



Начальный этап исследования стенда передачи Больших Данных по параллельным каналам данных (ПКС подход)

С. Хоружников¹ В. Грудинин¹ О. Садов^{1,2} А. Шевель^{1,2}
А. Каирканов¹ В. Титов^{1,3}

¹СПб НИУ ИТМО

²НИЦ «Курчатовский институт»

³СПбГУ

22.07.2014



- Источники Больших Данных
- Архитектура Больших Данных
- Технология передачи Больших Данных
- Наши исследования



Источники Больших Данных

- Экспериментальные установки
 - Широкоугольный обзорный телескоп-рефлектор (Large Synoptic Survey Telescope, **LSST**, ≈ 2020)
<http://www.lsst.org>, 15 ТБ/ночь (10 ПБ/год)
 - Радиointерферометр «Квадратная километровая решетка» (Square Kilometre Array, **SKA**, $\approx 2019-2024$)
<https://www.skatelescope.org/>, 300-1500 ПБ/год
 - **ЦЕРН**
<http://www.cern.ch>, 20 ПБ/год (обрабатывается 1 ПБ/день)
 - Международный экспериментальный термоядерный реактор (International Thermonuclear Experimental Reactor, **ITER**, ≈ 2020)
<http://www.iter.org>, 1 ПБ/год
 - Решетка черенковских телескопов Cherenkov Telescope Array, **СТА**, $\approx 2015-2020$
<http://www.cta-observatory.org/>, 20 ПБ/год
 - **ГАЙЯ**
Global Astrometric Interferometer for Astrophysics, **GAIA 2014**
<http://sci.esa.int/gaia/>, 200 ТБ/год – 1 ПБ/год



Рост сетевого трафика

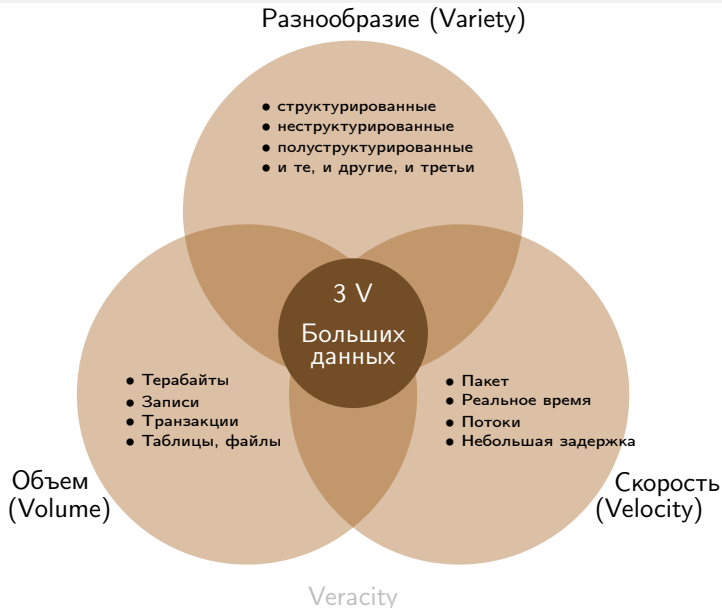
Тенденция роста трафика ESnet



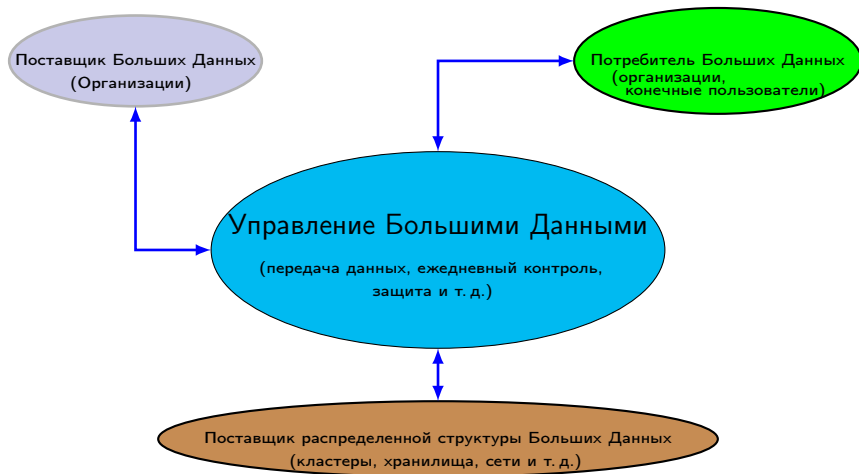
График ежемесячного входного трафика ESnet,
январь 1990 – декабрь 2012



3 V:



Архитектура Больших Данных



Особенности передачи Больших Данных

- Передача может занять много часов или дней
(при пропускной способности в 1 Гбит 100 ТБ/100 МБ=1 000 000 секунд или 11.6 дней).
- Обстановка в каналах может измениться: время прохождения сигнала (RTT), % потерянных сетевых пакетов, пропускная способность канала данных.
- Наконец, может случиться сбой в работе канала данных (часы?, дни?).
- Очевидно, полезно иметь доступ к двум или более каналам данных и механизмы динамического распределения потоков передачи по этим каналам.



Технологические особенности передачи Больших Данных

- Основные протоколы: все еще стек TCP/IP.
 - Число сетевых параметров в Linux около 500 (/proc)
-bash-4.1\$ /sbin/sysctl -a | grep "^net\" | wc -l
 - Важные параметры: размер блока, размер окна TCP и т. д.
Основной способ уменьшить время передачи (даже по одному каналу данных) — использовать мультипоточковую передачу.



Тестирование на первой стадии (программные инструменты)

- [BBCP](http://www.slac.stanford.edu/~abh/bbcp/), <http://www.slac.stanford.edu/~abh/bbcp/>
- [GridFTP](http://www.globus.org/toolkit/data/gridftp/), <http://www.globus.org/toolkit/data/gridftp/>
- [BBFTP](http://doc.in2p3.fr/bbftp/), <http://doc.in2p3.fr/bbftp/>
- [FDT](http://monalisa.cern.ch/FDT/), <http://monalisa.cern.ch/FDT/>
- [FTS3](http://fts3-service.web.cern.ch/), <http://fts3-service.web.cern.ch/>
- а также технологические компоненты для контроля состояния канала данных, [perfSONAR](#).



Сравнение инструментов передачи данных

- Доступность
- API
- Производительность
- Надежность
- Отслеживание операций
- Возможность предсказать время передачи данных на основе существующих отслеженных записей
- Требуемые ресурсы: память, время ЦПУ и т. д.
- Другое

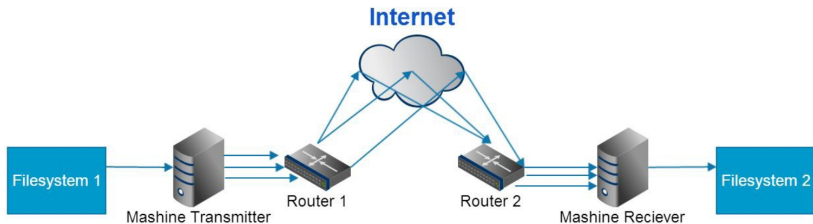
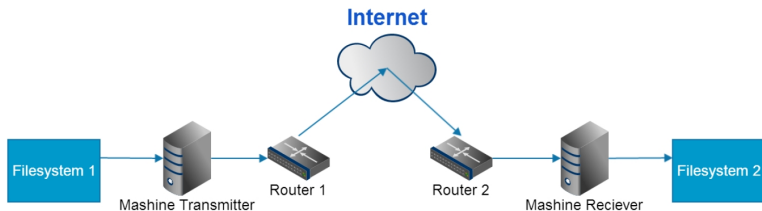


Характеристики утилит передачи данных

- Режим многопоточной передачи данных.
- Режим многоканальной передачи.
- Возможность устанавливать параметры нижнего уровня (например, размер окна TCP).
- Шифрование данных на лету.
- Сжатие данных на лету.
- Метод обхода сетевых проблем.



Процесс передачи данных



Из нескольких шагов передачи данных (чтение данных из хранилища, передача данных по сети, запись полученных данных) мы концентрируемся на втором, **передаче данных по сети**.



Тема исследования в НИУ ИТМО: передача Больших Данных

В лаборатории сетевых технологий, <http://sdn.ifmo.ru/>, Университета ИТМО, <http://www.ifmo.ru/>, выполняется новая научно-исследовательская работа «Передача Больших Данных по Интернету».

- Создается распределенный испытательный стенд (100 Тб дисковой памяти + сервер с 96 Гб оперативной памяти под управлением ОС ScientificLinux 6.5 с каждой стороны).
 - Сравнительное изучение существующих инструментов передачи данных (тестирование и измерение).
 - Использование стенда как инструмента для сравнения различных средств (отслеживание для измерений + результаты).
 - Расширенная информация автоматического трекинга об измерениях в процессе разработки.

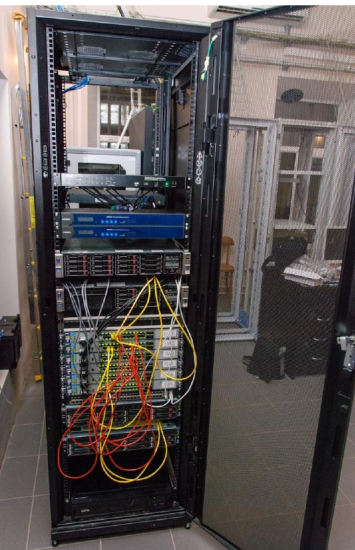


Планируемые измерения

- Локальные и удаленные сайты с существующими каналами данных (не только самые быстрые каналы).
- Замысел — использовать более, чем один, канал данных в параллель.
- В последнее время мы провели ряд исследовательских работ в области Программно-Конфигурируемых Сетей (ПКС-SDN) **OpenFlow** и теперь планируем использовать полученный опыт для совершенствования механизмов передачи Больших Данных.



Что сделано



- Используются:
 - Два сервера HP DL380p Gen8 E5-2609, Intel(R) Xeon(R) CPU E5-2640| 2.50GHz, 64 GB, ScientificLinux 6.5.
 - Шесть коммутаторов HP-3500-24G-PoE y1 (OpenFlow 1.0)
 - Pica8 P-3920 (OpenFlow 1.2)
 - Openstack Havana с подходящим набором виртуальных машин для тестирования упомянутых утилит. Openvswitch
 - PerfSonar
 - Скрипты для тестирования: <https://github.com/itmo-infocom/BigData>



Главные цели

- Объединить разработанные современные компоненты и методы с концепциями, разработками, опытом для достижения максимальной скорости передач Больших Данных на существующих каналах.
- Создать стенд, который использовался бы как место, где исследователи могли бы сравнить свои (новые) инструменты передачи данных с ранее полученными результатами измерений.
- Предложить сотрудничество с . . . (предложения?).
- Пригласить студентов из . . . (предложения?).



Партнеры (обмен замыслами)

- Лаборатория информационных технологий (ЛИТ)
<http://lit.jinr.ru/index.php?lang=lat>
Объединенный институт ядерных исследований (JINR.ru)
- Центр прикладных исследований компьютерных сетей,
Московский университет, <http://arccn.ru/>
- Начинаем сотрудничество с GENI, <http://www.geni.net/>

Работа поддержана Санкт-Петербургским национальным университетом информационных технологий, механики и оптики (www.ifmo.ru).



- Размер окна TCP;
- число потоков TCP;
- Размер буфера ввода/вывода;
- сжатие на лету;
- многонаправленное копирование;
- возобновление неудачного копирования;
- аутентификация по ssh;
- использование конвейеров, где источник и/или получатель могут быть конвейерами;
- специальная опция для передачи малых файлов;
- и много других опций, касающиеся многих практических деталей.



- зашифрованные при соединении имя и пароль;
- модули аутентификации SSH и Grid Certificate;
- многопоточковая передача;
- большие окна, как определено в RFC1323;
- сжатие данных на лету;
- автоматическое повторение;
- настраиваемые задержки;
- моделирование передачи;
- интеграция AFS аутентификации.



- две разновидности защиты: Globus GSI и SSH;
- файл с синонимами хостов: каждый следующий поток передачи данных будет использовать следующие синонимы хостов (полезно для вычислительных кластеров);
- конвейеры;
- специальный режим отладки для выяснения узких мест передачи данных;
- имя внутреннего модуля для сайтов источника и получателя;
- число параллельных потоков передачи данных;
- размер буфера;
- перезапуск неудачных операций и число перезапусков.



Другие утилиты

- Xdd — утилита, разработанная для оптимизации передачи данных и процессов ввода/вывода для систем хранения.
- fdp — Java-утилита для многопоточной передачи данных.
- FTS3.
- UDT.
- RDMA.
- MP TCP.



Скрипты

- Нужно рассматривать разные варианты данных для передачи: /dev/null (объемом 100 ТБ, без использования дисковой подсистемы), один файл >100 ТБ (с дисковой подсистемы), большое число небольших файлов, суммарным объемом 100 ТБ.
 - Генерация файлов: [create-test-file-dispersion.sh](#), [create-test-file-sigma.sh](#), параметры: директория, размер директории, средний размер файла, дисперсия, образец.
- Тестирование утилит:
 - [CopyData.bbcp](#), параметры: конфигурационный файл bbcp, директория для журналирования, исходная директория с файлом (файлами), удаленный хост, директория на удаленном хосте, комментарии.
- В процессе тестирования необходимо сохранять множества параметров, необходимых для дальнейшего анализа.
- Количественная оценка системы передачи данных включает время, затраченное на передачу данных, объем потребляемых ресурсов.
- Участвующие в испытаниях серверы:
 - Стенд ИТМО
 - Стенд ПИЯФ
 - Стенд МГУ



Спасибо...

Работа поддержана Санкт-Петербургским национальным университетом информационных технологий, механики и оптики (www.ifmo.ru).

