

Сравнительное изучение методов ускорения передачи данных через Интернет

Предварительные результаты измерений

Докладчик:

А.Е. Шевель

Авторы: В.А. Грудинин(*), Е. Корытько(*), А. Каирканов(*), О. Лазо(*),
А. А. Орешкин(**), О. Садов(*), В. Титов(***), С.Э. Хоружников(*),
А.Е. Шевель(**) (* – ИТМО, ** – ПИЯФ, *** – ПУ)

Краткий обзор

- Введение.
- Скорости передачи данных через Интернет и методы ускорения.
- Разработанная измерительная инфраструктура.
- Предварительные результаты измерений.
- Разработанный метод использования параллельных линий связи для ускорения передачи данных большого объёма.
- Тестовый сервер в ОФВЭ.

ICFA SCIC Report

Networking for High Energy Physics

Network bandwidth growth at «approximately a factor of 10 every 4.25 years. We can expect continued growth at a similar and perhaps a greater rate once the LHC resumes operation in 2015, with 13 TeV collisions, a higher luminosity and pileup, and anticipated greater trigger and data taking rates».

«The trend towards greater use of "location independent" data access is expected to continue, leading to an increasing need to manage the larger data flows, and match them to dynamic allocations of guaranteed bandwidth over long distances.»

Тема: копирование/реплицирование «больших» данных

- В лаборатория сетевых технологий <http://sdn.ifmo.ru/> в НИУ ИТМО <http://www.ifmo.ru/> сформировано направление исследований (среди других) «копирование/реплицирование данных большого объёма через Интернет».
- Создаётся распределённый Испытательный Стенд (100 ТБ дисковой памяти + сервер 96 GB оперативной памяти под OS RedHat/ScientificLinux с каждой стороны).
 - Разработка прототипа системы копирования/реплицирования данных большого объёма, которая может быть использована в конкретных экспериментах.
 - Основные методы ускорения передачи: настройка размеров передаваемых буферов, использование нескольких потоков передачи.

Тестирование на первом этапе (программные компоненты)

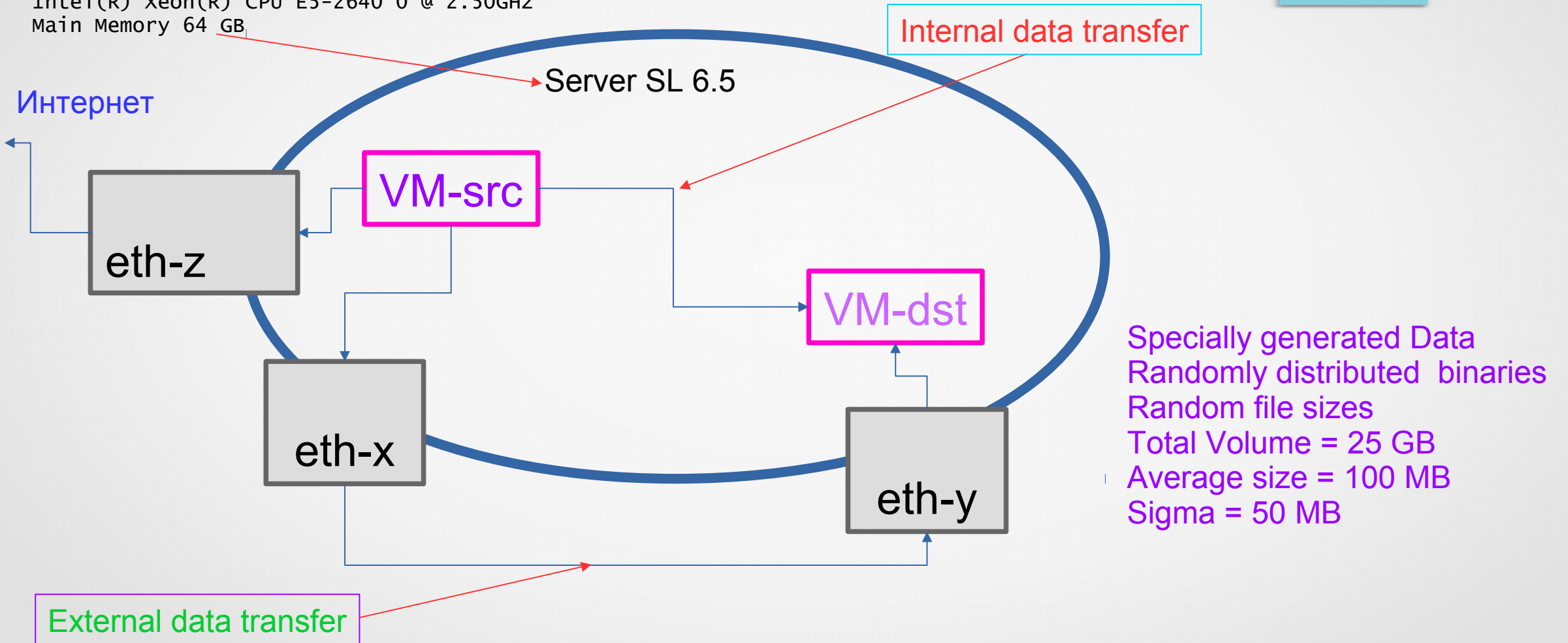
- BBSP - <http://www.slac.stanford.edu/~abh/bbcp/>
- BBFTP - <http://doc.in2p3.fr/bbftp/>
- FDT - <http://monalisa.cern.ch/FDT/>
- GridFTP - <http://www.globus.org/toolkit/data/gridftp/>
- Система наблюдения за состоянием линий связи: perfSONAR.

Организация тестирования и имеющаяся тестовая инфраструктура

- Необходим тестовый стенд, где условия измерений находятся под контролем исследователя.
- Каждое измерение должно выполняться с автоматической записью всех условий измерения (оборудование, параметры программ, сообщения, проч.) в специальной директории. Для каждой программы создано примерно 400 строк скриптов для выполнения измерений и около 300 строк скриптов для построения различных графиков.
- Поскольку скорость работы многих программ передачи данных зависят от содержания самих данных и размеров файлов, то на подготовку тестовых данных должно быть обращено специальное внимание.
- Все измерения выполняются с использованием виртуальных машин, которые организованы в рамках облачного инструментария Openstack (Icehouse).
- Подготовлена виртуальная машина с необходимой программной оснасткой, которую можно передать в любой центр обработки данных для проведения измерений.

Использованные схемы тестирования: ИТМО-ИТМО и ИТМО-Интернет

Intel(R) Xeon(R) CPU E5-2640 0 @ 2.50GHz
Main Memory 64 GB

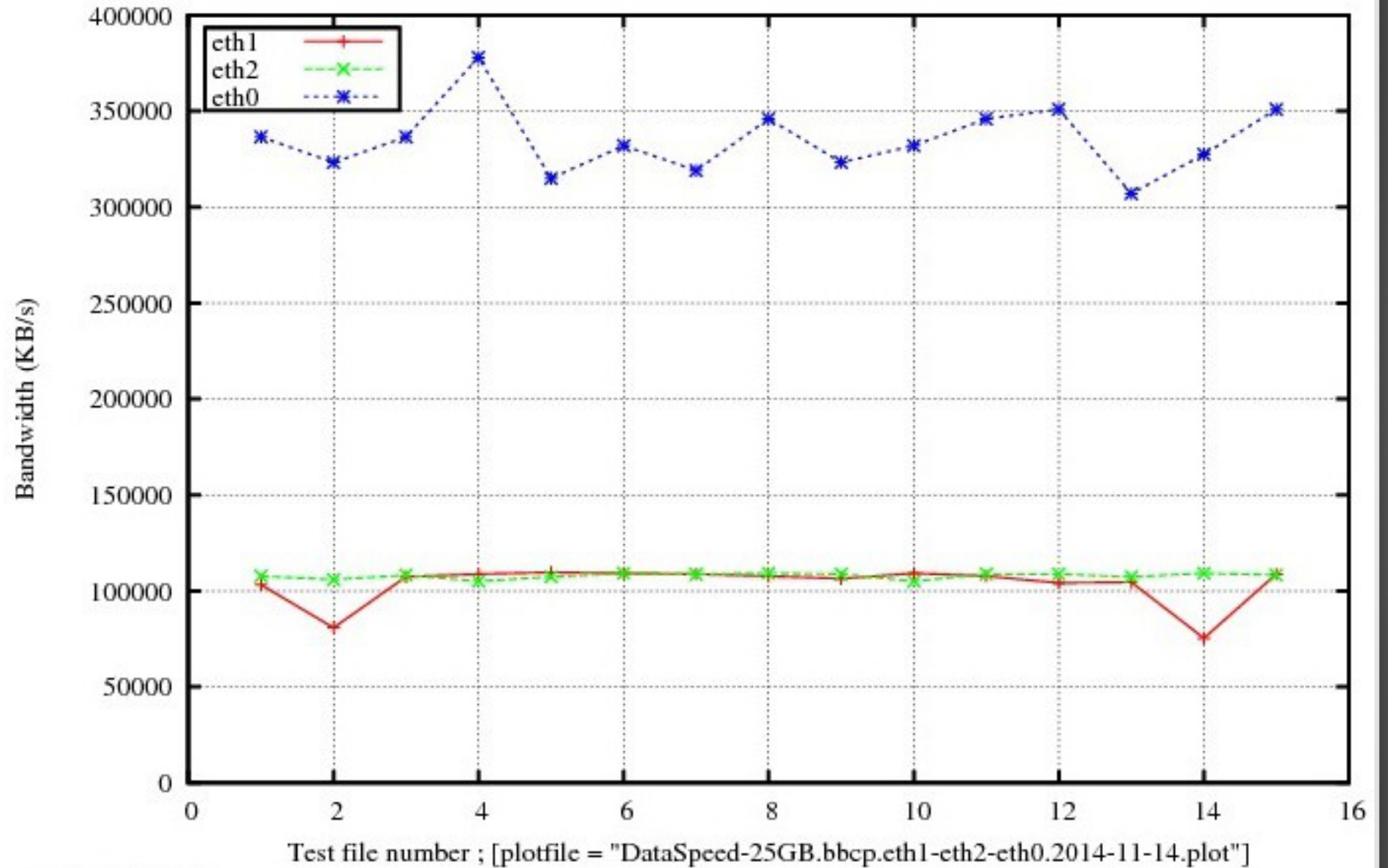


Скрипты для проведения тестов

- **create-test-file-sigma.sh DIRECTORY_NAME DIRECTORY_SIZE(KB) AVERAGE_SIZE(KB) DISPERSION SAMPLE_DESTINATION** — генерация случайных файлов в указанной директории
- **Комплект скриптов для каждой программы передачи данных:**
 - **CopyData.bbcp** — тестирование передачи тестовых данных программой BBSP
 - **Wrapper-CopyData.bbcp** — выполнении серии передач тестовых данных с разными параметрами программы BBSP
- **DataSpeed.plot.sh, gnuplot.sh** — скрипты построения графиков

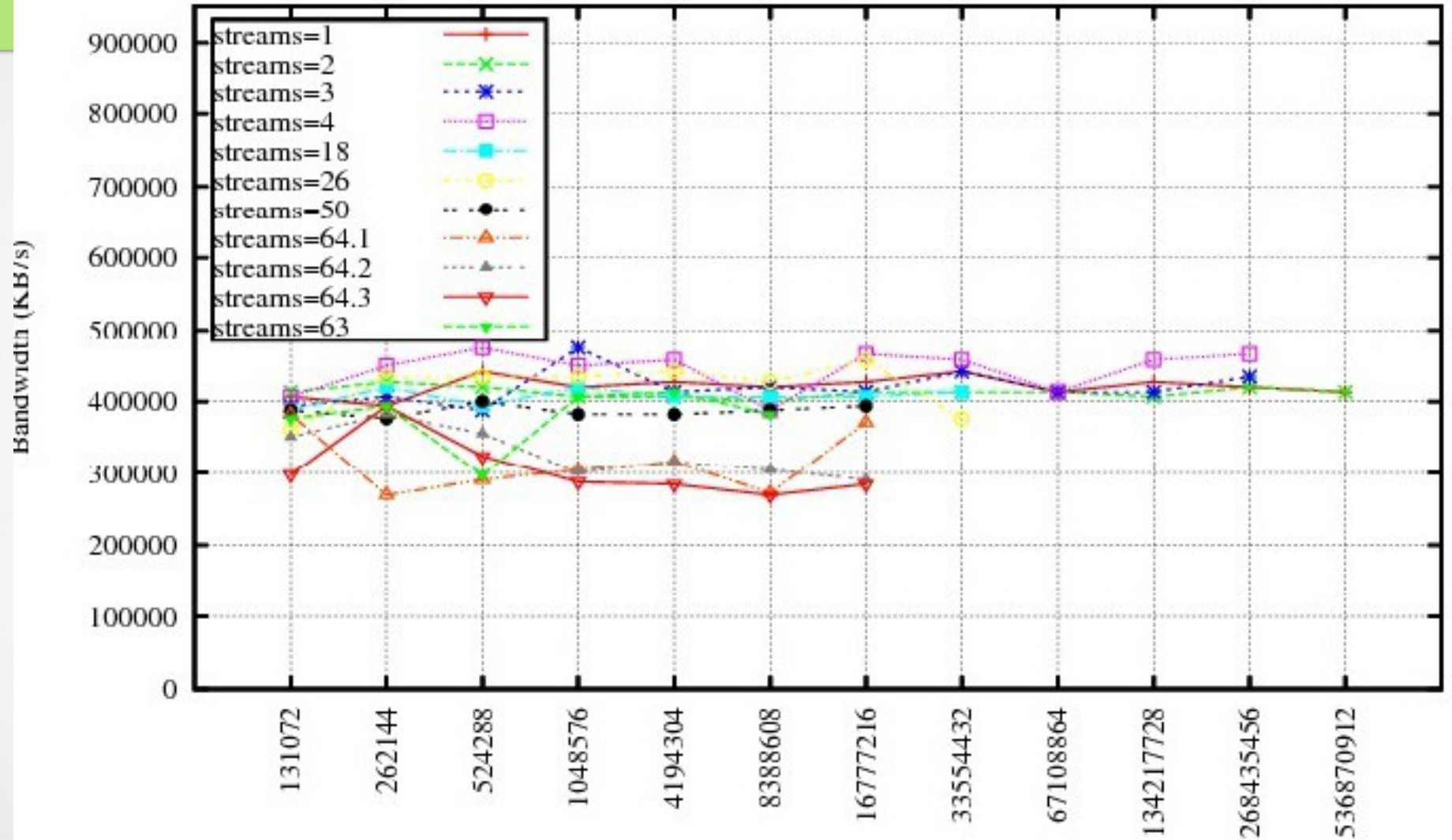
ITMO->ITMO(bbcpr)

Data Transfer Speed. Data Size 25 GB. BBCP.
Physical Connections eth1, eth2 and Virtual Connection eth0. 2014-11-14



ITMO -> ITMO(bbcP)

Data Transfer Speed. Data Size 25 GB. Different Number of Streams and TCP Window Size.
BBCP. VM ITMO. 2014-10-23 - 2014-10-28

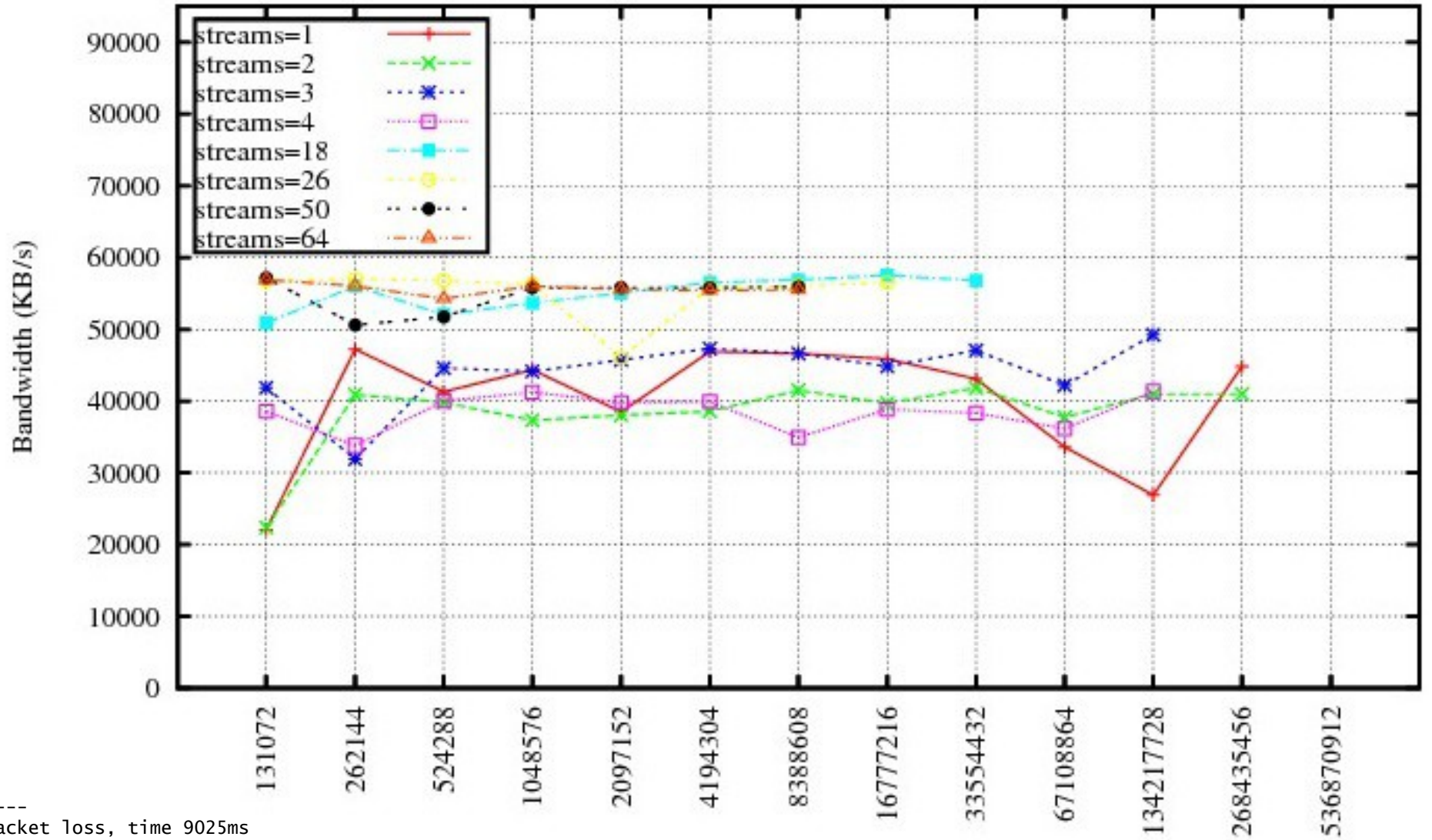


TCP Window Size; [plotfile = "DataSpeed-25GB.bbcP.itmo2itmo.tcp.2014-10-28.plot"]

ИИИИ ИИИИ (А. И.ИИИИ)

ITMO->PNPI(bbcp)

Data Transfer Speed. Data Size 25 GB. Different Number of Streams and TCP Window Size
BBCP from ITMO to PNPI. 2014-10-28

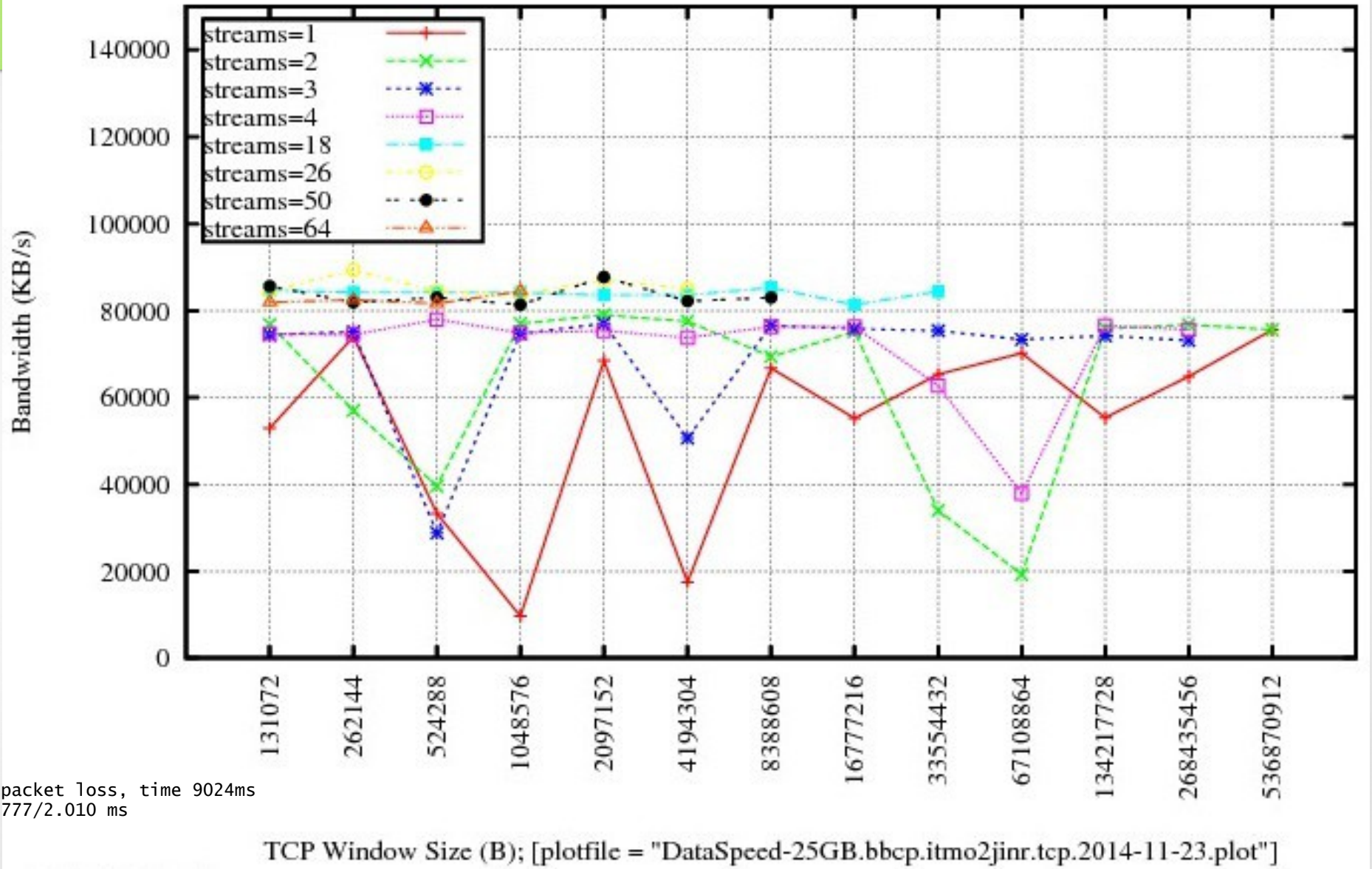


--- 1msys001.pnpi.spb.ru ping statistics ---
10 packets transmitted, 10 received, 0% packet loss, time 9025ms
rtt min/avg/max/mdev = 8.795/9.113/9.783/0.352 ms

TCP Window Size (B); [plotfile = "DataSpeed-25GB.bbcp.itmo2pnpi.tcp.2014-10-28.plot"]

HEPD PNPI (A. Y. Snevel)

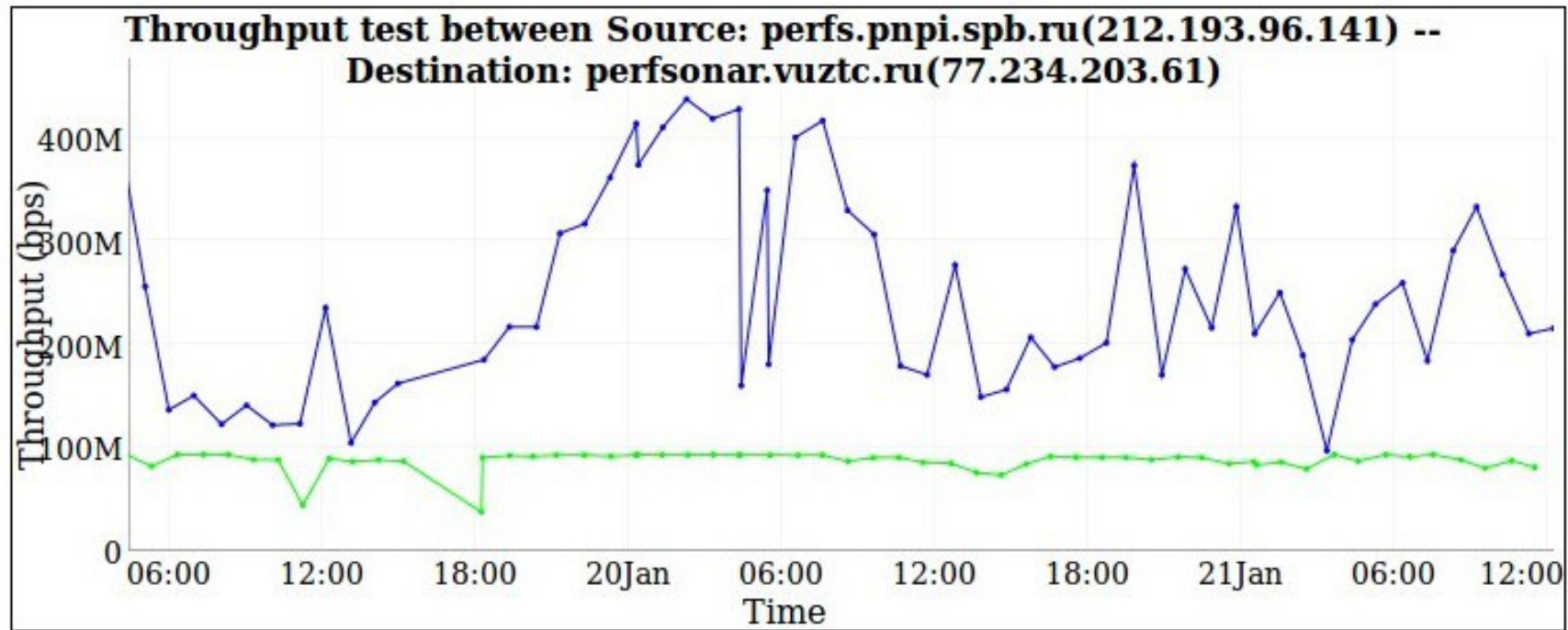
Data Transfer Speed. Data Size 25 GB. Different Number of Streams and TCP Window Size
BBCP from ITMO to JINR. 2014-11-23



--- 1xpub03.jinr.ru ping statistics ---
10 packets transmitted, 10 received, 0% packet loss, time 9024ms
rtt min/avg/max/mdev = 10.964/12.537/16.777/2.010 ms

TCP Window Size (B); [plotfile = "DataSpeed-25GB.bbcp.itmo2jinr.tcp.2014-11-23.plot"]
HEPD FNPI (A. Y. Sneyev)

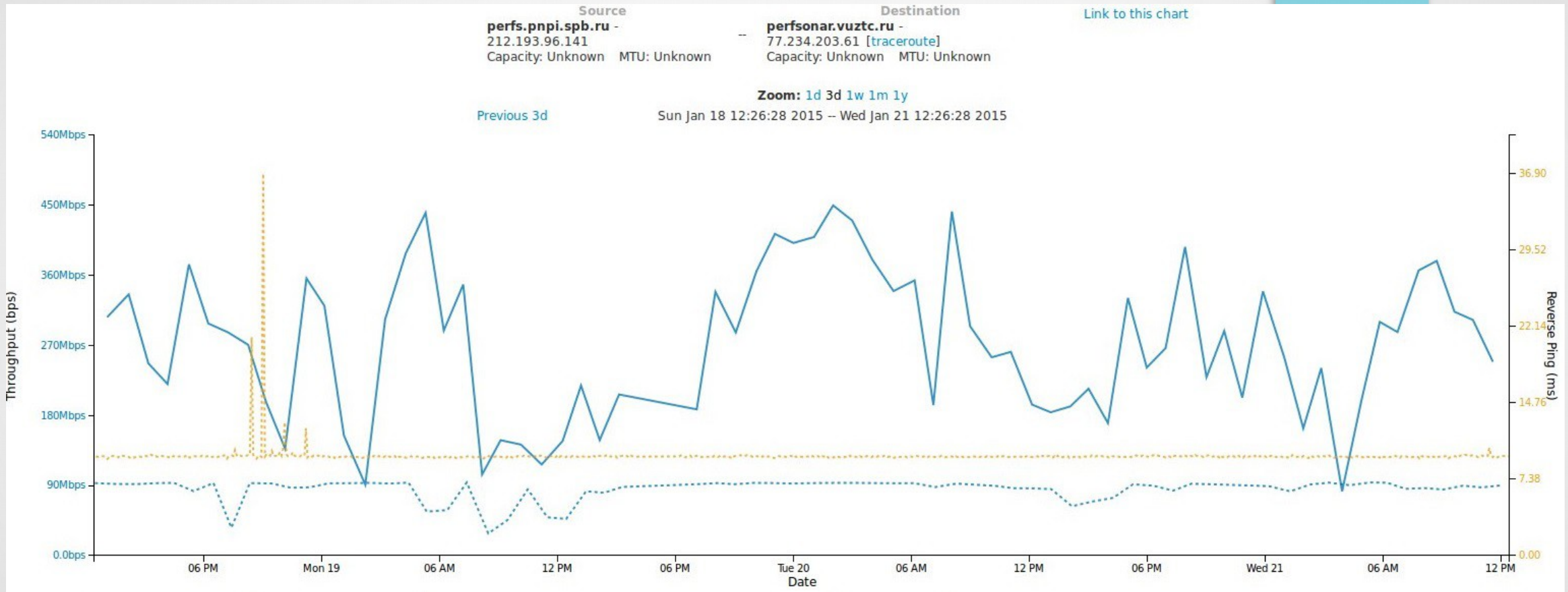
PerfSonar(PNPI)



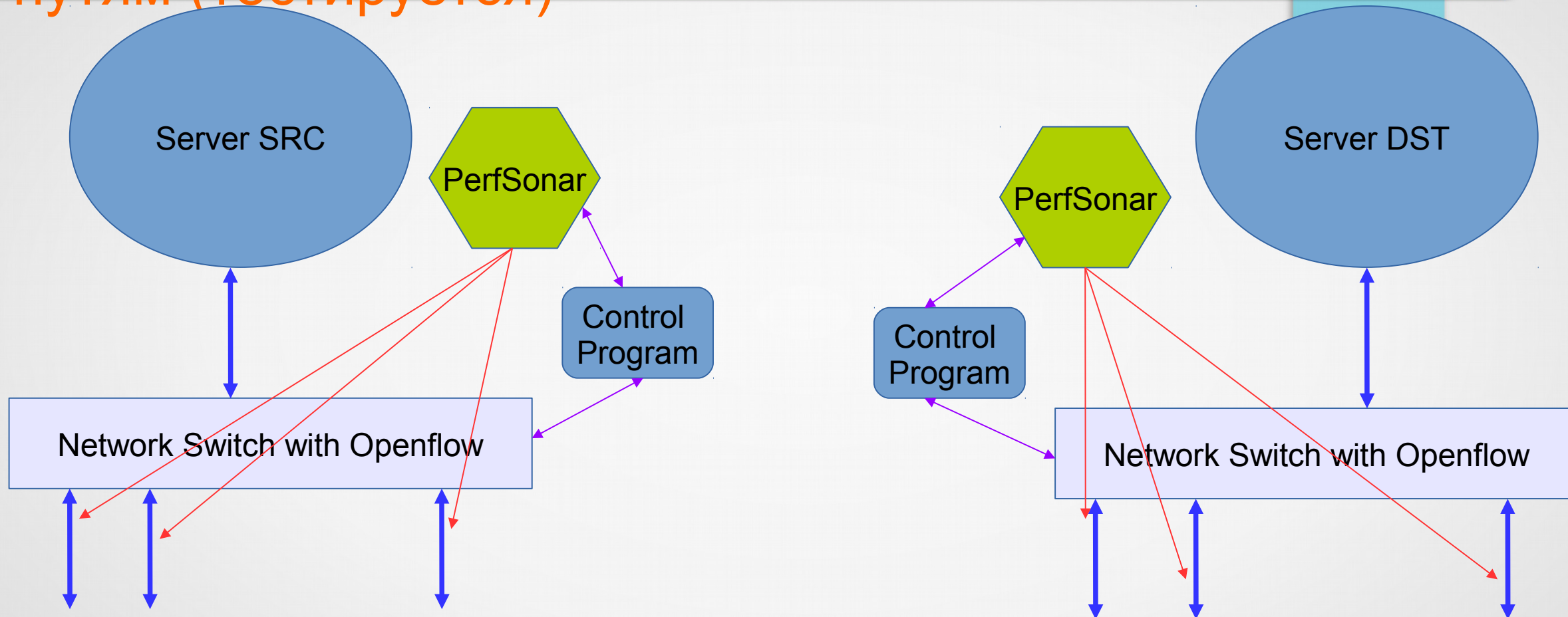
Graph Key

- Src-Dst throughput
- Dst-Src throughput

PerfSonar (ITMO)



Предлагается передача по параллельным путям (тестируется)



Сервер в ОФВЭ ПИЯФ

- Сервер вставлен в стойку, где кластер ОФВЭ, и находится в опытной эксплуатации
 - Дисковое пространство = 70 TB
 - Оперативная память = 96 GB
 - CPU Intel(R) Xeon(R) CPU E5-2650 v2 @ 2.60GHz, 4 порта eth
 - Сетевой коммутатор с возможностью ПКС – HP-3500-24G-POE-у1
- ОС – Scientific Linux 6.6
 - Облачный инструментарий – OpenStack (Icehouse)
- Можно ли использовать часть ресурсов сервера для сотрудников ОФВЭ после перевода в эксплуатацию? – Да, можно использовать.

Предварительные выводы по выполненным работам

- Разработан тестовый стенд для выполнения измерений (как локальных, так и на больших расстояниях).
- Разработаны необходимые технологические скрипты для проведения измерений и выполнена часть измерений для сравнения имеющихся программ передачи данных.
- Разработан набор тестовых виртуальных машин в рамках облачного инструментария **Openstack**.
- Проведена серия измерений (локальная и удалённая передача данных) по сравнению производительности различных программ передачи данных. Было установлено, что среди проверенных программ передачи данных наибольшей производительностью по передаче данных при схожих условиях показывает программа **ВВСП**.
- Продемонстрирована возможность передачи данных по параллельным линиям связи с использованием подхода **ПКС**.

Вопросы ?

Дополнительная информация по Большим Данным

- В физике высоких энергий
 - <http://icfa-scic.web.cern.ch/ICFA-SCIC/meetings.html>
 - <https://indico.cern.ch/contributionDisplay.py?sessionId=0&contribId=487&confId=214784>
 - <https://indico.cern.ch/contributionDisplay.py?sessionId=0&contribId=484&confId=214784>
- В других областях:
 - https://www.dropbox.com/s/3j7x7nnl0vpkcfb/Sokolov_27_06.pdf (Рус)