

Эксперименты по исследованию характеристик нового поколения системы метакомпьютинга X-Com

С.С. Тихонов

В статье описаны эксперименты по исследованию характеристик нового поколения системы метакомпьютинга X-Com в условиях использования географически удаленных вычислительных узлов. Рассматриваются поведение системы метакомпьютинга X-Com в условиях реальных прикладных задач, аспекты использования промежуточных серверов, влияние различных факторов на долю накладных расходов, возникающих при работе системы X-Com.

1. Введение

В настоящее время в НИВЦ МГУ разрабатывается система метакомпьютинга X-Com второго поколения [1]. Для исследования характеристик нового поколения разрабатываемого инструментария был проведен ряд экспериментов, позволивших оценить работоспособность системы X-Com, влияние различных факторов на характеристики вычислительной системы, а также исследовать поведение системы в условиях использования географически удаленных вычислительных узлов.

2. Описание вычислительной среды

2.1 Использованное оборудование

В экспериментах были задействованы вычислительные узлы кластеров российских университетов, таких как СКИФ Урал (ЮУрГУ), СКИФ Cyberia (Томск), кластер СКИФ МГУ "Чебышев" (МГУ). Узлы использовались в монопольном режиме, запуск клиента системы X-Com производился вручную. В том случае, когда узлы кластера не имели прямого доступа в сеть Интернет, на головной машине кластера запускалась программа SimpleProxu, позволявшая перенаправить запросы от клиентов на основной сервер системы X-Com. Сервер располагался на запасной головной машине кластера СКИФ МГУ "Чебышев". В таблице 1 показаны основные характеристики использованных вычислительных узлов.

Таблица 1. Характеристики вычислительных узлов

Организация	Город	Процессоры	Число узлов	Число ядер
НИВЦ МГУ	Москва	Xeon 5365 3 GHz	202	808
ЮУрГУ	Челябинск	Xeon E5472 3 GHz	208	832
УГАТУ	Уфа	Xeon 5345 2.333 GHz	90	360
СПбГТУ	Санкт-Петербург	Opteron 280 2.4 GHz	32	64
ССКЦ ИВМиМГ СО РАН	Новосибирск	Itanium 2 1.6 GHz	36	36
ТГУ	Томск	Xeon 5150 2.667 GHz	50	100

2.2 Система метакомпьютинга X-Com

Проект X-Com [2] нацелен на создание прозрачной, универсальной распределенной среды, в которой можно будет решать самые различные прикладные задачи. Система X-Com — это инструментарий, который позволяет после незначительной переработки существующей прикладной программы (возможно, уже параллельной) запускать ее распределенно, контролировать ход выполнения, балансировать нагрузку между узлами и отображать результат. Важным аспектом системы X-Com является ее работа через сеть Интернет, что позволяет

потенциально использовать все ресурсы, подключенные к ней. Другой важной особенностью является возможность использовать узлы самой различной конфигурации - от кластеров до обычных рабочих станций, работающих под управлением различных программно-аппаратных платформ. Система X-Com состоит из двух основных компонентов: клиента X-Com, написанного на языке Perl и сервера X-Com на Java. Сервер X-Com обеспечивает запуск прикладной задачи пользователя, координирование работы клиентов, выдачу клиентской части прикладной задачи клиентам и обмен данными с ними. Клиент X-Com соединяется с сервером X-Com, получает клиентскую часть прикладной задачи пользователя и обеспечивает запуск клиентской части, получение порций данных и отправку результатов работы серверу X-Com.

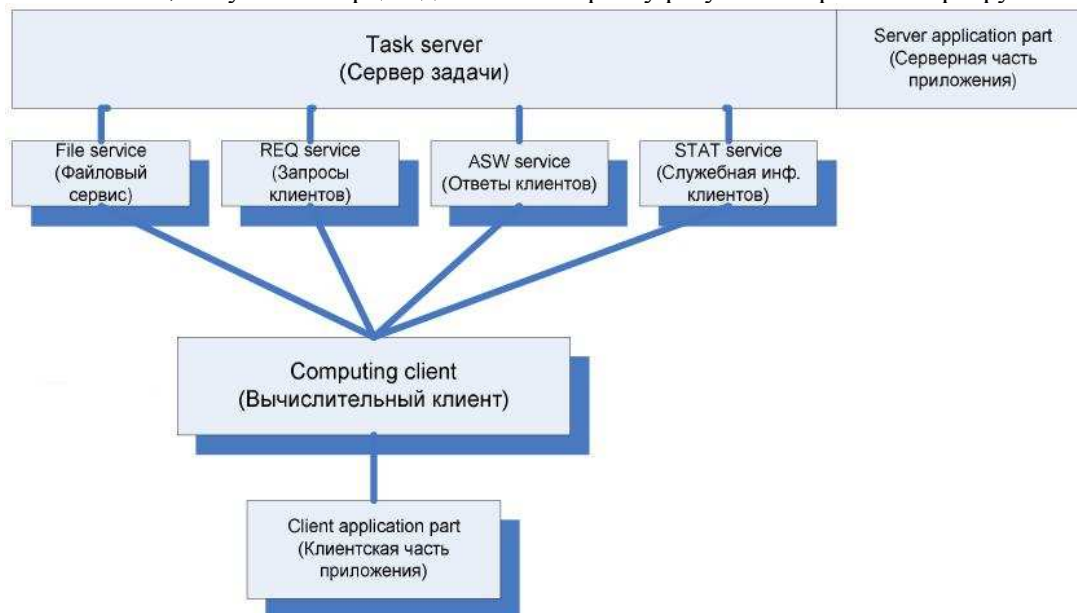


Рис 1. Архитектура системы X Com

Второе поколение системы метакомпьютинга X-Com разрабатывается с учетом достоинств и недостатков первой версии. Полностью сохранен интерфейс прикладной задачи пользователя и основная схема работы сервера. При этом основой нового поколения является язык Perl и стандартные модули для этого языка, обеспечивающие корректную передачу данных между клиентом и сервером, работу в несколько потоков, обработку ошибок. При этом использование языка Perl позволило сделать систему более переносимой и легкой в настройке. Значительно изменилась внутренняя архитектура сервера X-Com, что позволит увеличить производительность сервера за счет распределения различных функций сервера по разным вычислительным узлам.

3. Эксперименты с тестовой задачей

Задача «Test» предназначена для симуляции характеристик реальных вычислительных задач. Для этой задачи можно задать минимальное и максимальное время счета одной порции данных на клиенте, способ изменения этого времени. Также можно задать минимальные и максимальные размеры передаваемых порций данных и результатов, объем используемой оперативной памяти. Клиентская часть задачи реализует перемножение двух матриц различными способами.

Эксперименты с тестовой задачей были предназначены для определения работоспособности новой версии системы X-Com. Всего было проведено 2 расчета, в которых для задачи «Test» были заданы следующие параметры: Размер исходящих порций и результатов 100 байт, время счета одной порции 2 минуты. Клиентская часть не производила никаких вычислений. На рис. 2 показано количество запросов порций данных в минуту по разным кластерам. Верхний график соответствует кластеру СКИФ МГУ "Чебышев", средний соответствует кластеру СКИФ Урал (ЮУрГУ). Для того чтобы избежать перегрузки сервера запросами на получение клиентской части прикладной задачи, клиенты запускались с

достаточно большим интервалом, что отражено на диаграмме наклонной частью графика интенсивности запросов. Периодические пики с интервалом в 2 минуты на всех графиках коррелируют со временем счета одной порции данных.

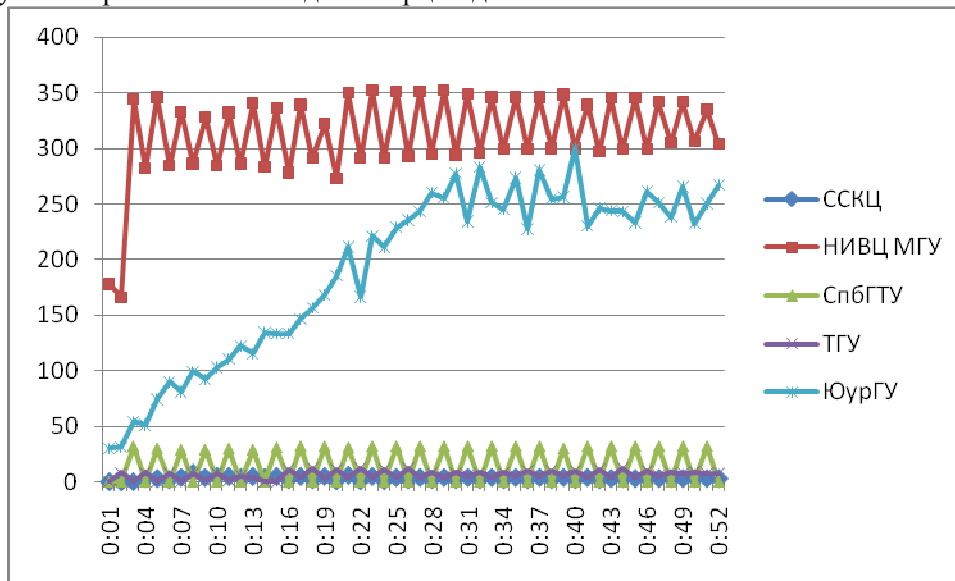


Рис 2. Эксперимент 1. Интенсивность запросов порций

На рис. 3 показаны те же графики для второго расчета, когда клиенты запускались с минимальным интервалом. Как видно из графиков, в случае одновременного запуска клиентов, общая производительность вычислительной среды оказывается ниже, так как из-за большого числа одновременных запросов к серверу накладные расходы на обработку запросов сильно возрастают. За 52 минуты второго эксперимента тем же количеством клиентов было посчитано всего лишь 16000 порций против 25000 за 52 минуты в первом эксперименте.

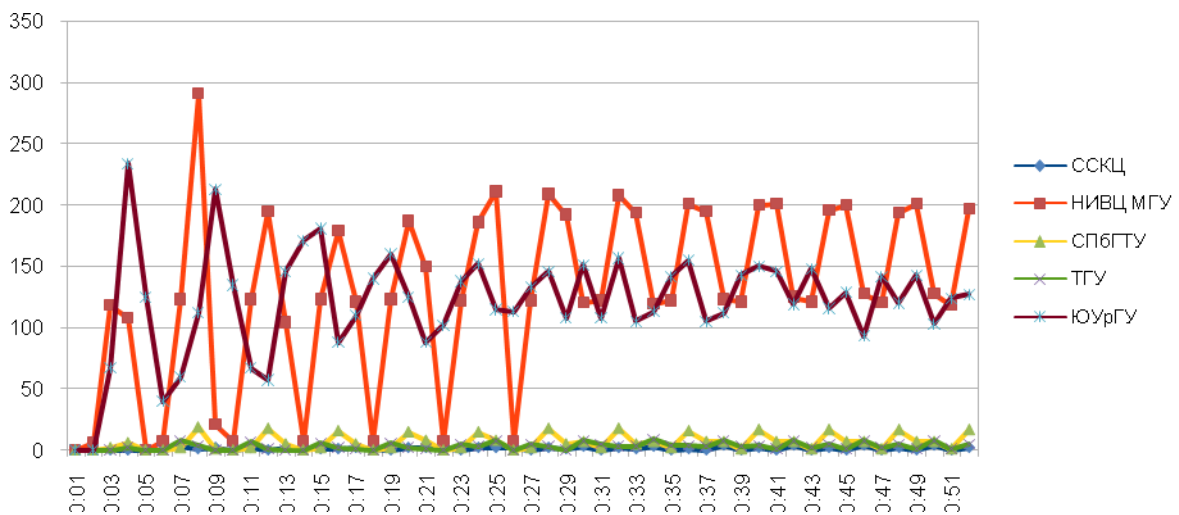


Рис 3. Эксперимент 2. Интенсивность запросов порций

4. Использование промежуточных серверов

Использование промежуточных серверов при проведении расчетов с использованием системы X-Com позволяет сократить объем передаваемых данных за счет уменьшения количества запросов к основному серверу на получение клиентской части прикладной задачи, а также снизить общее число запросов к серверу. Было проведено два расчета с использованием

задачи «Test» и промежуточных серверов, запущенных на отдельной машине на каждом кластере. В рамках каждого из этих расчетов было посчитано по 16000 порций данных. Клиенты X-Com подключались к своему промежуточному серверу вместо основного. Параметры задачи были те же, 100 байт порции данных и 2 мин. время счета одной порции. Промежуточные серверы были настроены в первом расчете на передачу 10 порций в одном запросе, во втором расчете на передачу 100 порций. На рис. 4 показана интенсивность запросов порций данных по каждому кластеру. Использование промежуточных серверов позволило существенно сократить общее время расчета (38 минут против 52). Однако на кластере СКИФ Урал (ЮУрГУ) применение промежуточного сервера только снизило общую производительность клиентских узлов, из-за неэффективной работы промежуточного сервера в условиях плохого канала связи с основным сервером.

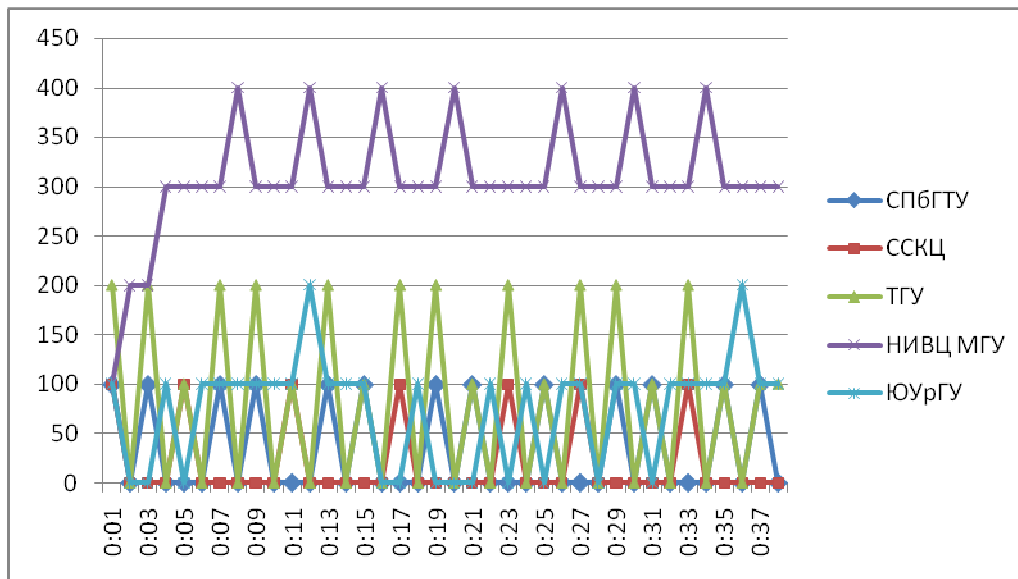


Рис 4. Использование промежуточных серверов. Интенсивность запросов порций

5. Вычислительные эксперименты

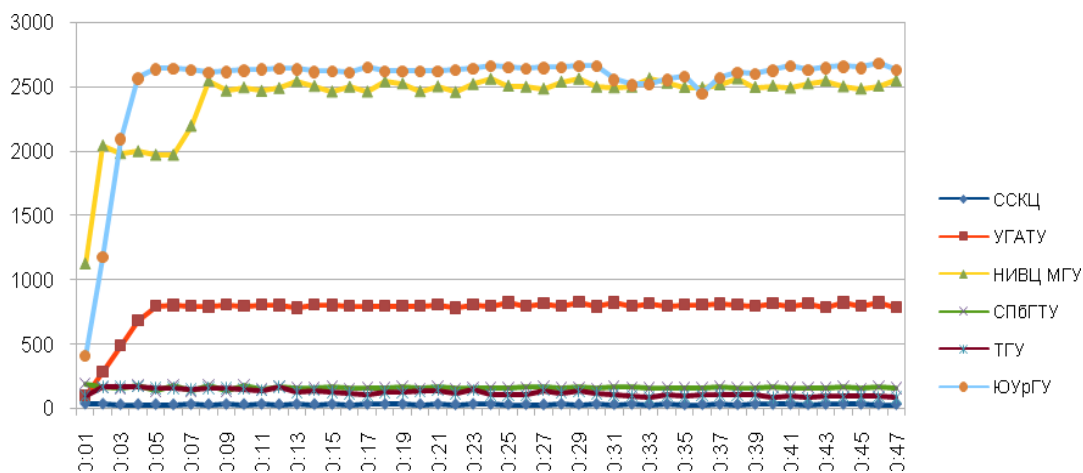


Рис 5. Расчет влияния магнитного поля. Интенсивность запроса порций

Было проведено два вычислительных эксперимента с реальными прикладными задачами. В первом эксперименте производился расчет элементов матрицы большой размерности для решения СЛАУ. Каждый элемент рассчитывался меньше секунды, поэтому в одну порцию

данных объединялись порядка тысячи элементов. В расчете принимали участие все доступные узлы, также для уменьшения времени расчета на кластере СКИФ МГУ "Чебышев" подключались дополнительные клиенты. Всего расчет занял 12 часов. На рис. 5 показаны графики количества запросов порций в минуту по каждому кластеру. Наклонные линии в начале графика отражают последовательный запуск клиентов с паузами между запуском клиентов на разных узлах.

Во втором эксперименте производился расчет взаимодействия молекул-лигандов с белком. Для расчета каждой порции требовалось от 16 секунд до 24 часов (в среднем 4,5 часа), размер входящих порций не превышал 30 КБ, размер результатов не превышал 300 КБ. Было рассчитано 7600 порций, объем полученных результатов составил 340 МБ. Суммарно в расчете приняли участие 1660 клиентов. Время расчета составило порядка 38 часов.

6. Эксперименты по измерению характеристик среды на основе системы метакомпьютинга X-Com

Было проведено две серии экспериментов по определению времени накладных расходов на обработку одной порции данных и пропускная способность сервера, т.е. объем данных, обрабатываемых сервером за единицу времени. Также в этих экспериментах исследовалась зависимость данных характеристик от интенсивности запросов к серверу. Время накладных расходов оценивалось при помощи задачи «Test» с минимальным объемом порций данных; пропускная способность оценивалась для типичных значений объемов порций данных, от 1 КБ до 100 КБ. Время счета одной порции уменьшалось со 100 сек. до 10 сек, что соответствует десятикратному увеличению интенсивности запросов в конце расчета относительно начала. Общее количество клиентов было подобрано таким образом, чтобы сервер не был перегружен запросами от клиентов сразу после старта расчета. Всего в экспериментах участвовало 220 клиентских узлов. В процессе экспериментов было рассчитано 38000 порций данных.

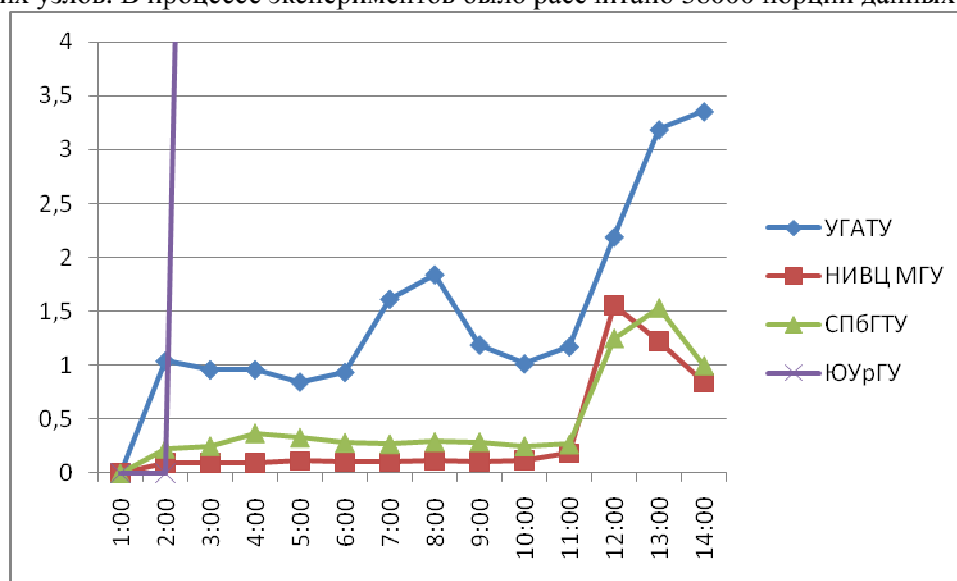


Рис 6. Определение зависимости эффективности сервера от интенсивности запросов. Время накладных расходов

На рис. 6 показана зависимость времени накладных расходов сервера на прием и обработку результата счета одной порции. Узлы кластеров СКИФ Урал (ЮУрГУ) и УГАТУ были подключены через слабый канал, обладающий малой пропускной способностью и большой латентностью. К концу расчетов средняя интенсивность запросов к серверу составляла порядка 400 запросов в секунду, из-за чего возник «провал» в производительности сервера, т.е. резкое увеличение времени накладных расходов, что и видно на рис. 6.

Варьирование размеров передаваемых порций данных показало, что изменение размеров порций с 1 до 100 КБ дает увеличение времени накладных расходов на 10%, что говорит о том, что основная часть накладных расходов связана с обработкой самого запроса, а не с обработкой больших объемов порций данных.

7. Заключение

Проведенные эксперименты показали готовность нового поколения системы метакомпьютинга X-Com к использованию в реальных расчетах. В ходе экспериментов были выявлены такие параметры вычислительной среды и прикладной задачи, при которых сервер перестает справляться с нагрузкой, были найдены узкие места сервера X-Com: падение производительности сервера при большом количестве основных запросов к серверу, неэффективная работа промежуточных серверов в условиях узкого канала связи с основным сервером. Их устранение станет важной задачей в ходе дальнейшей разработки. Эксперименты также показали эффективность использования промежуточных серверов в вычислительных средах с большим количеством вычислительных узлов.

Хочется выразить благодарность коллективу организаций-членов суперкомпьютерной Программы «СКИФ-ГРИД» Союзного государства за предоставленные вычислительные мощности и помощь в организации экспериментов.

Литература

1. Система метакомпьютинга X-Com [<http://x-com.parallel.ru/new/>]
2. Филимофитский М.П. Система поддержки метакомпьютерных расчетов X-Com: архитектура и технология работы // Вычислительные методы и программирование Т. 5, Раздел 2, с. 1-9, 2004 г.