Scalable Network Traffic Analytics workshop

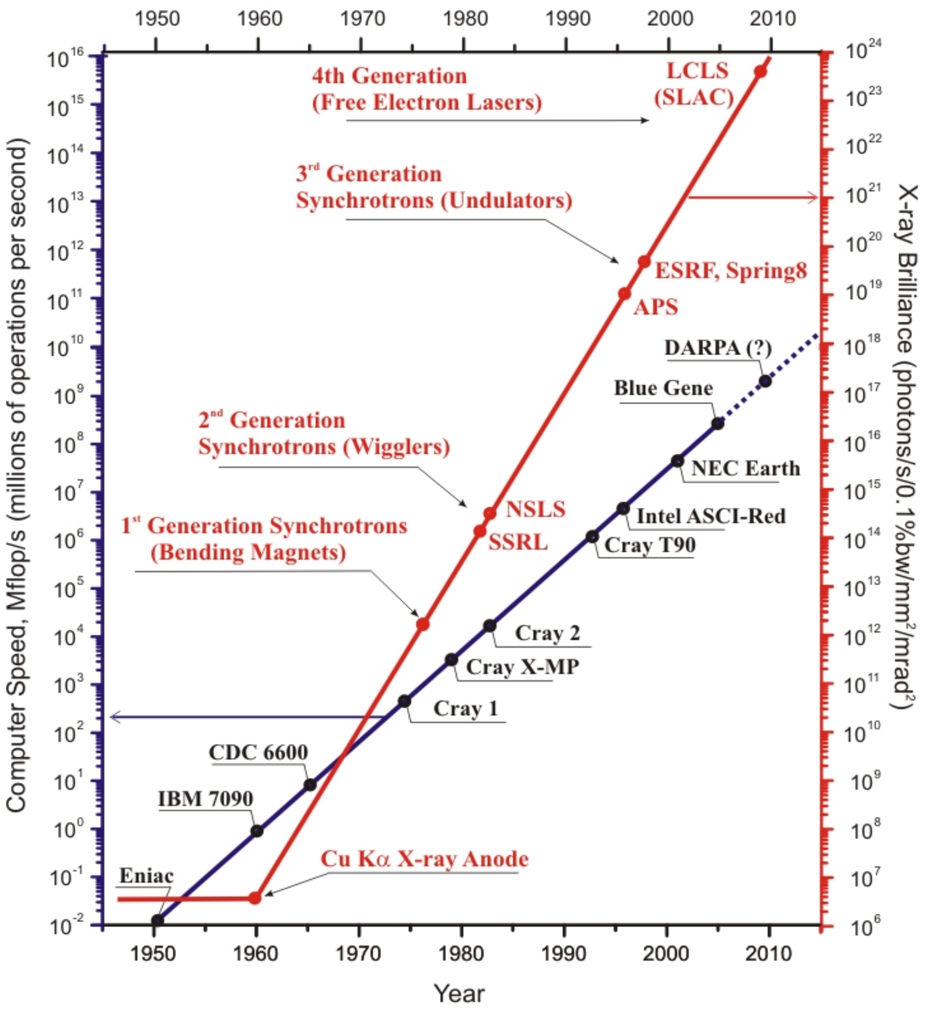
# A comprehensive study of wide area data movement at a scientific computing facility

Zhengchun Liu, Rajkumar Kettimuthu, Ian Foster and Yuanlai Liu

Presented by: Rajkumar Kettimuthu

**Vienna, Austria - July 2, 2018**

# X-ray sources produce a lot of photons, which translates to a lot of data

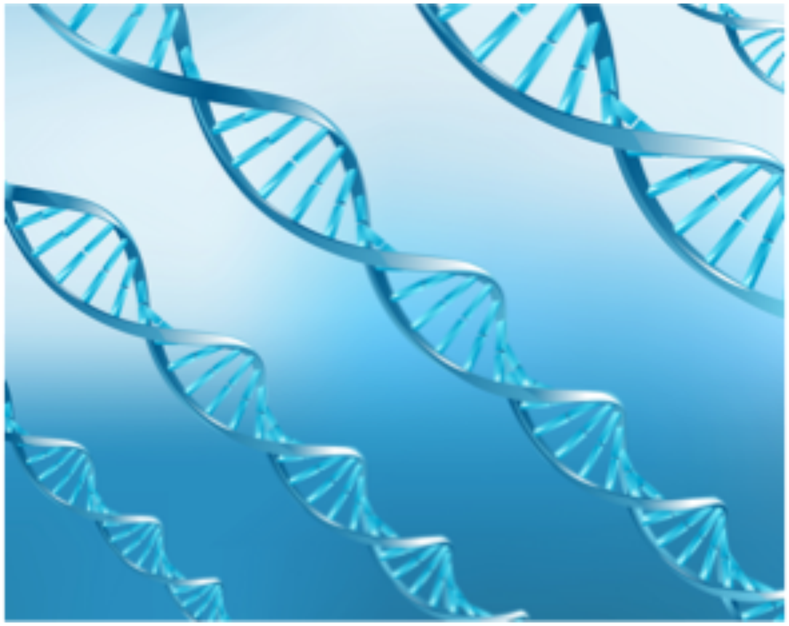


Computer speed:
12 orders
of magnitude
in 6 decades

X-ray source brilliance:
18 orders
of magnitude
in 5 decades!

# Data deluge

Genomics

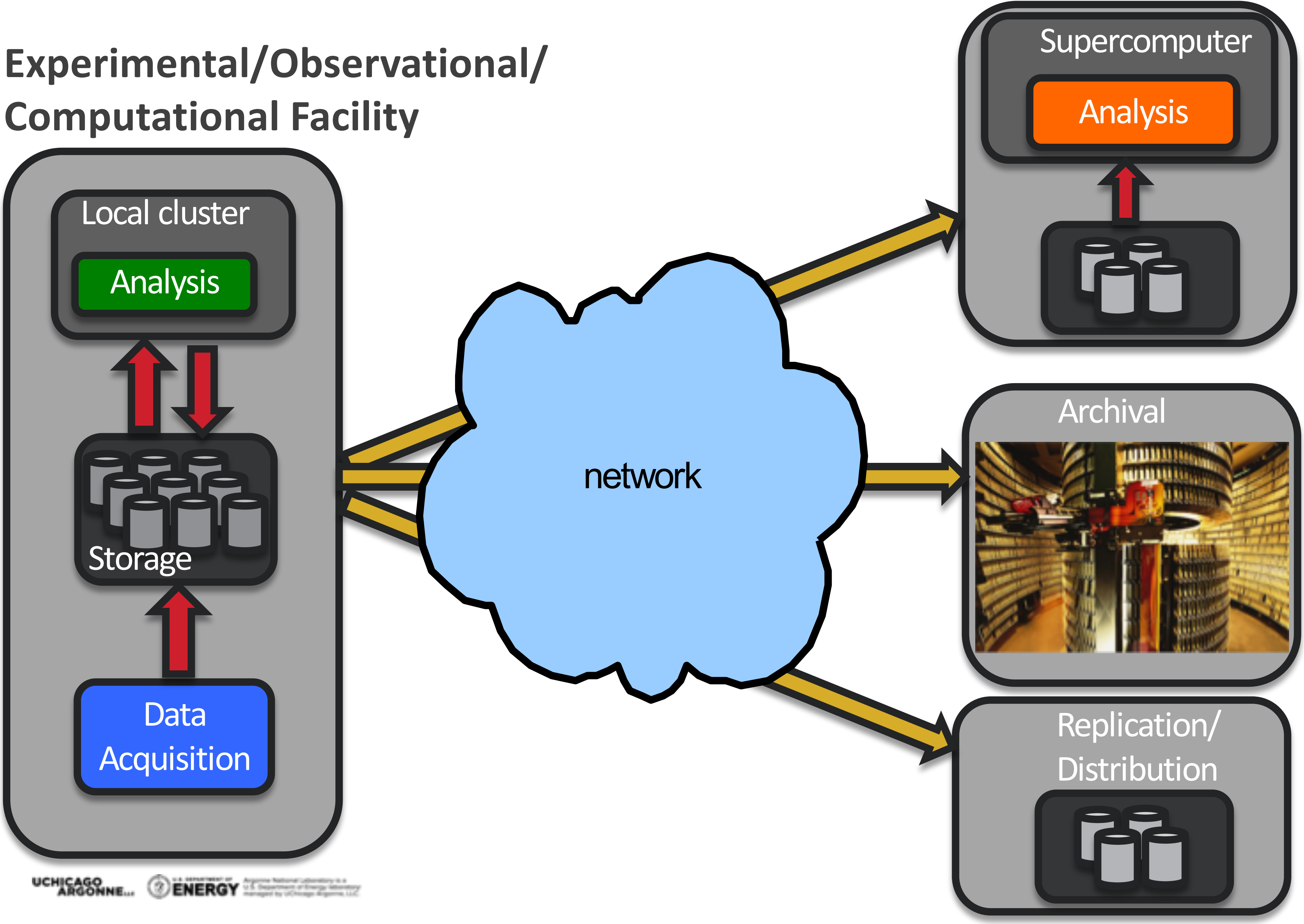


Cosmology

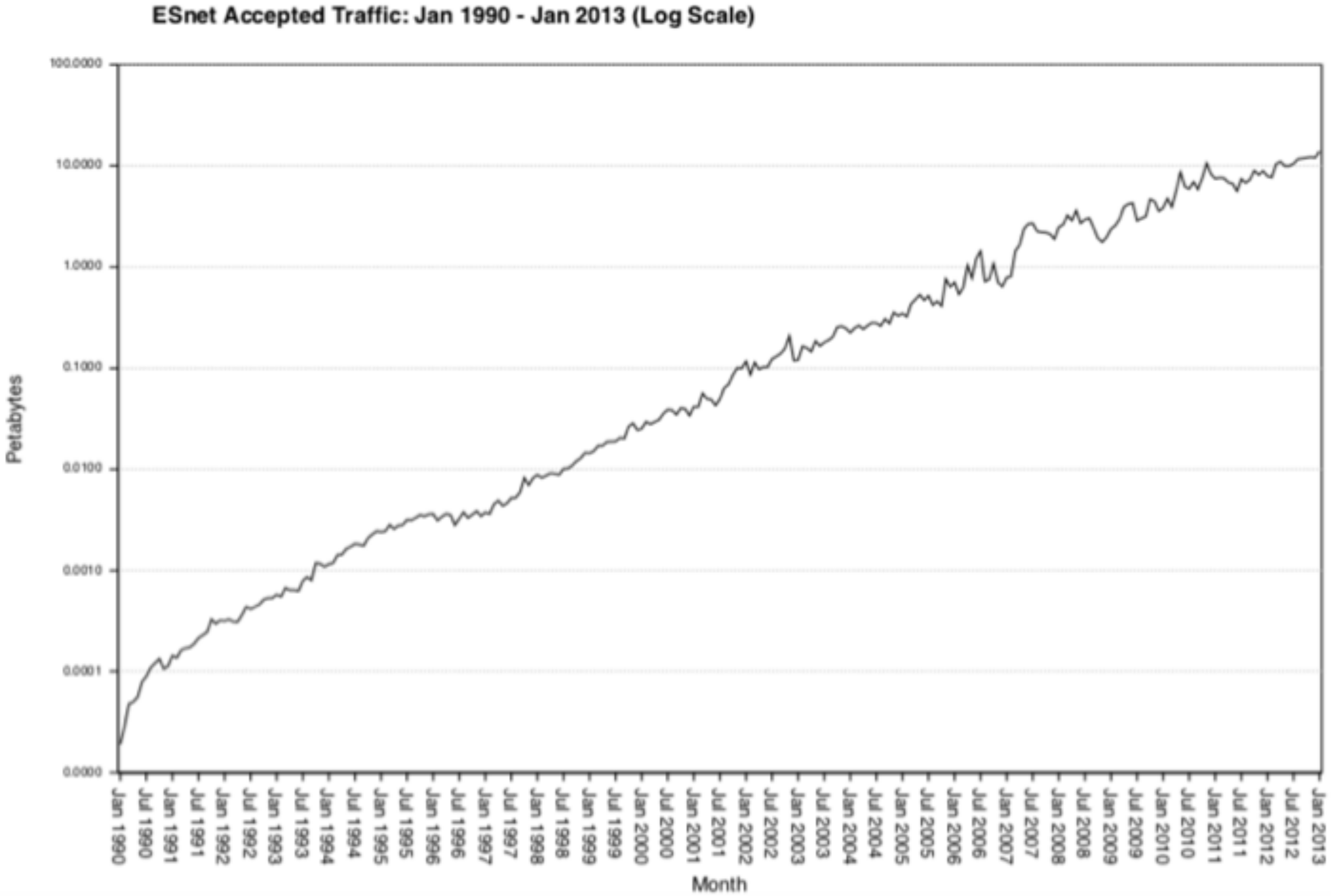
# Science workflows

**Remote Facility**

**Experimental/Observational/ Computational Facility**



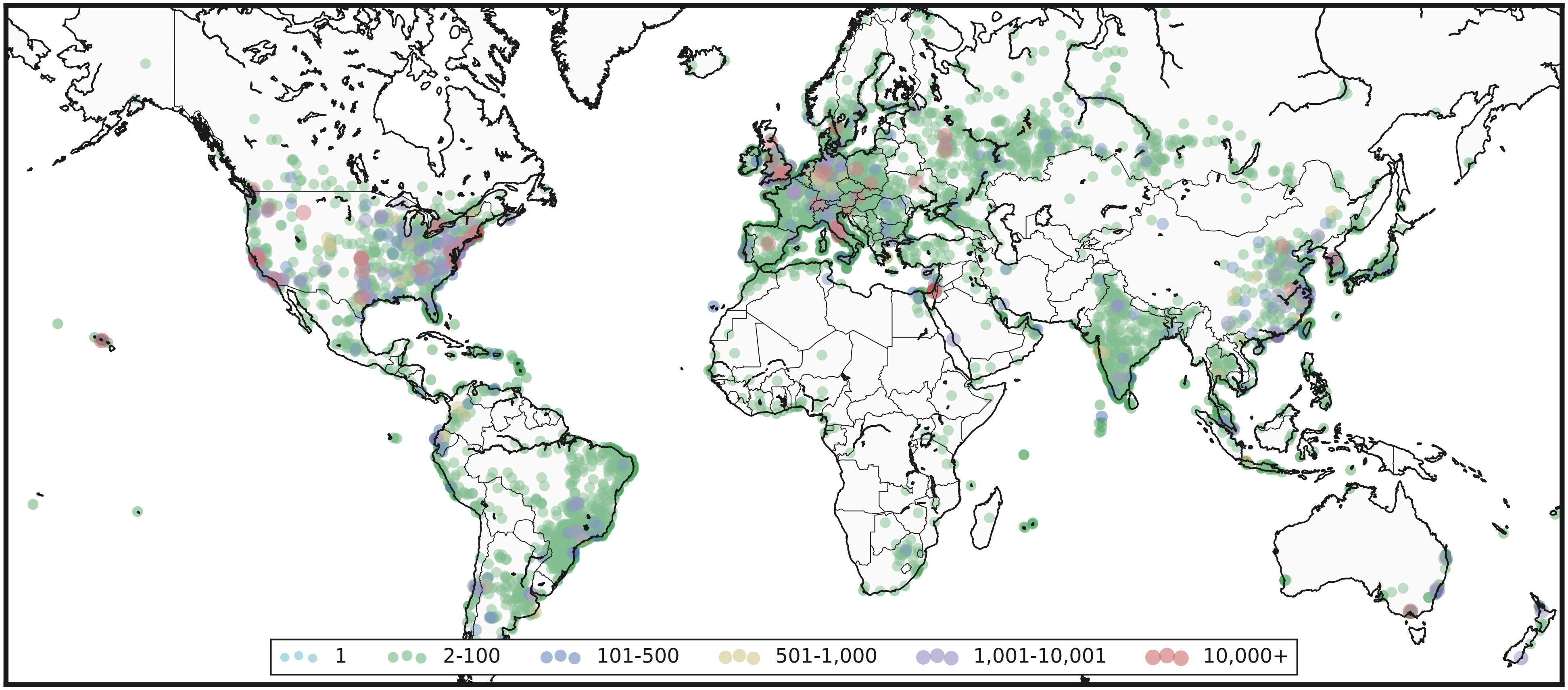
# Growth in wide area science data transfers



ESnet Accepted Traffic: Jan 1990 - Jan 2013 (Log Scale)

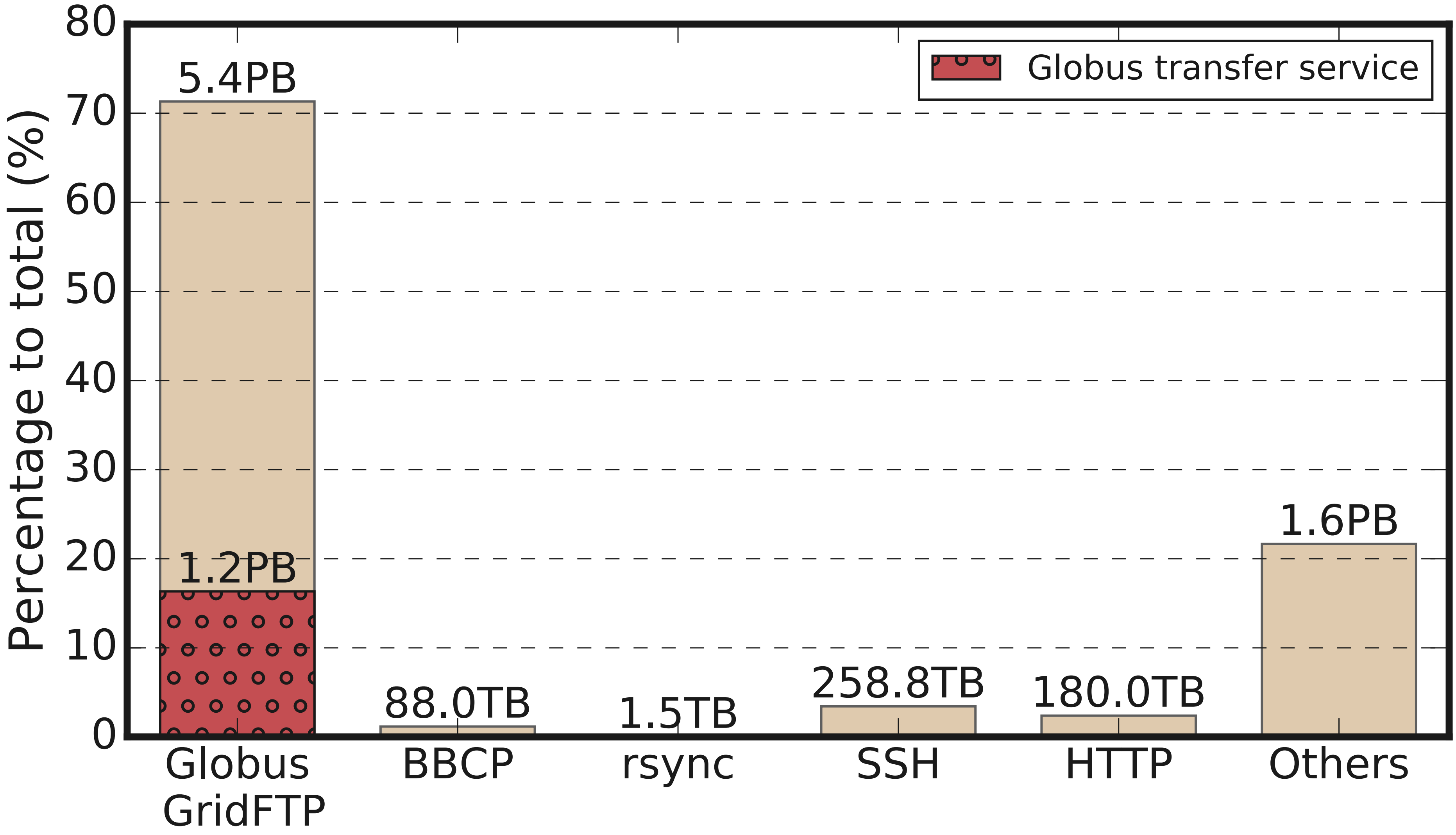
# Wide area data movement at a scientific computing facility (*BigSite*)

Geographical distribution of TCP flows to/from *BigSite* DTNs in 2017, with color used to code number per city.



# Data movement tools

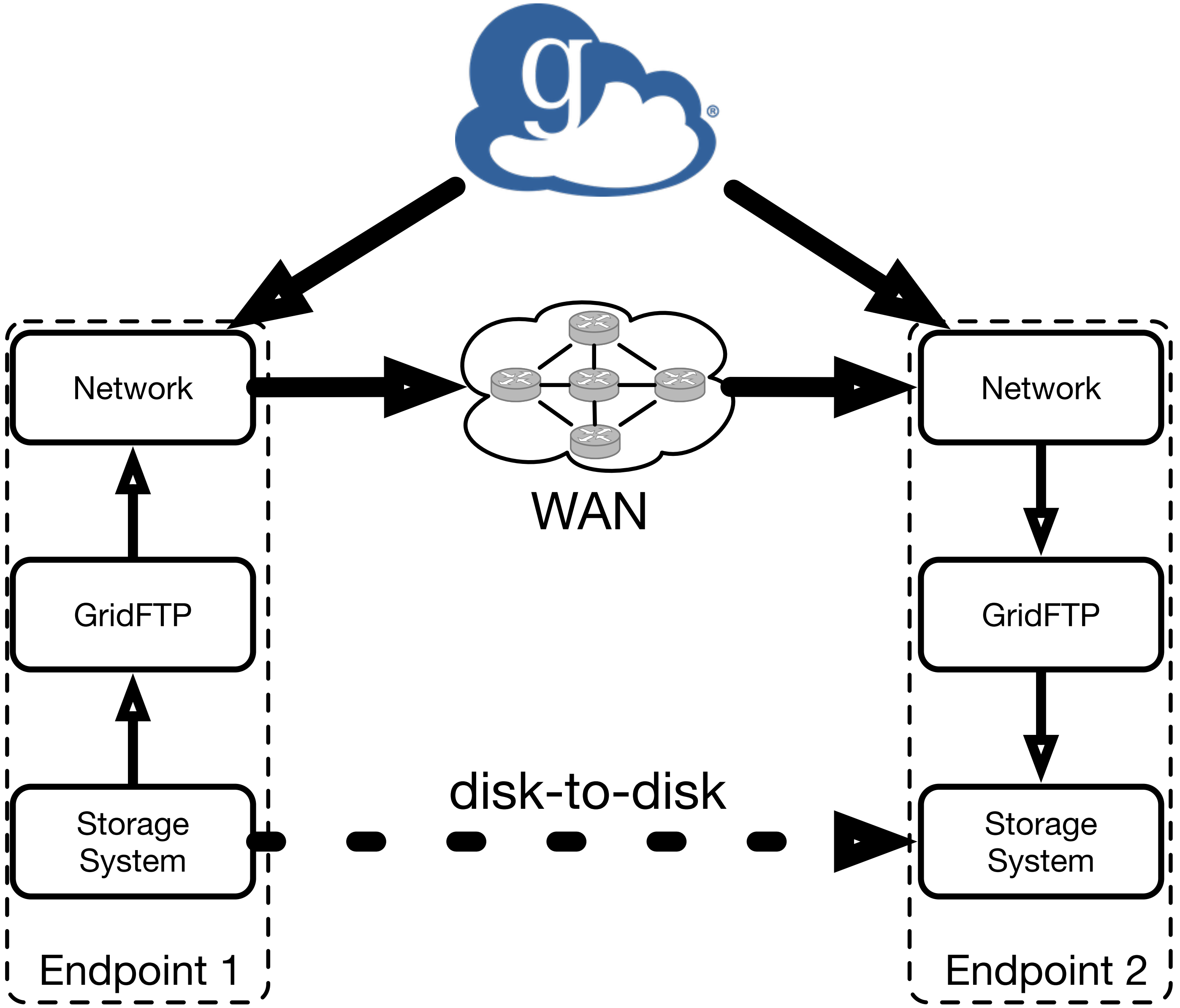
Data volumes transferred with different tools on *BigSite* during the five-month period 2017/08/01--12/31.



# GridFTP

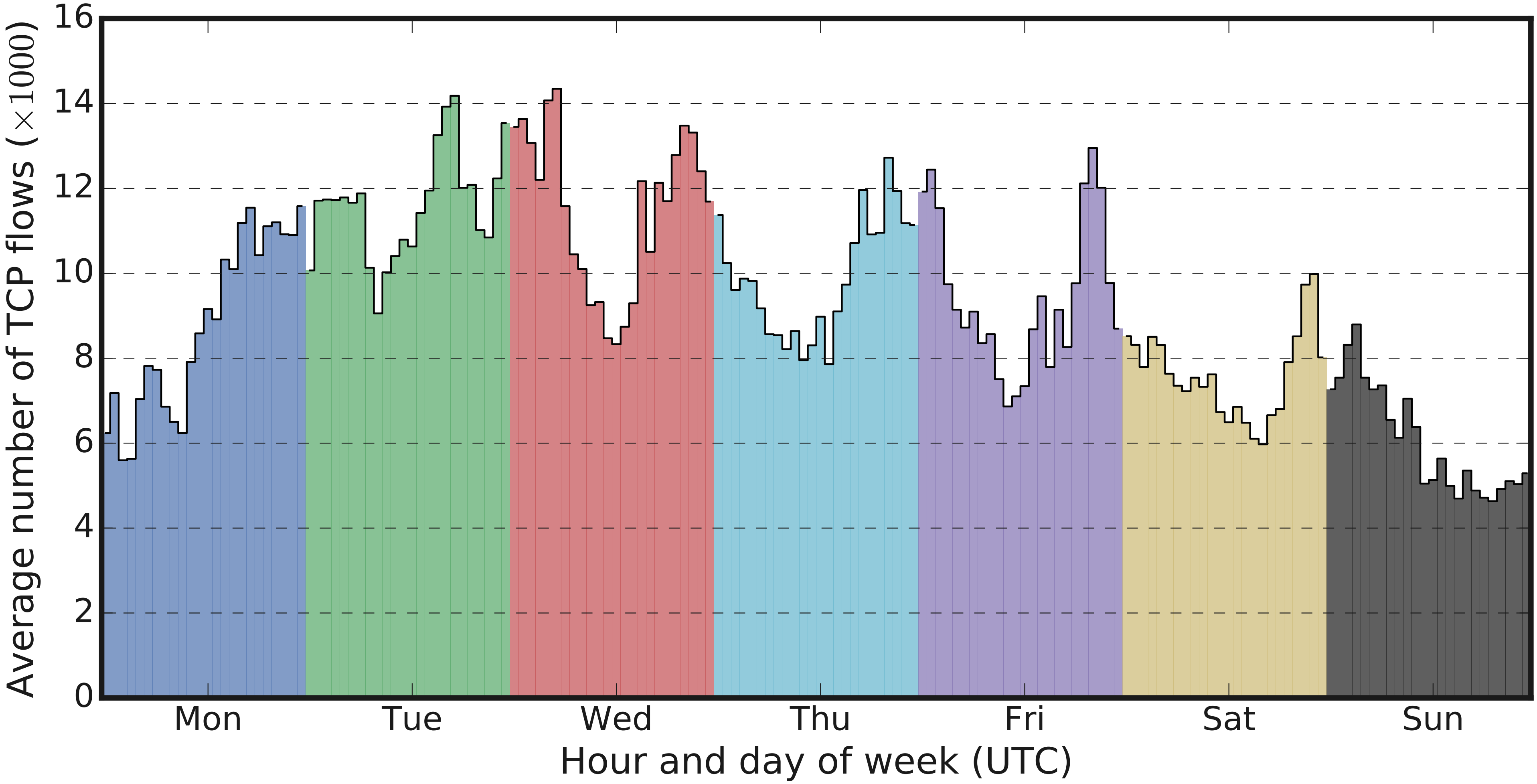
- High-performance, secure data transfer protocol optimized for high-bandwidth wide-area networks
  - Parallel TCP streams, PKI security for authentication, integrity and encryption, checkpointing for transfer restarts
- Based on FTP protocol - defines extensions for high-performance operation and security
- Globus implementation of GridFTP is widely used
- Globus GridFTP servers support usage statistics collection
  - Transfer type, size in bytes, start time of the transfer, transfer duration etc. are collected for each transfer

# Globus transfer service



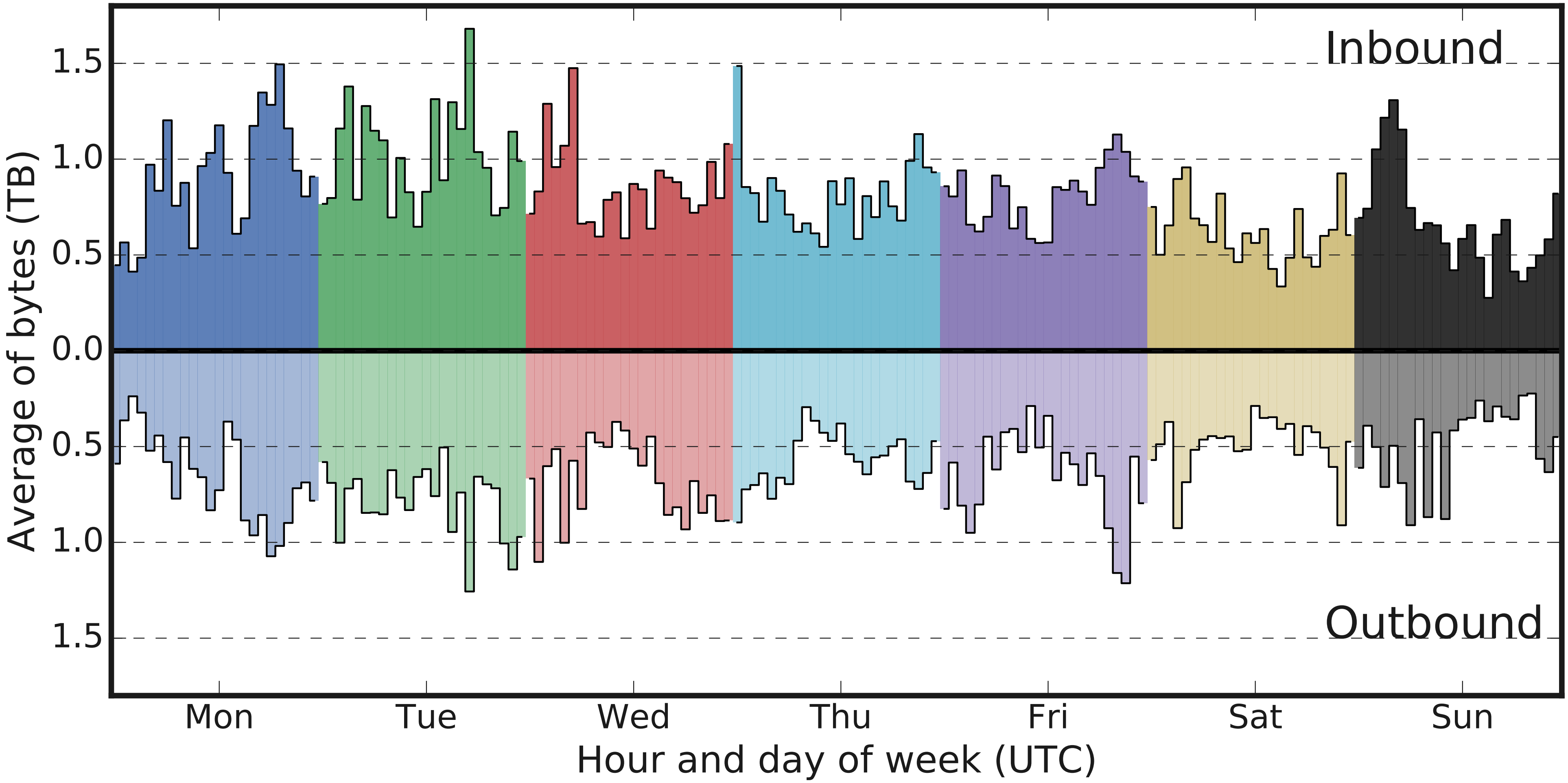
# Flow characteristics

Average number of TCP **flows**, to/from all DTNs, per hour and day of the week in 2017.  X axis is UTC time.



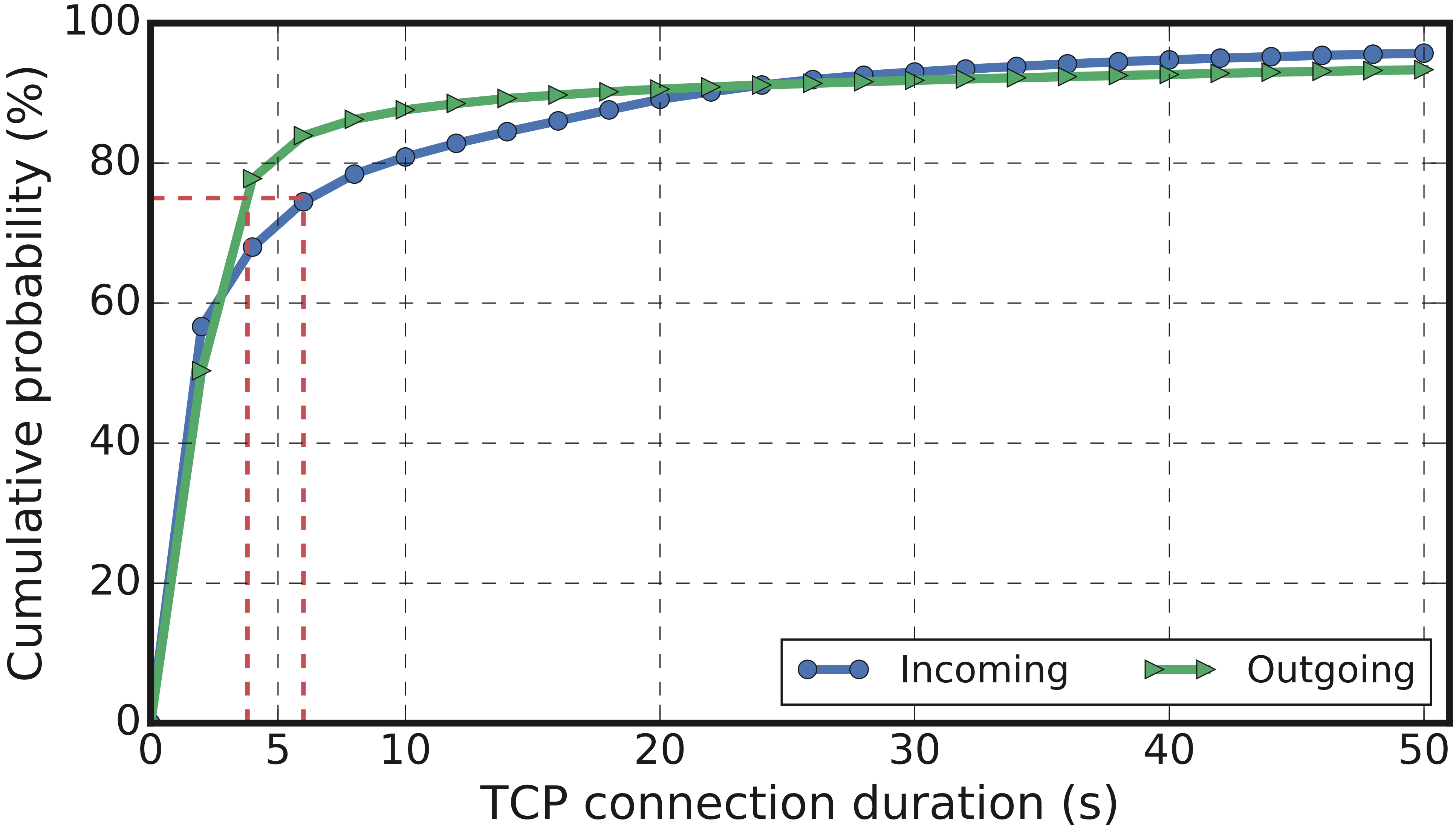
# Flow characteristics

Average number of **bytes** moved, to/from all DTNs, per hour of day of week in 2017.
X axis is UTC time.



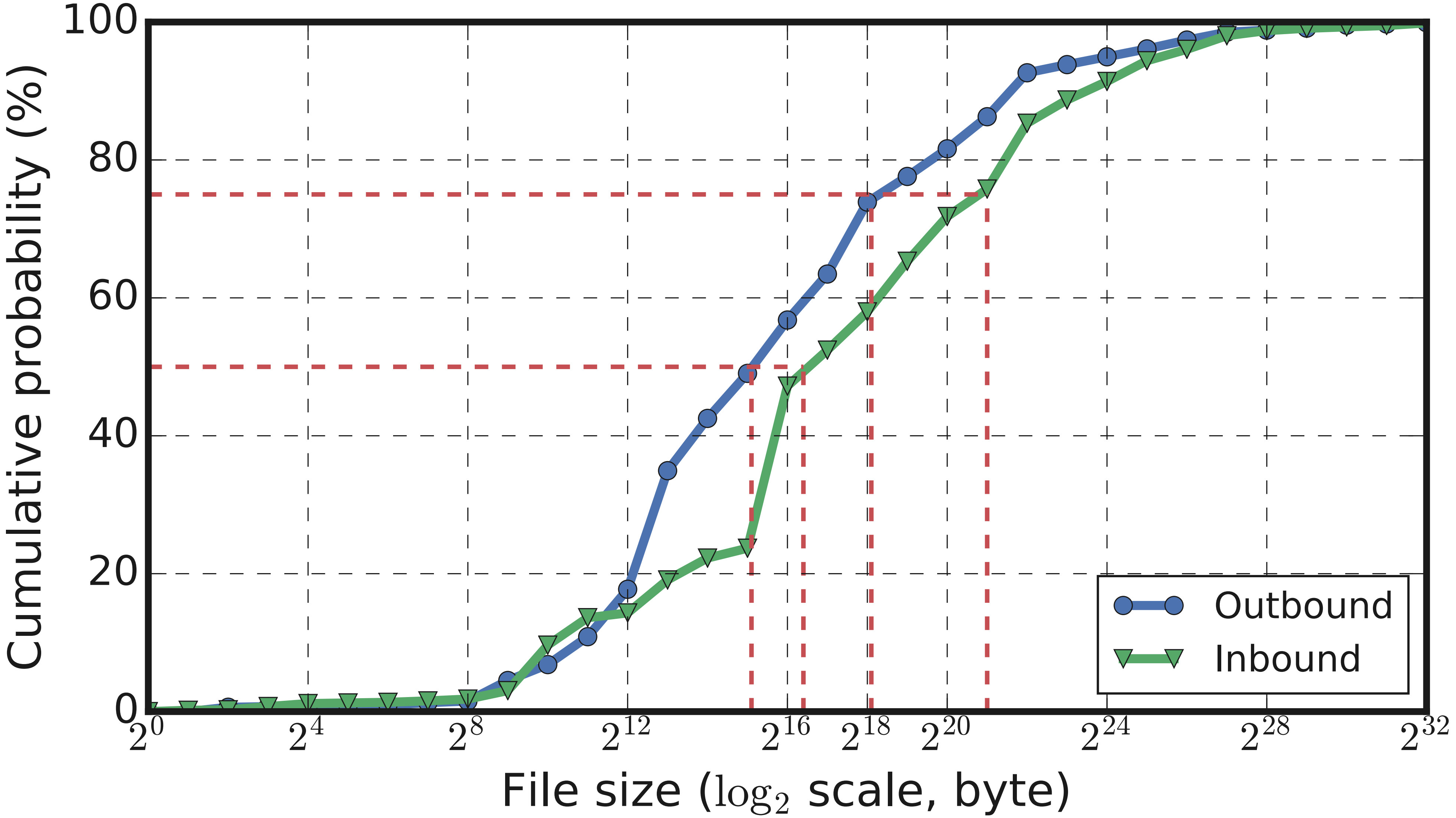
# Flow characteristics

Cumulative distributions of TCP flow **duration**, with 75th percentiles indicated by dashed red lines.



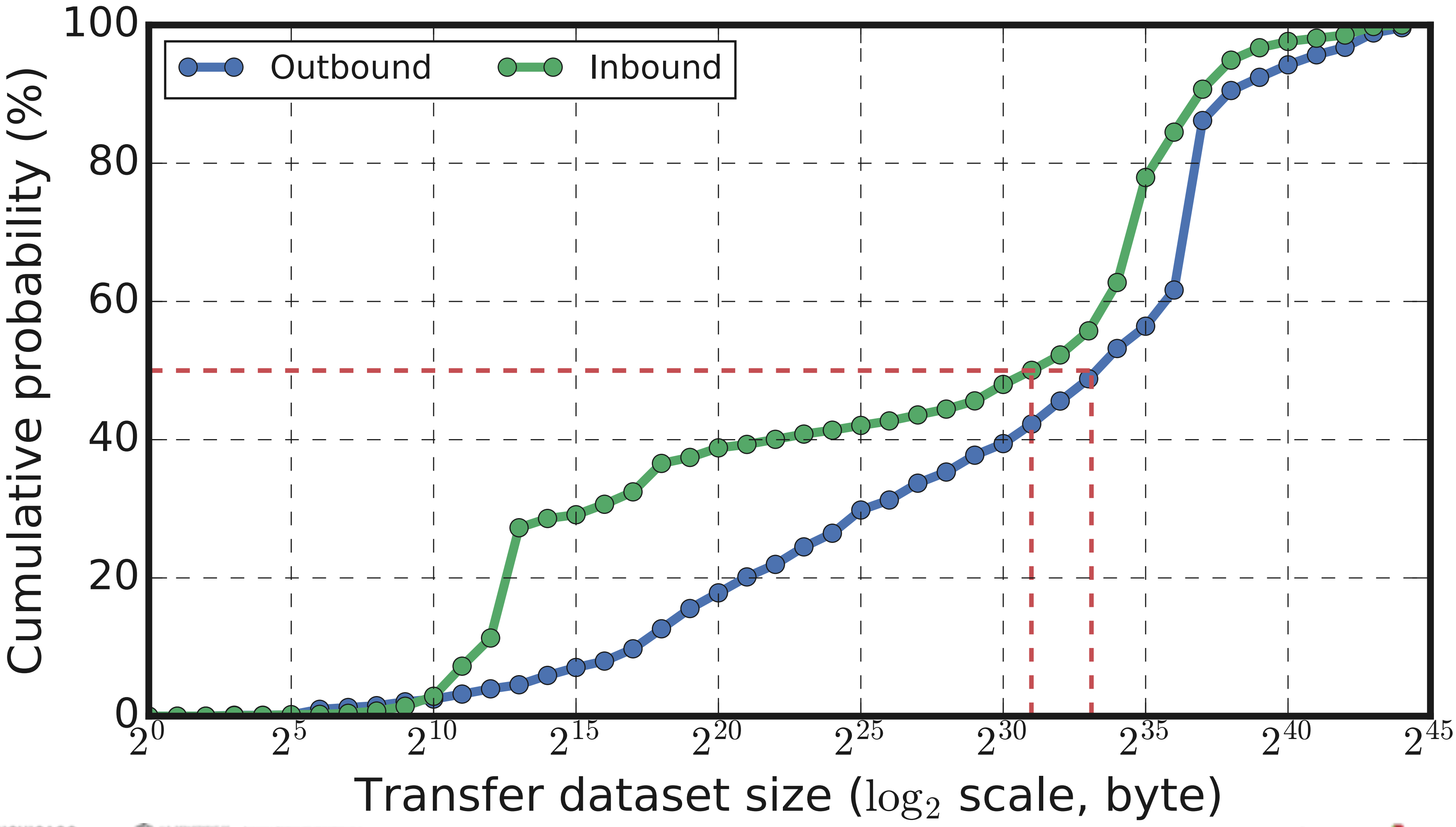
# File size characteristics

Cumulative distributions of GridFTP transfer **file sizes**, with 50th and 75th percentiles highlighted.



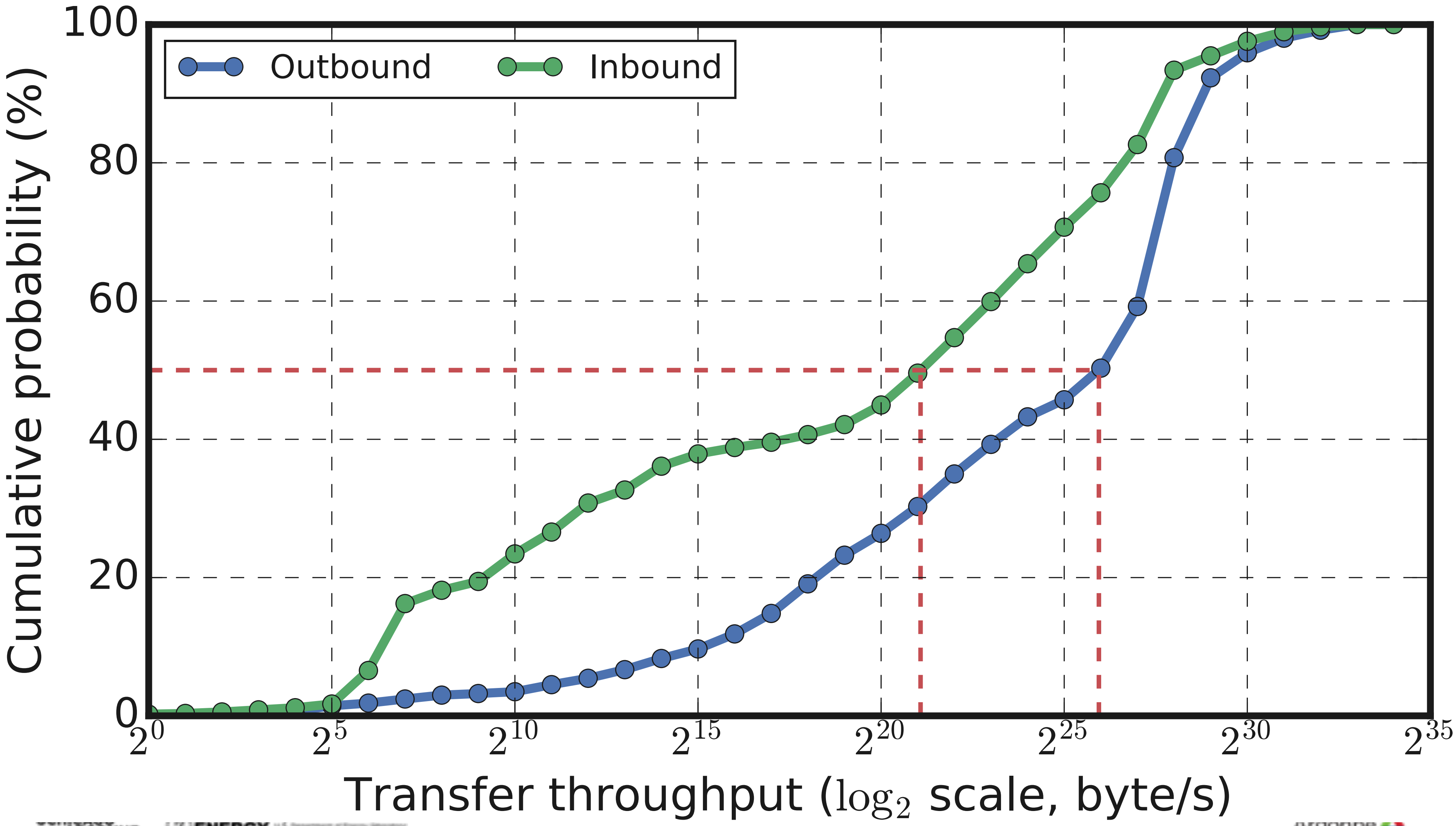
# Transfer size characteristics

Cumulative distributions of Globus **transfer sizes**, with 50th percentiles highlighted.



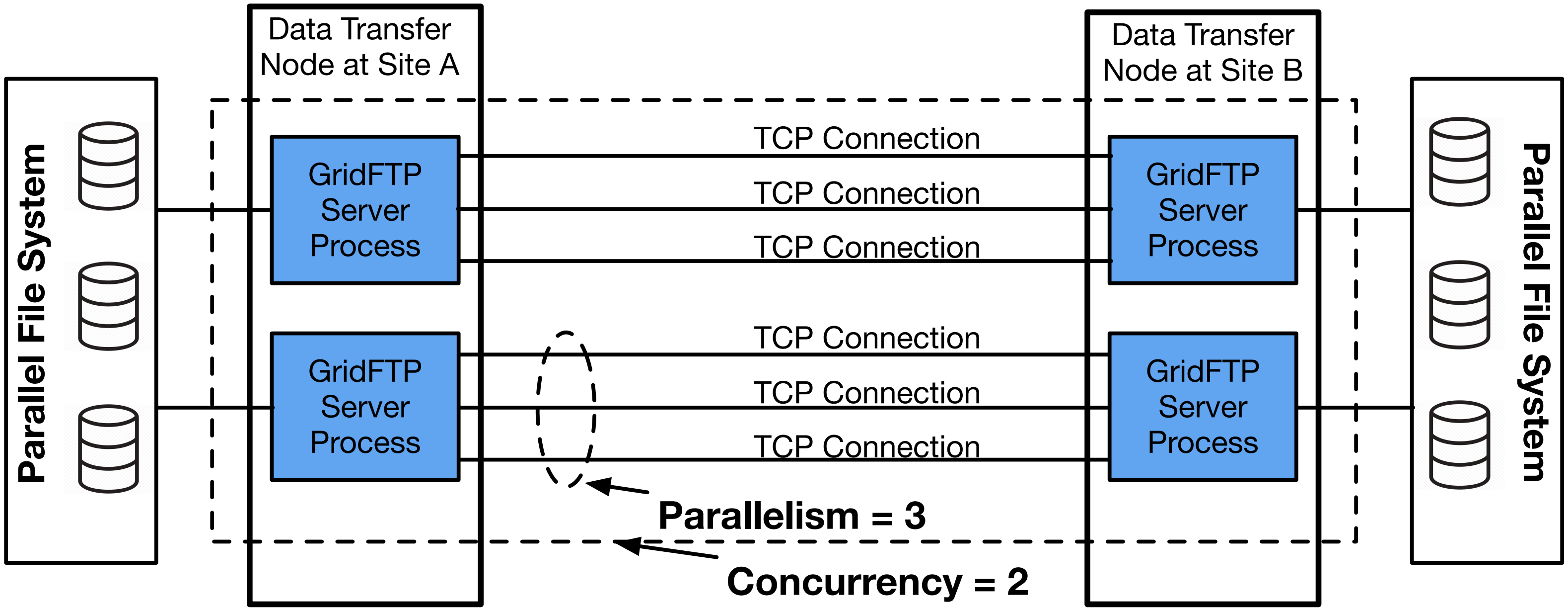
# Transfer rate characteristics

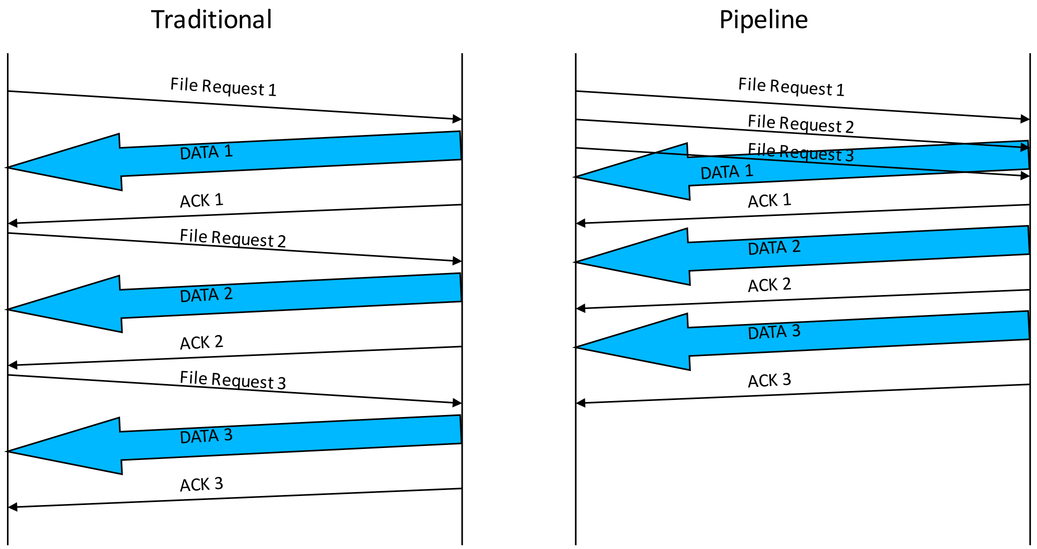
Cumulative distributions of Globus transfer **rate**, with 50th percentiles highlighted.



# A top-down view of workload distribution

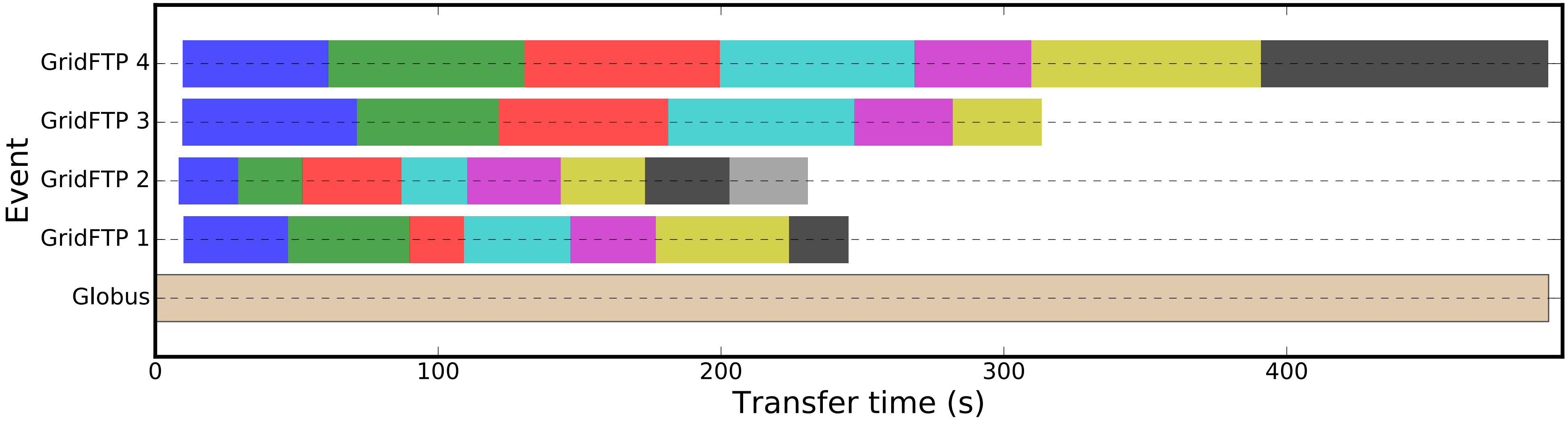
## GridFTP / Globus Concurrency, Parallelism and Pipeline





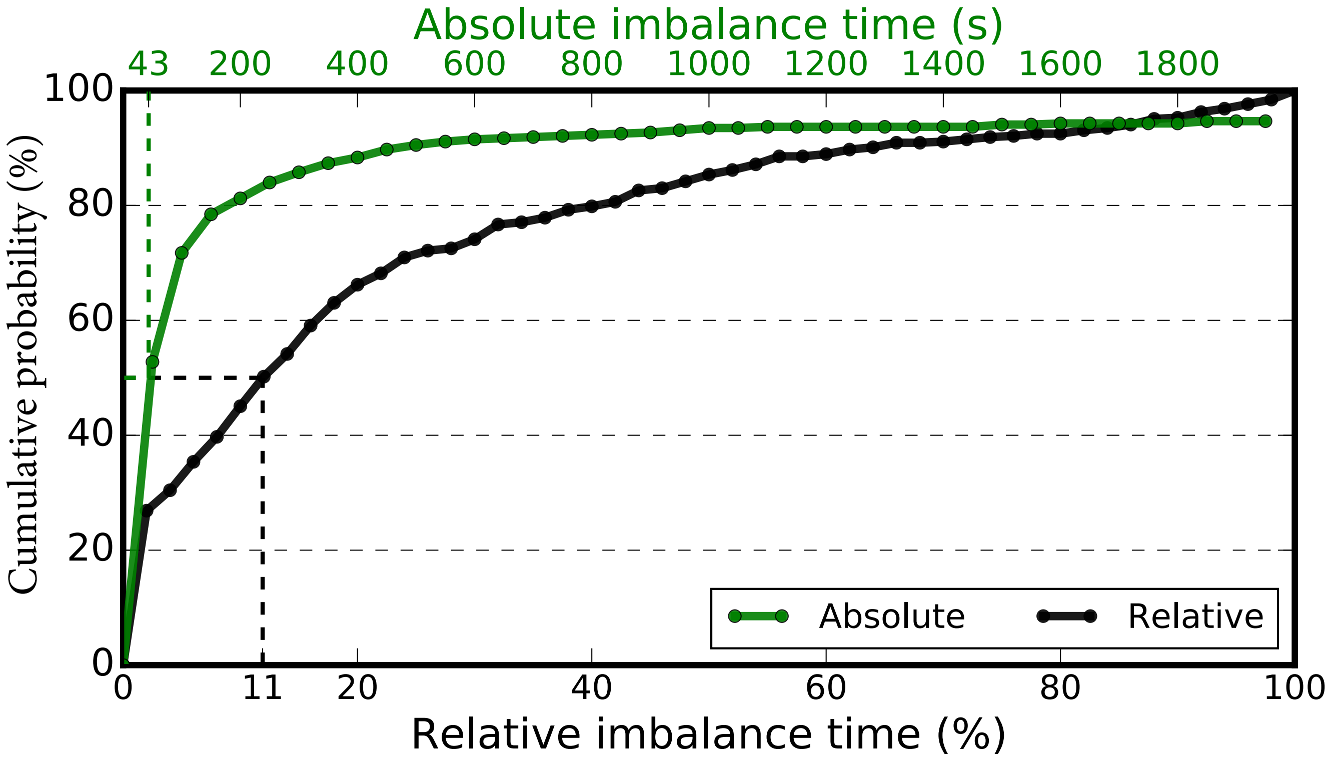
# Load imbalance among GridFTP server processes

Imbalanced GridFTP load due to pipelining. Each line represents activity at one of four GridFTP servers, with each rectangle corresponding to a single equisized file.



# Load imbalance among GridFTP server processes

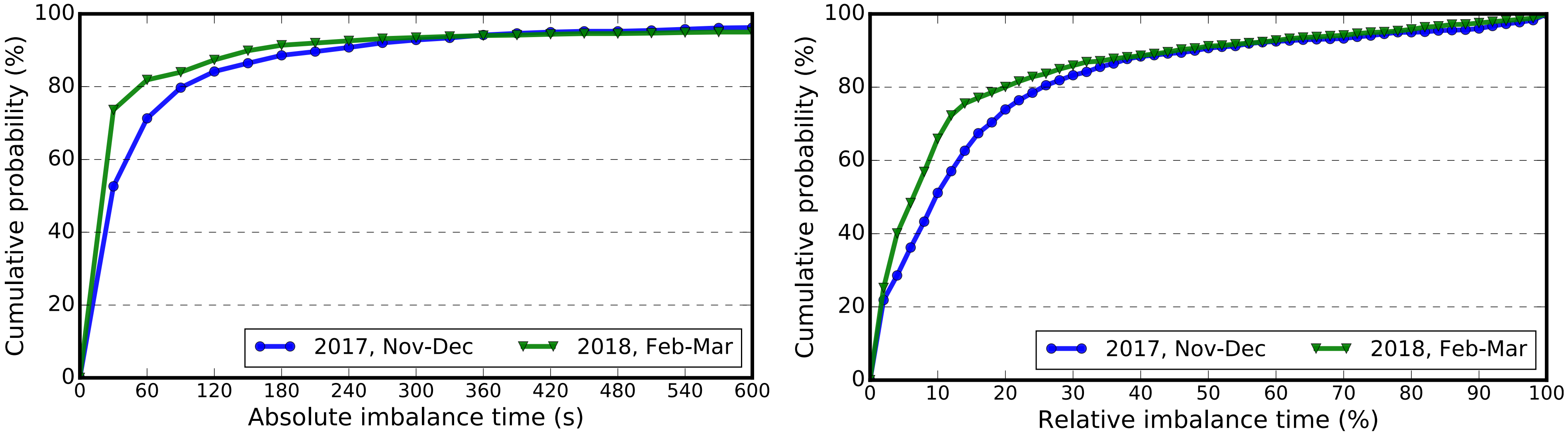
Cumulative distribution of imbalance (in concurrent GridFTP server processes) of Globus transfers.



**50% of the transfers have an absolute imbalance time > 43 seconds. In terms of relative imbalance, 50% of the transfers have a relative imbalance > 11%. The imbalance is significant.**

# Optimization to reduce imbalance

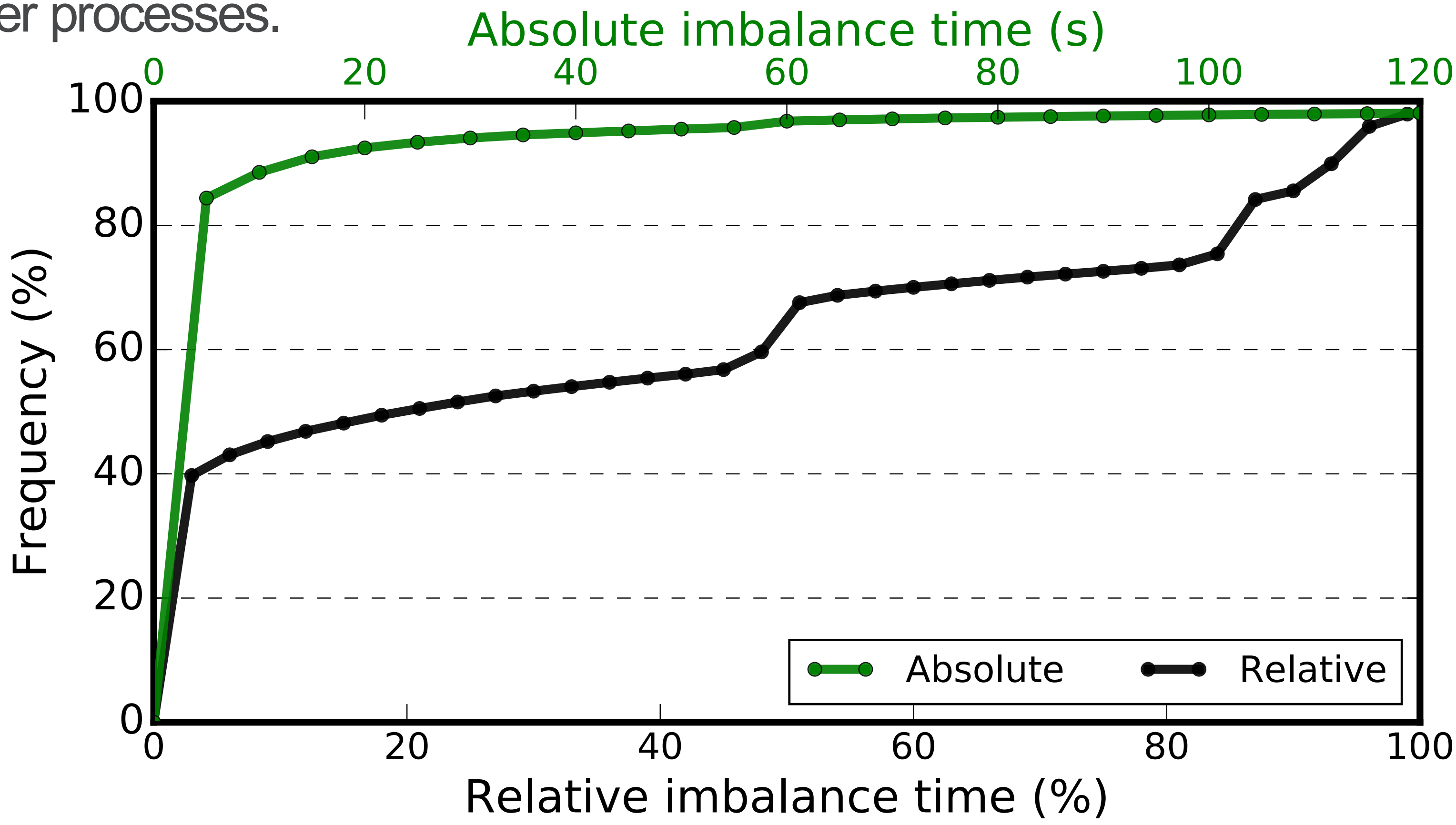
Cumulative distributions of absolute (left) and relative imbalance (right), **before** and **after** the Globus transfer service improvement.



- **Both absolute and relative imbalance have decreased.**
- **20% of the Globus transfers still experience an absolute imbalance of more than 20 seconds**
- **An equal percentage of transfers experience a relative imbalance of 25%.**

# Load imbalance among TCP flows

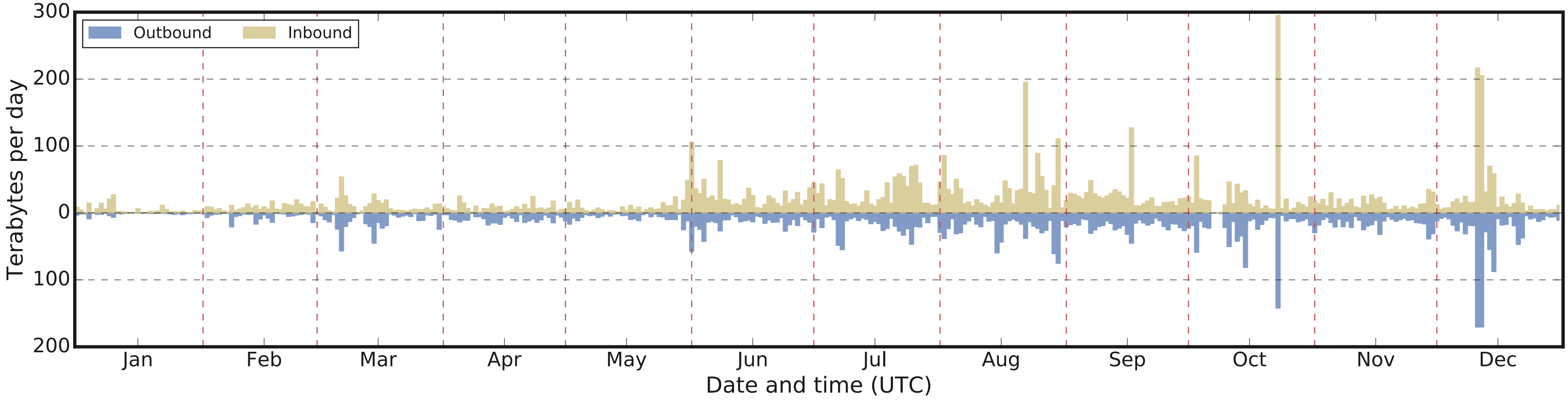
Cumulative distribution of the imbalance (in parallel TCP streams) of GridFTP server processes.



- **70th percentile values are less than 0.3 second and 0.4%**
- **Parallel TCP streams in 70% of server processes had little imbalance**
- **Parallel TCP streams in 20% of GridFTP server processes had an absolute imbalance time between 1 and 2 seconds.**

# Gap between peak and average usage

Terabytes per day in 2017



- The peaks are 170TB and 295TB for outbound and inbound transfers
- Averages are only 15.0TB and 19.6TB for outbound and inbound
- 75% of days have outbound and inbound volumes less than 18.7TB and 22.0TB

# Summary

✓ We characterized the network traffic of a computer facility's DTNs at multiple levels, from user transfer requests down to TCP flows.

✓ Combining the logs from different layers allowed us to identify load imbalances and opportunities for improvement in wide area data movement.

✓ The case study provides valuable insights into the design, operation, and management of data transfer nodes and data transfer tools.

✓ These insights are useful not only for optimizing existing systems and tools but also for planning system upgrades and future investments.

# THANKS FOR YOUR ATTENTION !

# QUESTIONS ?

Argonne National Laboratory is a
U.S. Department of Energy laboratory
managed by UChicago Argonne, LLC.