The 26th International ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC)

Explaining Wide Area Data Transfer Performance

By: Zhengchun Liu, Prasanna Balaprakash, Rajkumar Kettimuthu and Ian Foster

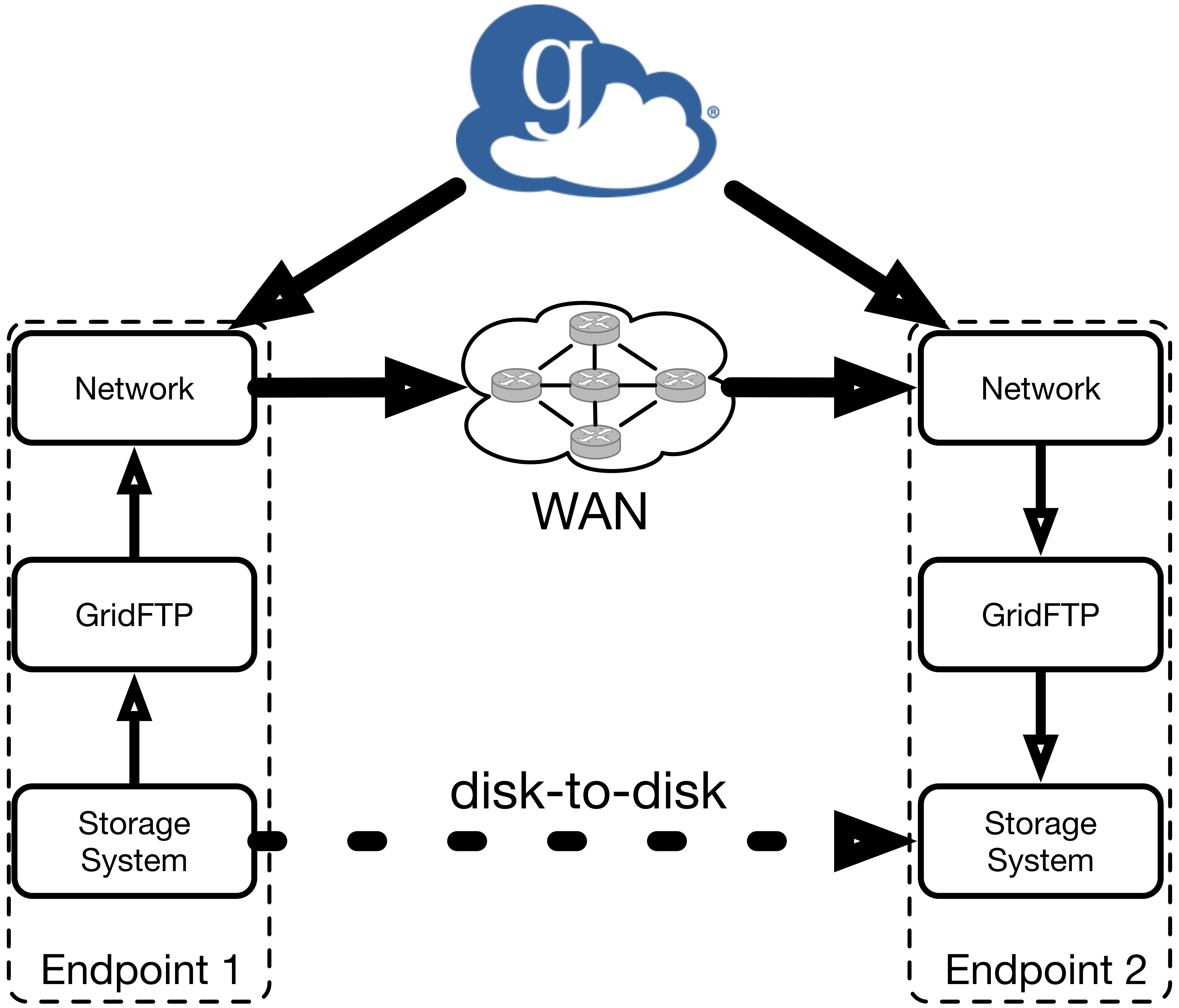
Presented by: Zhengchun Liu

Jun 29, 2017, Washington D.C.

26th HPDC

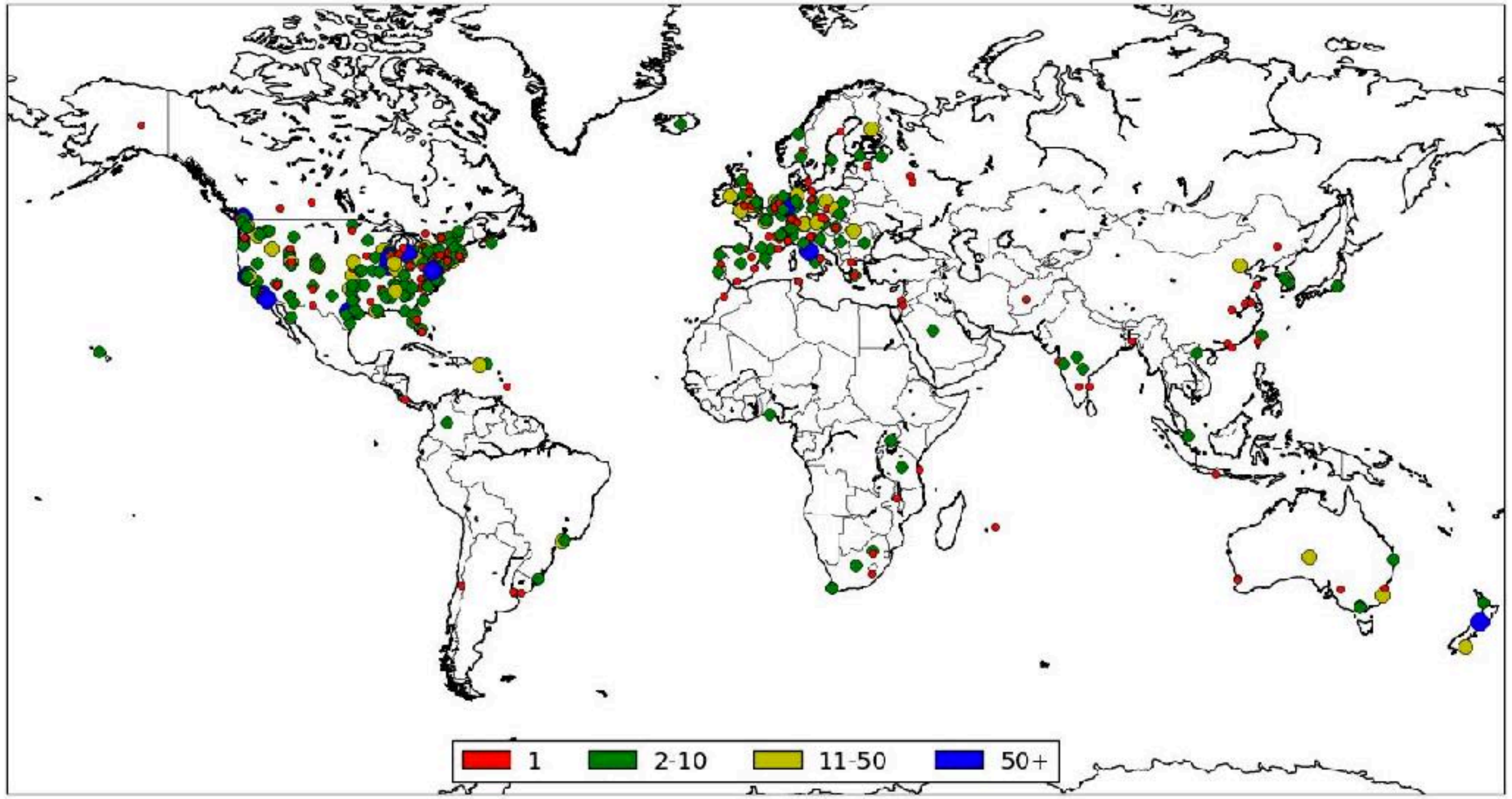
Argonne NATIONAL LABORATORY

Introduction - globus.org



The Globus transfer service is a cloud-hosted software-as-a-service, to provide convenient, reliable and secure file transfers service between pairs of storage systems

$$R^{max} \leq min(DR^{max}, MM^{max}, DW^{max})$$

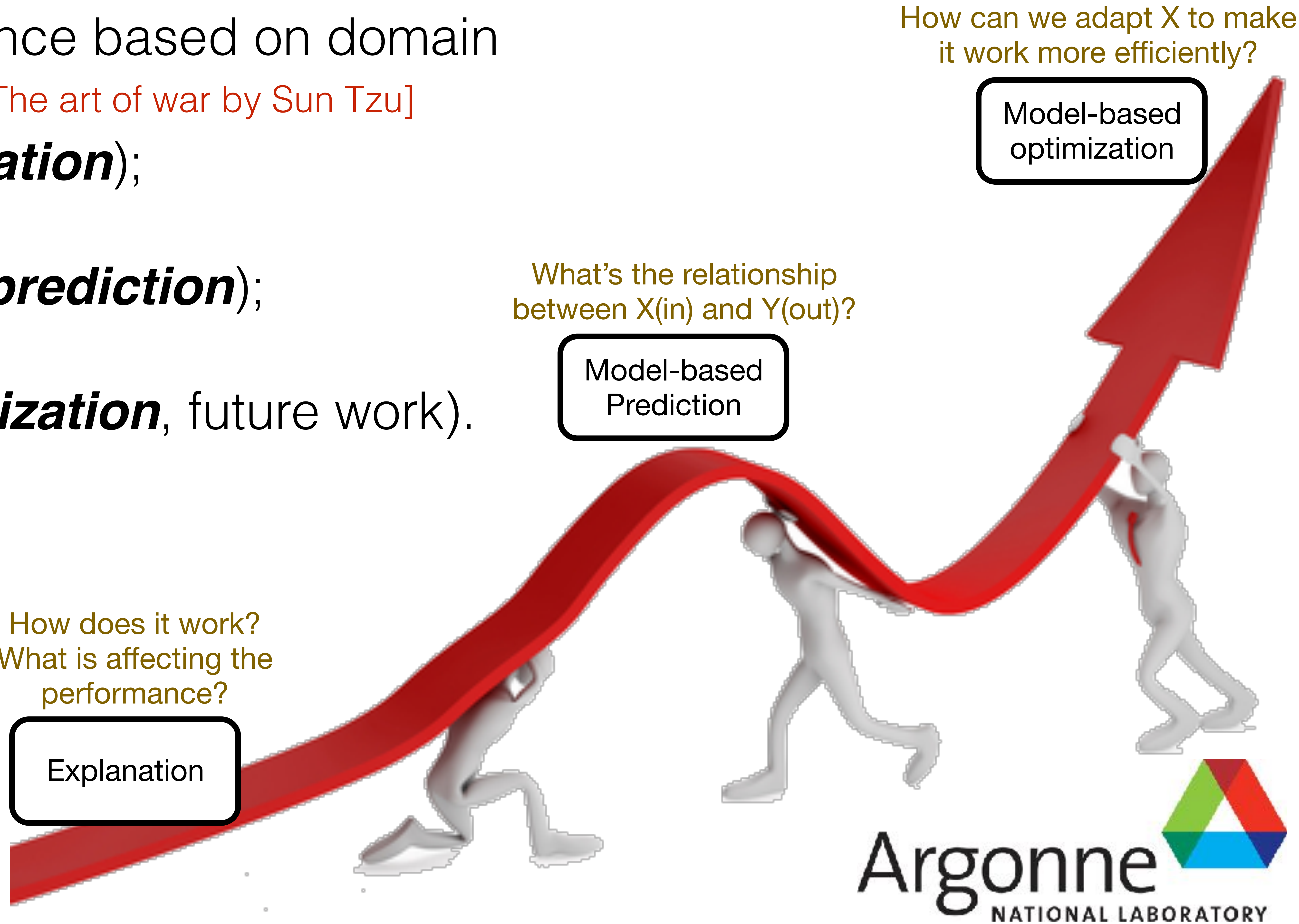


Globus endpoints, grouped by number of deployments in a single location. (Some endpoints geolocate erroneously to the center of countries.)

Motivation

Armed with a large collection of Globus wide-area file transfer records, and experiments performed in the ESnet testbed environment, we want to:

- Extract factors that affect the transfer performance based on domain knowledge, and study their importance (***explanation***);
  [If you know yourself and your enemy, you'll never lose a battle. — The art of war by Sun Tzu]
- Build models to predict transfer performance (***prediction***);
- Model based performance optimization (***optimization***, future work).



Outline

- *Background & Motivation;*
- Which factors are affecting the transfer performance (*qualitatively*)?
- Deriving features based on domain knowledge, to explain transfer performance (*quantitively*).
- Building models to make prediction by using derived features (*validate feature explainability*).
- Conclusion and future work.

What affect transfer performance?

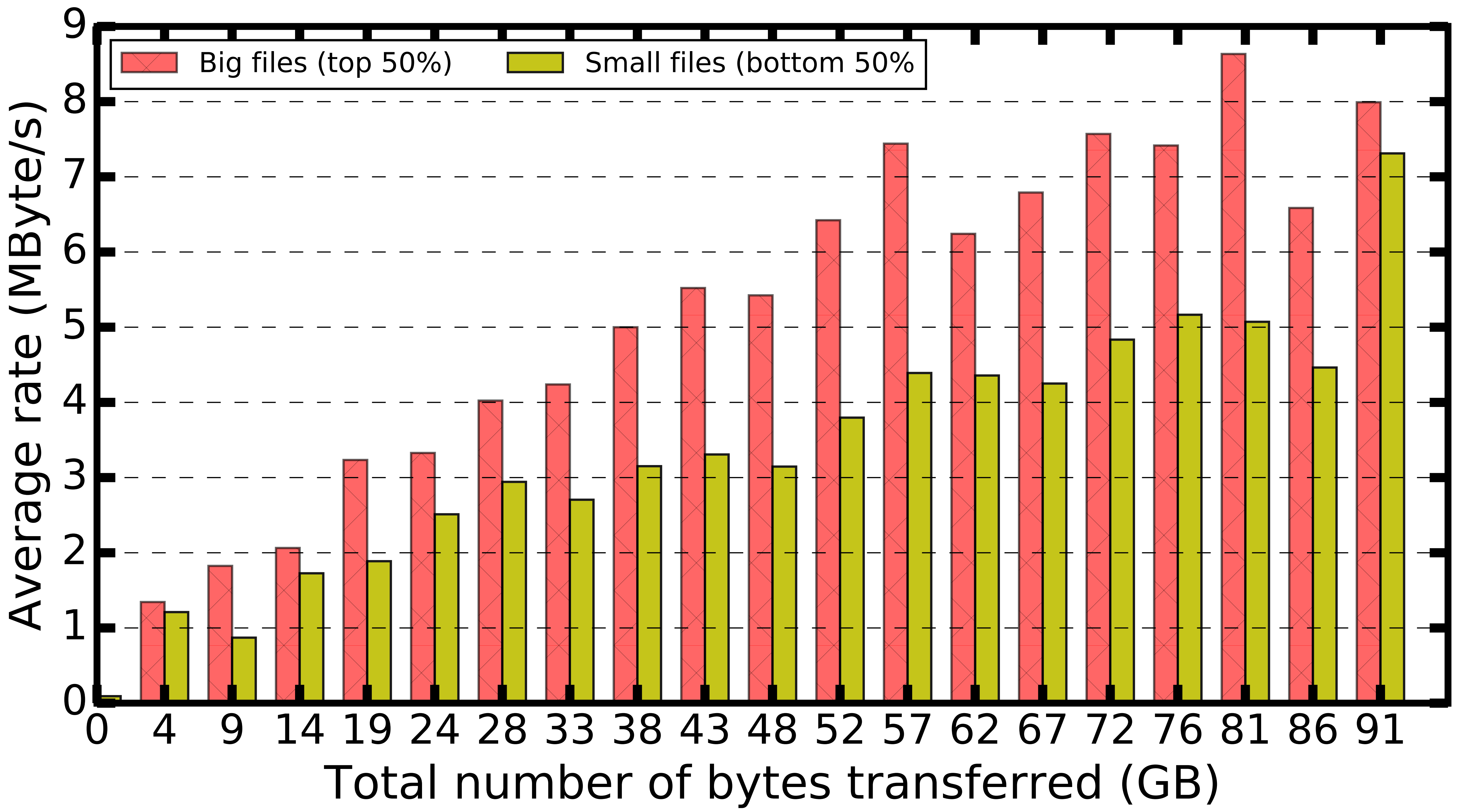
4 kinds (3 known and 1 unknown):

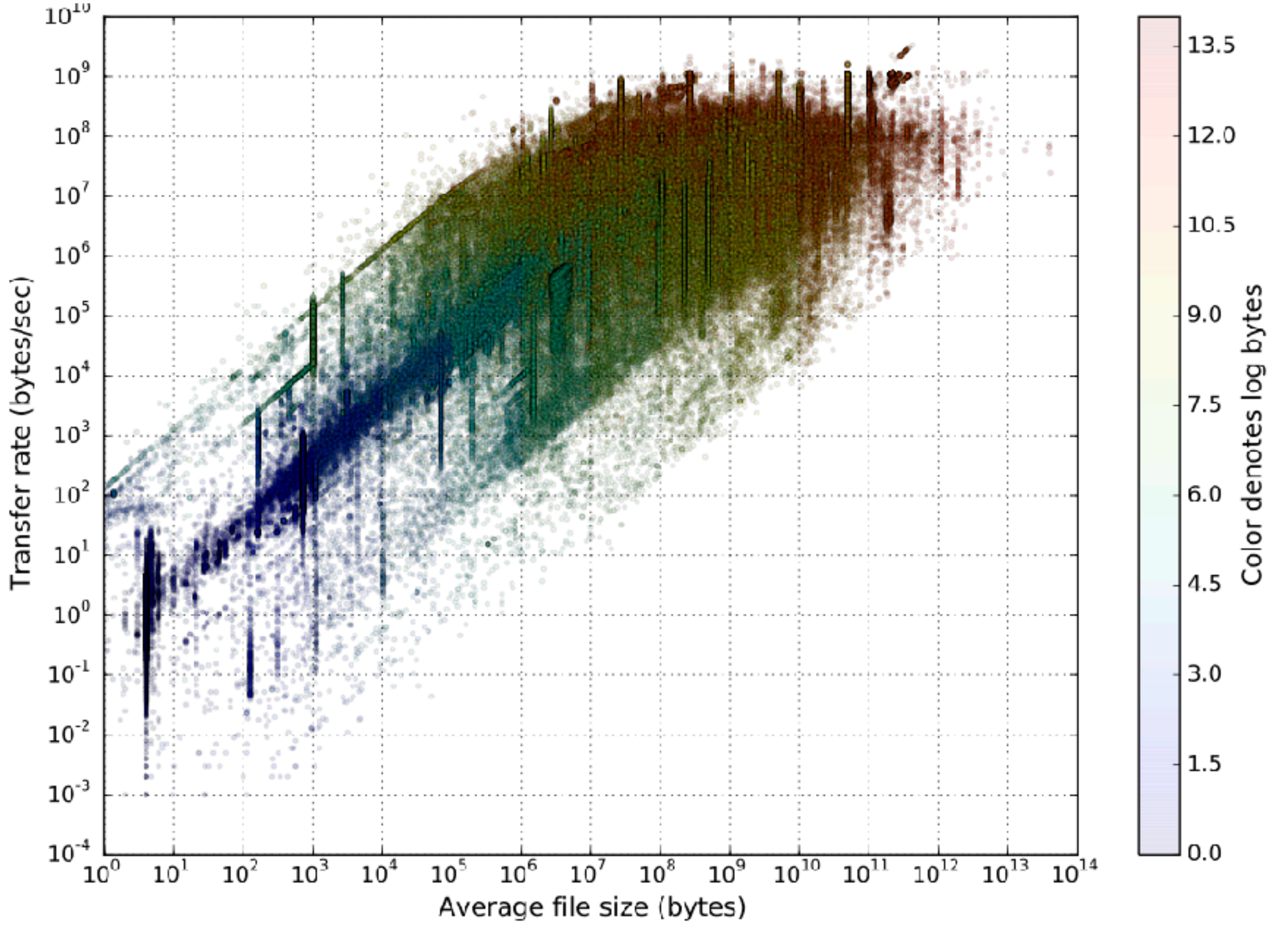
For a given endpoint pair:

1) Transfer file characteristic, e.g., file size;
2) Tunable transfer parameters, e.g., concurrency (flying files), parallelism;
3) Contentions from other simultaneous Globus transfers (known to us) and,
4) Contentions from other programs (unknown to us), e.g., sharing PFS with SC, network.
   (a way to clean the data)

What affect transfer performance? -1

File characteristics:

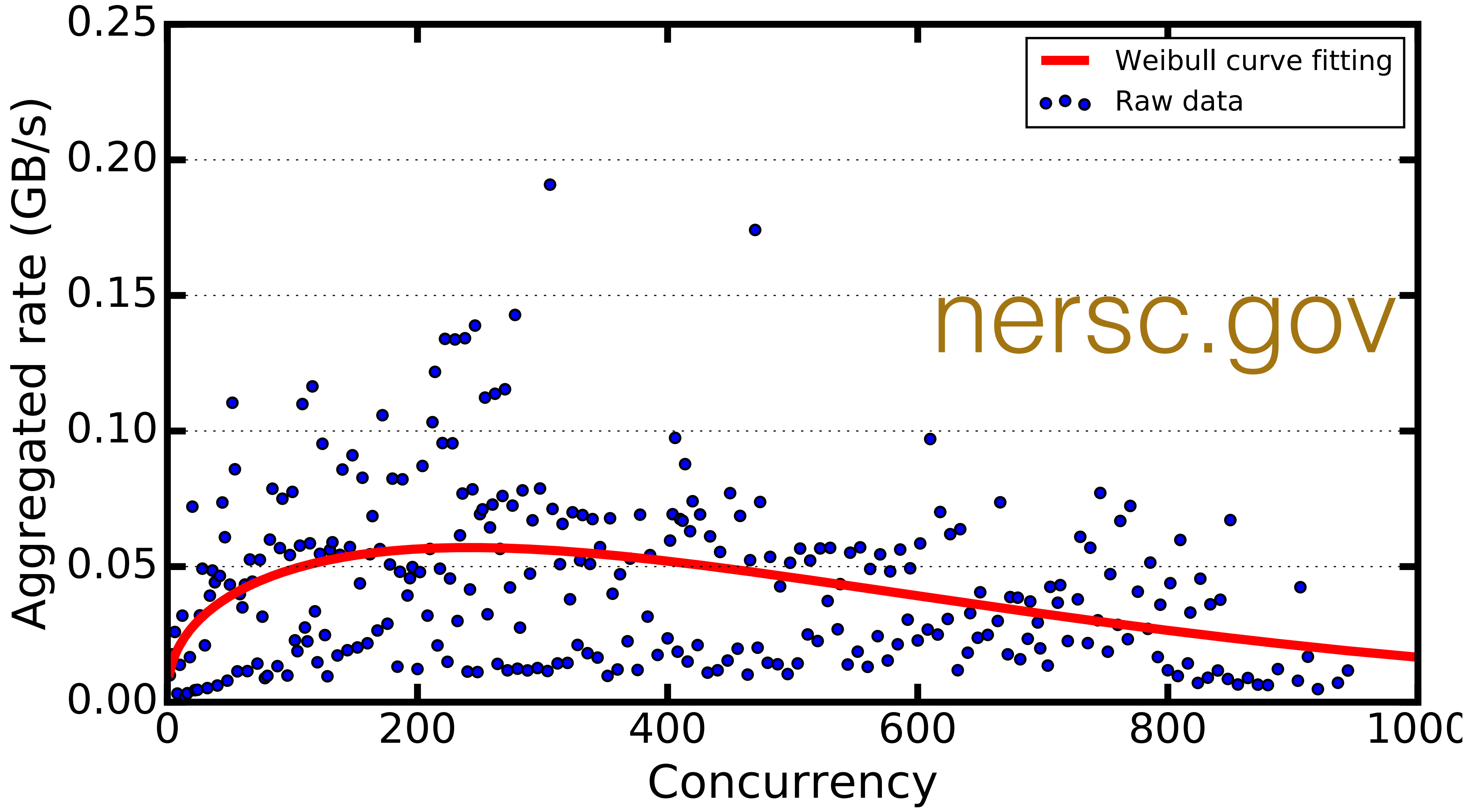


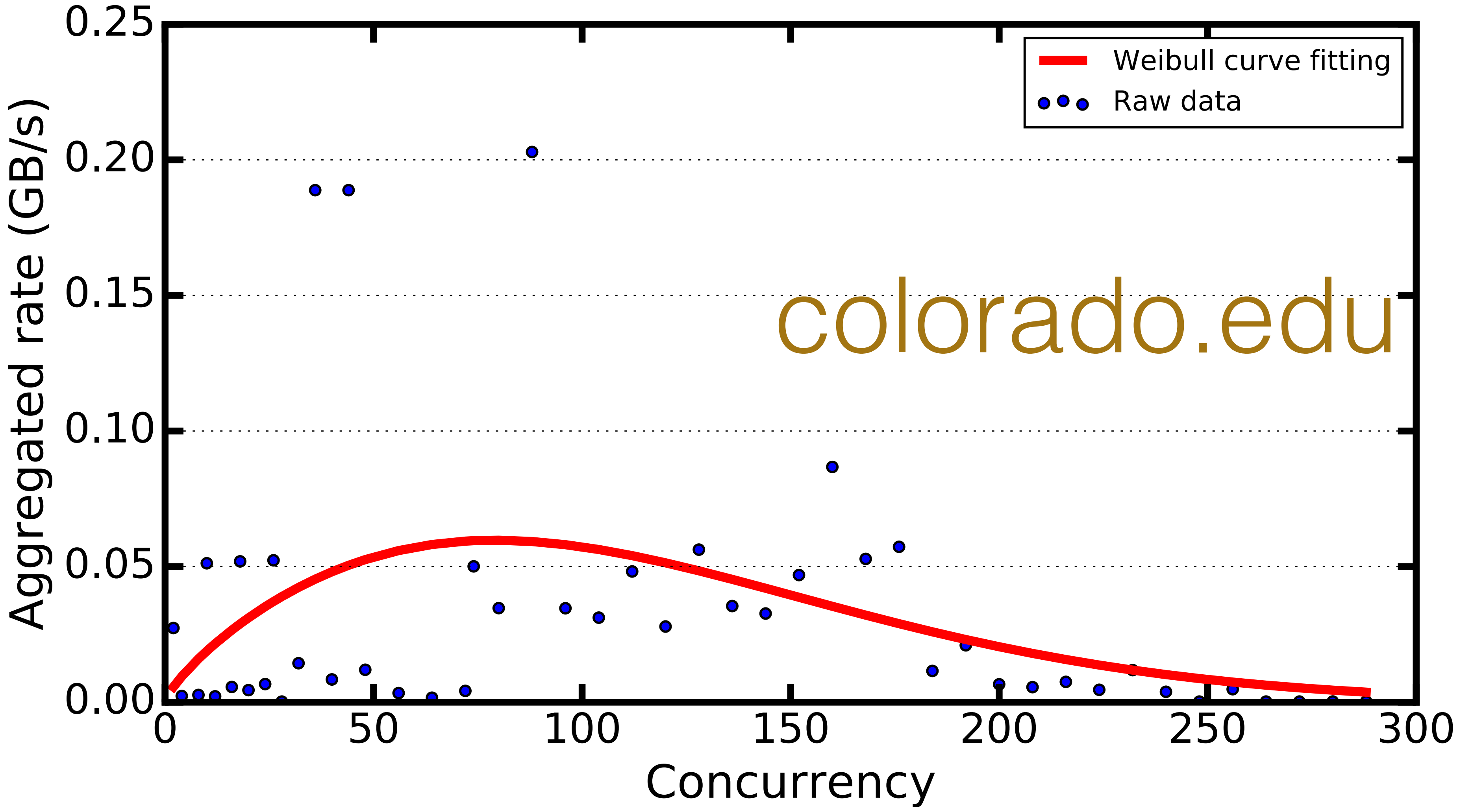


Large transfers with big average file size are more likely to have better performance.
I.E, The startup cost is high.

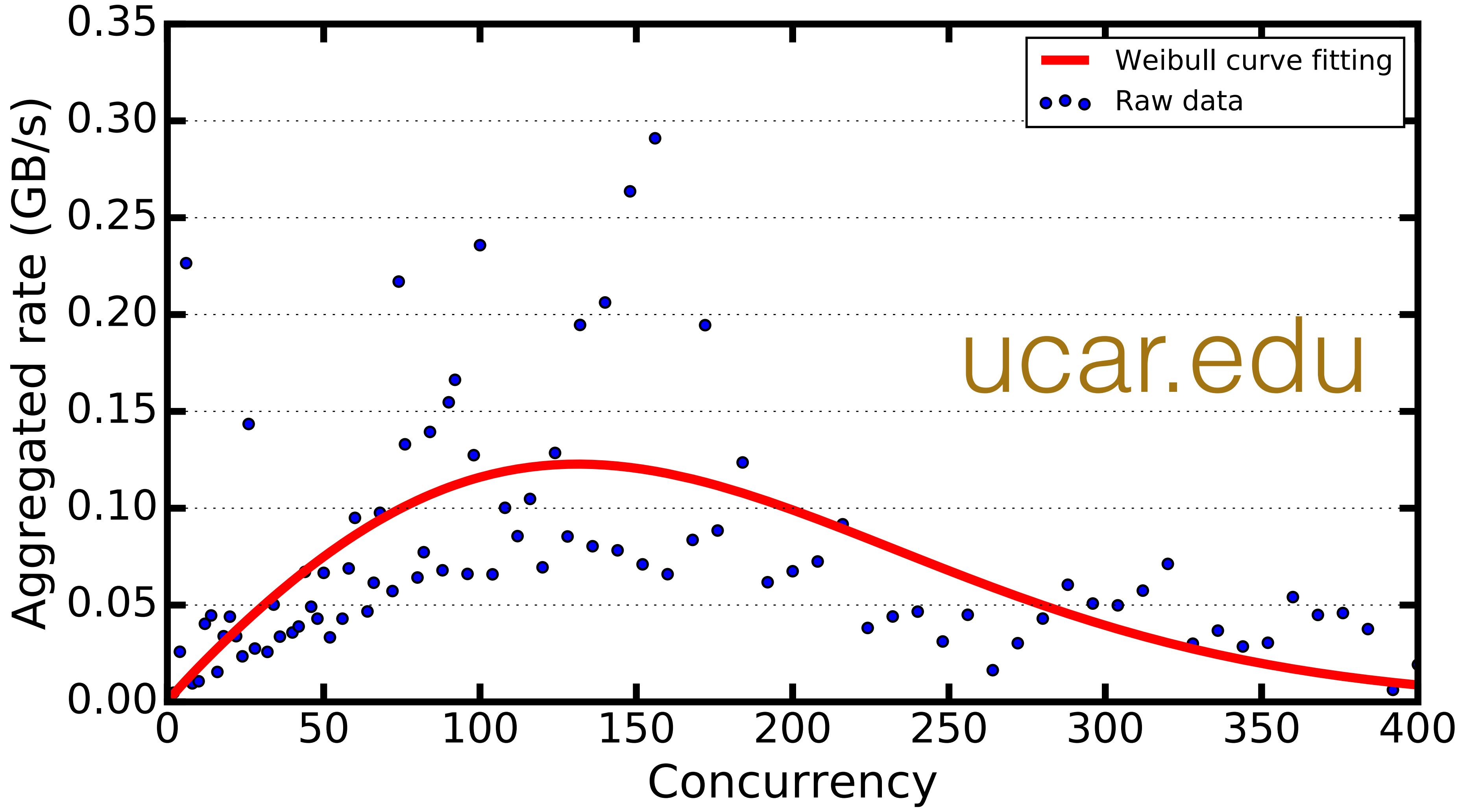
What affect transfer performance? -2

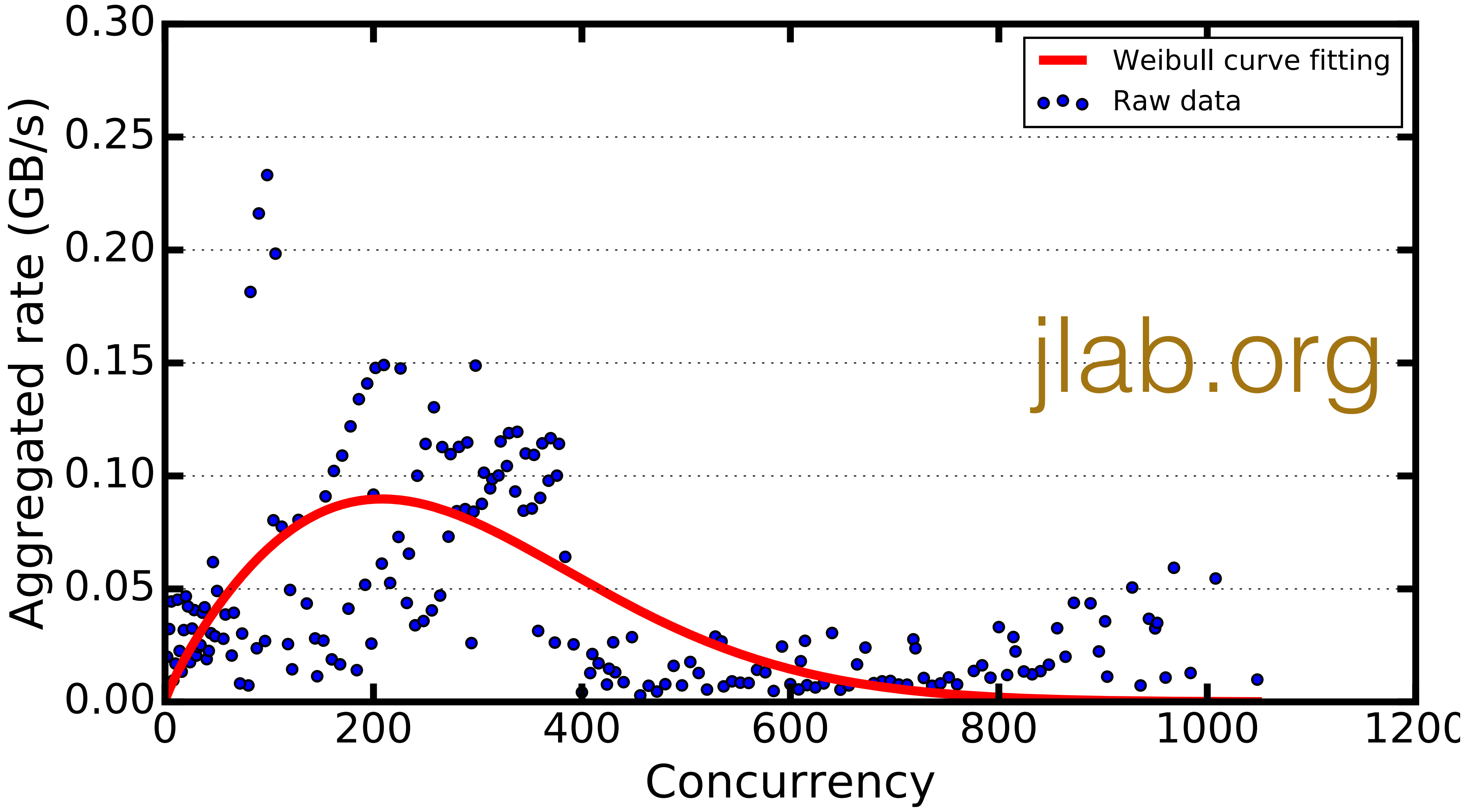
Tunable transfer parameters





Aggregated concurrency ***versus*** aggregated throughput on the data transfer node.



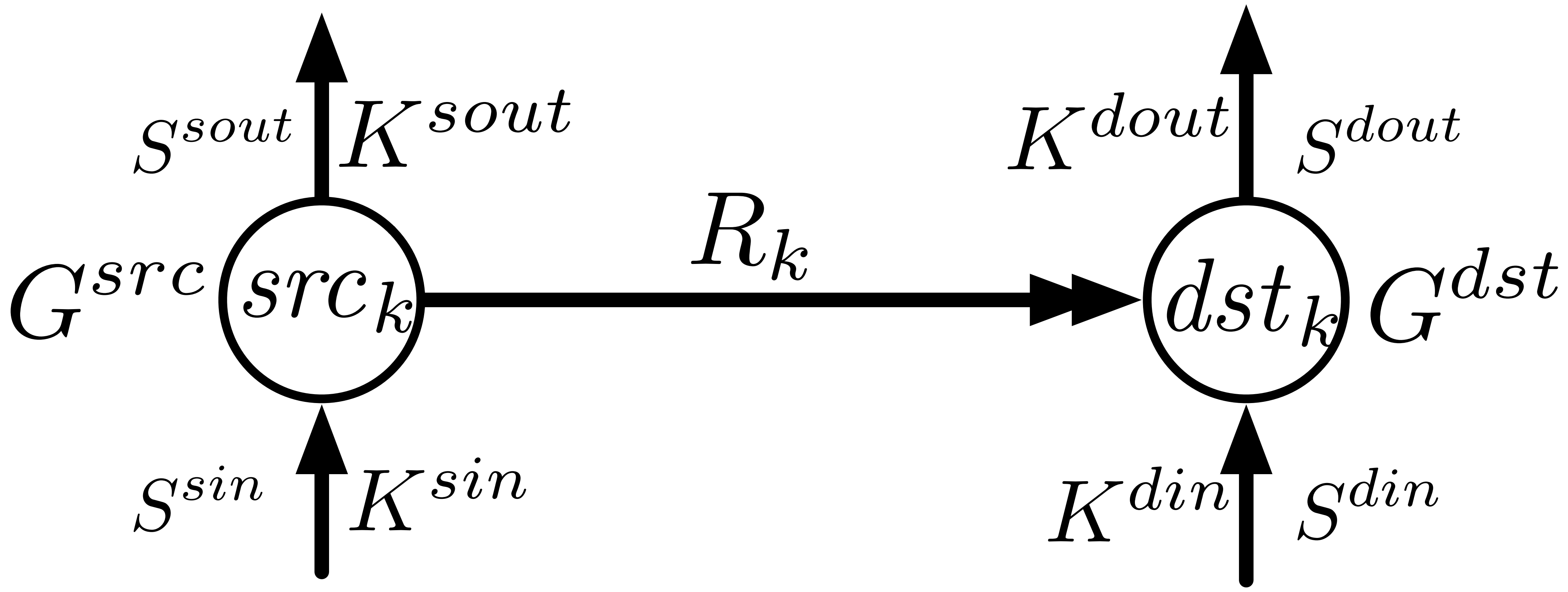


What affect transfer performance?

- [x] *Transfer file characteristic, e.g., file size;*
- [x] *Tunable transfer parameter, e.g., concurrency, parallelism and pipeline;*
- [ ] **Contentions from other Globus transfers (known to us).** 👍
- [ ] **Contentions from non-globus programs (unknown to us), e.g., sharing storage, network.** 👎

What affect transfer performance -3?

Contention from simultaneous globus transfers (I/O, NIC, CPU & RAM):



Load experienced by a Globus transfer *k* from src_k to dst_k with rate R_k

E.G.,

The Globus **contending transfer rate** for a transfer k at its source (src^k) and destination (dst^k) endpoints is

$$K^{x\in\{sout,sin,dout,din\}}(k) = \sum_{i\in A_x} \frac{\mathcal{O}(i,k)}{Te_k - Ts_k} R_i, \tag{1}$$

where A_x is the set of transfers (excluding k) with src_k as source when $x = sout$; src_k as destination when $x = sin$; dst_k as source when $x = dout$; and dst_k as destination when $x = din$. $\mathcal{O}(i,k)$ is the overlap time for the two transfers:

$$\mathcal{O}(i,k) = \max(0, \quad \min(Te_i, Te_k) - \max(Ts_i, Ts_k)).$$

Features to explain a transfer

K^{sin}	Contending incoming transfer rate on src_k.
K^{sout}	Contending outgoing transfer rate on src_k.
K^{din}	Contending incoming transfer rate on dst_k.
K^{dout}	Contending outgoing transfer rate on dst_k.
C	Concurrency: Number of GridFTP processes.
P	Parallelism: Number of TCP channels per process.
S^{sin}	Number of incoming TCP streams on src_k.
S^{sout}	Number of outgoing TCP streams on src_k.
S^{din}	Number of incoming TCP streams on dst_k.
S^{dout}	Number of outgoing TCP streams on dst_k.
G^{src}	GridFTP instance count on src_k.
G^{dst}	GridFTP instance count on dst_k.
Nf	Number of files transferred.
Nd	Number of directories transferred.
Nb	Total number of bytes transferred.

What affect transfer performance? -3

$$ReL = max\left(\frac{K^{sout}}{R_k + K^{sout}}, \frac{K^{din}}{R_k + K^{din}}\right)$$

Transfers over ESnet testbed

(less likely to have non-globus load on endpoints)

Transfer over production DTN

(more likely to have non-globus load on endpoints)

-4 Contention from other non-globus program also matter!!!

Machine learning models to predict performance

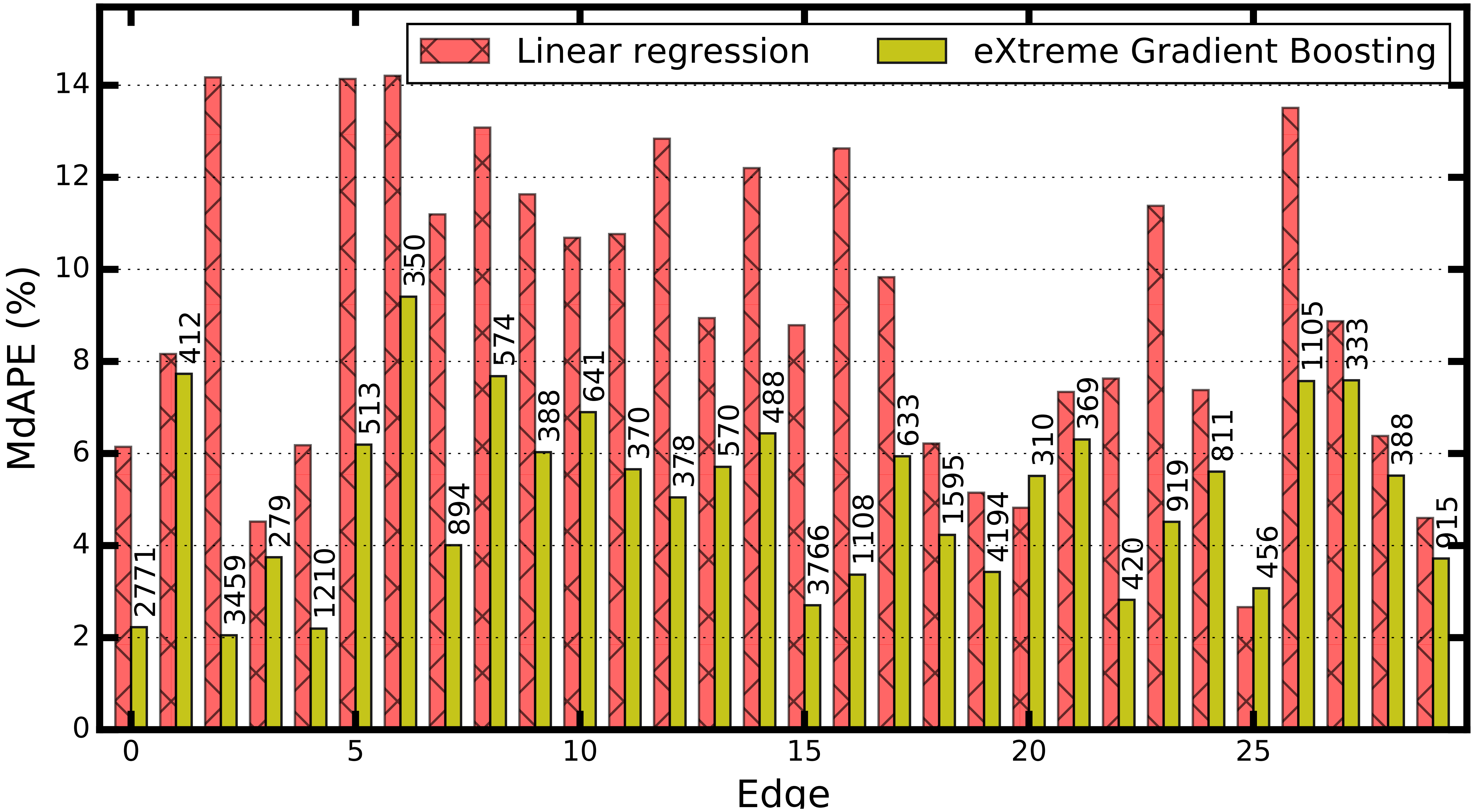
- One model for one (source-to-destination) edge;
- Linear model and nonlinear model (Extreme Gradient Boosting *);
- 70% for training and 30% for testing;
- Data cleaning: remove transfers that are likely to have unknown load;
- One general model for all endpoint pairs (with two extra features to characterize endpoint);
- A representative set of 30k transfer over 30 heavily used edges.

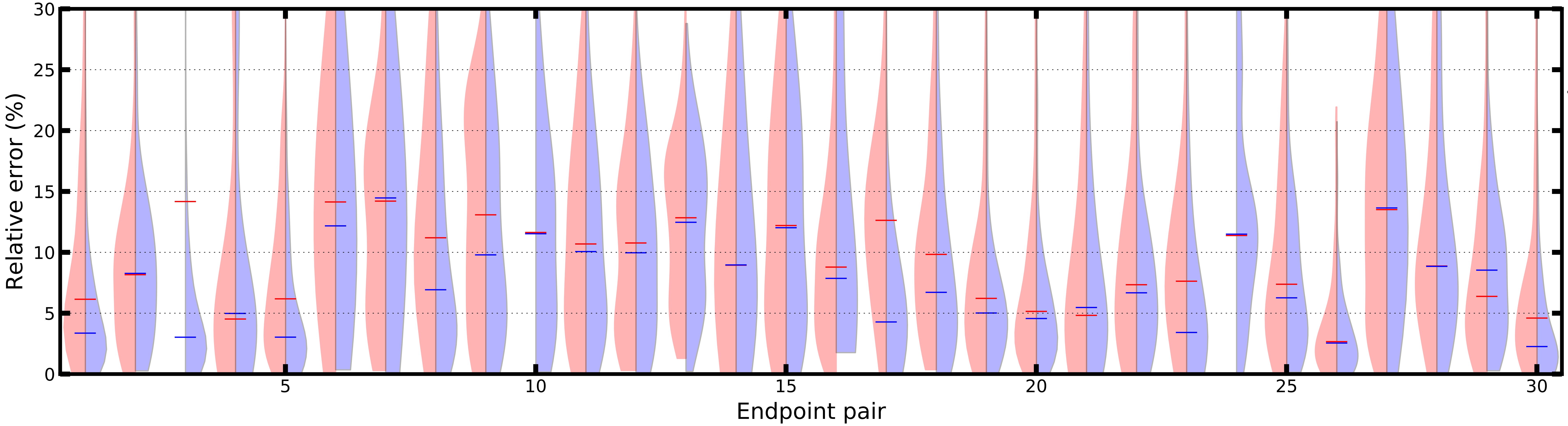
* https://xgboost.readthedocs.io/en/latest/

Data driven models to predict transfer performance

Linear versus nonlinear

Obviously, nonlinear model outperforms linear model, implies that the relationship is nonlinear.

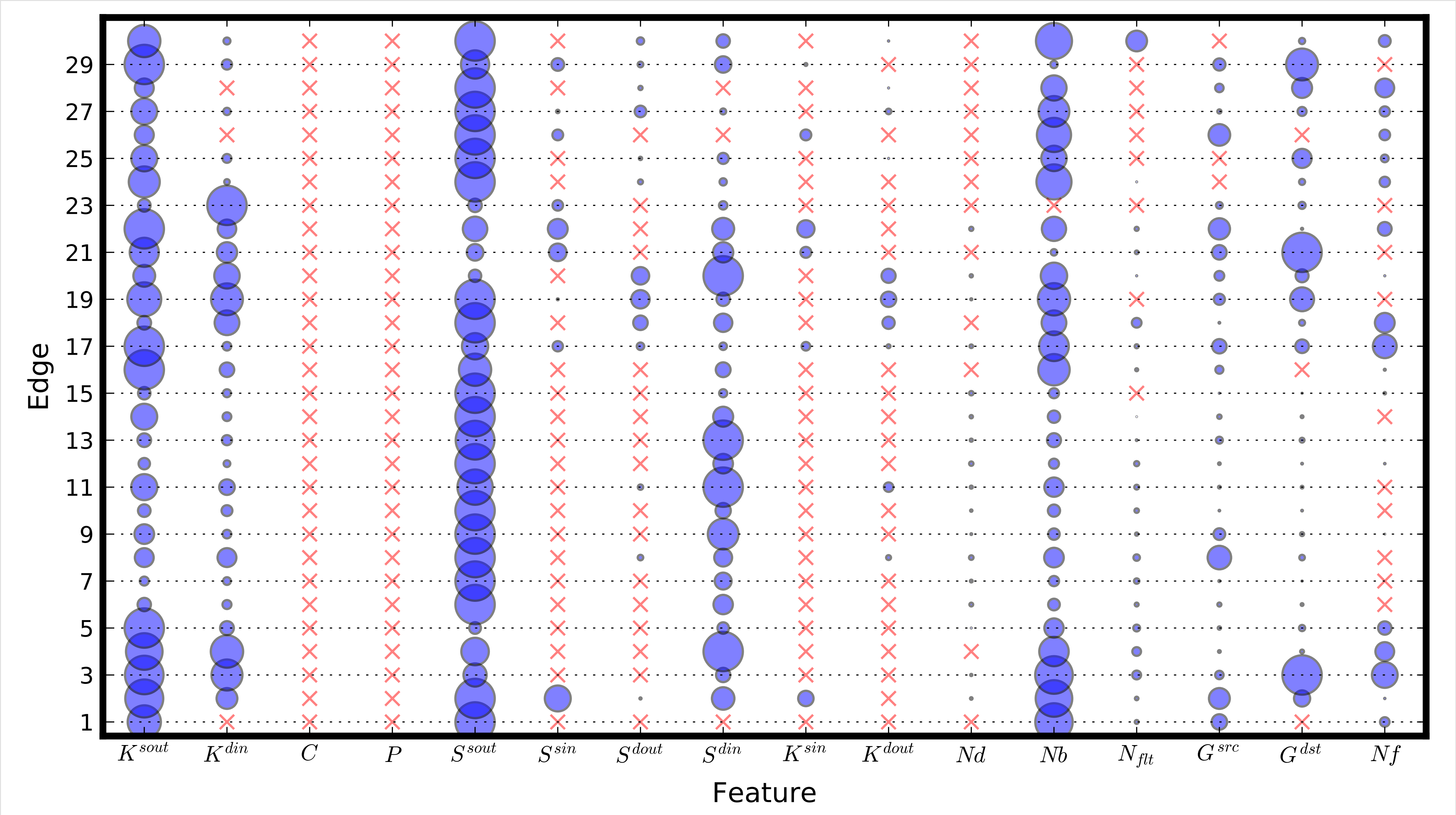




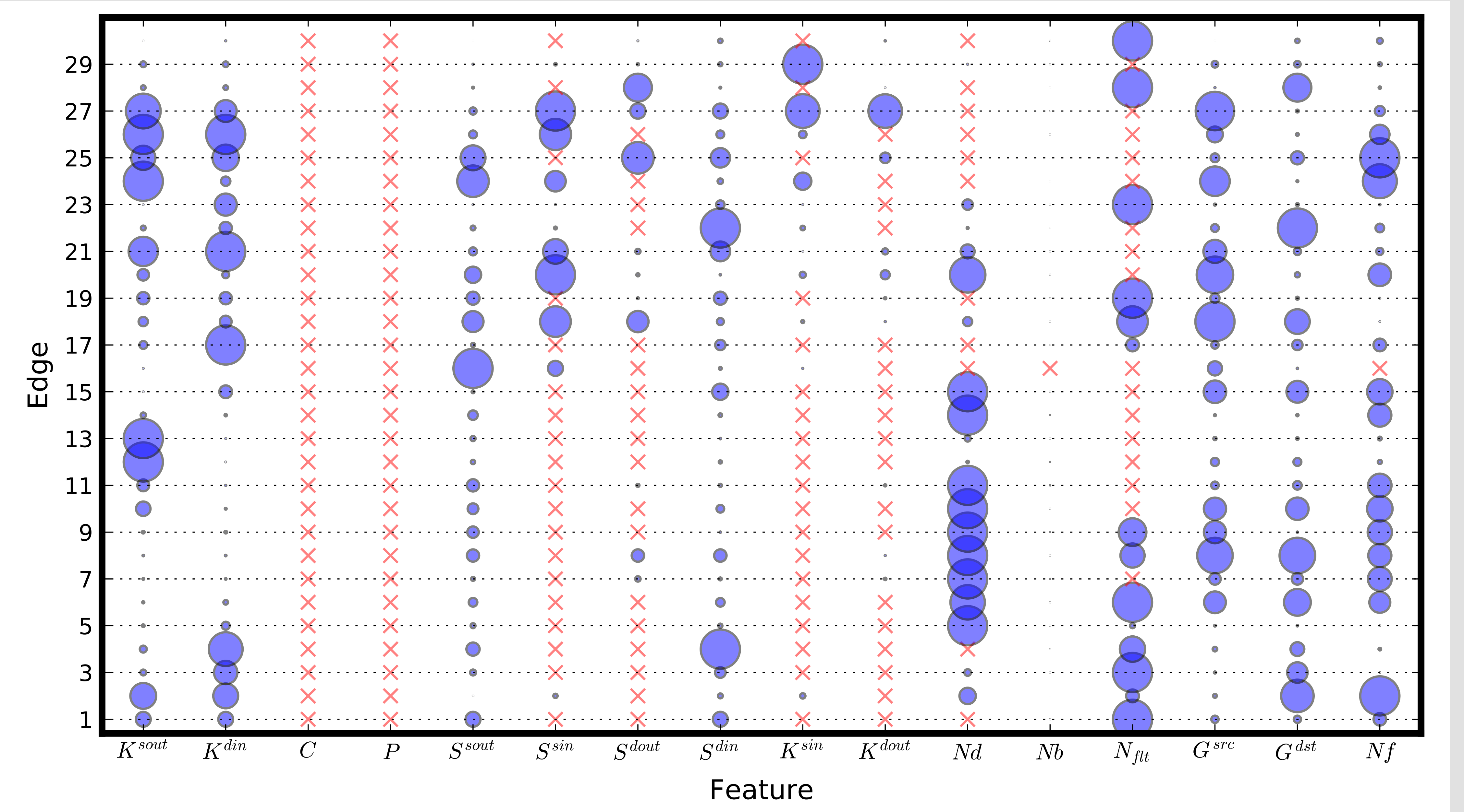
The left Violin plot gives the prediction error of the linear regression model and the right the prediction error of eXtreme Gradient Boosting model.

Model-based feature importance

Circle size indicates the relative significance of features in the linear model, for each of 30 edges. A red cross means that the corresponding feature is eliminated because of low variance.



Linear regression model based feature significance



Nonlinear regression model (XGBoost) based feature significance

What we learned:

Resource contention at endpoint is clear, K^{sout}, K^{din}, S^{sout} and S^{din} are significant in the models.
Total transfer bytes also matters, means that the startup cost is high.

Derived feature data: http://dx.doi.org/11466/globus_A4N55BB

Argonne NATIONAL LABORATORY

Prediction

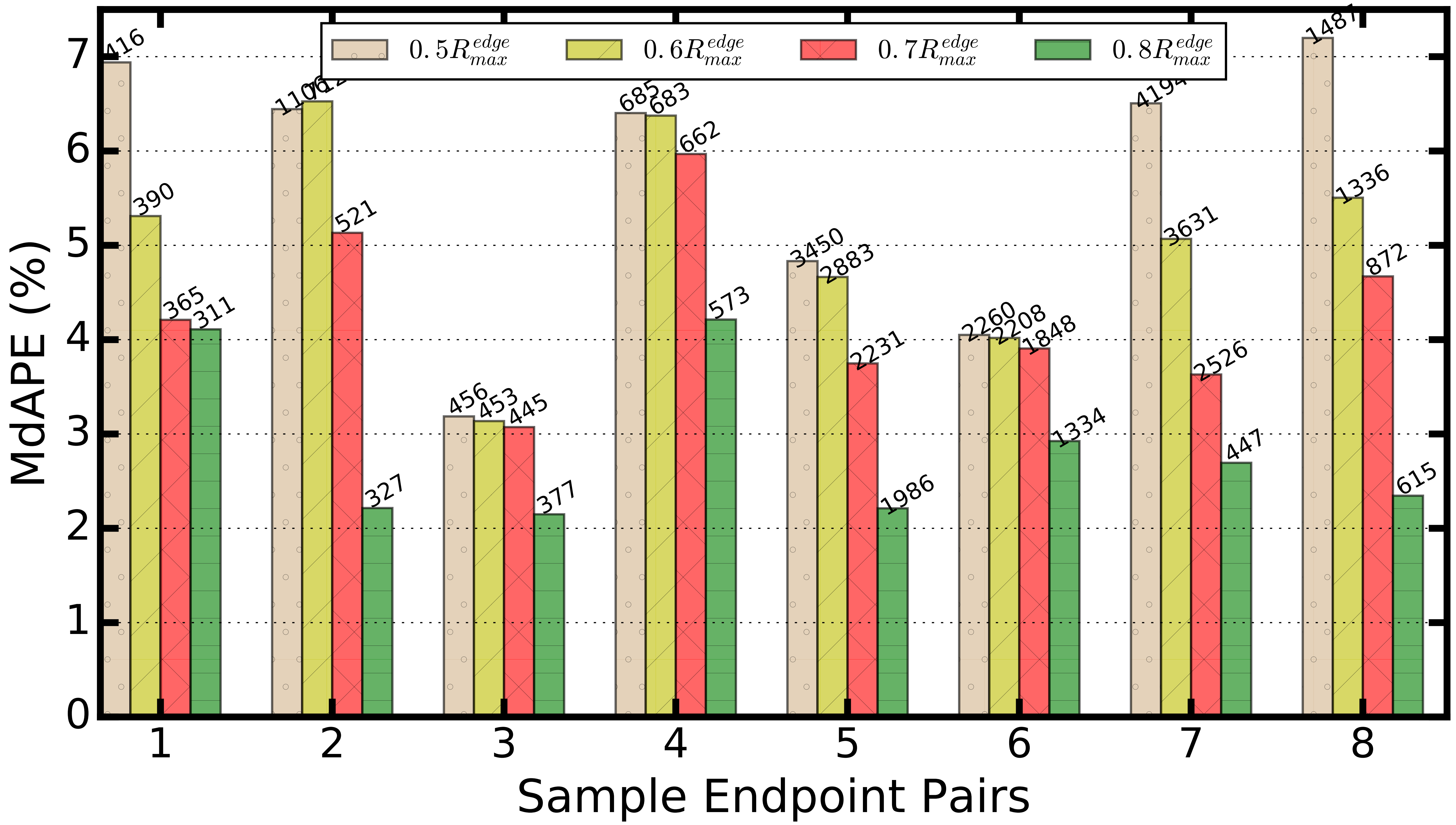
Influence of unknown load:

Select transfers with:

$$\frac{R_k + K^{sout}(k)}{ROmax} \geq \eta \quad and \quad \frac{R_k + K^{din}(k)}{RImax} \geq \eta$$

$\eta \in \{0.5, 0.6, 0.7, 0.8\}$

large η means less likely to have unknown load because the max is fixed.



Applicability to other tools

Although we performed this work using Globus data, we believe that our methods and conclusions are applicable to all wide area data transfers.

Because:

- The data we used (e.g. *number of TCP connections, number of concurrent transferring files, size of the data transfer, number of files*) to derive features are **generic features** that impact the performance of any wide area data transfer, irrespective of the tool employed.

- The raw data to derive our features can be obtained in a **straightforward** fashion for other data transfer tools such as FTP, rsync, scp, BBCP, FDT, and XDD.

Conclude and Future work

- Gain insights into the behavior of wide area data transfers.
- We derived features from Globus transfer log and studied their importance.
- We tried to make prediction based on the features we derived.
- Our models achieve good accuracy when there is less unknown load.

- Unknown load coming from non-globus program is “unavailable”;
- Can cutting edge methods, e.g deep learning, help for more accurate prediction?

Thank you for your attention!

We also want to THANK:

- U.S. Department of Energy, Office of Science, ASCR, and the program manager *Richard Carlson*;
- *Nagi Rao* for useful discussions, *Brigitte Raumann* for help with Globus log analysis;
- *Glenn Lockwood* for help with experiments at NERSC.

Q & A

Argonne NATIONAL LABORATORY