

Моделирование системы распределённой обработки данных

А. С. Родионов¹, С. В. Рудометов², В. В. Окольнішников²

¹*Институт вычислительной математики и математической геофизики СО РАН
Новосибирск, Россия*

²*Конструкторско-технологический институт вычислительной техники СО РАН
Новосибирск, Россия
e-mail: alrod@sscc.ru*

Описана адаптивная имитационная модель системы распределённой обработки данных (СРОД), предназначенная для анализа производительности СРОД и оказываемой ею нагрузки на вычислительные сети. Модель реализована в визуально-интерактивной среде имитационного моделирования MTSS.

Ключевые слова: имитационное моделирование, распределённая обработка данных.

Введение

Целью работы являлось построение гибкой имитационной модели системы распределённой обработки данных (СРОД), позволяющей с минимальными затратами переконфигурировать её для моделирования различных конфигураций сети и параметров как потоков данных, так и средств их обработки.

Рассматриваются СРОД, предназначенные для выделения значимой информации из большого потока данных. Примером таких систем может быть метеорологический мониторинг с целью предсказания погодных аномалий. Для решения таких задач применяются распределённые вычислительные сети с большим количеством узлов, каждый из которых представляет собой сложный вычислительный центр с многоядерными вычислителями или даже суперкомпьютерами.

Основные требования к имитационной модели — это эффективность моделирования, наглядность представления модели, хода моделирования и результатов, а также освобождение потенциального пользователя от участия в разработке и сопровождении программы модели.

Подобные модели удобно создавать, используя системы с проблемно-ориентированным графическим интерфейсом. Существует достаточно много систем имитационного моделирования информационно-вычислительных и телекоммуникационных сетей [1, 2], однако они не позволяют отображать все особенности функционирования СРОД, в частности использование различных программ обработки данных.

Авторами использовалась система моделирования MTSS [3], позволяющая создавать проблемно-ориентированные интерфейсы требуемого качества и обеспечивающая выполнение перечисленных выше требований к имитационной модели.

1. Описание СРОД

Моделируемая СРОД представляет собой четырёхуровневую распределённую вычислительную сеть сбора и обработки больших объёмов данных мониторинга окружающей среды, состоящую из узлов сбора и обработки информации (рис. 1).

Узлы сбора информации представляют собой разнородные устройства с различными протоколами работы и различной скоростью передачи данных (спутник, метеостанция). Каждый узел подключён к одному, нижнему, вычислительному узлу в иерархии СРОД. Все узлы в СРОД имеют полную связность. Информация между ними передаётся с различными задержками. Узлы первого уровня принимают данные из различных источников мониторинга. Закон распределения входного потока каждого источника близок к нормальному с высокой интенсивностью.

Обработка входных данных в каждом узле СРОД осуществляется различными программами. Результаты обработки сохраняются на текущем узле обработки, а также передаются на узлы верхнего уровня СРОД для дальнейшей обработки. Данные могут “спускаться” на нижние уровни для повторной обработки. Конечным результатом обработки является сигнал возможной опасности, передаваемый на один из узлов верхнего уровня.

При проектировании СРОД должны решаться следующие задачи: минимизация вероятности потери сигналов об опасной ситуации; минимизация времени обнаружения и обработки сигналов об опасной ситуации; минимизация ресурсов вычислительных узлов (память, вычислительные каналы); оценка эффективности используемых программ обработки данных. В соответствии с этими задачами и создавалась рассматриваемая имитационная модель.

2. Описание имитационной модели СРОД

Имитационная модель СРОД реализована с использованием механизма графической сборки в графической среде системы имитационного моделирования MTSS [4, 5], основанной на понятии *элементарных моделей* (ЭМ) этой системы.

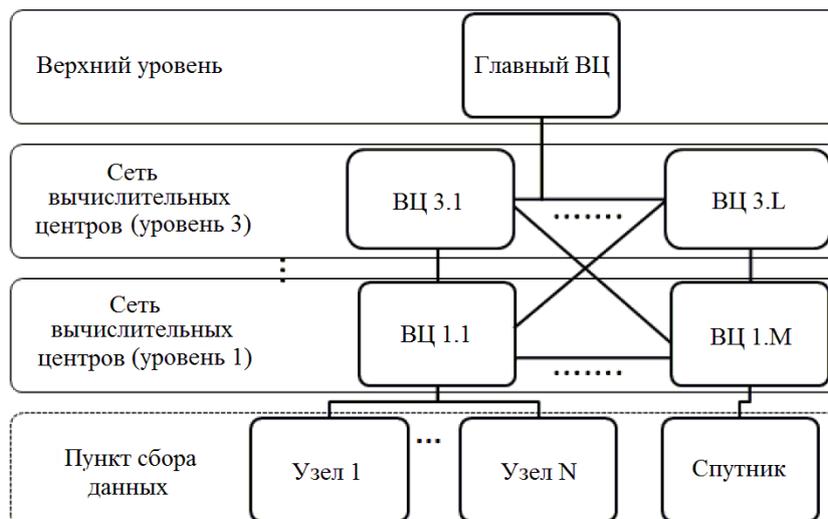


Рис. 1. Архитектура СРОД

Система MTSS выбрана прежде всего из-за своей идеологии, которая требует полностью скрывать все аспекты имитационного моделирования от пользователя имитационной модели, позволяя ему, однако, неявно использовать методологию имитационного моделирования без привлечения специалистов в имитационном моделировании. Также важным аспектом является её настраиваемость в части применения различных подходов имитационного моделирования, реализованных в MTSS как “машины” имитационного моделирования.

Подход MTSS заключается в разработке специализированной для некоторой предметной области библиотеки ЭМ. Такой подход успешно применён для разных предметных областей [6–8]. Имитационная модель в MTSS строится графически из ЭМ, определенных на этапе декомпозиции, путём расположения на экране нужного количества ЭМ и соединения их между собой. Модель допускает произвольную модификацию набора ЭМ и связей между ними.

Декомпозиция исходной системы. Для моделирования СРОД, состоящей из N узлов обработки (вычислительных центров), соединённых между собой линиями связи, реализована библиотека, включающая такие библиотечные ЭМ, как “Вычислительный центр”, “Генератор данных”, “Линия связи”.

Элементарная модель “Вычислительный центр” (ЭМ ВЦ) имитирует работу вычислительного центра по сбору, хранению и обработке информации, которая поступает в неё по входящим линиям связи. Сама ЭМ ВЦ также может посылать информацию по исходящим линиям связи.

Элементарная модель “Генератор данных” (ЭМ ГД) требуется для имитации появления пакетов данных в системе. Пакеты данных появляются по заданному распределению (уникальному для каждого экземпляра ЭМ ГД).

Элементарная модель “Линия связи” (ЭМ ЛС) служит для задания топологии в модели СРОД. Наличие этой ЭМ между любыми двумя вычислительными узлами или между генератором данных и вычислительным узлом означает возможность посылки данных между такими объектами.

Графическая часть этих ЭМ служит для отображения состояния каждого экземпляра ЭМ. Также для каждого экземпляра ЭМ задаются наборы индикаторов и вычисляемых значений, которые могут собираться системой MTSS как статистические данные для последующей обработки и отображаться в модели как графические индикаторы.

“Продуктом” в такой имитационной модели служит пакет данных. Пакет данных имеет тип, метку времени генерации, размер. Имитируется следующее поведение пакета (в процессе его прохождения по имитационной модели): порождение нового пакета данных реализуется в ЭМ ГД, передача пакета по линии связи реализуется в ЭМ ЛС, возможная задержка определяется типом данных и их размером. Обработка пакета реализуется в ЭМ ВЦ. При этом возможен как “вывод” пакета из системы (окончание его жизни), так и порождение нового пакета (после обработки исходного). Пакеты могут копироваться при передаче по нескольким линиям связи.

Элементарная модель ВЦ является подробной моделью стандартного вычислительного центра.

Пакет данных, поступивший в ЭМ ВЦ, проходит следующие этапы обработки: хранение в подсистеме хранения данных, определение вида обработки пакета (выбор программы обработки данных), постановка в очередь на вычисление, вычисление, посылка результатов вычисления по выходным линиям связи.

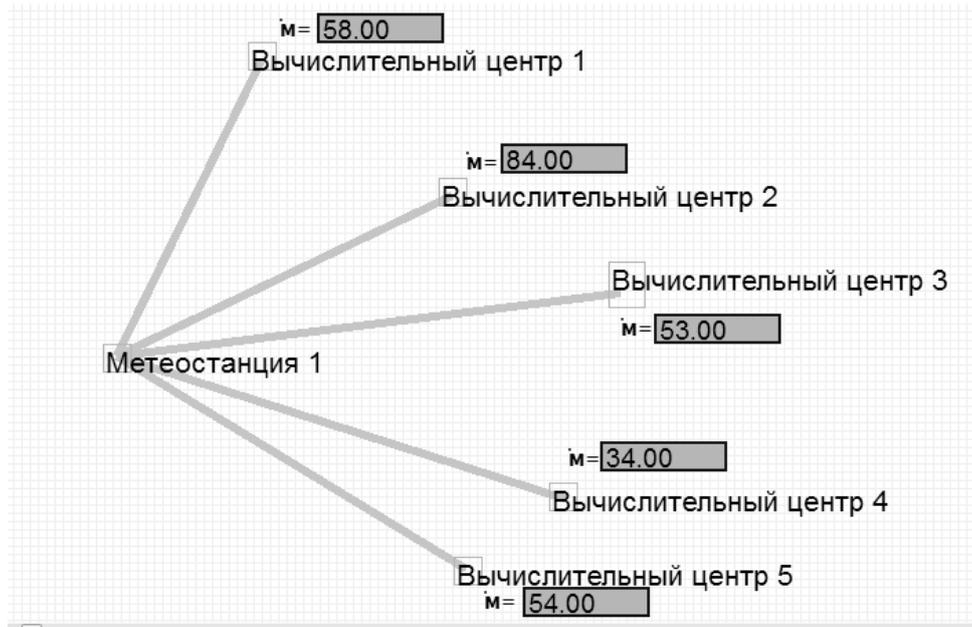


Рис. 2. Фрагмент имитационной модели СРОД в системе MTSS

ЭМ ГД позволяет имитировать входящий поток данных в виде пакетов данных для их последующей обработки. Пакеты данных имеют нормальное распределение по времени генерации и равномерное — по размеру данных.

ЭМ ЛС позволяет визуально соединить два объекта в модели: ЭМ ГД и ЭМ ВЦ или ЭМ ВЦ и ЭМ ВЦ. Элементарная модель ЛС имитирует однонаправленную связь.

С использованием построенной имитационной модели были решены задачи: контроль загрузки всей вычислительной сети и каждого узла в отдельности, контроль количества обработанных пакетов данных, имитация отказа части сети, модификация линий связи и контроль скорости обработки информации после такой модификации, количество и качество применяемых программ обработки данных на каждом из узлов, контроль расхода памяти в подсистеме хранения данных каждого узла.

Фрагмент имитационной модели СРОД, реализованной в системе MTSS, представлен на рис. 2.

В процессе работы имитационной модели выполняются анализ и вывод статистических данных. Эти статистические данные используются для быстрого определения глобальных параметров модели, таких как средняя загрузка вычислительных узлов, количество обработанных пакетов данных, расход электроэнергии и других ресурсов и т. д.

3. Валидация имитационной модели

Валидация имитационной модели СРОД осуществлялась путём сравнения с результатами аналитического решения задачи для упрощённой сети распределённой обработки данных.

Параметры СРОД для аналитического решения:

- одноуровневая сеть;
- количество вычислительных центров 5;
- количество устройств сбора данных 1;

- заявки с устройства сбора поступают на каждый из ВЦ по мере своего появления. Выбор узла для посылки очередной заявки определяется посредством вектора вероятностей, заданного в устройстве сбора для каждого ВЦ;
- задачи на каждом ВЦ начинают исполнение сразу (отсутствует очередь исполнения);
- вычислительные центры связаны друг с другом посредством матрицы переходов, в которой записана вероятность отправки заявок после их обработки между любыми двумя ВЦ;
- поток данных от устройства сбора представляет собой гамма-распределение с $D = 400$ и $M = 100$ (высокая интенсивность).

В задаче требуется получить вектор *максимального количества заявок* на каждом из ВЦ. Также требуется получить вектор средних и ковариационную матрицу количества заявок.

Аналитическое решение выполнено другим коллективом авторов [9] и является самостоятельным научным результатом. Для валидации потребовалось существенное упрощение элементарных моделей. Результаты, полученные с помощью упрощённой модели, совпадают с точностью 95 % с указанным аналитическим решением на данных из [9] (сеть 5 узлов).

Альтернативный вариант имитационной модели, реализованной с помощью среды моделирования AGNES [10, 11], показал сходные (в рамках статистической погрешности) результаты, однако при несколько меньшем времени счёта. Это объясняется тем, что AGNES не предоставляет присущих MTSS интерфейсных возможностей и модели AGNES исполняются параллельно на нескольких узлах вычислительной сети.

Таким образом, проведённое сравнение с альтернативными вариантами моделей свидетельствует о корректности представленной имитационной модели в среде MTSS.

Заключение

Подход и архитектурно-программные решения, реализованные в системе MTSS, позволяют моделировать широкий класс сложных вычислительных сетей обработки данных. Библиотека ЭМ СРОД в системе MTSS может быть переиспользована или использована в качестве прототипа для создания сложных имитационных моделей распределённых систем обработки данных.

Список литературы

- [1] КАРПОВ Ю.Г. Имитационное моделирование систем. Введение в моделирование с AnyLogic 5. СПб.: БХВ-Петербург, 2006. 400 с.
- [2] MODELING and Tools for Network Simulation / K. Wehrle, M. Günes, J. Gross (Eds). Springer, 2010. 256 p.
- [3] РУДОМЕТОВ С.В. Система имитационного моделирования MTSS. Новосибирск: ФАП СО РАН, 2011, URL: <http://fap.sbras.ru/node/2325>
- [4] РУДОМЕТОВ С.В. Создание системы имитационного моделирования технологических систем // Имитационное моделирование. Теория и практика (ИММОД-2011): Тр. Пятой Всерос. науч.-практ. конф. СПб., 2011. Т. 1. С. 383–387.

- [5] РУДОМЕТОВ С.В. Визуально-интерактивная система имитационного моделирования технологических систем // Вест. СибГУТИ. 2011. № 3. С. 14–27.
- [6] АНДРЮШКЕВИЧ С.К., ЖУРАВЛЕВ С.С., ОКОЛЬНИШНИКОВ В.В. и др. Разработка системы мониторинга с использованием имитационного моделирования // Пробл. информатики. 2010. № 4. С.65–75.
- [7] OKOLNISHNIKOV V., RUDOMETOV S., ZHURAVLEV S. Using simulation for development of process control system in coal mining // Proc. 3th Intern. Conf. on Appl. Inform. and Comp. (AICT'12), Barcelona, Spain, Oct. 17–19, 2012. P. 141–144.
- [8] OKOLNISHNIKOV V., RUDOMETOV S., ZHURAVLEV S. Simulation environment for development of automated process control system in coal mining // Intern. J. of Systems Appl., Eng. and Development. 2013. Issue 5, Vol. 7. P. 255–262.
- [9] ГРАЧЁВ В.В., МОИСЕЕВ А.Н., НАЗАРОВ А.А., ЯМПОЛЬСКИЙ В.З. Многофазная модель массового обслуживания системы распределённой обработки данных // Докл. ТУСУРа. 2012. № 2 (26), ч. 2. С. 248–251.
- [10] ГЛИНСКИЙ Б.М., РОДИОНОВ А.С., МАРЧЕНКО М.А. и др. Агентно-ориентированный подход к имитационному моделированию суперЭВМ экзафлопсной производительности в приложении к распределённому статистическому моделированию // Вест. ЮУрГУ. 2012. № 18 (277), вып. 12. С. 93–106.
- [11] ПОДКОРЫТОВ D., RODIONOV A., CHOO H. Agent-based simulation system AGNES (AGent NEtwork Simulator) for networks modeling // Proc. 6th Intern. Conf. on Ubiquitous Information Management and Communication, ICUIMC'12, 2012. Article No 115 ACM New York, NY, USA, 2012.

Поступила в редакцию 29 ноября 2013 г.