

Потоковая обработка больших массивов экспериментальных данных на удаленном суперкомпьютере*

А.Г. МАСИЧ

Институт механики сплошных сред УрО РАН

e-mail: mag@icmm.ru

Г.Ф. МАСИЧ

e-mail: masich@icmm.ru

В.А. ЩАПОВ

e-mail: schapov@icmm.ru

Р.А. СТЕПАНОВ

e-mail: rodion@icmm.ru

Рассмотрены разработанные в ИМСС УрО РАН информационная модель прямого ввода в вычислительные узлы удаленной суперЭВМ интенсивного потока экспериментальных данных и созданные на ее основе протокол и программное обеспечение. Приводятся результаты апробации этих решений в рамках проекта “Распределенный” PIV на участке Пермь-Екатеринбург, с использованием канала связи, построенного по технологии GE.

1. Общие сведения об эксперименте

Экспериментальная установка (ЭУ) использует метод PIV (Particle Image Velocimetry) [1] – оптический метод измерения полей скорости жидкости или газа в выбранном сечении потока. Принцип метода: импульсный лазер создает тонкий световой нож и освещает мелкие частицы, взвешенные в исследуемом потоке. Положения частиц в момент двух последовательных вспышек лазера регистрируются на два кадра цифровой камеры. Скорость вихревого потока определяется расчетом перемещения, которое совершают частицы за время между вспышками лазера. Измерительная часть установки генерирует поток данных 1-10-100 Гбит/с в зависимости от разрешения, частоты и режимов работ камер (моно/стерео/томография).

Ограниченность доступной вычислительной мощности в местах проведения экспериментов сдерживает развитие математического аппарата и возможности проведения экспериментов. Перенос вычислений на удаленные суперкомпьютеры позволит использовать ресурсоемкие, но высокоточные алгоритмы, избегать хранения гигантских объемов избыточной информации, обрабатывать измерения “на лету” и проводить эксперименты с обратной связью. Эта идея реализуется в рамках проекта “Распределенный PIV” [2].

*Работа выполнена при поддержке РФФИ (грант №11-07-96001-р_урал_a) и Региональной целевой программы УрО РАН (№РЦП-11-П10).

2. Разработанные архитектурные решения

2.1. Сетевая инфраструктура проекта

Проект “Распределенный PIV” [3] использует высокоскоростную сеть передачи данных, создающуюся в рамках проекта GIGA UrBRAS [4]. Схема используемого в настоящем проекте участка сети приведена на рисунке 1.

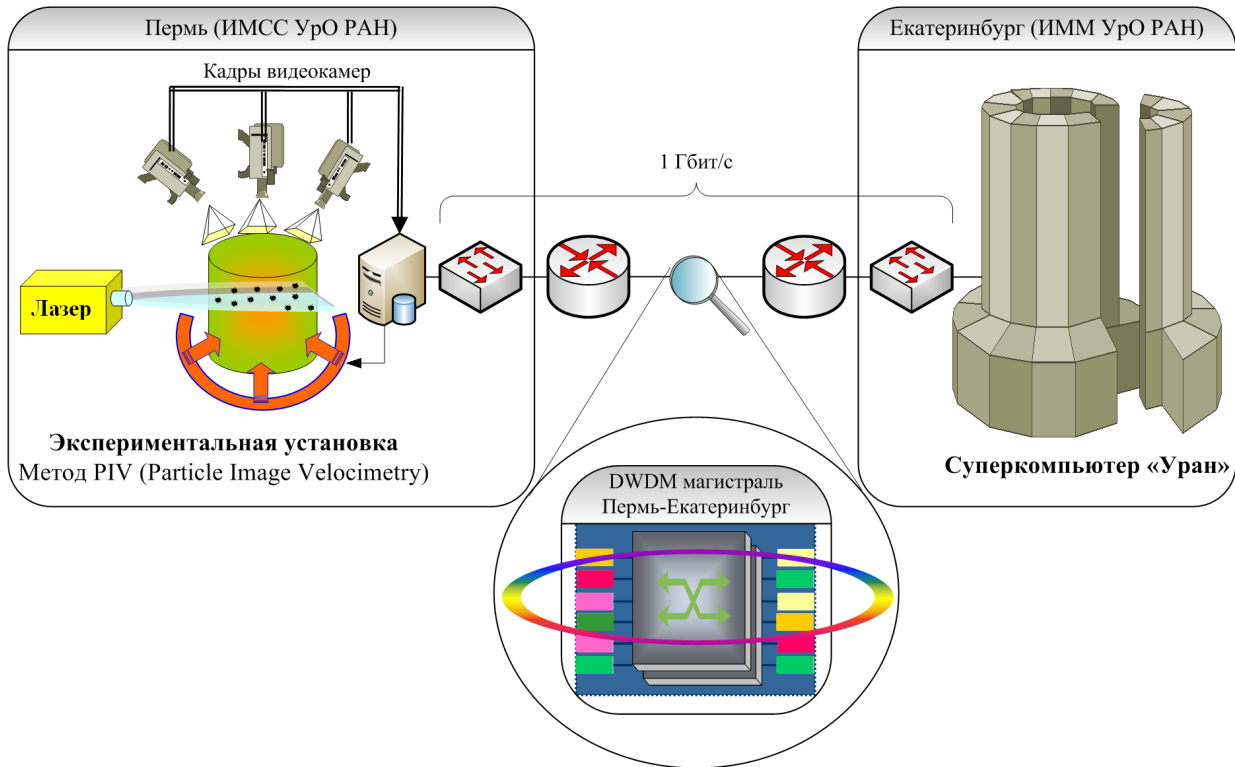


Рис. 1. Схема сети проекта “Распределенный PIV”

2.2. Модель прямого доступа к вычислительным узлам

Информационная модель прямого доступа к вычислительным узлам была разработана в ИМСС УрО РАН и представлена в публикации [4]. Суть данной модели заключается в передаче данных между ЭУ и вычислительными узлами суперкомпьютера минуя сохранение данных на дисках, с промежуточным хранением только в оперативной памяти (рисунок 2).

Предложенная информационная модель базируется на LambdaGrid технологиях, суть которых заключается в объединении систем высокоскоростными оптоволоконными каналами связи.

2.3. Протокол PIV

Особенностями разработанного программного обеспечения являются: клиент-серверная архитектура, возможность управления числом задействованных вычислительных узлов, передача данных непосредственно между памятью территориально распределен-

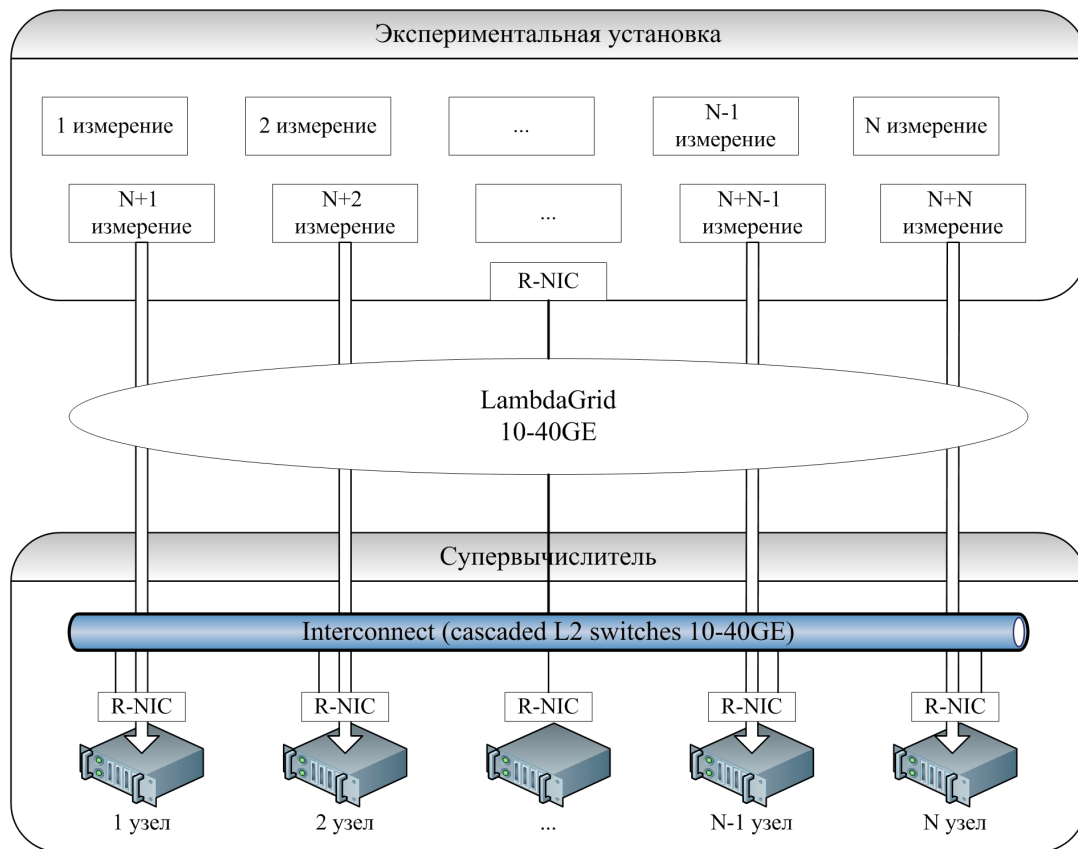


Рис. 2. Модель обработки интенсивных потоков данных

ных систем без промежуточного хранения на дисках, возможность управления экспериментом в квазиреальном времени на основе получаемых результатов.

Разработанный протокол PIV базируется на идее отказа от однозначного отображения измерений на вычислительные узлы [5]. В случае наличия единого источника данных, который соответствует экспериментальной установке, отказ от однозначного отображения позволит:

- отказаться от синхронизаций вычислительных узлов перед обработкой каждого измерения;
- изменять число задействованных узлов непосредственно во время эксперимента, добавляя в случае необходимости вычислительные мощности;
- минимизировать объем потерянной информации в случае выхода из строя одного или нескольких вычислительных узлов.

Протокол PIV является протоколом прикладного уровня, работает по схеме запрос-ответ и базируется на протоколе TCP. Текущая архитектура позволяет использовать вместо протокола TCP любой протокол, работающий в потоковом режиме и гарантирующий доставку данных.

Протокол PIV рассчитан на передачу нескольких блоков бинарных данных. В одном пакете протокола можно передать от нуля до 65535 блоков, каждый из которых может иметь размер до 4 Гбайт. Поля заголовка пакета протокола кодируются в сетевом порядке байт. Формат пакета протокола PIV приведен на рисунке 3.

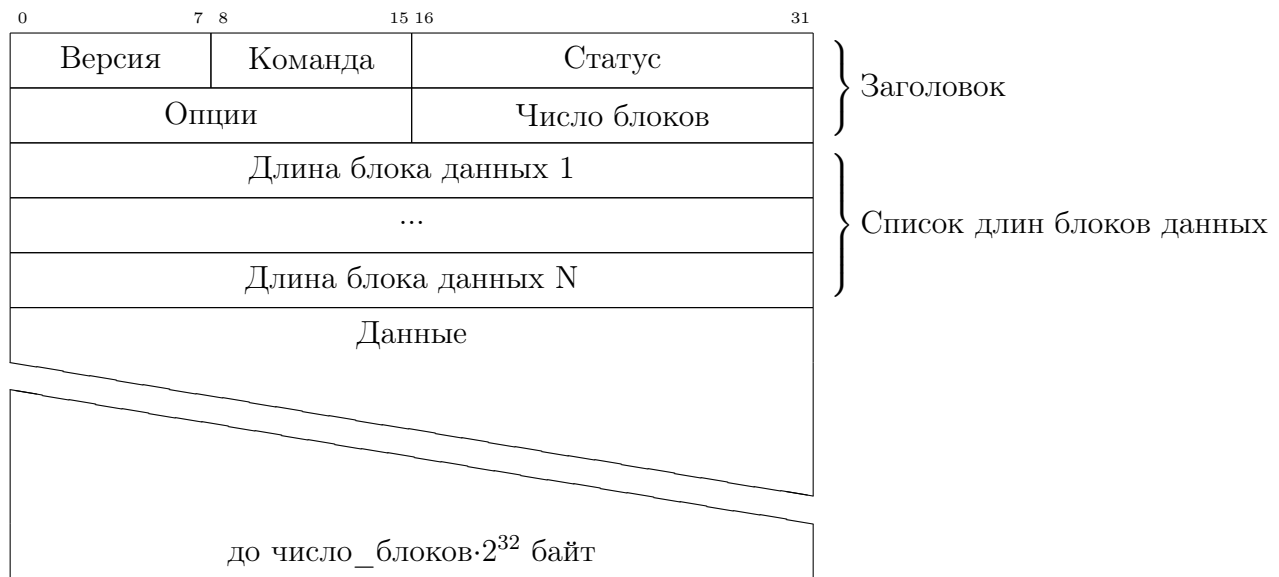


Рис. 3. Формат пакета протокола PIV

Протокол поддерживает три типа пакетов (поле “Команда” протокола PIV):

- GET – предназначен для запроса данных вычислительными узлами у ЭУ;
- POST – предназначен для передачи данных с вычислительных узлов на ЭУ;
- RESPONSE – пакет ответа на запрос вычислительного узла.

Информация о подтипе RESPONSE-пакета кодируется в поле статуса. Через поле статуса вычислительным узлам сообщается успешно ли был обработан запрос или нет. В случае неуспеха конкретное значение поля статуса определяет произошедшую ошибку.

3. Анализ производительности разработанного решения

Исследования производительности разработанного решения проводились на экспериментальном фрагменте сетевой инфраструктуры, состоящей из экспериментальной установки PIV (ИМСС УрО РАН, Пермь) и суперкомпьютера “УРАН” 20 Tflops (ИММ УрО РАН, Екатеринбург), соединенных по технологии GE по двум оптическим волокнам длиной 410 км.

Прямой доступ к вычислительным узлам суперкомпьютера “УРАН” осуществлялся через фронтальную машину посредством технологии RAT. График зависимости суммарной скорости передачи данных между эмулятором ЭУ и узлами суперкомпьютера приведен на рисунке 4. В качестве эталонной пропускной способности канала использовались величины, полученные с помощью программного обеспечения iperf.

Из графика следует, что потери скорости, вносимые логикой работы протокола PIV и служебными данными не превышают 10%, когда число параллельных соединений превышает 10. Необходимо отметить, что реальный эксперимент потребует задействования десятков и сотен вычислительных узлов, что еще приблизит скорость передачи данных протокола PIV к предельно возможной в данном канале связи.

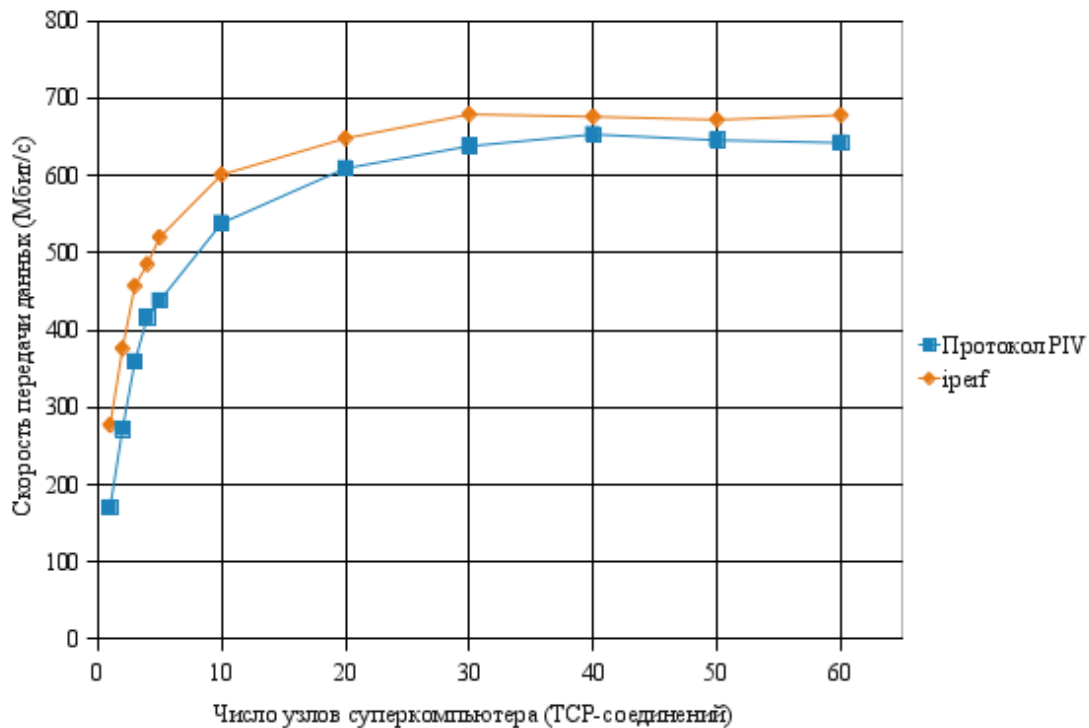


Рис. 4. Суммарная скорость передачи данных между ЭУ и узлами суперкомпьютера

4. Заключение

Спроектирован протокол и разработано программное обеспечение для передачи данных с экспериментальной установки на узлы вычислительного кластера.

Исследование производительности показало, что при использовании более 10 параллельных соединений скорость передачи данных по протоколу PIV превышает 90% от доступной скорости канала.

Дальнейшее развитие протокола будет связано с использованием транспортных протоколов, ориентированных на высокоскоростные сети, такие как UDT и высокоскоростные варианты TCP.

Список литературы

- [1] ADRIAN R.J. Scattering particle characteristics and their effect on pulsed laser measurements of fluid flow: speckle velocimetry vs. particle image velocimetry // Appl. Opt. 1984. Vol. 23. Pp. 1690-1691.
- [2] СТЕПАНОВ Р.А., МАСИЧ А.Г., МАСИЧ Г.Ф. Инициативный проект «Распределенный PIV» // Научный сервис в сети Интернет: масштабируемость, параллельность, эффективность: труды Всероссийской суперкомпьютерной конференции – М.: Изд-во МГУ, 2009. – С. 360–363. (ISBN 978-5-211-05697-8)
- [3] МАСИЧ А.Г., (и др) Инфраструктура распределенного эксперимента // Сб. тез. докл. XVI конференции представителей региональных научно-образовательных сетей «RELARN-2009». - М.-СПб, 2009. – С. 58–60.

- [4] Масич А.Г., Масич Г.Ф. GIGA UrB RAS подход к LambdaGrid парадигмам вычислений // Научный сервис в сети Интернет: суперкомпьютерные центры и задачи: труды Международной суперкомпьютерной конференции – М.: Изд-во МГУ, 2010.
- [5] Масич А.Г., Масич Г.Ф., Степанов Р.А., Щапов В.А. Скоростной I/O-канал супервычислителя и протокол обмена интенсивным потоком экспериментальных данных // Сб. тез. докл. X международной конференции высокопроизводительные параллельные вычисления на кластерных системах “НПС-2010” – Пермь: Изд-во ПГТУ, 2010. - Т. 2. С. 119–128. (ISBN 978-5-398-00506-6)