ST.PETERSBURG NATIONAL RESEARCH UNIVERSITY
OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
Laboratory of the
Network Technologies in Distributed Computing Systems

Big Data transfer over computer networks
Initial

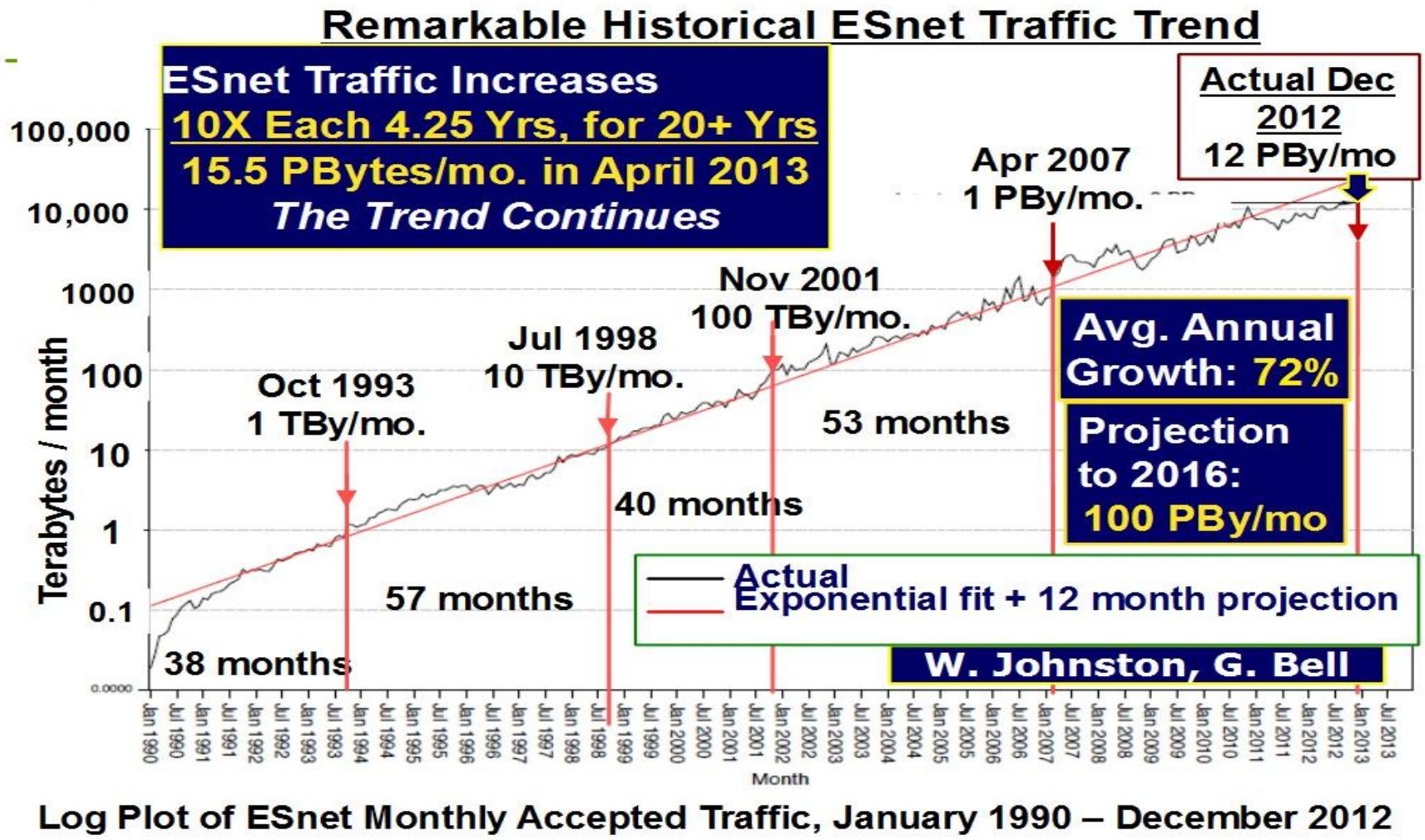
Sergey Khoruzhnikov
Vladimir Grudinin
Oleg Sadov
Andrey Shevel
Anatoly Oreshkin
Elena Korytko
Alexander Shkrebets
Vladimir Titov
Oleg Lazo
Arsen Kairkanov (presenter)

Faculty of Infocommunication Technologies
ITMO University St.Petersburg, Russian Federation

Outlook

- Sources of the Big Data.
- Ecosystem of the Big Data.
- Technology of the Big Data transfer.
- Our recently started research.

Network traffic growth

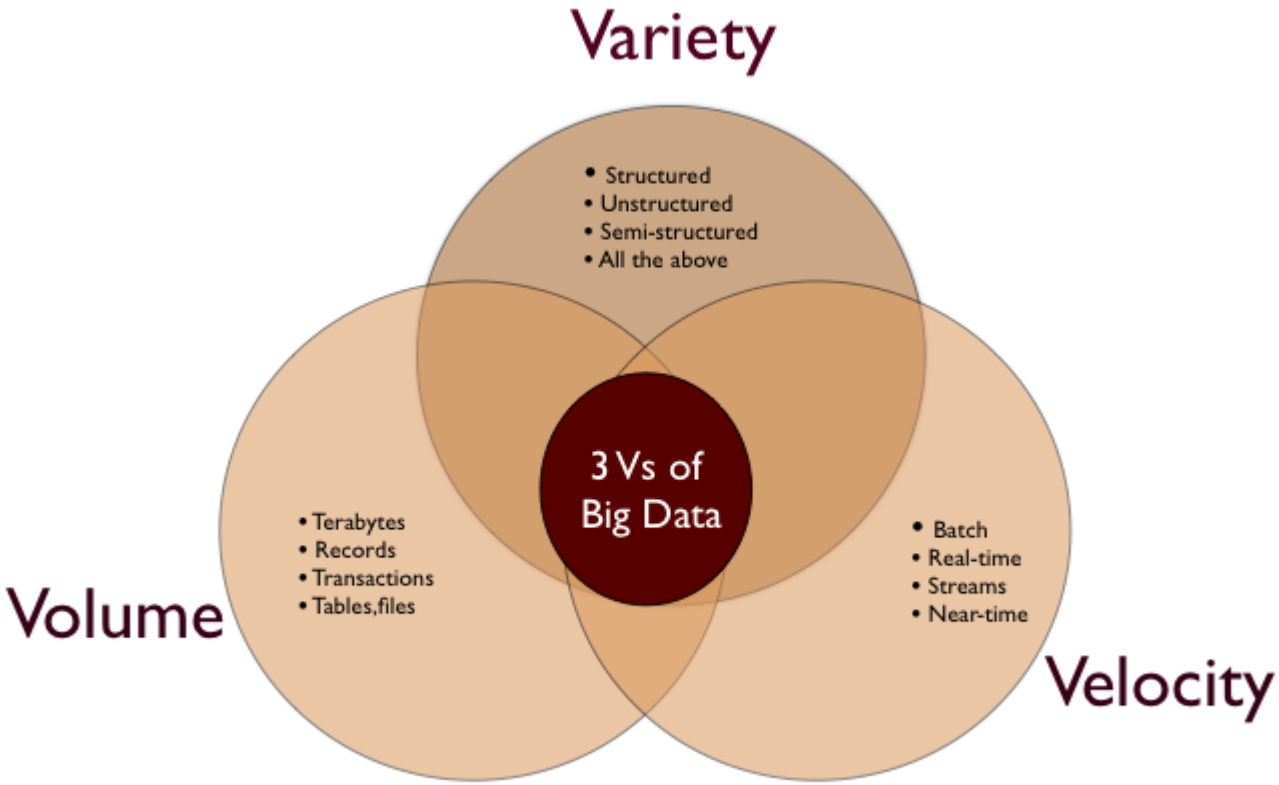


Log Plot of ESnet Monthly Accepted Traffic, January 1990 – December 2012

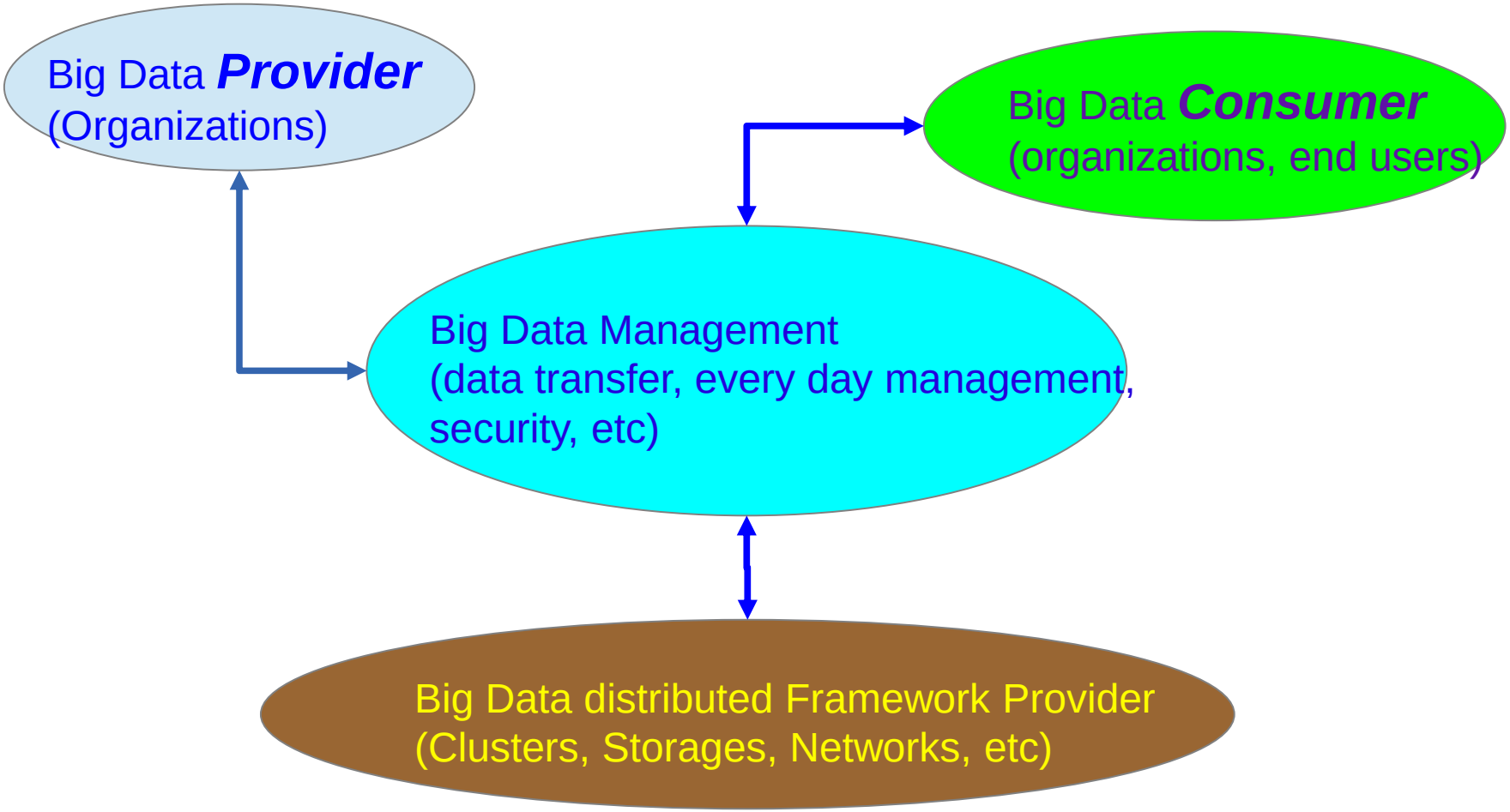
Scientific sources of Big Data

- **Scientific experimental installations**
  - http://www.lsst.org - **Large Synoptic Survey Telescope**
    - **15 TB per night (may be 10 PB/year)**
  - https://www.skatelescope.org/ - **Square Kilometre Array**
    - **300-1500 PB/year**
  - http://www.cern.ch — **CERN**
    - **~ 20PB/year (FAIR ~ same)**
  - http://www.iter.org - **International Thermonuclear Experimental Reactor**
    - **~ 1 PB/year**
  - http://www.cta-observatory.org/ - **CTA - The Cherenkov Telescope Array**
    - **~ 20 PB/year**

3 Vs of Big Data



Big Data ecosystem



Peculiarity of the Big Data transfer

- Big Data transfer might consume many hours or days.
- The situation in channel might be changed: RTT, % of lost network packages, data link bandwidth).
- Finally, it might occured the interruption (hours?, days?) in operation of data link .
- Obviously it is useful to have access to two or more independent data links.

Technology peculiarities with Big Data transfer

- Still main protocols — stack of TCP/IP.
  - Number of network parameters in Linux (around ½ thousand). /proc
  - -bash-4.1$ /sbin/sysctl -a | grep "^net\." | wc -l
    - Important parameters: e.g. size of block, size of TCP Window, etc. Main method to decrease the transfer time (even over one data link) is using the multi-stream data transfer.

Testing on the first stage (program tools)

- BBCP - http://www.slac.stanford.edu/~abh/bbcp/
- GridFTP - http://www.globus.org/toolkit/data/gridftp/
- BBFTP - http://doc.in2p3.fr/bbftp/
- FDT - http://monalisa.cern.ch/FDT/
- FTS3 - http://fts3-service.web.cern.ch/
- Also technology components to watch the data links status, e.g. perfSONAR.

Ideas to compare the data transfer tools

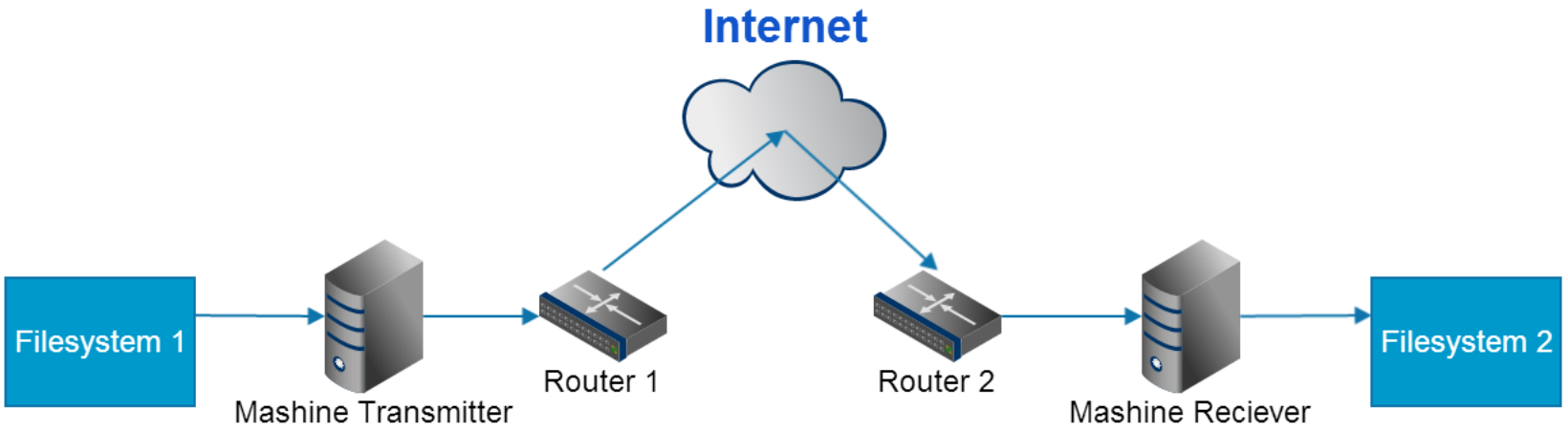
- Availability.
- API.
- Performance.
- Reliability.
- Operation tracking.
- Ability to predict the time to transfer the data on the base of existing tracking records.
- Required resources: memory, CPU time, etc.
- Others.

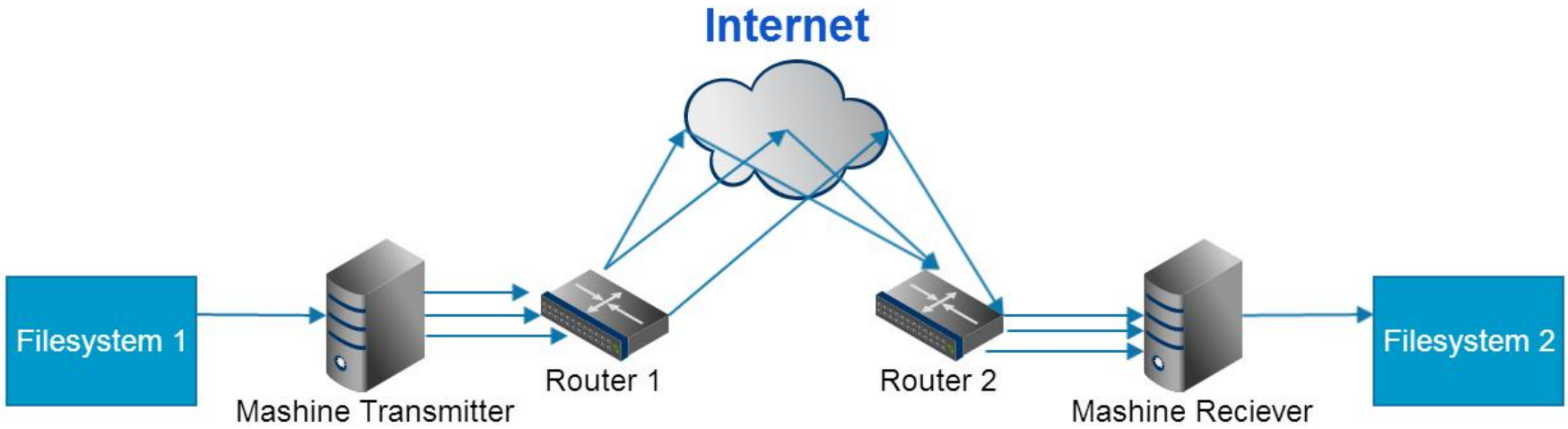
Research topic at ITMO University: the transfer of Big Data

In laboratory of network technologies http://sdn.ifmo.ru/ at ITMO University http://www.ifmo.ru/ the new research «Big Data transfer over Internet» has been formed .

- It is planned to implement the special testbed (100 TB of disk storage + server 96 GB of main memory under OS RedHat/ScientificLinux on each side).
  - Comparative study of the existing tools of the data transfer (testing and measurements).
  - To use the testbed as instrument to compare various tools (tracking for the measurements + results).
  - Extended automatic tracking information about measurements is under development.

Process of the Data transfer

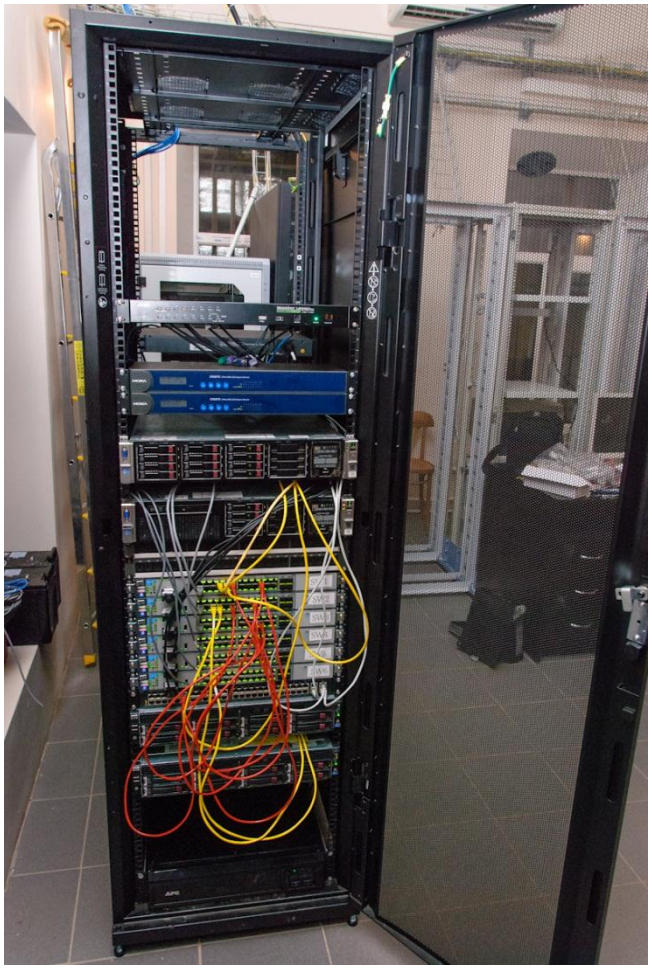




Planned measurements

- Local and long distant sites with existing data links (not only most advanced links).
- The idea is to use more than one data link in parallel.
- Recently we obtained some experience with Software Defined Networks (SDN) approach (*protocol Openflow*) and now we plan to use it in the Big Data transfer.

What was done until now



- **There were deployed**
  - Two servers HP DL380p Gen8 E5-2609, Intel(R) Xeon(R) CPU E5-2640 @2.50GHz, 64 GB under Scientific Linux 6.5.
  - Six HP-3500-24G-PoE yl (OpenFlow 1.0)
  - Pica8 P-3920 (OpenFlow 1.2)
  - Openstack Havana with appropriate set of Virtual Machines to test a number of mentioned utilities.
  - PerfSonar
  - Scripts for testing https://github.com/itmo-infocom/BigData

Main goals

- Combining the developed contemporary components and methods with ideas, developments, experience to achieve maximum speed for Big Data transfer on existing links.
- To create the testbed which would be used as place where researchers might compare theirs (new) tools for data transfer with earlier recorded measurement results.
- To sugggest the collaboration with … (suggestions?)
- To invite students from … (suggestions?)

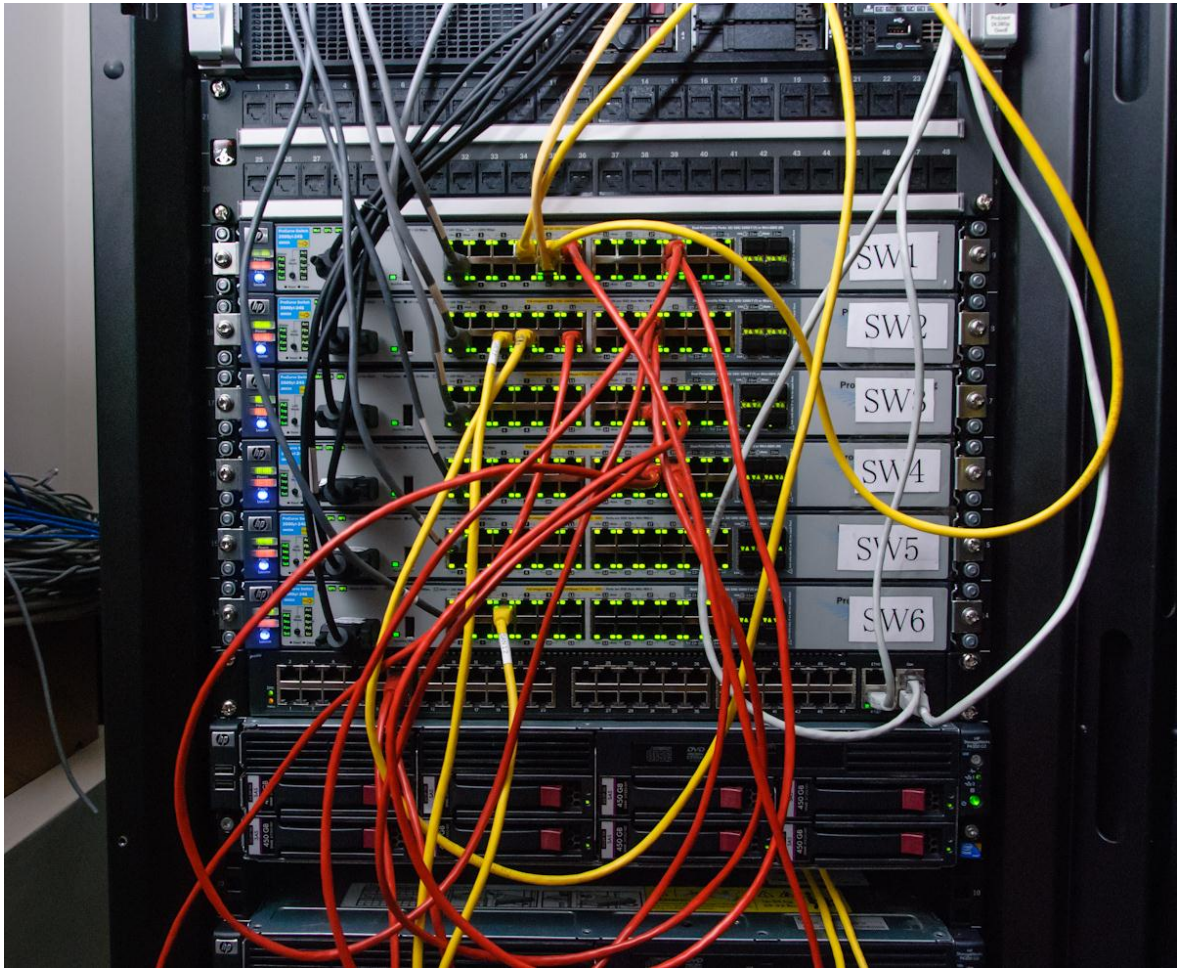
Partners (ideas exchange)

- Laboratory of Information Technology (LIT)
  http://lit.jinr.ru/index.php?lang=lat
  @ Joint Institute for Nuclear Research (JINR.ru)
- The Application Research Center for Computer Networks at Moscow University http://arccn.ru/
- We are starting to collaborate with GENI http://www.geni.net/

The work is supported by the Saint-Petersburg University of Information Technology, Mechanics & Optics (ITMO University www.ifmo.ru)

Questions?

OF Switches



bbcp

- TCP Window size;
- number of TCP streams;
- I/O buffer size;
- compression on the fly;
- multi-directory copy;
- resuming failed copy;
- authentication with ssh;
- using pipes, where source or/and destination might be pipe;
- special option to transfer small files;
- and many other options dealing with many practical details.

bbftp

- encoded user name and password at connection;
- SSH and Grid Certificate authentication modules;
- multi-stream transfer;
- big windows as defined in RFC1323;
- on-the-fly data compression;
- automatic retry
- customizable time-outs;
- transfer simulation;
- AFS authentication integration.

gridFTP

- two security flavors: Globus GSI and SSH;
- the file with host aliases: each next data transfer stream will use next host aliases (useful for computer cluster);
- pipes;
- special debugging mode to find bottleneck in data transfer;
- backend module name for source and destination sites;
- number of parallel data transfer streams;
- buffer size;
- restart failed operations and number of restarts.

Other utilities

- Xdd – utility developed to optimize data transfer and I/O processes for storage systems.
- fdp – Java utility for multi-stream data transfer;
- FTS3
- UDT
- RDMA
- MP TCP