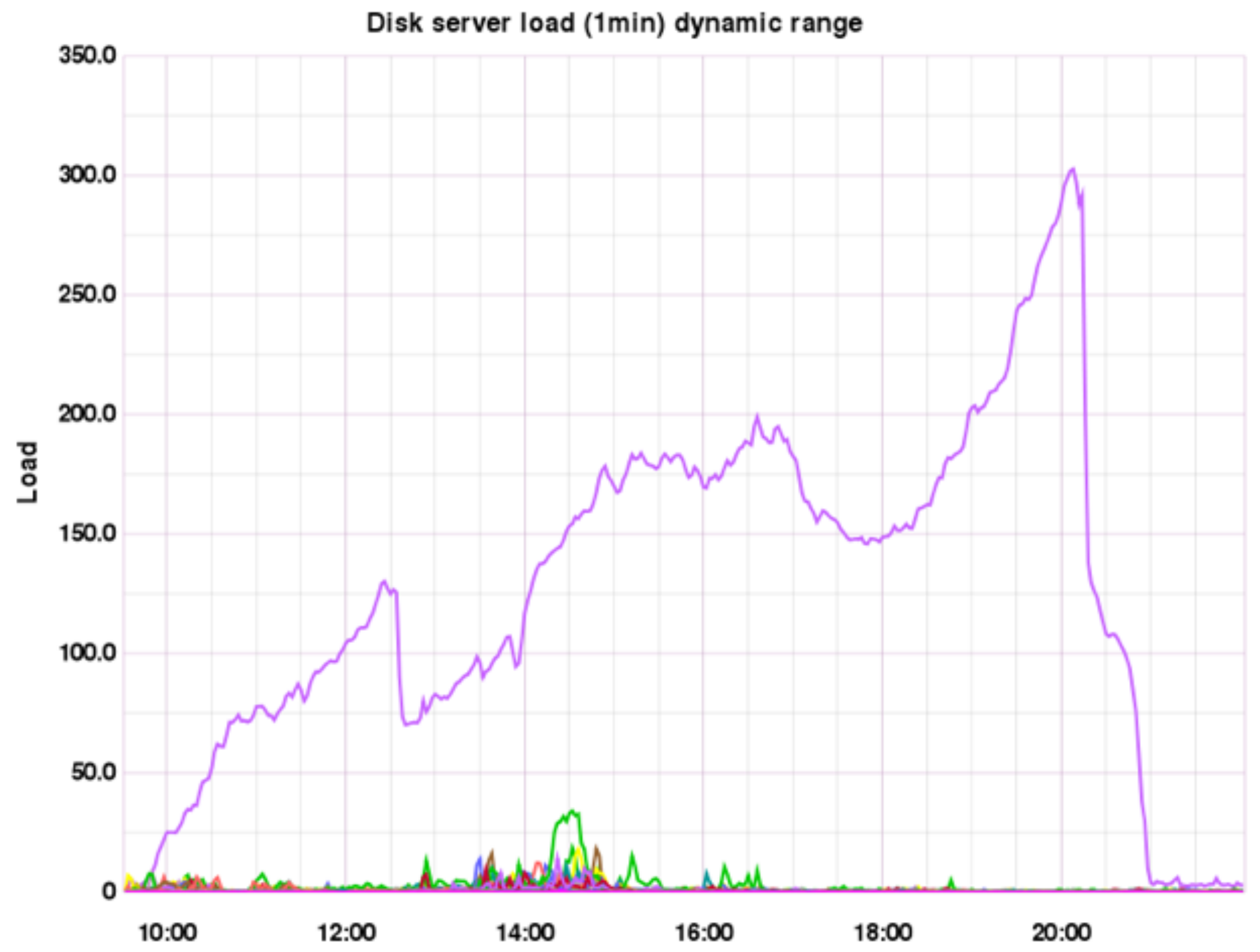
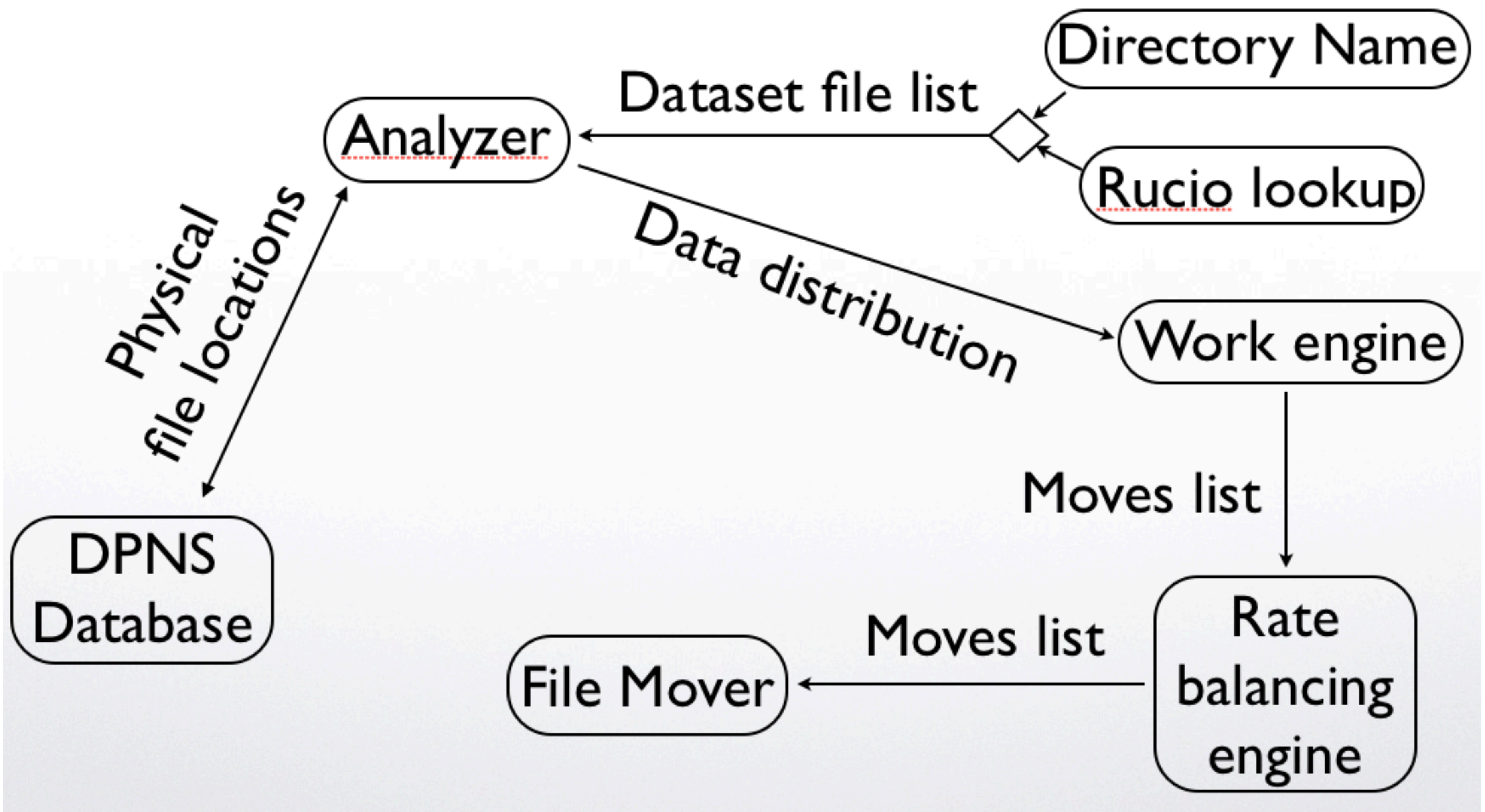
Analysis and improvement of data-set level file distribution in Disk Pool Manager

The widely used DPM storage system distributes files across component servers with a naive round-robin algorithm (modifiable by fixed weightins per filesystem). While this balances total file distribution well, it cannot take into account the distribution of files within a given dataset.
Poor distribution of files at this level is a contributory factor to anomalous load on DPM based sites, as large numbers of analysis jobs working on a single dataset overload disk servers with higher than average shares.



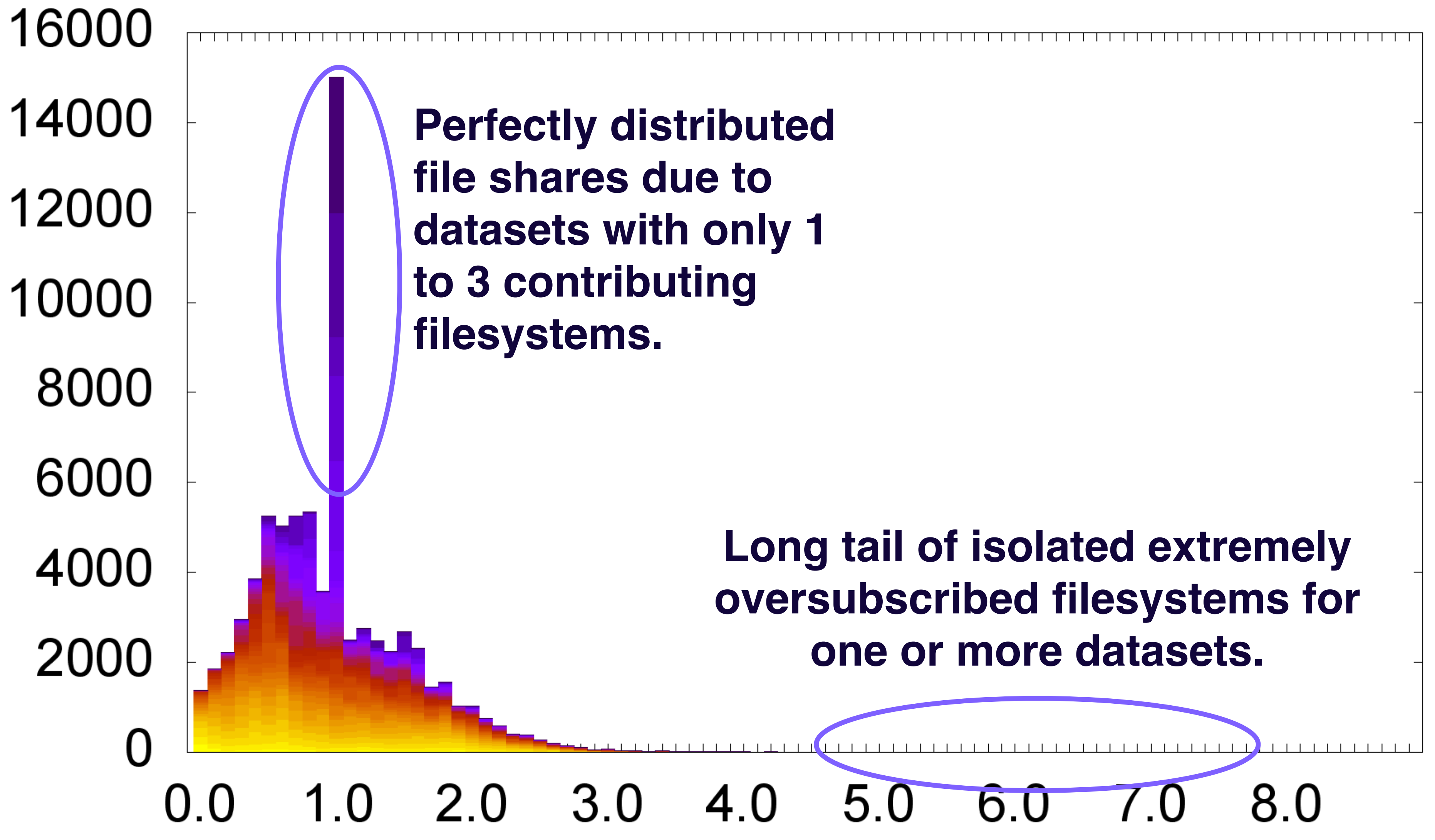
Example of anomalous load on isolated disk server, as a result of unbalanced dataset.



Schematic of algorithm for full tool workflow. Actual implementation in python, with Rucio lookup pending (and File mover pending JIRA LCGDM-1007)

We constructed a tool to analysis and fix file distribution at the dataset level within a DPM. The assumption can be made that a dataset is represented by a directory within the DPNS; this is broken by Rucio where we need to add a query to Rucio to get the paths for all the associated files.
After analysis of the file distribution, work is planned as a list of moves of files from more to less occupied filesystems, server-location aware. A rate balancing engine optimises the workflow to distribute network and disk load over the storage cluster.

Analysis phase shows that file distribution on (for example) UKI-SCOTGRID-GLASGOW ATLASDATADISK spacetoken is extremely poor. Application of the suggested move list to the site is still awaiting exposure of the ability to request given files be replicated to specific filesystems in the python API (conventional replication does not allow the destination to be selected).



Perfectly distributed file shares due to datasets with only 1 to 3 contributing filesystems.

Long tail of isolated extremely oversubscribed filesystems for one or more datasets.

Stacked histogram of normalised relative filesystem share of dataset distribution, for datasets striped over decreasing numbers of filesystems.

Samuel Cadellin Skipsey[1], Stuart Purdie[2]
David Britton[1], Mark Mitchell[1], Wahid Bhimji[3]

1 University of Glasgow 2 University of St Andrews
3 University of Edinburgh