных платформ // Вычислительные методы и программирование. – 2003. – Т. 4. Раз. 2. – С. 37-43.
6. *Андреев А.Н., Воеводин Вл.В*. Методика измерения основных характеристик программно-аппаратной среды. http://www.parallel.ru/testmpi/.

Статья представлена к публикации членом редколлегии О.В. Абрамовым.