

ИССЛЕДОВАНИЕ И РАЗРАБОТКА ПРОЦЕССА ФОРМИРОВАНИЯ РЕГИОНАЛЬНОЙ КИБЕРИНФРА- СТРУКТУРЫ НА ОСНОВЕ λ -GRID-ТЕХНОЛОГИЙ*



Г.Ф. Масич,
*Институт механики сплошных
сред УрО РАН*



А.Г. Масич,
*Институт механики сплошных
сред УрО РАН*



Р.А. Степанов,
*Институт механики сплошных
сред УрО РАН*



А.Н. Сухановский,
*Институт механики сплошных
сред УрО РАН*



А.В. Бобров,
*Институт механики сплошных
сред УрО РАН*



Д.П. Чугунов,
*Институт механики сплошных
сред УрО РАН*



Г.Г. Тирон,
Институт экономики УрО РАН



Р.А. Файзрахманов,
*Пермский национальный
исследовательский
политехнический университет*



В.А. Шапов,
*Пермский национальный
исследовательский
политехнический университет*

Приведены сведения о создаваемой на Урале киберинфраструктуре, призванной использовать все усовершенствования в сфере высокопроизводи-

* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований и Пермского края в рамках гранта РФФИ № 11-07-96001-р_урал_a.

тельных вычислений и состоящей из объединенных скоростной оптической сетью вычислителей, систем хранения данных и установок. Показаны проблемы организации соединений между оконечными системами по скоростным протяженным линиям связи и пути их решения посредством параллелизма соединений. Особое внимание уделено архитектурному решению обработки интенсивного потока экспериментальных данных на удаленном суперкомпьютере в реальном времени, демонстрирующему принципиально новые возможности проведения экспериментальных исследований.

Ключевые слова: *киберинфраструктура, оптическая сеть, соединение, суперкомпьютер, экспериментальная установка.*

Понятие «киберинфраструктура» в начале XXI столетия емко сформулировано Национальным научным фондом (NSF) США [Cyberinfrastructure (CI) Vision for 21st Century Discovery NSF, March 2007]. Всестороннюю инфраструктуру, которая должна объединять все усовершенствования в сфере информационных технологий (ИТ), назвали киберинфраструктурой (Cyberinfrastructure – CI). CI в интерпретации NSF объединяет в себе аппаратные средства для вычислений, данные и сети, цифровые датчики, обсерватории и экспериментальные установки, взаимодействующий набор программного обеспечения

и услуг микропрограммных средств и инструментов. Из всего многообразия аспектов CI в рамках данного проекта (рис. 1) решается задача построения *скоростной сети*, разрабатывается программное обеспечение, обеспечивающее *соединение* экспериментальной установки с суперкомпьютером для измерения характеристик потоков и течений при проведении *экспериментальных исследований*.

Скоростные сети – существенная компонента CI. Появившаяся в начале 21-го столетия возможность приобретения свободной емкости оптических волокон (darkfiber – темное волокно) у компаний

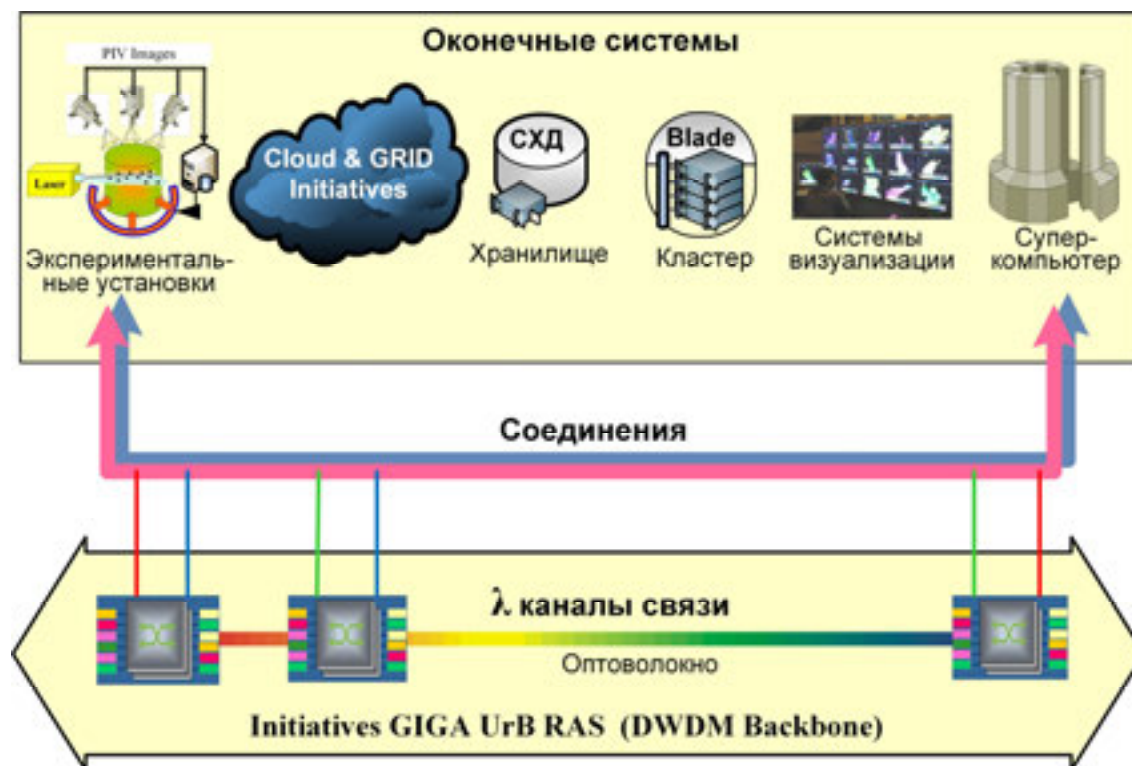


Рис. 1. Территориальная киберинфраструктура Урала

по протяженным линиям связи привела к построению в мире собственных научно-образовательных (R&E – Research and Education) оптических инфраструктур со спектральным уплотнением каналов. Это обстоятельство стимулировало работы по построению глобальной λ -системы, продвигающей парадигму глобальных λ -сетей (GLIF – Global Lambda Integrated Facility, Рейкьявик, Исландия, август 2003, <http://www.glif.is>). Сообщество GLIF разделяет взгляды в построении LambdaGrid парадигм грид-вычислений, в которых центральный архитектурный элемент – это оптические сети, параллелизм которых заключен в многочисленных длинах волн света (λ) в одном оптическом волокне.

Соединения между оконечными системами. Одним из основных источников ухудшения совокупной производительности территориально распределенных высокоскоростных приложений является низкая скорость передачи данных по прокладываемому соединению между оконечными системами по протяженным скоростным линиям связи. Поэтому вторая задача проекта – исследование, анализ и решение проблем взаимосвязи (межплатформенное ПО) подключаемых к скоростной сети оконечных систем. Эти системы характеризуются большой и сложной памятью, процессорами и подключенными к сети каналами ввода/вывода с иерархиями и наследуемым параллелизмом в различных подсистемах, и требуют разработки новых методов и технологий эффективного использования архитектур и ресурсов оконечных систем.

Экспериментальные исследования являются третьей целью проекта, заключающейся в апробации, тестировании и использовании разрабатываемых архитектурных решений на научных задачах (приложения), требующих предельной производительности компонент создаваемой СИ. Одной из таких задач является обработка в реальном времени потока изображений, генерируемых экспериментальными установками на скорости 1–10 Гбит/с.

I. Скоростная сеть.

Следуя мировому опыту, в Уральском отделении РАН реализуется проект «Инициатива GIGA UrB RAS», направленный на построение оптической магистрали, связывающей научные центры Отделения [8, 11]. Целесообразность Инициативы заключается в следующем.

В Научных центрах УрО РАН *созданы свои оптические сети*, объединение которых решит проблему скоростей Отделения.

Удачная организационно-территориальная структура УрО РАН позволила Президиуму УрО РАН выступить в роли заказчика проекта, удовлетворяющего не только Научные центры Отделения, но и вузы, а также предприятия промышленного сектора экономики, с которыми традиционно Уральское отделение выполняет совместные научно-исследовательские работы. К тому же охватываемая территория от Архангельска до Екатеринбурга способствует использованию технологий оптоволоконных систем передачи данных со спектральным уплотнением каналов, поскольку чем больше протяженность ($L \sim 2340$ км) и чем выше скорость передачи, тем более заметны преимущества оптических сетей.

Экономическая эффективность показана на примере равенства разовых затрат на построение собственной скоростной DWDM магистрали «Архангельск – Сыктывкар – Ижевск – Пермь – Екатеринбург» на скорости 40 Гбит/с и возможных ежегодных затрат в случае аренды каналов связи между Научными центрами на скорости 0,2 Гбит/с.

Инвестиционная привлекательность заключается в возможности поэтапного наращивания производительности DWDM-системы до терабитных скоростей и согласуется с мировой тенденцией развития R&E сетей.

Следует также отметить, что судьба пропускной способности и динамика разветвления сети оказывается на «кончиках пальцев» у институтов УрО РАН, а не у операторов связи, что позволяет опера-

тивно настраивать λ -коммуникационный сервис для разнообразных потребностей вместо одного арендуемого потока, приспособляемого для всех приложений/проектов.

Соучастие в проекте образовательных учреждений и инновационной промышленности Уральского региона и прилегающих территорий формируется в рамках Соглашений и Договоров с УрО РАН.

Предоставлена возможность *создания принципиально новых инструментов* проведения научных исследований, показывающих влияние скоростных сетей на распределенные вычисления.

Инициатива GIGA UrB RAS является фрагментом национальной научно-образовательной сети передачи данных. Основные цели Инициативы: поддержка исследований по сетевой тематике, продвижение приложений и услуг следующего поколения, создание простого интерфейса взаимодействия между научно-образовательными сетями.

1.2. DWDM тракт Пермь – Екатеринбург.

Сформулирован следующий класс услуг передачи данных [13], охватывающий 3 уровня эталонной модели взаимодействия открытых систем (L1-L3 OSIRM):

а) **λ -услуга** – предоставление фиксированного диапазона длин волн («лямбда»).

б) **FrameNet-услуга** – представление двухточечного или многоточечного Ethernet канала связи между Ethernet портами DWDM магистрالی.

в) **PacketNet-услуга** – присоединение к Public Интернет.

λ -услуга обеспечивает соединение точка-точка на физическом уровне эталонной модели взаимодействия открытых систем (L1 OSIRM).

FrameNet-услуга предполагает использование доступных EthernetGE (LX или на ZX) интерфейсов, таких как 1–10 GE порты оконечного оборудования магистрالی. Обеспечиваются точка-точка и многоточечная доставка данных на Ethernet (канальном) уровне (L2 OSIRM). Доступны два базовых типа услуг FrameNet:

– специализированная FrameNet – Ethernet-услуга, гарантирующая заказанную пропускную способность;

– неспециализированная FrameNet – Ethernet-услуга, не гарантирующая постоянную пропускную способность.

PacketNet-услуга предоставляет базовую услугу IP маршрутизации для доступа к public Интернет, обеспечивая соединения на сетевом уровне (L3 OSI RM). Прием анонсов IP-сетей, как и анонсирование IP-сетей, выполняется согласно политикам маршрутизации, отраженным в RIPENCC.

Практическим результатом этих работ явилась разработка технического задания, которое было использовано при заключении Государственного контракта № 59/11 от 27 мая 2011 года «На поставку оборудования для построения инфраструктуры опорной высокоскоростной сети для проведения научно-технических расчетов и экспериментов на участке «Пермь – Екатеринбург». Технические решения основаны на использовании двух свободных («темных») волокон в существующем кабеле с известными характеристиками и технологии спектрального уплотнения каналов (DWDM).

В мировой практике системы связи с длинными пролетами обычно применяются при создании соединений между островами/городами и OADM ответвлений к транснациональным сетям. В нашем случае применение технически самых совершенных, но дорогостоящих решений неоправданно. На рис. 2 показано размещение трех регенерационных пунктов, равноудаленно охватывающих точки присутствия территорию и обеспечивающих длину между пунктами регенерации не более 130 км.

Общая оптическая длина тракта Пермь – Екатеринбург равна 456 км. Используется стандартное одномодовое волокно с дисперсией, соответствующей Рек. ITU-T G.652. Вышеперечисленные классы услуг сети со спектральным уплотнением каналов (DWDM) реализованы на оборудовании компании ECI-



Рис. 2. DWDM тракт Пермь – Екатеринбург (2012)

Telecom. При расчете регенерационного участка DWDM системы связи учитывалась величина требуемого OSNR, которая является ключевым параметром транспондера и от которого напрямую зависит максимальная длина однопролетной линии связи (и стоимость). OpticalSignal-to-NoiseRatio (OSNR) – *отношение сигнал/шум для оптического сигнала*.

Оконечные узлы DWDM тракта Пермь-Екатеринбург на платформах XDM-2000 установлены на площадках Президиум ПНЦ УрО РАН (Пермь, Ленина, 13а) и суперкомпьютерном центре ИММ УрО РАН (Екатеринбург, Софьи Ковалевской, 16). Оптические мультиплексоры (MUX) на 16 λ -каналов обеспечивают возможность использования транспондеров со скоростью передачи 10–40 Гбит/с в каждой λ .

Регенерационное оборудование XDM-40 промежуточных узлов размещено в необслуживаемых регенерационных пунктах (НРП). Спецификация промежуточных узлов обеспечивает только оптическое усиление сигнала, однако закладываемое DWDM оборудование предполагает возможность установки ROADM мультип-

лексоров для выделения и маршрутизации λ -каналов не только на оконечных, но и на промежуточных узлах. Система мониторинга и управления магистральной ВОЛС (SunFire V245 и программное обеспечение LightSoft) располагается в центре управления сетью в ИМСС УрО РАН (Пермь).

1.3. Структура каналов связи (рис. 3).

Установленные транспондеры используют три λ -канала (каналы C21, C22 и C23 ITU-T-частотного плана) со скоростью потоков по 10,7 Гбит/с (ITU-TG.709) в каждой λ , поддерживают механизм упреждающего исправления ошибок (FEC – Forward Error Correction) и имеют следующие характеристики портов ввода/вывода для подключения оконечного оборудования: два Ethernet-порта по 10 Гбит/с ($2 \times 10\text{GE}$), выделенных из λ -каналов C21 и C23, и 8 Ethernet-портов по 1 Гбит/с ($8 \times 1\text{GE}$), сформированному из λ -канала C22 [21, 22]. Такая совокупность портов образует слой гарантированных каналов связи по схеме точка-точка, позволяет использовать распространенные в оконечных системах 1GE порты и осуществить постепенный переход на скорость 10 Гбит/с.

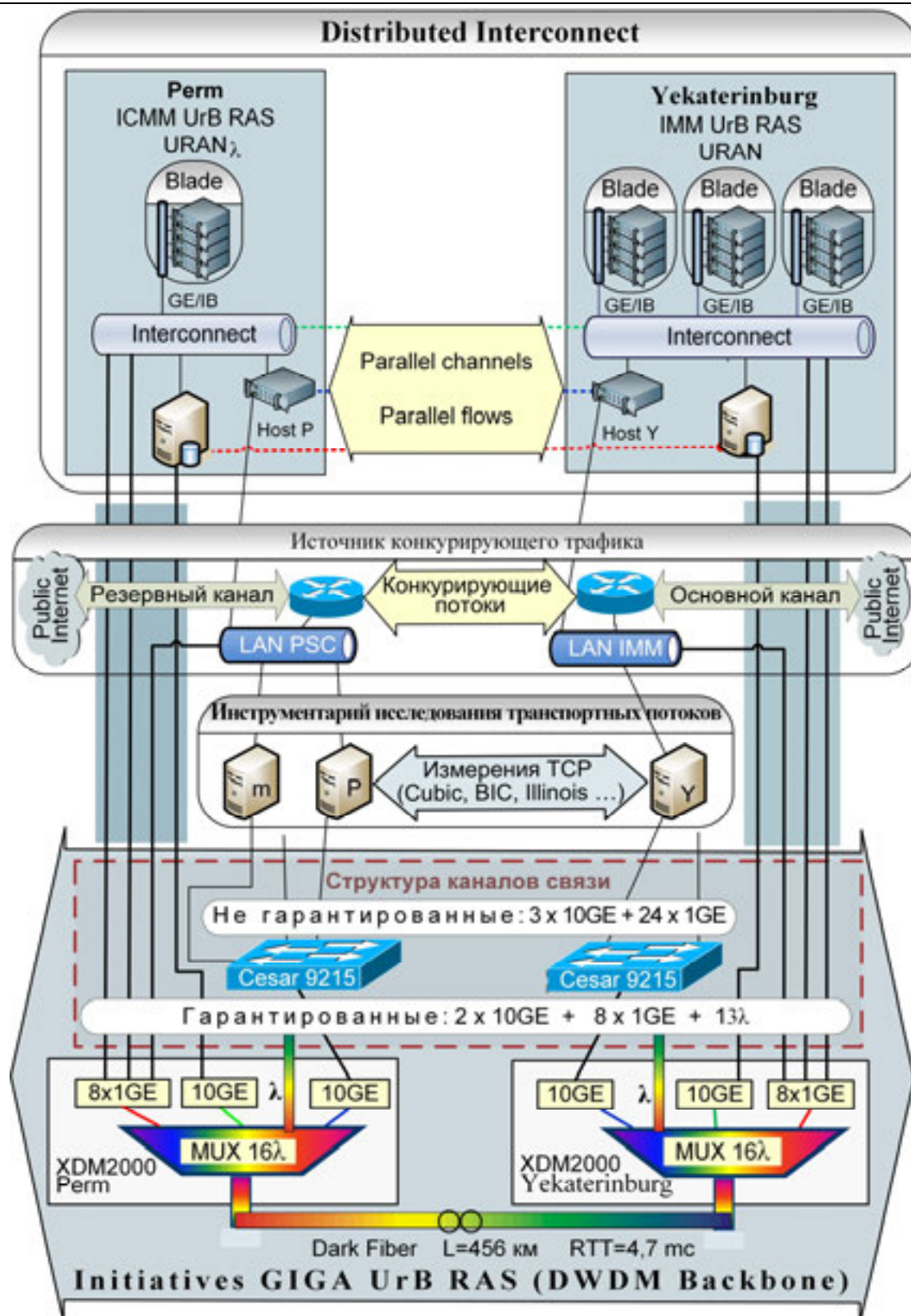


Рис. 3. Структура каналов связи

Оставшиеся 13 λ -каналов – потенциал масштабирования по 10–40 Гбит/с или когерентной DWDM 100 Гбит/с. Для большей гибкости проводимых исследо-

ваний на площадках установлены L2 коммутаторы Cesar, имеющие по 24 порта 1GE и 4 порта 10GE, один из которых подключен к 10GE порту XDM2000. Эта

совокупность портов образует слой негарантированных каналов связи при нагрузке больше 10 Гбит/с.

Аналогичные по производительности и сетевым сервисам региональные научно-образовательные сети в Российской Федерации отсутствуют.

II. Соединения.

Соединения между оконечными системами (соединения end-to-end) реализуются посредством транспортных протоколов, расположенных на 4-м уровне OSI/RM. Транспортными протоколами являются TCP (надежный), UDP (ненадежный) и UDP (надежный).

Несоответствие между вычислительной производительностью и компонентами ввода/вывода высокопроизводительных систем текущего поколения сделал ввод/вывод наиболее узким местом. Одним из основных источников ухудшения совокупной производительности территориально распределенных высокоскоростных приложений является плохая end-to-end производительность протокола TCP [22]. В RFC5681 (M. Allman, 2009) определены 4 связанных между собой алгоритма управления перегрузкой (congestion control algorithms) для протокола TCP: медленный старт (SS – slow start), предотвращения перегрузки (CA – congestion avoidance), быстрой повторной передачи (FR – fast retransmit) и быстрого восстановления (fast recovery).

Произведение полосы пропускания на время ретрансмиссии ($BW \cdot RTT$) называется в англоязычной литературе параметром BDP (bandwidth-delay product), большая величина которой характерна для скоростных сетей большой протяженности. Для участка DWDM магистрали Пермь-Екатеринбург с оптической длиной 456 км и временем распространения сигнала по оптическому волокну 5 мкс/км параметр RTT равен 4,7 мс. С учетом задержек обработки сегментов в оконечных системах примем $RTT = 5$ мс. Тогда для скорости передачи 1 Гбит/с значение $BDP = 625\,000$ байт, а для скорости 10 Гбит/с $BDP = 6\,250\,000$ байт, или 4 130 пакетов длиной 1 514 байтов.

Для TCP в фазе медленного старта чтобы увеличить окно от 1 до 4 130 пакетов потребуется приблизительно 16 RTTs (12 + 3 на троекратное рукопожатие) ~ 90 мс. И если TCP-соединение завершится раньше или будут одиночные ошибки с вероятностью $p > 5,8 \cdot 10^{-7}$, полоса пропускания канала будет существенным образом недоиспользована. Это делает передачу данных в рамках протокола TCP, используемого в большинстве операционных систем, включая Windows и Linux, достаточно вялой, не использующей всю имеющуюся полосу пропускания. На рис. 4 иллюстрируется фаза медленного старта протокола TCP и темп увеличения скорости передачи.

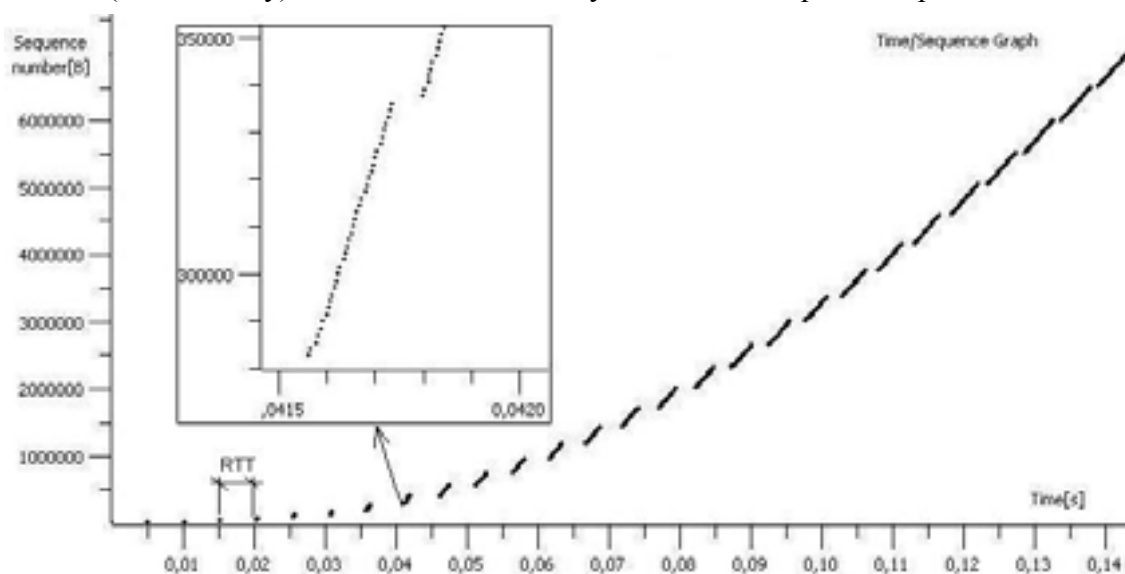


Рис. 4. Медленный старт TCP, $RTT = 5$ мс, скорость канала 1 Гбит/с

Классический способ использования доступной физической скорости в канале связи – тонкая настройка параметров управления соединением и организация параллелизма соединений. Версии TCP (RENO, Cubic, BIC и т.д.) направлены на увеличение пропускной способности в различных условиях функционирования: наличие или отсутствие конкурирующих потоков, сетевая конфигурация, протяженность и скорость линий связи. По индикации перегрузки различные варианты TCP делятся на три группы: по сигналу потери (loss-based), по задержке (delay-based) и использующие оба показателя (hibrid). Поэтому требуется выполнять измерения и производить тонкую настройку TCP при построении алгоритмов и программного обеспечения для организации end-to-end соединений.

2.1. Инструментарий измерения.

Наиболее удачным подходом к решению проблемы тонкой настройки TCP является проект Web10G, разработанный и поддерживаемый командой Питтсбургского Центра Супервычислений. Web10G предоставляет доступ к расположенным в ядре ОС переменным протокола TCP, позволяя считывать (и изменять) их. Знание динамики изменения параметров, таких как CurCwnd (текущее окно перегрузки), NonRecvDA (количество дубликатов подтверждений), RTTVar (значение RTT) и многих других, позволяет выявить первопричину недостаточной производительности и скорректировать стратегию управления перегрузкой. Измерения проводились между серверами Р (Perm) и Y (Yekaterinburg) с зеркалированием трафика на компьютер m (mirgor), изображенных на рис. 3. Сетевое конфигурирование позволяет добавлять или отключать существующий фоновый (конкурирующий) Интернет-трафик, интенсивность которого оценена по MRTG и имеет величину не более 100 Мбит/с по причине его ограничения. Доступная на момент измерения скорость канала связи – 1 Гбит/с.

На рис. 5 приведен импульсный трафик, характерный для структурированно-

го потока экспериментальных данных. Цель измерений – определение возможной потери темпа передачи при переходе от одного кванта времени передачи к другому. Наглядно видно уменьшение средней скорости передачи последовательности одинаковых порций данных, которое может быть объяснено хорошим дружелюбием Cubic к фоновому Интернет-трафику. Проиллюстрировано сохранение темпа передачи в конце текущего и начале следующего интервалов передачи (макроанализ на графике не показан).

На рис. 6 показана измеренная зависимость производительности RENO и Cubic от числа параллельных TCP-соединений. Для измерения пропускной способности использована утилита Iperf.

Видно, что Cubic эффективнее использует канал, нежели Reno, в диапазоне 2–16 параллельных потоков. Количество параллельных потоков для Cubic рекомендуется делать не более 8, оставив полосу пропускания ~ 100 Мбит/с для Интернет-трафика.

На рис. 7 показана скорость одного потока на фоне Интернет-трафика. Для генерации трафика и измерения средней пропускной способности использована утилита Iperf. Параметр CurCwnd снимался с помощью Web10G. Результаты измерений: Cubic – 143 Мбит/с; Reno – 97,2 Мбит/с. Видно, что стратегия управления перегрузкой Reno ведет себя более «осторожно» и не имеет большого числа перегрузок по сравнению с Cubic. Однако Cubic лучше Reno по скорости передачи данных. При этом один поток Cubic не доиспользует доступную полосу пропускания канала: 143 Мбит/с (Cubic) + 100 Мбит/с (Internet) < 1 Гбит/с (канал связи).

На рис. 8 представлено сравнение работы алгоритмов управления перегрузкой BIC, Cubic и Illinois для непрерывного трафика по гарантированному каналу связи 1 Гбит/с. Используемый инструментарий: утилита Iperf для генерации трафика и Web10G для анализа и интерпретации. Условия: одно TCP-соединение, гарантированный канал связи 1 Гбит/с с RTT = 5 мс (BDP 625 000 байт), фоновый интер-

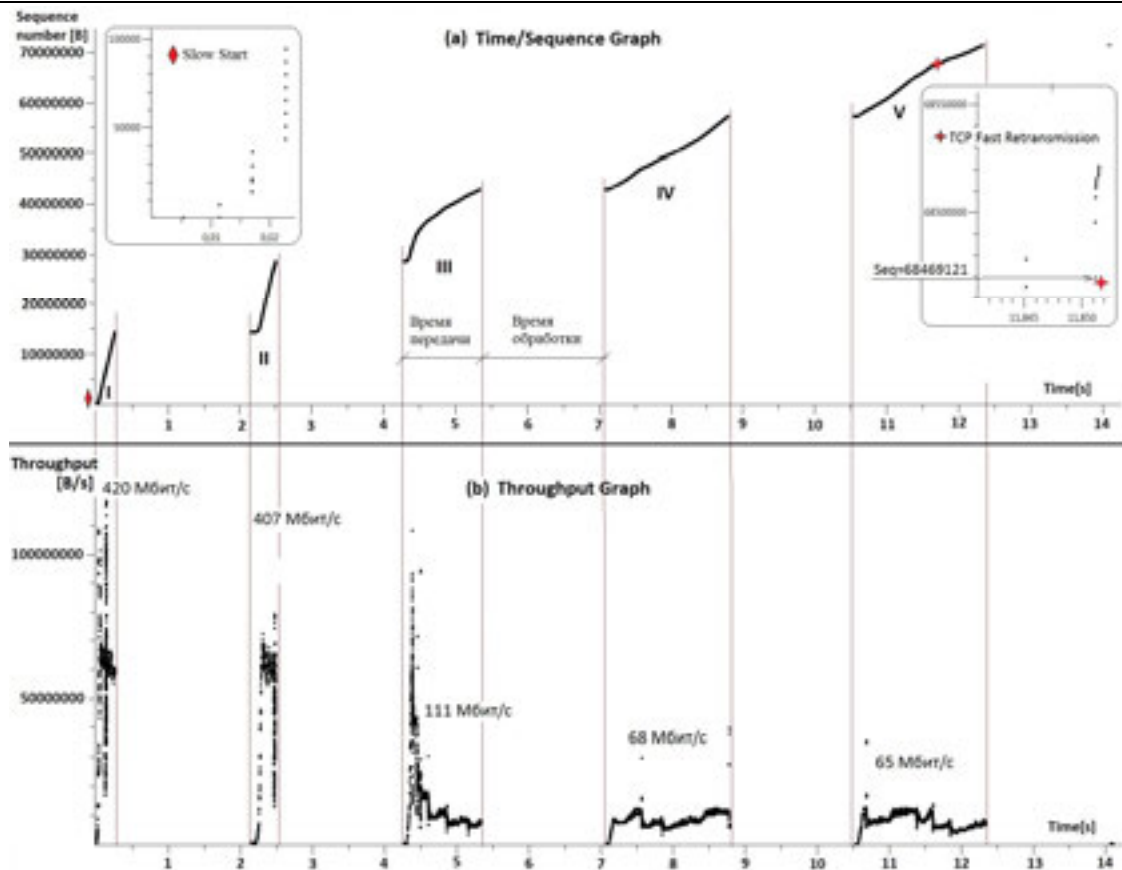


Рис. 5. Wireshark, импульсный трафик (Sequence number – порядковый номер байта в потоке)

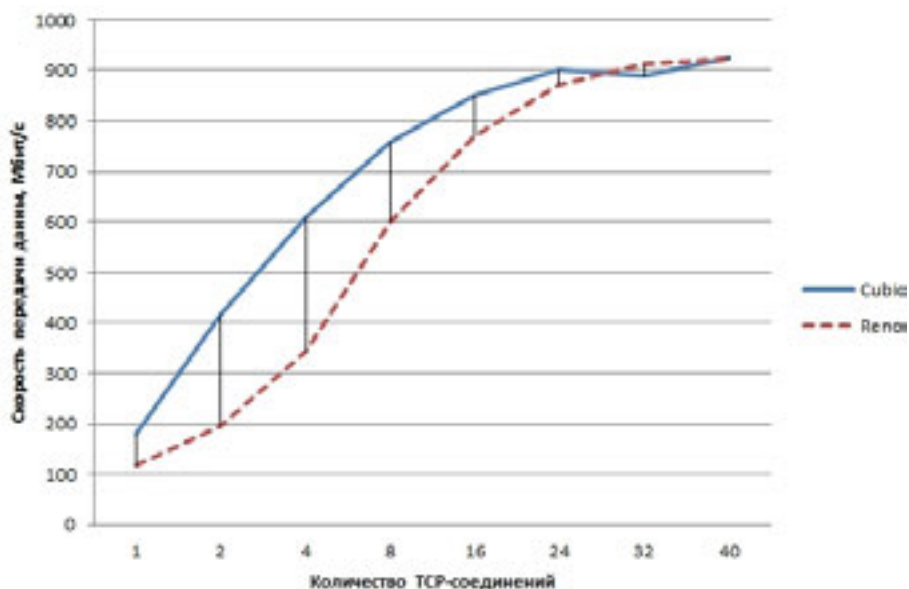


Рис. 6. Зависимость скорости передачи данных от количества соединений

нет-график отсутствует, периодичность снятия параметров 1 мс, время измерения 1 с, (на графике отображено 0,12 с). *Описание параметров.* CurCwnd – текущее значение окна перегрузки. SegsOut – общее количество переданных сегментов, наклон этой кривой является текущей ско-

ростью передачи. *Интерпретация.* Как видно из графиков, все стратегии достигают максимальной скорости передачи после достижения $CurCwnd \geq BDP$, что соответствует теории протоколов с обратной связью. Однако для ВIC и Illinois это происходит в фазе SS в точке А на 40 мс, а

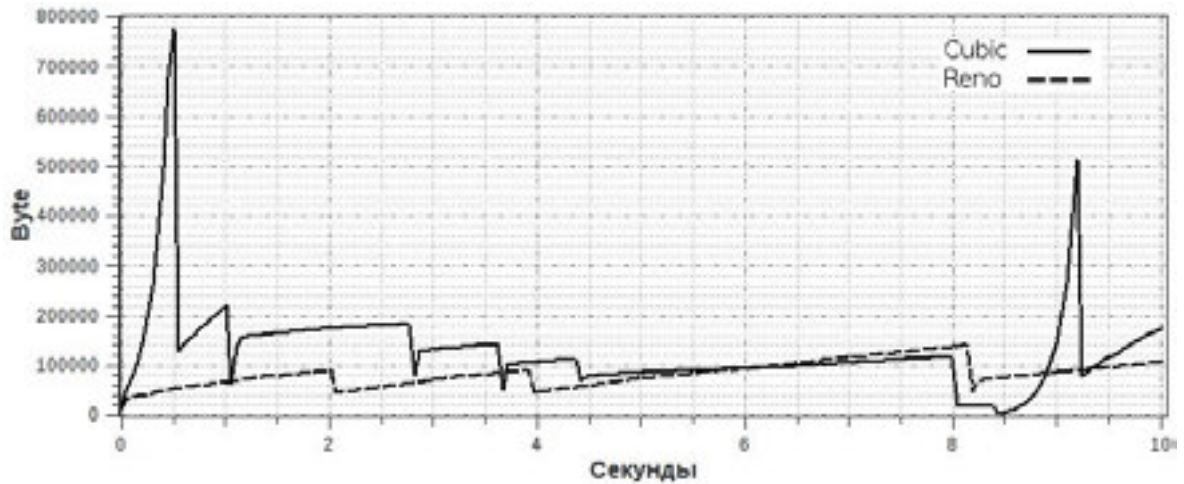


Рис. 7. Сравнение динамики изменения параметра CurCwnd для Cubic и Reno

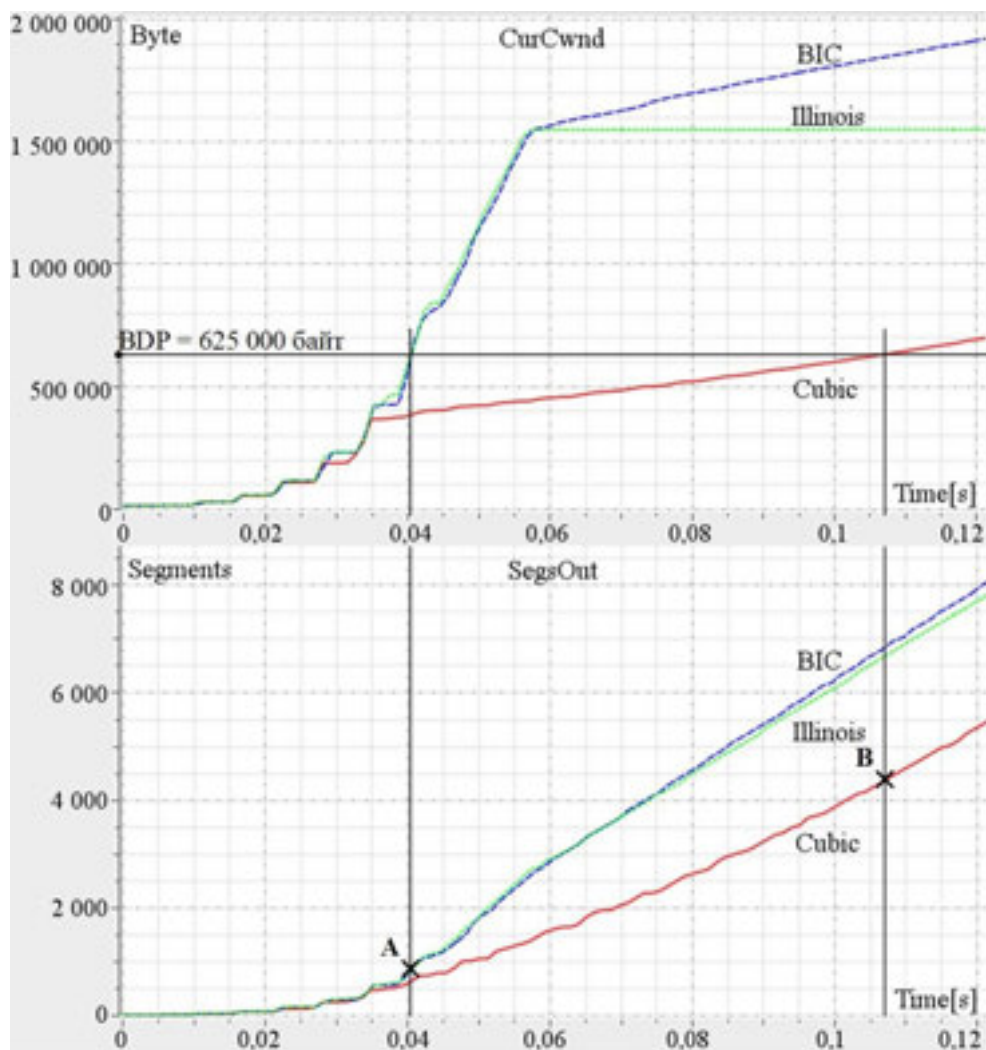


Рис. 8. Сравнение динамики изменения параметров TCP в области медленного старта

для Cubic – в фазе SA в точке B на 110 мс, поскольку MaxSsCwnd у Cubic на 35 мс завершил фазу SS и перешел в фазу SA, в которой темп увеличения CurCwnd происходит медленнее.

Подтверждается негативное влияние фазы SS, которое лучше преодолевается в BIC и Illinois. Размер передаваемых данных при коротких TCP-сессиях должен быть существенно больше BDP.

Основные выводы по измерениям.

Принципиальным решением проблемы медленного старта при изначально известной гарантированной скорости в end-to-end канале связи и отсутствии фонового трафика в нем является установка начального значения параметра *cwnd* в оконечных системах равным BDP. Верхнюю границу числа параллельных конкурирующих потоков на фоне интернет-трафика в канале 1 Гбит/с с $RTT = 5$ мс для Cubic нецелесообразно делать более 8. Как и ожидалось, в L2-каналах связи порядок следования пакетов не нарушается и не приводит к появлению ложных SACK. При передаче импульсного трафика в TCP-соединении темп передачи в конце текущего и начале следующего интервалов передачи сохранятся.

Методология оценки эффективности параллелизма потоков данных должна содержать следующие компоненты: (1) список и описание системных параметров оконечных систем; (2) описание сетевой инфраструктуры, связывающей оконечные системы; (3) перечень и параметры нагрузки; (4) оцениваемые метрики параллельных потоков; (5) измерительные средства; (6) метод анализа и интерпретации данных; (7) способы представления результатов.

2.2. Модель потоковой обработки структурированного потока данных.

Структурированным называется поток, компоненты которого допускают независимую обработку. Таковым, например, является поток экспериментальных данных, получаемых по методу PIV (Particle Image Velocimetry) – оптическому методу измерения полей скорости жидкости или газа в выбранном сечении потока. Метод основан на обработке пар фотографий трассеров – мелких частиц, взвешенных в потоке, в моменты, когда они подсвечиваются импульсным лазером, создающим тонкий световой нож. Скорость потока определяется расчетом перемещения трассеров за время между вспышками лазера. Интенсивность порождаемого потока данных зависит от числа, разрешения и частоты ра-

боты камер и может достигать нескольких гигабит в секунду [1, 6, 12]. Обработка таких потоков данных с использованием только локального компьютера экспериментальной установки (ЭУ) чрезвычайно затруднена. В то же время развитие математического аппарата и появление новых высокоточных алгоритмов расчетов увеличивают требования к необходимой вычислительной мощности. Поэтому перенос вычислений на удаленные суперкомпьютеры достаточной производительности позволит применять новые высокоточные алгоритмы, обрабатывать данные в реальном времени, уменьшить объемы сохраняемых на ЭУ данных. Сопутствующей задачей является 4 K-визуализация в 24 кадра в секунду (FPS) в формате RGBA (32 бит на пиксель: с красными, зелеными, синими и альфа-каналами по 8 бит в каждом), которая требует 6,4 Гбит/с сетевой полосы пропускания к потоку от источника рендеринга до панелей отображения. Идеальная модель обработки структурированного потока экспериментальных данных на удаленном суперкомпьютере в реальном времени представлена на рис. 9 [9].

Суть модели – прямой ввод потока данных в оперативную память вычислительных узлов суперкомпьютера без предварительного сохранения на внешних носителях ЭУ или СХД суперкомпьютера [2, 9, 15]. Такой способ передачи данных позволит минимизировать задержки на двойную запись/чтение данных и отказаться от копирования данных между хранилищами с использованием классических протоколов SCP, FTP, NFS или CIFS.

Благодаря тому, что каждое измерение в эксперименте PIV может обрабатываться независимо от других [4] и существует один источник данных на ЭУ, было предложено отказаться от однозначного отображения измерений на вычислительные узлы.

Новый принцип передачи и распределения данных по вычислительным узлам суперкомпьютера базируется на трех идеях:

- прямой ввод данных в вычислительные узлы, минуя промежуточное хранение на дисках;

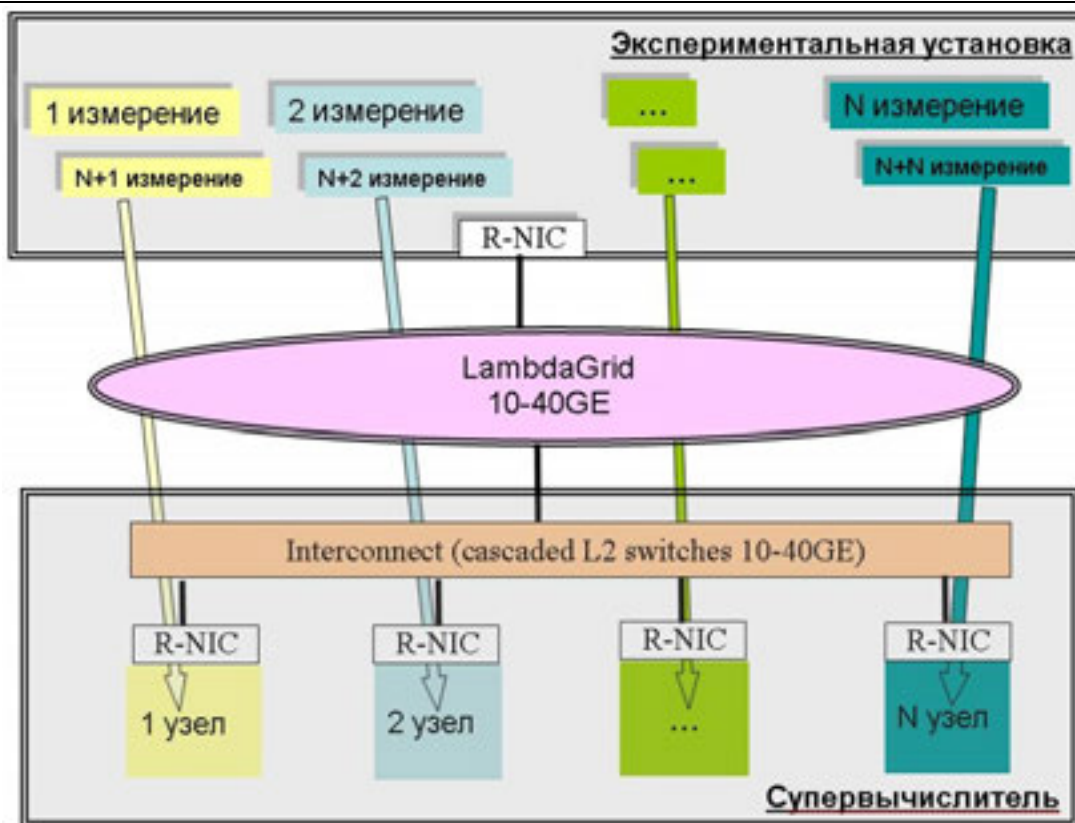


Рис. 9. Идеальная модель обработки структурированного потока данных

– отказ от однозначного отображения сообщений на вычислительные узлы;

– представление потока данных в виде единой очереди, распределяемой по вычислительным узлам по запросам.

На экспериментальной установке организуется очередь измерений, которые раздаются вычислительным узлам по принципу First-In-First-Out (FIFO) [14, 15], как показано на рис. 10. Подсистема программного обеспечения, решающая задачу распределения данных по вычислительным узлам, названа диспетчером. Несмотря на то что извлечение элемента из очереди является блокирующей операцией, после извлечения измерений из очереди они могут передаваться клиентам параллельно. В случае наличия большого количества конкурентных запросов число параллельных передач определяется доступными ресурсами на компьютере ЭУ, при этом остальные вычислительные узлы будут простаивать в ожидании появления доступных ресурсов. Это позволяет использовать канал связи максимально эффек-

тивным образом. В настоящий момент реализован подход, когда при наличии данных в очереди они будут отдаваться сервером в ответ на все поступающие запросы при наличии доступных ресурсов на компьютере ЭУ.

Для взаимодействия ЭУ и вычислительных узлов разработан «Протокол PIV». При этом ЭУ является сервером, а на вычислительных узлах работают клиенты.

Применение предложенной модели дает следующие преимущества:

– Отсутствует необходимость в синхронизации между вычислителями на стороне суперкомпьютера.

– Автоматическая балансировка нагрузки по вычислительным узлам.

– Возможность изменять число используемых вычислительных узлов во время проведения процесса обработки данных.

– Возможность использовать вычислительную мощность нескольких суперкомпьютеров.

– Минимизация потери данных в случае выхода из строя одного или несколь-

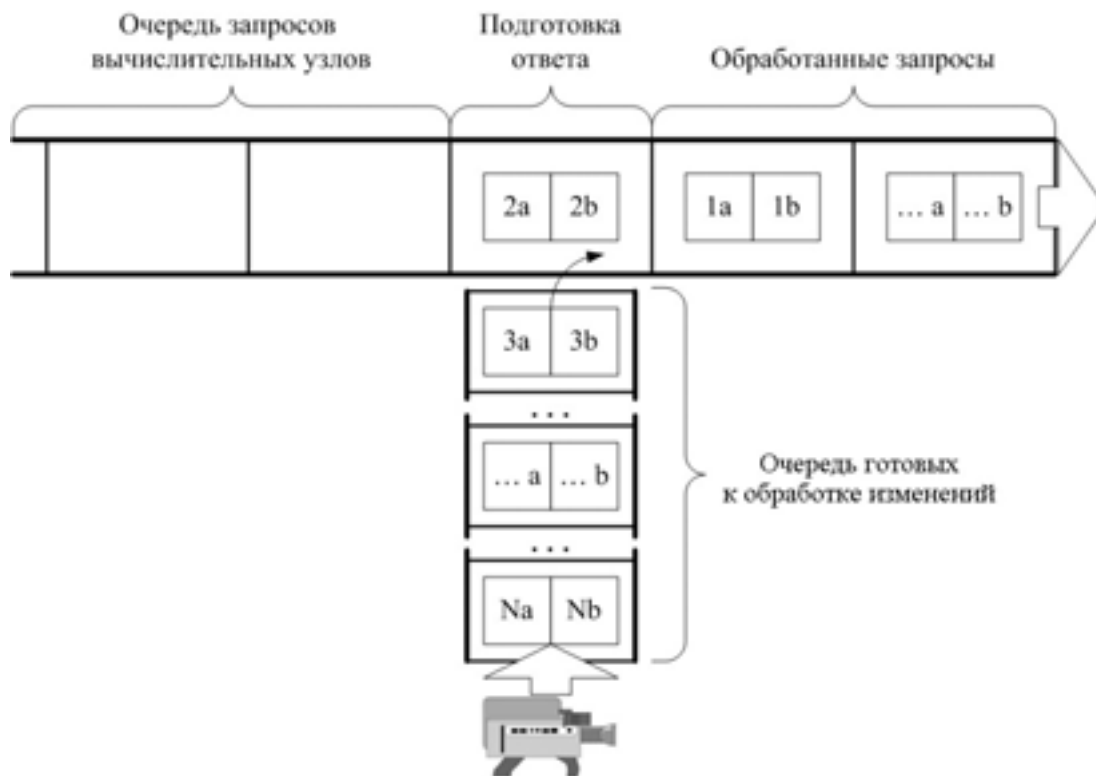


Рис. 10. Распределение данных в ответ на запросы вычислительных узлов

ких вычислительных узлов (в худшем случае будет потеряно только текущее обрабатываемое измерение сбойного узла). В ситуации, когда потеря данных недопустима, на время расчета сообщение необходимо сохранять в памяти сервера и в случае выявления сбоя повторно добавлять эти сообщения в очередь готовых к обработке сообщений.

Необходимо отметить, что предлагаемый подход не ограничивает нас в выборе транспорта для передачи данных. Текущая реализация использует для передачи данных протокол PIV. Однако возможно применение общей СХД, подключенной и к ЭУ, и к супервычислителю. В этом случае задачей диспетчера будет распределение по вычислительным узлам не самих данных, а путей к файлам с данными на СХД, при этом вычислительные узлы будут считывать данные непосредственно из файлов на СХД. Это позволит анализировать и сравнивать поведение как классической схемы с общим дисковым пространством, так и предлагаемой схемы без промежуточного хранения данных на дисковых массивах.

2.3. Протокол взаимодействия оконечных систем [3, 7].

Протокол PIV является протоколом прикладного уровня, реализующим идею модели RPC. Протокол работает по схеме запрос-ответ и может использоваться в качестве протокола транспортного уровня любой надежный потоковый протокол передачи данных.

Протокол PIV рассчитан на передачу нескольких блоков бинарных данных. В одном пакете протокола можно передать от нуля до 65 535 блоков, каждый из которых может иметь размер до 4 Гбайт. Формат пакета протокола PIV приведен на рис. 11. Поля заголовка пакета протокола кодируются в сетевом порядке байт.

Текущая реализация протокола PIV поддерживает транспортные протоколы TCP и UDT [2]. Положение протокола PIV в стеке сетевых протоколов показано на рис. 12. Протокол UDT – это основанный на UDP протокол передачи данных для высокоскоростных сетей. Он был разработан в Университете штата Иллинойс в Чикаго. Функциональные возможности протокола UDT аналогичны протоколу TCP. UDT яв-

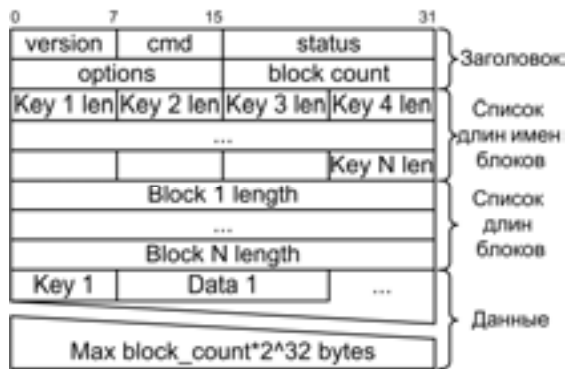


Рис. 11. Формат пакета протокола SciMP

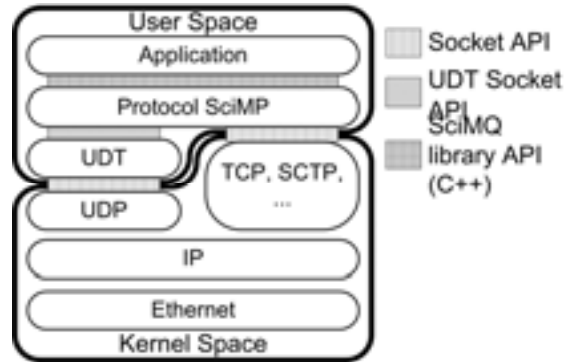


Рис. 12. Положение протокола SciMP в стеке сетевых протоколов

ляется дуплексным протоколом передачи потока данных с предварительной установкой соединения. Особенностью протокола UDT является оригинальная архитектура и реализация, а также оригинальный алгоритм управления перегрузкой. При этом протокол UDT позволяет программисту реализовать и использовать свой алгоритм управления перегрузкой.

Программное обеспечение [20], реализующее передачу данных с помощью предложенной модели, разработано с использованием языка программирования C++ и библиотек Boost. Язык C++ был выбран по причине удобства разработки и его распространенности на суперкомпьютерах. Использование библиотек Boost позволило упростить кроссплатформенную реализацию асинхронной работы с сетевыми соединениями. Программное обеспечение состоит из трех логических компонентов: подсистема загрузки данных; диспетчер (менеджер очередей); клиентская библиотека. Клиентская библиотека предоставляет API для написания расчетных приложений, запускаемых на суперкомпьютере, и приложений загрузки данных для экспериментальной установки. В дополнение к этому реализовано расчетное приложение с использованием клиентской библиотеки, реализующее кросскорреляционный алгоритм расчета PIV-данных.

Сервер работает непосредственно на экспериментальной установке и отвечает за загрузку данных и их распределение по вычислительным узлам. Для минимизации затрат ресурсов сетевое взаимодейст-

вие реализовано по асинхронной модели, что позволяет эффективно обрабатывать большое количество параллельных соединений в небольшом числе процессов. Клиентская библиотека предоставляет API, которое позволяет запрашивать у сервера новый блок исходных данных и отправлять на сервер результаты расчетов (или подтверждение об успешном расчете, если отправка полных результатов на ЭУ не требуется).

2.4. Оценка эффективности применения предложенной модели.

Устойчивая работа приведенной на рис. 9 модели распределения данных по вычислительным узлам суперкомпьютера может быть достигнута при следующих условиях. Во-первых, сквозная пропускная способность тракта передачи данных (v) не должна сдерживать генерируемого экспериментальной установкой потока данных. Во-вторых, число вычислительных узлов суперкомпьютера (n) должно быть достаточным для обработки потока входных в темпе их поступления. И если v_i – достижимая производительность одноканального тракта передачи данных, в системе из m параллельных каналов ожидаемым результатом будет: $v = m \cdot v_i \cdot \partial_v$, где ∂_v – эффективность параллельной передачи данных. В идеальной параллельной системе параллельная эффективность стремится к единице ($\partial_v \rightarrow 1$).

Оценка эффективности проводилась с использованием управляющего компьютера PIV-системы (ИМСС УрО РАН, Пермь) и суперкомпьютера «URAN»

(ИММ УрО РАН, Екатеринбург), соединенных по выделенному каналу пропускной способностью 1 Gbps. Вычислительные узлы получали доступ к PIV-системе через головной узел суперкомпьютера с использованием технологии RAT.

В качестве тестового набора данных использовались 1 000 ранее полученных измерений размером 13,6 Мбайт каждое. На рис. 13 приведен график зависимости суммарного времени расчета всего множества данных от числа задействованных вычислительных узлов для двух алгоритмов: «А» и «В». Исследуется ситуация, когда число параллельных потоков m равно числу вычислительных узлов n , т.е. $m = n$. Рассматриваемым отличием алгоритмов является время расчета t , при этом детальное описание алгоритмов выходит за рамки этой статьи. Алгоритм «А» производит расчет одного измерения за время порядка 1,4 с, а алгоритм «В» – 2,5 с.

На графике (см. рис. 13) можно выделить три области:

I – уменьшение времени расчета пропорционально числу узлов;

II – уменьшение времени расчета изменяется медленнее, чем число узлов;

III – время расчета не зависит от числа узлов.

На участке «I» пропускная способность системы ограничивается только

доступными вычислительными ресурсами. Такая ситуация возникает, когда темп генерации данных много меньше пропускной способности сети и параллельные потоки протокола PIV мало конкурируют между собой, т.е. $\partial_v = 1$.

На участке «II» пропускная способность системы ограничивается и доступными вычислительными ресурсами, и пропускной способностью сети ($\partial_v < 1$). Активные параллельные потоки начинают подходить к пределу пропускной способности сети и оказывают негативное влияние друг на друга. Уменьшение времени обработки массива данных меньше, чем на участке «I».

На участке «III» пропускная способность системы ограничивается только пропускной способностью сети, а ее эффективный параллелизм для алгоритма «А» равен, как будет показано далее, $\partial_v = 0,65$. Нахождение системы в этой области приводит к частичному простоя вычислительных узлов.

Из графика следует, что границы областей зависят от времени расчета одного измерения t и пропускной способности сети. Для скорости сети 1 Гбит/с теоретическая пропускная способность тракта передачи данных $v = 8,75$ измерений в секунду. Тогда в идеальном случае число вычислительных узлов должно быть не

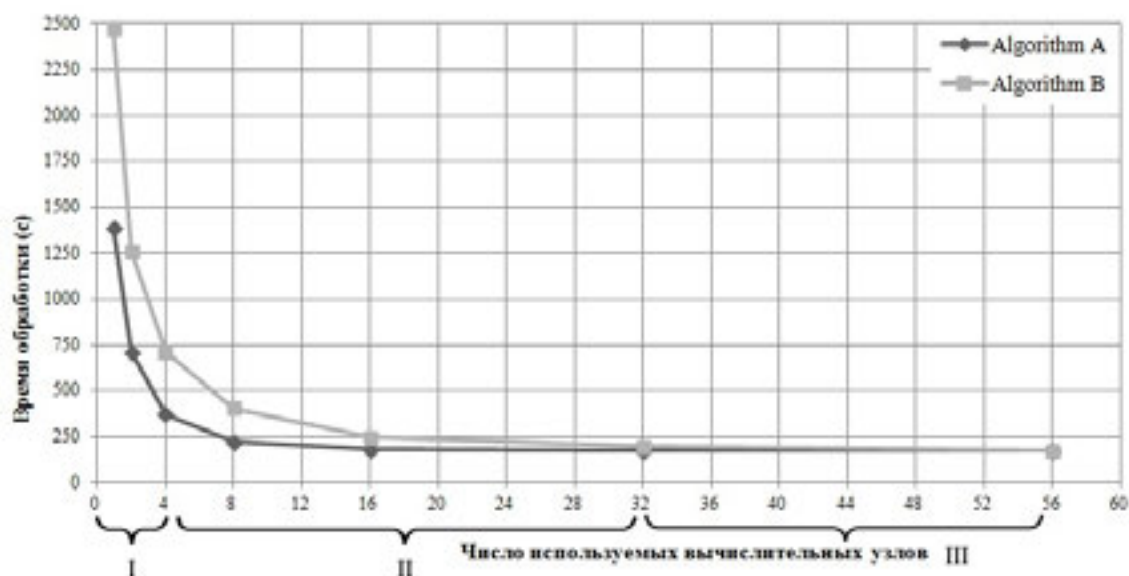


Рис. 13. Суммарное время расчета множества данных

менее $n = v \cdot t = 8,75 \cdot 1,4 \sim 13$. Однако, как видно из рис. 13 для графика «А», уменьшение времени счета происходит при увеличении числа узлов до 16. Время расчета одного измерения получается равным 2,8 с (увеличение вызвано тем, что часть времени вычислительные узлы простаивают). Реальная пропускная способность системы составляет $v = n/t = 16/2,8 = 5,71$ измерений в секунду. Эффективность параллельной передачи данных $\delta_v = 5,71/8,75 = 0,65$. Это обстоятельство объясняется механизмами управления перегрузками протокола TCP в каналах связи с большим BDP и наличием конкурирующих потоков.

Эмпирически можно установить следующие закономерности. Для расчета уже существующего набора данных число вычислительных узлов лучше выбирать из области «II». Эта область даст разумный баланс между используемыми вычислительными ресурсами и временем расчета. В случае расчета экспериментальных данных в реальном времени число вычислительных узлов должно находиться на участке «III».

III. Приложения.

В качестве научной задачи [10, 17, 18, 23], требующей сочетания больших скоростей передачи данных и высокопроизводительных систем обработки данных, сформулирована задача обработки интенсивного потока экспериментальных данных на удаленном суперкомпьютере в реальном времени. Одной из таких задач является обработка экспериментальных данных, получаемых по ранее описанному методу PIV (Particle Image Velocimetry) – оптическому методу измерения полей скорости жидкости или газа в выбранном сечении потока.

3.1. Экспериментальная установка.

Измерительная часть установка PIV состоит из следующих компонент [1, 2]. Мощный импульсный лазер Quantel используется для освещения измерительной области лазерным ножом регулируемой толщины. Две цифровые камеры ВИДЕО-СКАН-11002-2001 на базе ССД матрицы

разрешением 4004×2671 (пикселей) и разрядностью АЦП 12 бит/пиксель регистрируют мгновенные изображения с кадровой частотой от 1,2 до 5,8 Гц. Синхронизатор и персональный компьютер с программным обеспечением Actual Flow управляют экспериментом и сохраняют полученные с камер данные.

Камеры подключены к компьютеру через карту видеозахвата (PCI-контроллер VS2001). Компьютер работает под управлением операционной системы Windows XP, оснащен четырехъядерным процессором (Phenom 2X4) с тактовой частотой 3,2 ГГц, имеет объем оперативной памяти 4 ГБ и дисковой памяти 1 ТБ, подключен двумя портами 1 GE к коммутирующему оборудованию ИМСС УрО РАН. ПО Actual Flow возможна также последующая обработка данных и визуализация результатов обработки, однако большой объем получаемых данных, а также ресурсоемкость расчетных программ не позволяют проводить обработку в реальном времени.

Экспериментальная установка (рис. 14) представляет собой прямоугольную кювету ($70 \times 70 \text{ см}^2$) с цилиндрической вставкой. Такая конфигурация позволяет обеспечить осевую симметрию и избежать оптических искажений в вертикальных сечениях. Конвективное течение создается набором распределенных нагревателей, что позволяет задавать структуру течения. Кювета располагается на вращающемся столе, который обеспечивает стабильное вращение в широком диапазоне угловых скоростей Ω . В качестве рабочих жидкостей используются силиконовые масла с различным значением кинематической вязкости. Трассерами являются стандартные PIV-частицы диаметром 20 мкм.

3.2. Алгоритм обработки структурированного потока данных.

Алгоритм параллельного расчета мгновенных полей скорости по изображениям, поступающих с измерительной установки, реализуется в двух вариантах. *Первый вариант* состоит в распределении

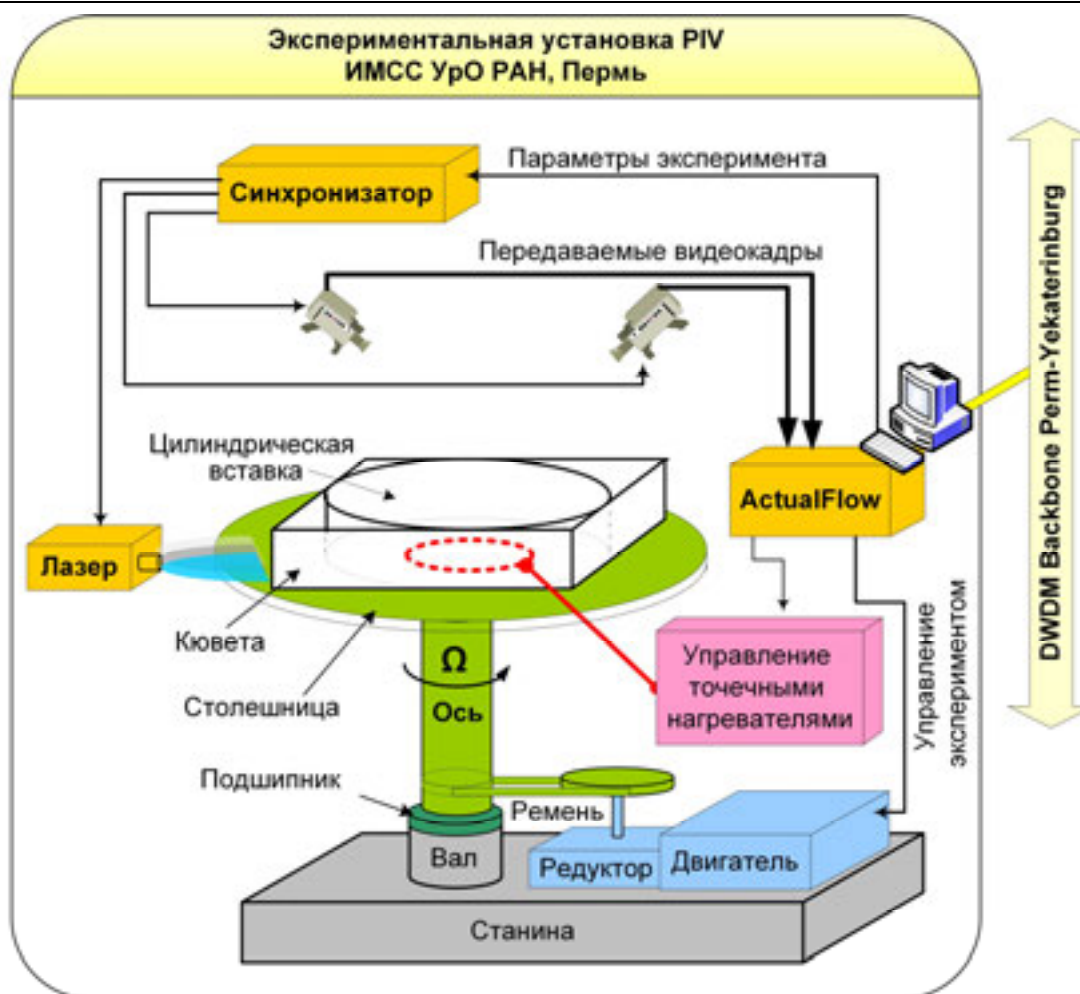


Рис. 14. PIV-экспериментальная установка

непрерывного потока данных по процессорам. При этом каждый процессор выполняет последовательный алгоритм определения поля скорости по одной паре изображений во всей расчетной области. *Второй вариант* заключается в параллельном расчете на нескольких процессорах поля скорости для одной пары изображений. В этом случае целесообразна передача одной пары изображений от удаленной экспериментальной установки до внутренней сети кластера и последующее дублирование этой пары в оперативную память каждого процессора средствами межзвеновой сети кластера. Непосредственное распараллеливание расчетного алгоритма делается стандартным способом деления расчетной области на подобласти. Итерации циклов кросскорреляционного алгоритма являются информационно независимыми. Минимальный межпроцессор-

ный обмен необходим лишь на стадии адаптационного уточнения.

Оба подхода параллельной обработки данных имеют свои преимущества и недостатки. Основное удобство использования первого подхода состоит в непосредственном применении любого последовательного: стандартного, адаптивного или вейвлет кросскорреляционного алгоритма. Межпроцессорный обмен отсутствует. Недостатком является ограничение минимального времени обработки одной пары изображений временем работы последовательного алгоритма. Это может быть критично при использовании высокоточных и ресурсозатратных алгоритмов, а также при обработке данных с высоким разрешением. Второй способ требует доработки в плане организации межпроцессорного обмена, но обладает практически неограниченными возможностями по минимизации

времени обработки одной пары изображений. Окончательный выбор метода имеет смысл делать при всех известных условиях проведения эксперимента.

3.3. Экспериментальное и численное исследование формирования вихревых течений различных масштабов в цилиндрическом слое.

В предлагаемой вниманию работе исследуются гидродинамические процессы в цилиндрическом слое с локализованным нагревом в центральной части. Локализованный нагрев приводит к появлению крупномасштабного течения в виде тороидальной конвективной ячейки. В пограничном температурном слое в области нагрева формируются мелкомасштабные вторичные течения, подобные тем, что подробно описаны в [24]. В результате возникает очень интересная для изучения гидродинамическая система с локализованными вертикальными струями, вихревыми структурами различных масштабов и крупномасштабной циркуляцией. Данная работа является первым этапом детального анализа распределения завихренности в описываемой системе, в ходе которого будут рассматриваться: структура полей завихренности при взаимодействии сдвигового течения с вертикальными струями; формирование и эволюция вихревых структур в пограничном слое адвективного течения в неподвижном цилиндрическом слое с локализованным нагревом.

В пограничном слое основного потока формируются вторичные течения в виде горизонтальных валов (рис. 15, а) и всплывающих конвективных струй (рис. 15, б),

являющиеся основным источником пульсаций поля скорости. В результате проведенной работы было показано, что радиальная завихренность возникает вследствие образования радиально ориентированных конвективных валов. Распределение азимутальной компоненты завихренности носит более сложный характер. Имеется ее переменная составляющая, обусловленная поперечными валами, которые сносятся к центру основным течением.

На среднем поле эти структуры не видны, и распределение завихренности определяется сдвигом основного радиального течения, а знак завихренности – вертикальным градиентом радиальной компоненты скорости. Вертикальная компонента завихренности имеет значения, заметно меньшие значений горизонтальных компонент. Источники генерации вертикальной завихренности также находятся в местах образования и всплытия вторичных структур. Подробное описание результатов данного раздела можно найти в статье [25].

3.4. Экспериментальное исследование характеристик двухфазного потока в факеле форсунки.

Разработан уникальный алгоритм с использованием вейвлет-функций для экспериментального исследования структуры двухфазного потока в факеле форсунки при помощи системы PIV «Полис», реализация которого ориентирована для параллельного счета на суперкомпьютере.

Один из простейших вариантов обработки изображений состоит в следующем. Исходя из средней засветки выбирается

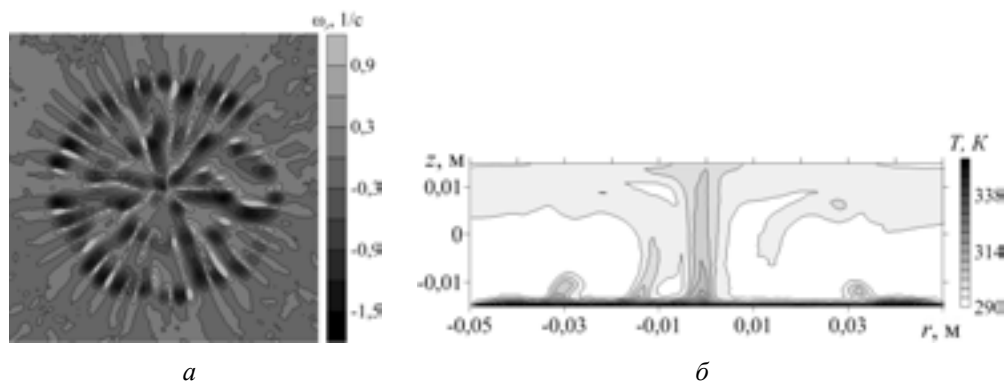


Рис. 15. Поле радиальной компоненты завихренности на высоте $h = 3$ мм (а), всплывающие конвективные струи в центральном вертикальном сечении (б)

порог интенсивности, что позволяет бинаризовать изображения. Далее соседние ненулевые пиксели объединяются в группы, которые при превышении определенного числа пикселей в группе рассматриваются как блик. После это осуществляется поиск парных бликов, ближайших друг к другу. Такой алгоритм позволяет находить пары бликов («двойные блики») и восстанавливать размеры капель.

Задача становится более сложной, если изображение характеризуется высоким уровнем шума, связанного с наличием крупных засвеченных объектов. Блики могут иметь неправильную форму, а интенсивность света в области блика может быть сильно неоднородна. Поэтому в случае изображений с высоким уровнем контраста и низким уровнем шумов удается получить хорошие результаты, однако в случае наличия заметных оптических дефектов нельзя использовать предложенную выше методику поиска пар бликов без предварительной фильтрации изображения от шумов и крупных оптических дефектов. Оптимальным способом оказалось применение вейвлет-преобразования с использованием вейвлет-функции

$$\psi(x, y) = (2 - (x^2 + y^2)) e^{-\frac{(x^2 + y^2)}{2}}$$
 («мексиканской шляпы»). Масштаб вейвлет-функции подбирался по среднему размеру бликов.

После вейвлет-преобразования изображения получаем плоскость вейвлет-коэффициентов (рис. 16, б), на которой блики имеют ряд отличительных признаков. В

центре блика – максимум интенсивности (темное пятно), а на периферии – область отрицательных значений вейвлет-коэффициентов (светлое кольцо). Объекты такой структуры достаточно надежно определяются по величине коэффициента кросс-корреляции с шаблоном. Затем по парам соседних бликов определяется размер капли (рис. 16, б).

Ключевой особенностью предложенной методики является процедура валидации, то есть отбора только тех парных бликов, которые действительно сформированы на одной капле. В условиях достаточно плотных потоков или при высоком уровне шума корректная валидация является необходимым условием для применения того или иного метода измерений. Для процедуры валидации предлагается использовать два изображения потока, полученных с малой (заданной) задержкой по времени. Подобным образом производится съемка изображений для восстановления поля скорости методом PIV. Затем анализируется смещение пар бликов на втором изображении относительно первого. Если оба блика смещаются синхронно, то предполагается, что данная пара бликов валидна и может быть использована для измерения размеров капли, в противном случае данные пары бликов не валидны и исключаются из рассмотрения. Пример использования процедуры валидации показан на рис. 17. Валидные капли соединены прямыми линиями.

Другие капли, обнаруженные на первом этапе валидации и отмеченные окружностями, но для которых не нашлось

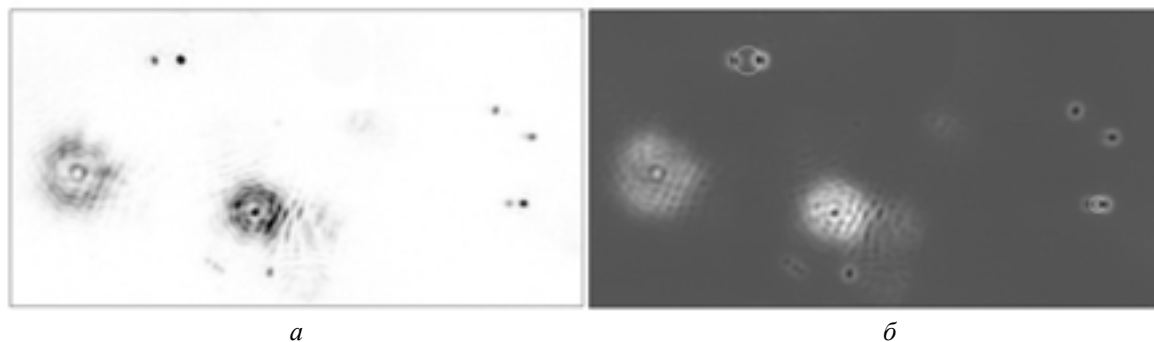


Рис. 16. Применение вейвлет-обработки для анализа изображений: а – исходное изображение; б – результат обработки (распределение вейвлет-коэффициентов). Окружности показывают пары бликов, по которым производится восстановление размеров капель



Рис. 17. Пример использования процедуры валидации по смещению парных бликов. Показано наложение результатов обработки двух последовательных изображений. Линиями показано смещение капель

пары на втором кадре, не используются в дальнейшем. Надо подчеркнуть, что в результате использования процедуры валидации помимо получения достоверной информации о размерах капель, восстанавливается также и скорость капель. Таким образом, появляется возможность исследовать зависимость скорости капель от их размера, восстанавливать пространственное распределение расхода (в плос-

кости лазерного ножа). Детальное описание разработанной методики можно найти в [26]. Распределение размеров, восстановленное при помощи предлагаемой методики, показано на рис. 18.

Измерения производились в факеле распыла пневматической форсунки, в области размерами $14 \times 9 \text{ мм}^2$, для двух режимов распыливания – режима «холостого хода» (рис. 18, а) и рабочего режима (рис. 18, б).

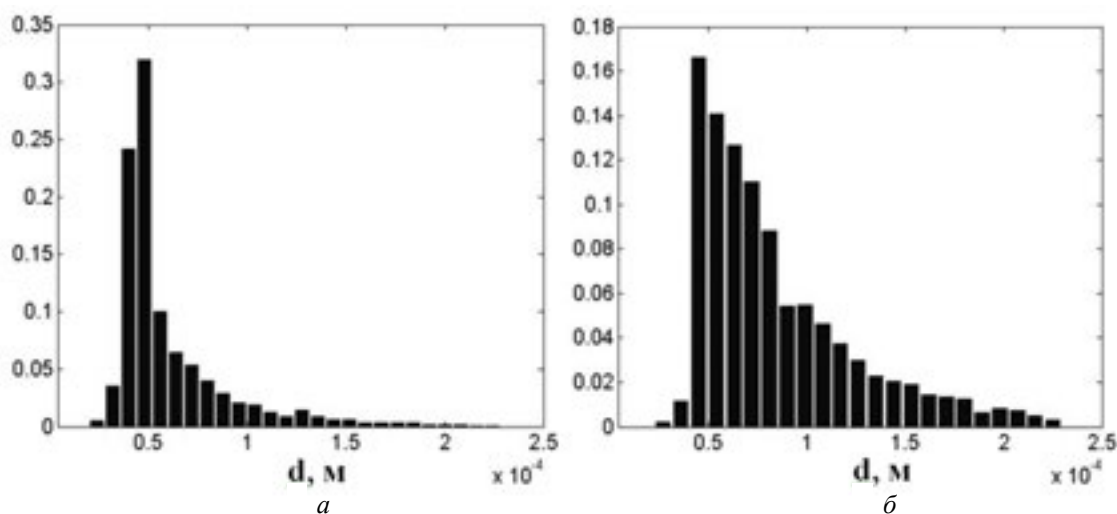


Рис. 18. Распределение капель по размерам, а – для $p = 0,1 \text{ атм.}$, б – для $p = 10 \text{ атм.}$, по вертикали отложена относительная доля капель

Хорошо видно, что с ростом расхода увеличивается размер капель, в режиме «холостого хода» заутеровский диаметр $d_{32} = 95$ мкм, а средний диаметр $d = 60$ мкм, в рабочем режиме размер капель увеличивается: $d_{32} = 124$ мкм, $d = 84$ мкм. Интересно, что с увеличением расхода заметно меняется структура распределения капель в плоскости скорость-диаметр (рис. 19), хорошо видно, что большая часть крупных капель движется с высокой скоростью.

Возможность распараллеливания созданных алгоритмов позволяет выполнять обработку потока измеряемых данных от установки PIV (ИМСС) на удаленном суперкомпьютере (ИММ) через скоростную сеть. Используется для оптимизации конструкций форсунок газотурбинных авиационных двигателей.

3.5. Экспериментальное исследование крупномасштабной циркуляции в кубической полости.

В настоящее время большой интерес представляют гидродинамические системы, которые, несмотря на достаточно простую геометрию, демонстрируют сложное динамическое поведение. Примером такой системы является турбулентная конвекция Рэлея–Бенара в прямоугольной полости. Показано, что в данной конвективной системе на фоне развитой турбулентности возникает крупномасштабный поток с нетривиальным характером эволюции. Для турбулентной

конвекции в рассматриваемом классе прямоугольных полостей измерения динамики поля скорости показали, что в зависимости от значений числа Рэлея и аспектного отношения могут реализоваться три различных режима крупномасштабной циркуляции (КМЦ) (рис. 20).

Первому режиму свойственна устойчивая циркуляция, интенсивность которой демонстрирует стохастические колебания, но направление циркуляции остается неизменным. *Второй режим* – режим с инверсиями, имеет чередование интервалов времени с КМЦ в том или ином направлении. Длительность этих интервалов случайна, а в пределах каждого интервала КМЦ ведет себя подобно режиму 1. *В третьем (смешанном) режиме* наблюдаются многочисленные смены направления КМЦ, которые не разделяются интервалами с квазиустойчивой циркуляцией в одном направлении.

В ходе проведенной работы было показано, что задача о конвекции Рэлея–Бенара в прямоугольных полостях с различным аспектным отношением совмещает простоту постановки с широким диапазоном реализующихся в ней режимов конвекции при высоких значениях числа Рэлея. Задача может служить одним из бенчмарков для CFD-кодов, обеспечивающих, в том числе, расчеты турбулентного теплообмена в атомной энергетике. Благодаря канонической геометрии расчетной области можно

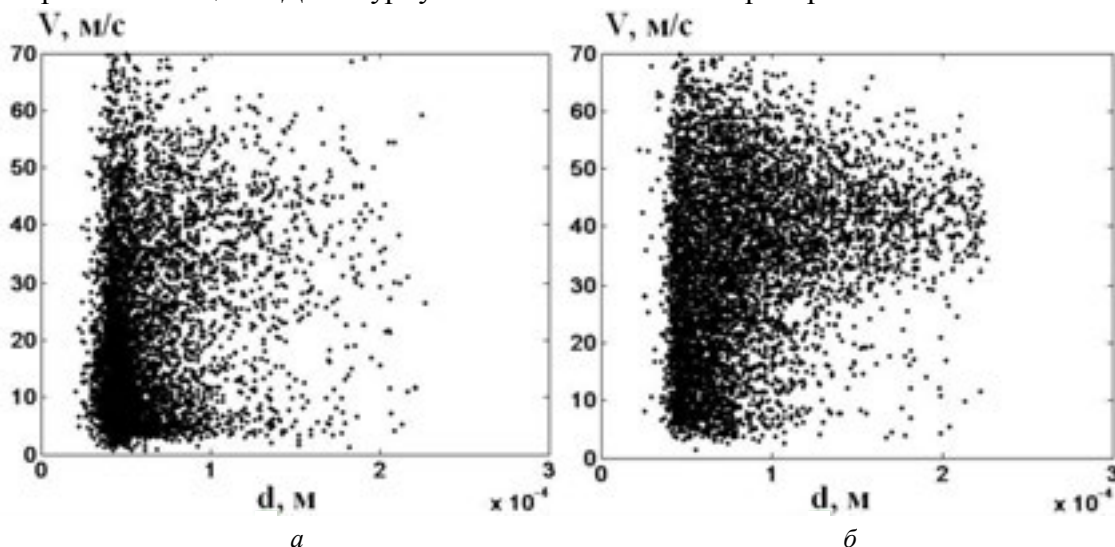


Рис. 19. Распределение капель в плоскости V - d : а – для $p = 0,1$ атм., б – для $p = 10$ атм.

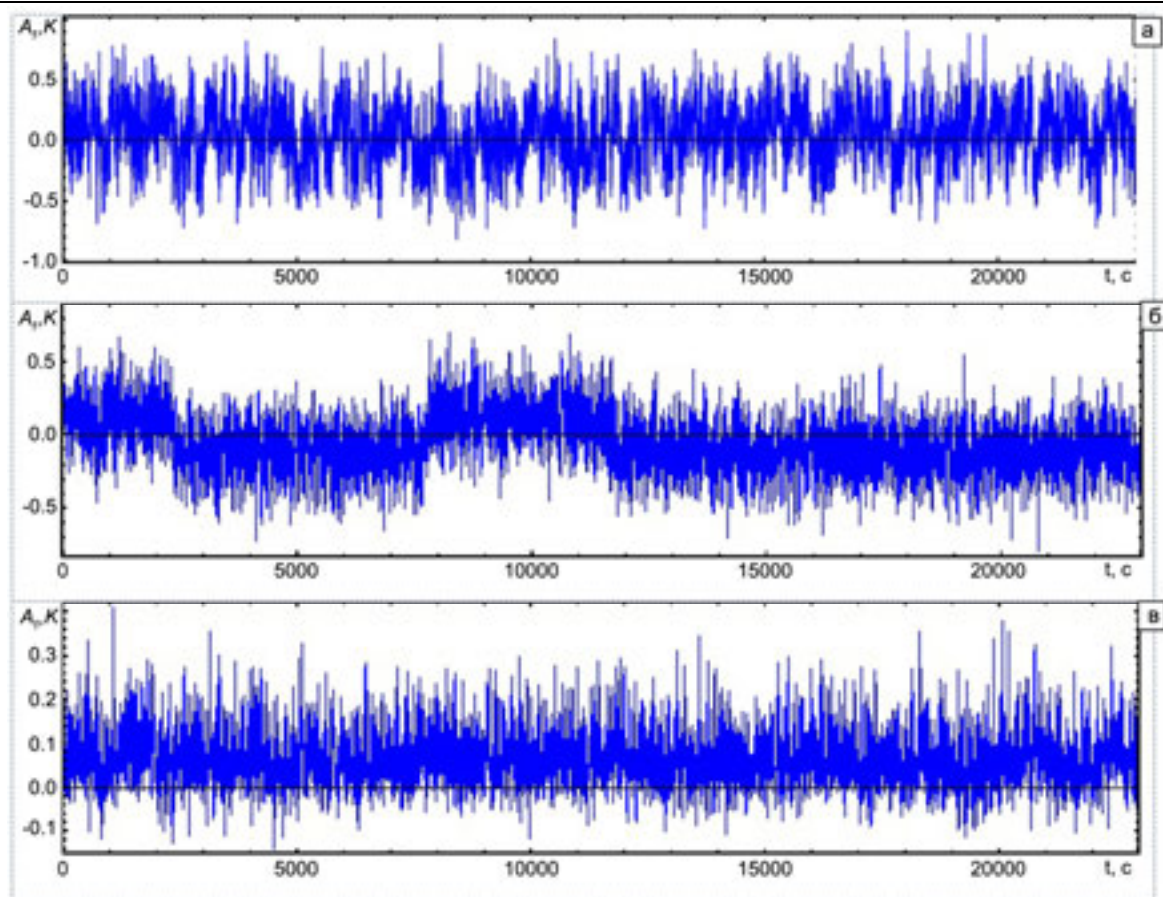


Рис. 20. Вариации амплитуды крупномасштабных пульсаций температуры при $Ra = 4,4 \cdot 10^9$ в полостях с различной геометрией: $\Gamma = 0,1$ (смешанный режим) (а), $\Gamma = 0,2$ (режим с инверсиями) (б), $\Gamma = 1,0$ (режим без инверсий) (в)

надеяться на воспроизведение экспериментов различными исследовательскими группами с целью получения достоверной экспериментальной базы данных. Более подробно результаты работы изложены в [27].

Особенностью данных исследований является проведение длительных серий измерений (более суток). Время обработки данных на персональном компьютере, входящем в состав установки PIV, занимает на порядок большее время, чем реальное время эксперимента. Значительные временные затраты затрудняют проведение исследований. Применение предложенных в проекте решений позволяет существенно ускорить время обработки результатов и обеспечить полномасштабное исследование всех режимов КМЦ.

Выводы.

Отличительной особенностью созданной Lambda Ggrid технологии является параллелизм передачи и обработки данных:

- прикладные задачи в оконечных системах – уникальные параллельные алгоритмы обработки данных;
- технологии передачи данных – темное оптоволокно и система спектрального уплотнения каналов;
- соединения между оконечными системами – эффективная диспетчеризация параллельных потоков данных.

Область применения. Создание интегрированной платформы измерительного комплекса (изображенный на рис. 21 Distributed PIV), включающего высокотехнологичное измерительное оборудование и производительную вычислительную технику, обеспечит условия для проведения уникальных физических экспериментов в исследовательских лабораториях, а на производстве – для организации диагностических стендов с возможностью real-time-мониторинга. Скоростные каналы передачи данных позволяют связать место

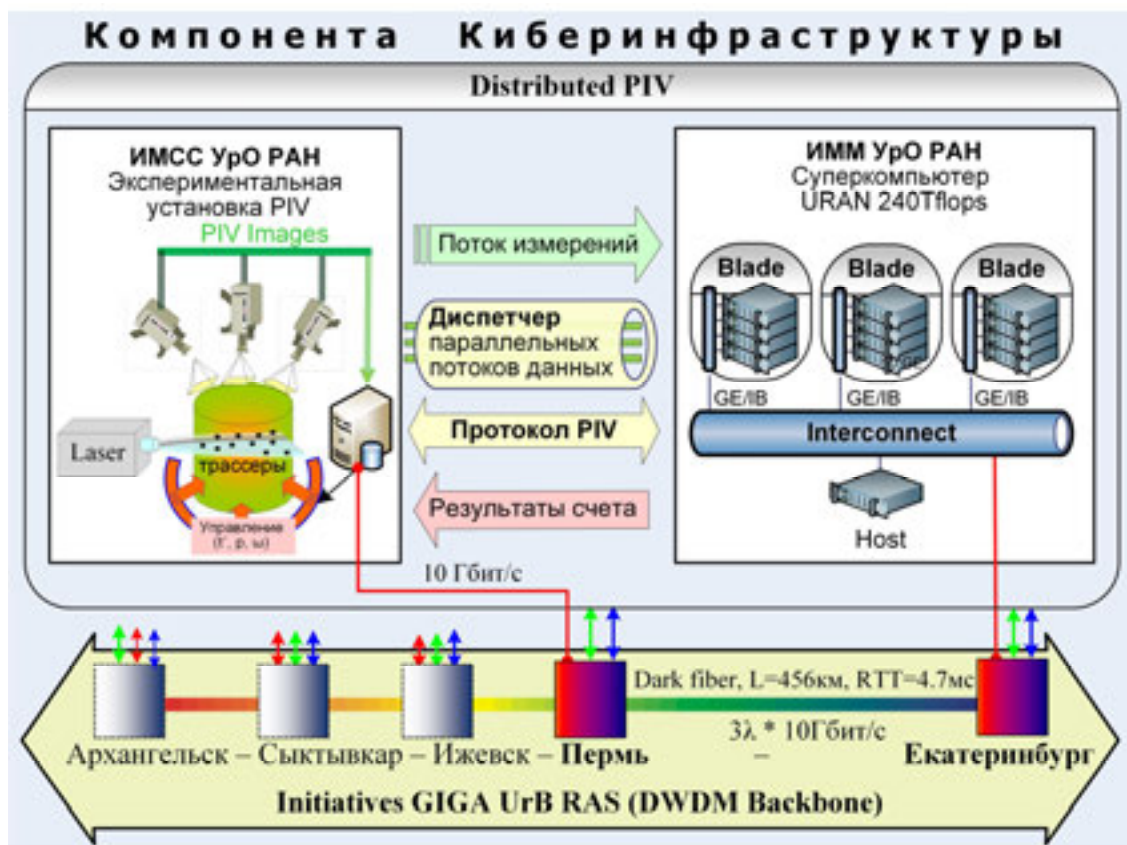


Рис. 21. Прототип киберинфраструктуры «УРАЛ»

проведения измерений с суперкомпьютерами, которые, как правило, находятся в центрах коллективного пользования. В результате отпадает необходимость наращивания локальных вычислительных ресурсов и тем самым снижаются финансовые затраты на проведение исследований.

Планируемое направление дальнейших работ: исследование различных аспектов параллельных, распределенных, сетевых и облачных вычислений, в том числе проект «Распределенный интерконнект» для

создания территориально распределенных вычислительных сред и систем хранения данных [5, 16, 19].

Уместно также отметить, что успешность разработки новых парадигм распределенных вычислений возможна только при тесном взаимодействии архитекторов многопроцессорных систем и сетевых технологий, профессионалов по системному и параллельному программированию, ученых и экспериментаторов разных предметных областей.

Библиографический список

1. Андрейчиков Ю.С., Воронова Л.А., Токмакова Т.Н., Тендрякова С.П., Онорин А.А. А.с. 529162 (СССР). Бюл. изобрет. – 1976. – № 35.
2. Степанов Р.А., Масич А.Г., Сухановский А.Н., Щапов В.А., Игумнов А.С., Масич Г.Ф. Обработка на супервычислителе потока экспериментальных данных // Вестник УГАТУ, Уфа, 2012. – Т. 16. – № 3 (48). – С. 126–133.
3. Щапов В.А., Масич А.Г., Масич Г.Ф. Модель потоковой обработки экспериментальных данных в распределенных системах // Вычислительные методы и программирование. – 2012. – Раздел 2. – С. 139–145. (<http://num-meth.srcc.msu.ru/>).
4. Shchapov V., Masich A. Protocol of High Speed Data Transfer from Particle Image Velocimetry System to Supercomputer // Proc. of The 7th International Forum on Strategic Technology (IFOST 2012) September 18–21, 2012, Vol. 1. National Research Tomsk Polytechnic University, Tomsk. – P. 653–657.
5. Большухин М.А., Васильев А.Ю., Будников А.В., Патрушев Д.Н., Романов Р.И., Свеишников Д.Н., Сухановский А.Н., Фрик П.Г. Об экспериментальных тестах (бенчмарках) для программных пакетов

- обеспечивающих расчет теплообменников в атомной энергетике // Вычислительная механика сплошных сред. – 2012. – Т. 5. – № 4. – С. 469–480.
6. *Созыкин А.В., Гольдштейн М.Л., Масич Г.Ф., Масич А.Г.* Территориально-распределенная система хранения данных среды высокопроизводительных вычислений УрО РАН // Научно-технический вестник Поволжья. – Казань, 2013. – № 5. – С. 291–293.
 7. *Степанов Р.А., Масич А.Г., Сухановский А.Н., Щапов В.А., Изумнов А.С., Масич Г.Ф.* Обработка на супервычислителе потока экспериментальных данных // Научный сервис в сети Интернет: экзафлопсное будущее: Труды междунар. суперкомпьютерной конференции (19–24 сентября 2011 г., Новороссийск). – М.: Изд-во МГУ, 2011. – С. 168–174.
 8. *Масич А.Г., Масич Г.Ф., Щапов В.А., Степанов Р.А.* Поточковая обработка больших массивов экспериментальных данных на удаленном суперкомпьютере // Тезисы междунар. конференции «Математические и информационные технологии MIT-2011», (27.08.–31.08.2011, Врнячка Баня, Сербия, 31.08. – 5.09.2011, Будва, Черногория). – Белград, 2011. – С. 105.
 9. *Масич А.Г., Масич Г.Ф., Тирон Г.Г.* Аспекты реализации первого этапа Инициативы GIGA UrB RAS // Сборник докладов XVIII конференции представителей региональных научно-образовательных сетей «RELARN-2011» (26 сентября – 2 октября 2011 г, Ростов-на-Дону – Казань). – М.: Изд-во Ассоциации РЕЛАРН, 2011. – С. 27–31.
 10. *Масич А.Г.* Модель обмена интенсивными потоками данных в оптических сетях // Там же. – С. 31–34.
 11. *Баталов В.Г., Колесниченко И.В., Степанов Р.А., Сухановский А.Н.* Применение полевых методов измерений для исследования двухфазных потоков // Вестник Пермского университета. Сер.: Математика. Механика. Информатика. – 2011. – № 5 (9). – С. 21–25.
 12. *Масич А.Г., Масич Г.Ф., Матвеев В.П., Тирон Г.Г.* Инициатива GIGA UrB RAS: методология построения и архитектура научно-образовательной оптической магистрали Уральского отделения РАН // Труды Межд. конференции «Математические и информационные технологии», MIT-2011. – Белград, 2012. – С. 257–265.
 13. *Масич А.Г., Масич Г.Ф., Щапов В.А., Степанов Р.А.* Поточковая обработка больших массивов экспериментальных данных на удаленном суперкомпьютере // Там же. – С. 266–270.
 14. *Масич А.Г., Масич Г.Ф.* Аспекты реализации научно-образовательной оптической магистрали УрО РАН // Труды XIX Всерос. науч.-метод. конференции «Телематика'2012» (25–28 июля 2012 г., Санкт-Петербург). – СПб.: изд-во ИТМО, 2012. – С. 247–250.
 15. *Щапов В.А., Масич А.Г., Масич Г.Ф., Хохлов И.А.* Алгоритм распределения потока экспериментальных данных по вычислительным узлам суперкомпьютера // Доклады междунар. конференции «Интеллектуализация обработки информации» (16–22 сентября 2012 г., Черногория, г. Будва). – С. 699–702.
 16. *Щапов В.А., Масич А.Г., Масич Г.Ф.* Модель потоковой обработки экспериментальных данных в распределенных системах // Научный сервис в сети Интернет: поиск новых решений: Труды междунар. суперкомпьютерной конференции (17–22 сентября 2012 г., Новороссийск). – М.: изд-во МГУ, 2012. – С. 227–232.
 17. *Goldshstein M.L., Sozykin A.V., Masich G.F., Masich A.G.* Distributed High Performance Computing Environment of UrB RAS // Distributed Computing and Grid-Technologies in Science and Education: Book of Abstr. of the 5th Intern. Conf. – Dubna: JINR, 2012. – P. 71.
 18. *Евграфова А.В., Левина Г.В., Сухановский А.Н.* Исследование влияния вторичных структур на генерацию спиральности в адвективном потоке // Тезисы Всерос. науч. школы молодых ученых «Волны и вихри в сложных средах». – Москва, 3–5 декабря, 2012 г. – С. 81.
 19. *Sukhanovsky A., Batalov V., Stepanov R.* Measurements of the droplets size and velocity in two-phase flows // Abstracts of the 9th European Fluid Mechanics Conference. – Rome, 9–13 September, 2012.
 20. *Гольдштейн М.Л., Созыкин А.В., Масич Г.Ф., Масич А.Г.* Вычислительные ресурсы УрО РАН. Состояние и перспективы // Параллельные вычислительные технологии (ПаВТ'2013): труды междунар. науч. конференции (1–5 апреля 2013 г., Челябинск). – Челябинск: изд. центр ЮУрГУ, 2013. – С. 330–337.
 21. *Щапов В.А.* Программная архитектура системы передачи интенсивного потока данных в распределенных системах // Там же. – С. 566–576.
 22. *Масич А.Г., Масич Г.Ф.* Методология построения территориальной киберинфраструктуры // Сб. докл. XX конференции представителей региональных научно-образовательных сетей «RELARN-2013» (1–6 июня 2013 г., Санкт-Петербург). – М.: изд-во Ассоциации РЕЛАРН, 2013. – С. 58–67.
 23. *Масич А.Г., Масич Г.Ф., Сидоров И.А.* Распределенный интерконнект: структура параллельных каналов и инструментарий исследования транспортных потоков в них // Научный сервис в сети Интернет: все грани параллелизма: Труды междунар. суперкомпьютерной конференции (23–28 сентября 2013 г., Новороссийск). – М.: изд-во МГУ, 2013.

24. Баталов В.Г., Колесниченко И.В., Степанов Р.А., Сухановский А.Н. Два метода измерения размеров капель // Сборник тезисов XVIII Зимней школы по механике сплошных сред (18–22 февраля 2013 г., Пермь) – С. 40.
25. Sukhanovsky A., Batalov V., Teymurazov A., Frick P. Horizontal rolls in convective flow above a partially heated surface // Eur. Phys. J. B. – 2012. – Vol. 85. – № 1. – P. 1–12.
26. Евграфова А.В., Левина Г.В., Сухановский А.Н. Исследование распределения завихренности и спиральности при взаимодействии адвективного потока с вторичными структурами // Вычислительная механика сплошных сред. – 2013. – Т. 6. – № 4. – С. 451–459.
27. Баталов В.Г., Степанов Р.А., Сухановский А.Н. Применение прямых оптических методов измерений для исследования характеристик двухфазного потока // Электронный журнал «Труды МАИ». – 2014. – № 76. – 20 с.
28. Большуихин М.А., Васильев А.Ю., Будников А.В., Патрушев Д.Н., Романов Р.И., Свешиников Д.Н., Сухановский А.Н., Фрик П.Г. Об экспериментальных тестах (бенчмарках) для программных пакетов обеспечивающих расчет теплообменников в атомной энергетике // Вычислительная механика сплошных сред. – 2012. – Т. 5. – № 4. – С. 469–480.

RESEARCH AND DEVELOPMENT OF THE PROCESS OF FORMATION OF A REGIONAL CYBERINFRASTRUCTURE ON THE BASIS OF LAMBDA-GRID TECHNOLOGIES

G.F. Masich, A.G. Masich, R.A. Stepanov, A.N. Sukhanovskii, A.V. Bobrov,
D.P. Chugunov, G.G.Tiron, R.A. Fayzrakhmanov, V.A. Shchapov

The paper provides information about the cyberinfrastructure created in the Urals that is designed to use all of the improvements in the field of high performance computing and consists of high-speed integrated optical network calculators, data storage systems and installations. The paper shows the problems of the organization of connections between end systems for high-speed extended communication lines and solutions to them through parallelism of their compounds. Particular attention is paid to the architectural solution that involves processing intensive flow of experimental data on a remote supercomputer in real time, demonstrating fundamentally new opportunities for experimental studies.

Keywords: cyberinfrastructure, optical network, connection, supercomputer, experimental stand.

Сведения об авторах

Масич Григорий Федорович, кандидат технических наук, заведующий лабораторией, Институт механики сплошных сред (ИМСС УрО РАН), 614013, г. Пермь, ул. Академика Королева, 1; e-mail: masich@icmm.ru

Масич Алексей Григорьевич, младший научный сотрудник, ИМСС УрО РАН; e-mail: mag@icmm.ru

Степанов Родион Александрович, доктор физико-математических наук, ведущий научный сотрудник, ИМСС УрО РАН; e-mail: rodion@icmm.ru

Сухановский Андрей Николаевич, кандидат физико-математических наук, старший научный сотрудник, ИМСС УрО РАН; e-mail: san@icmm.ru

Бобров Артем Викторович, младший научный сотрудник, ИМСС УрО РАН; e-mail: bobrov@icmm.ru

Чугунов Денис Петрович, ведущий инженер, ИМСС УрО РАН; e-mail: chugunov@icmm.ru

Тирон Георгий Георгиевич, кандидат экономических наук, старший научный сотрудник, Институт экономики УрО РАН, 614990, г. Пермь, ул. Ленина, 13а, e-mail: tiron@tiron.ru

Файзрахманов Рустам Абубакирович, доктор экономических наук, заведующий кафедрой, Пермский национальный исследовательский политехнический университет (ПНИПУ), 614013, г. Пермь, ул. Профессора Поздеева, д. 7, ауд. 227; e-mail: itas@pstu.ru

Щапов Владислав Алексеевич, ассистент, ПНИПУ; e-mail: shchapov@icmm.ru

Материал поступил в редакцию 26.06.2014 г.