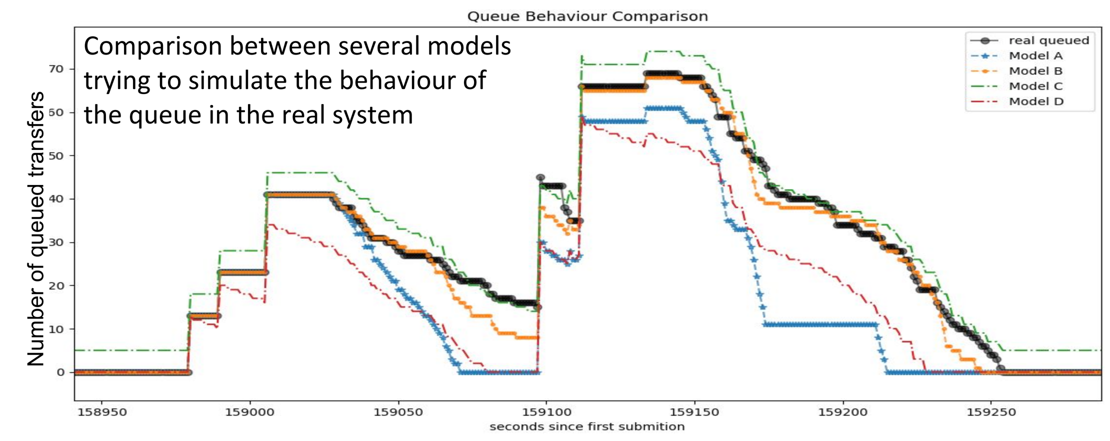
Modelling High Energy Physics Data Transfers in ATLAS

Joaquin Bogado, Fernando Monticelli, Mario Lassnig, Ilija Vukotic, Javier Díaz
on behalf of the ATLAS Collaboration
Contact: *rucio-dev@cern.ch*



Introduction

The study is based on data from the distributed data management system Rucio, in charge of managing all ATLAS data on the grid. Rucio is free and open source and in use by other collaborations.

Objective of this work (part of an ongoing PhD study)

- to understand the interactions between the different systems in the scientific data management environment (e.g., Rucio, FTS, storage, network, …)
- to identify problematic scenarios that could lead to delays in transfer times
- to make predictions about how much time a transfer will take at submission time
- to allow decision makers to improve the performance of the architecture using this knowledge

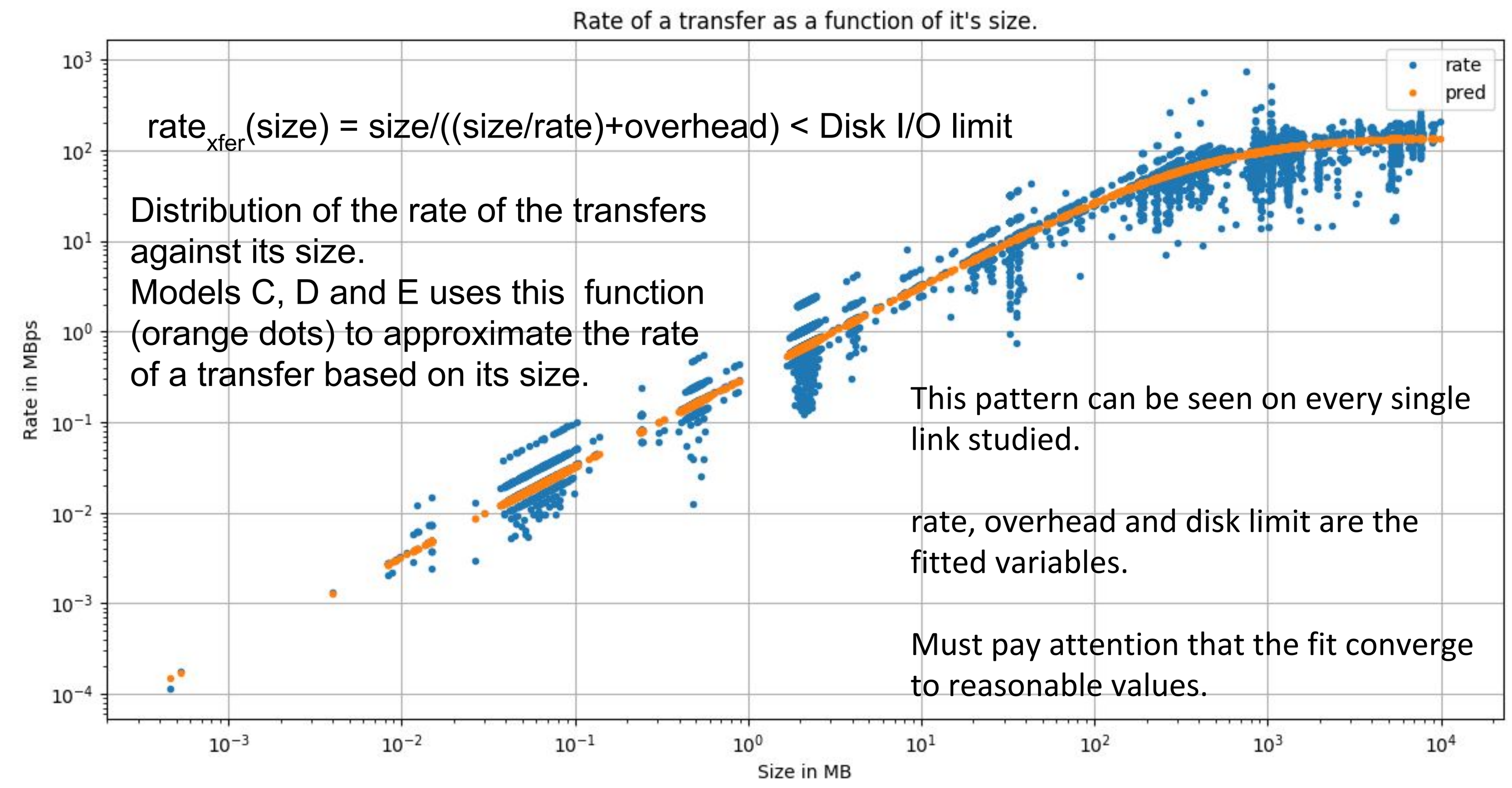
About Rucio

- The main purpose of the system is to help the collaboration to store, manage, and process experiment data (detector, simulation, user) in a heterogeneous distributed environment.
- Typical tasks are:
  - Transfer data to and from sites
  - Delete data at sites
  - Enforce the experiment computing model and policies
- The system manages 400+ Petabytes of physics data across more than 120 data centres globally with 1 billion files.

Simulations process

- Five models were studied, the difference between them relies on how the number of active transfers and the available bandwidth is calculated.
- Every second in the lifetime of the transfer is simulated.
- Creation and submission times are taken from Rucio's database.
- Transfers sent to a simulated FTS Queue until they can be removed.
- The number of transfers that can be served (the number of concurrent active transfers) is calculated based on real data and depends on the model.
- When a transfer exits FTS Queue, it's Network Time is simulated dividing the available bandwidth among the current amount of active transfers.
- The available bandwidth calculation depends on the model.

Bandwidth approximation

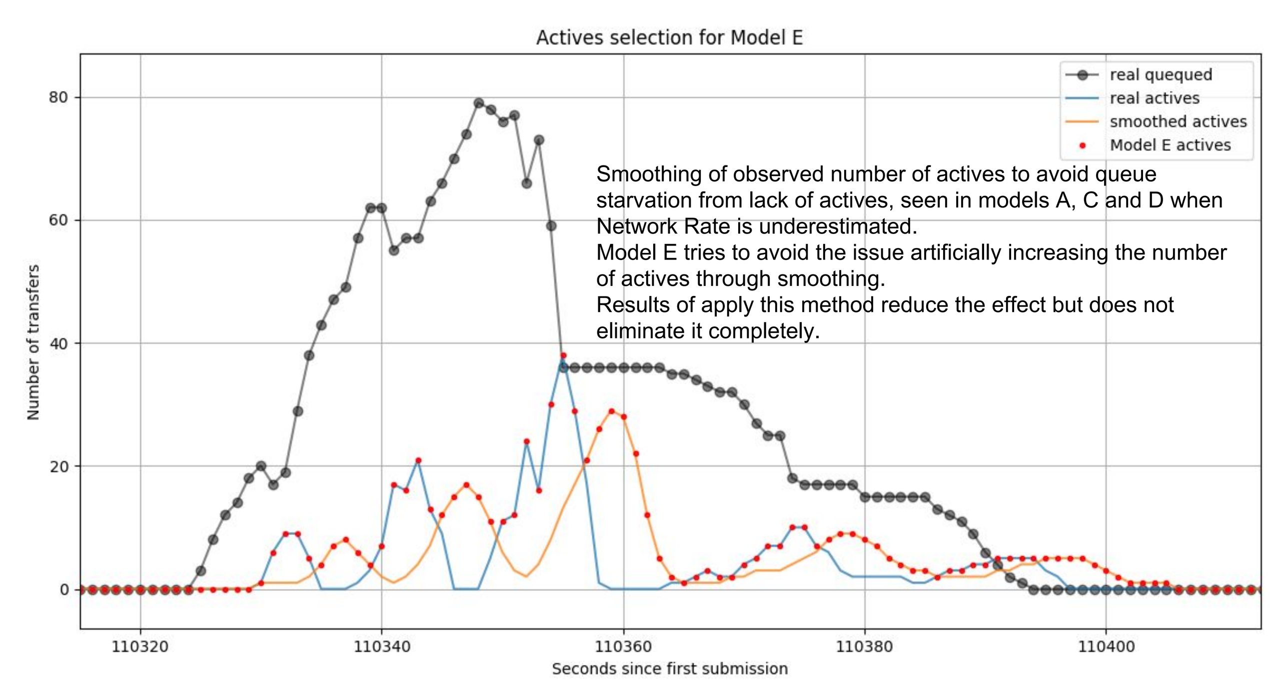


Challenges

- Dealing with real, unfiltered, unlabeled data, directly extracted from Rucio's database.
- Data is difficult to filter. Can't discard anything without being sure is not related with the variables we need to study.
- The amount of data is huge, about 3 million transfers per day. Dataset using to study is 2.5 million transfers for 5 days.
- Overall system is rich in complex interactions with other tools (in other words: tape systems are weird.)
- Rucio is composed of several demons, each with several instances.
- Different instances of FTS servers, each of which is using different configurations.

Number of concurrent active transfers selection

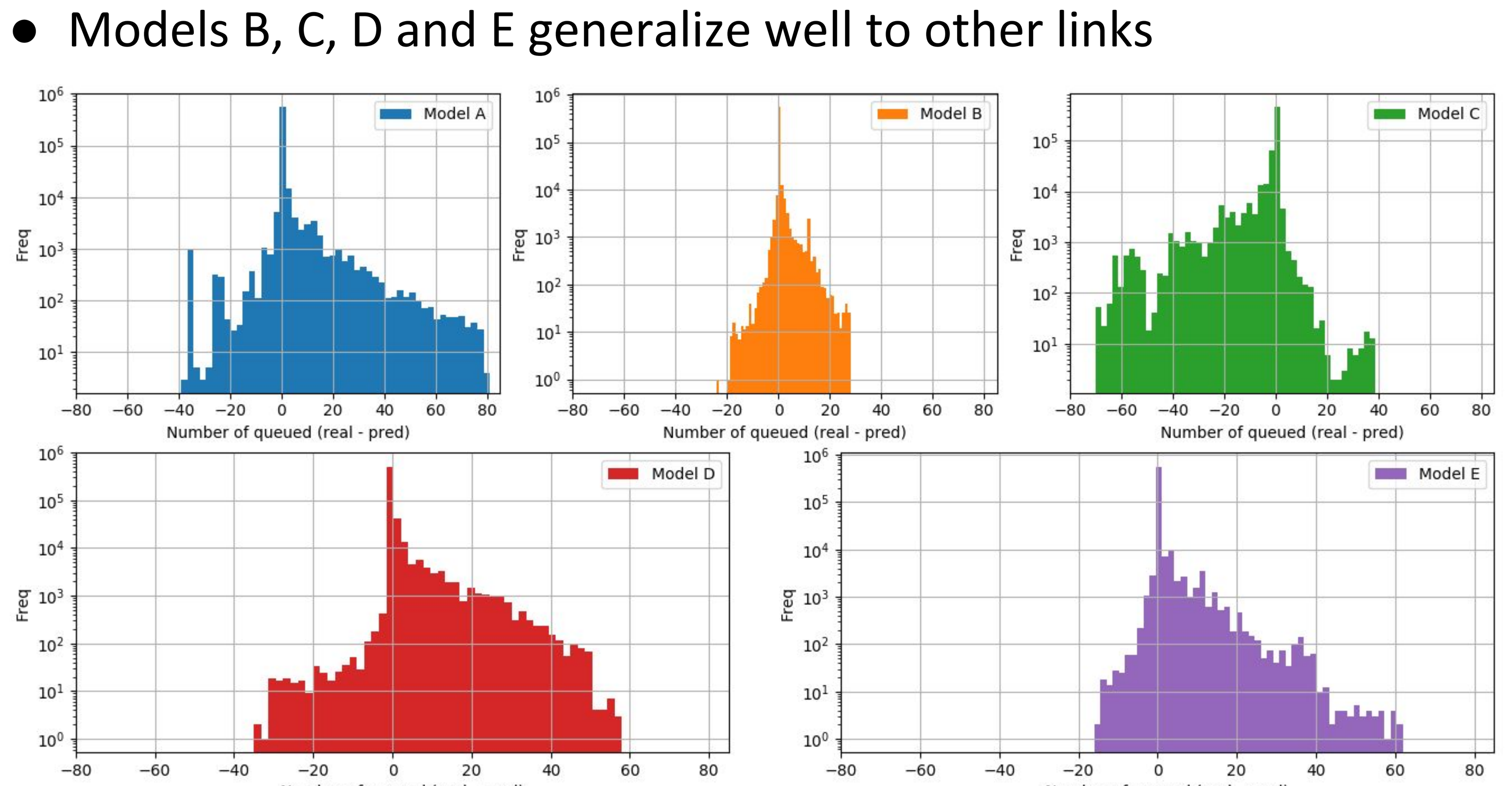
The number of concurrent active transfers can't be know in real time.



This is one of the most important limitations for the models and represent the number of transfers the model can take out of the simulated queue.

Results

- Model B got the better results, but is the model that also have more information about the system.
- Model D and E have similar performance, yet model D is simpler and thus, preferred. Model D underestimate the number of queued.
- Model A does not generalize well and thus, other models should be preferred in favor.
- Models B, C, D and E generalize well to other links



Next steps

- More knowledge is being gathered about FTS internals. Optimizer algorithm is being studied to predict the number of concurrent active transfers.
- Active limits are imposed by FTS configuration over Links (source and destination Sites) but also over at storage level on the sites.
- ATLAS experiment also use Activity Shares to avoid a group of transfers saturate the link impeding other transfers to progress.
- This variables play an important role determining how FTS Queues behave, and therefore will be included in future work.