

АВТОМАТИЗИРОВАННАЯ НАСТРОЙКА ПАРАМЕТРОВ БИБЛИОТЕКИ S-MPI ДЛЯ КЛАСТЕРНЫХ СИСТЕМ С РАЗЛИЧНОЙ АРХИТЕКТУРОЙ

Д.В. Донцов

*ООО «Центр компетенций и обучения», Саров
e-mail: dvdontsov@compcenter.org*

Представлен инструмент, предназначенный для настройки параметров библиотеки S-MPI, реализующей стандарт MPI-2. Данный инструмент позволяет автоматизировать подбор оптимальных значений параметров библиотеки S-MPI с учетом как специфики архитектуры кластерной системы, так и специфики параллельного приложения. Доклад посвящен рассмотрению подходов, применяемых для подбора оптимальных значений параметров, включая возможность анализа зависимости производительности параллельного приложения от значений настраиваемых параметров без анализа его исходного кода.

Введение

Библиотека S-MPI [1], являющаяся отечественной реализацией стандарта MPI-2 [2], предназначена для запуска параллельных приложений на кластерах с различной архитектурой. Она имеет большое количество параметров, которое контролирует ее поведение и влияет на ее производительность. Оптимальные значения этих параметров могут сильно варьироваться в зависимости от среды исполнения и от архитектуры параллельного приложения. Еще одной важной особенностью является наличие фактора взаимного влияния параметров библиотеки друг на друга. Поэтому процедура поиска оптимальных значений параметров библиотеки S-MPI является достаточно сложной, долгой и ресурсоемкой.

Представляемый инструмент предназначен для автоматизации поиска оптимальных значений параметров библиотеки S-MPI с учетом как специфики кластера, так и специфики приложения.

Инструменты автоматической настройки

В настоящее время уже существуют инструменты автоматической настройки MPI-реализаций:

- MPI Tuner, являющийся коммерческим продуктом, входящим в поставку библиотеки Intel MPI [3];
- ОТПО (Open Tool for Parameter Optimization [4]), являющийся свободно распространяемым продуктом и предназначенный для настройки OpenMPI [5].

Обе реализации инструментов настройки обладают рядом недостатков, которые не позволяют их использовать для настройки библиотеки S-MPI. Если MPI Tuner предназначен только для настройки Intel MPI, то ОТПО достаточно ограничен в его функциональности: отсутствие возможности настройки с учетом специфики кластера, отсутствие возможности настройки параметров с композиционными значениями, изначальная поддержка ограниченного набора тестов. В связи с этим для библиотеки S-MPI был разработан собственный инструмент, получивший название MPIBoost, позволяющий проводить настройку как с учетом специфики кластера, так и с учетом специфики па-

раллельного приложения. Каждый из режимов работы разработанного инструмента настройки обладает рядом своих преимуществ и не только не исключает их совместное использование, а позволяет добиться максимальной эффективности библиотеки S-MPI в ходе их совместного использования.

Режим настройки на специфику кластера

Режим настройки на специфику кластера предназначен для поиска оптимальных значений параметров S-MPI с учетом среды запуска (совокупности программных и аппаратных средств, включая коммуникационное оборудование). Этот режим работы базируется на использовании IMB (Intel® MPI Benchmarks [6]), который позволяет получать необходимые данные о выполнении MPI-операций для различных размеров сообщений. Получаемые с его помощью данные позволяют провести оценку производительности MPI-операций, независимо от архитектуры параллельного приложения, и на ее основе определить оптимальные значения параметров библиотеки S-MPI.

Параметры S-MPI можно разделить на три группы:

- могут принимать значения из predetermined списка допустимых значений;
- могут принимать значения из заданного диапазона;
- могут принимать композиционные значения.

На рис. 1 показан результат подбора параметра библиотеки S-MPI, отвечающего за правило выбора алгоритма коллективной операции Allreduce и принимающего композиционное значение.

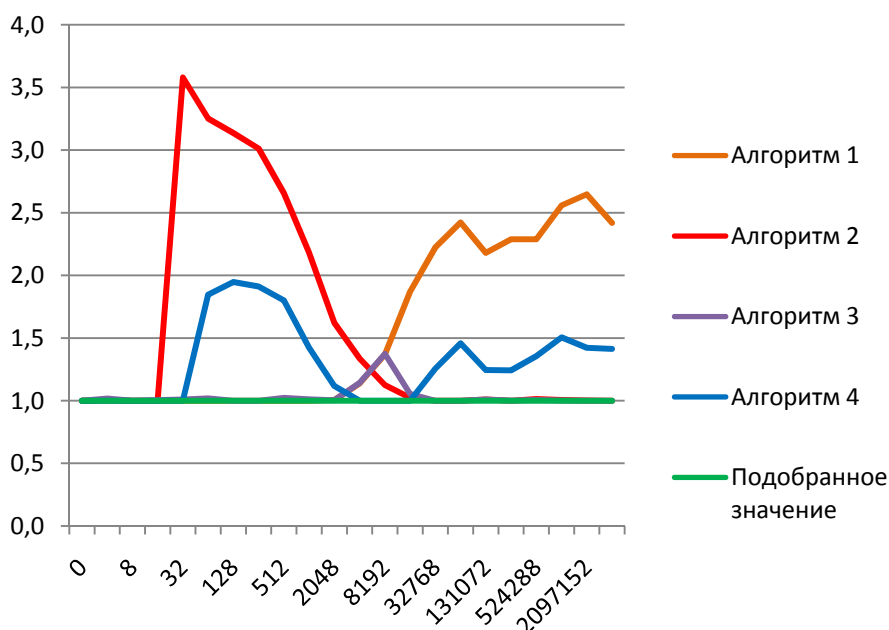


Рис. 1. Оптимизация MPI-операции Allreduce. N1:P8

Режим настройки на специфику приложения

Основной особенностью режима настройки на специфику приложения является задача определения необходимого набора параметров, подлежащих настройке, в условиях отсутствия сведений о том, как оно реализовано. MPIBoost не требует от пользователя каких-либо сведений об архитектуре приложения (набор используемых MPI-операций, размеры передаваемых сообщений и др.). В качестве информационной базы инструмент использует статистическую информацию, предоставляемую библиотекой S-MPI. Такая информация позволяет не только определить эффективность выполнения MPI-операций, но и исключить из процесса настройки параметры, которые не могут

повлиять на производительность. Настройка каждого параметра S-MPI производится с учетом возможности влияния его значения на другие параметры библиотеки.

Запуск MPIBoost с настройкой на тест IS (класс B) из набора NPB (NAS Parallel Benchmarks [7]) позволил получить настройки, улучшающие его производительность на 3–5%.

Процедура автоматической настройки параметров библиотеки S-MPI

Процедура автоматической настройки параметров библиотеки S-MPI является многоуровневой. Она позволяет достичь оптимального соотношения между затраченным на настройку параметров временем и достигнутой производительностью приложений:

- Настройка умалчиваемых значений параметров библиотеки S-MPI для наиболее распространенных архитектур кластеров и параметров запуска приложений. Этот уровень настройки позволяет библиотеке эффективно работать сразу после ее установки на большинстве кластерных систем;
- Настройка параметров библиотеки S-MPI на специфику конкретной кластерной системы. Предполагается, что эту настройку будет производить администратор кластера при установке пакета библиотеки S-MPI и при изменениях архитектуры кластера. Этот уровень настройки позволяет учесть специфические особенности конкретной кластерной системы и обеспечить высокую производительность большинства приложений, которые будут выполняться на данном кластере. Наибольший эффект этот уровень настройки может дать на кластерах со специфической архитектурой, настройки параметров для которой не были включены в библиотеку S-MPI как умалчиваемые значения;
- Настройка параметров библиотеки S-MPI на специфику конкретного приложения. Этот уровень настройки позволяет достичь максимальной производительности выполнения конкретного приложения на конкретной кластерной системе.

Литература

1. Воронов Г.И., Трущин В.Д., Шумилин В.В., Ежов Д.В. Программный комплекс S-MPI для обеспечения разработки, оптимизации и выполнения высокопараллельных приложений на суперкомпьютерных кластерных системах // Вопросы атомной науки и техники, сер. "Математическое моделирование физических процессов", 2013, №3, с.55-60.
2. Message Passing Interface Forum, "MPI: A Message Passing Interface" // In Proceedings of Supercomputing '93. IEEE Computer Society Press, November 1993. P. 878–883.
3. Библиотека Intel MPI. – [Электронный ресурс]: <http://software.intel.com/en-us/intel-mpi-library>.
4. Инструмент настройки OpenMPI (OTPO). – [Электронный ресурс]: <http://www.openmpi.org/projects/otpo/>.
5. Gabriel E., Fagg G.E., Bosilca G., et al. Open MPI: Goals, concept, and design of a next generation MPI implementation // Proceedings of the 11th European PVM/MPI Users' Group Meeting, Budapest, Hungary, September 2004. P. 97–104.
6. Тест на производительность Intel MPI Benchmarks. – [Электронный ресурс]: <http://software.intel.com/en-us/articles/intel-mpi-benchmarks>.
7. Тест на производительность NAS Parallel Benchmark. – [Электронный ресурс]: <http://www.nas.nasa.gov/publications/npb.html>.