



Globus Toolkit

## **Globus XIO Pipe Open Driver: Enabling GridFTP to Leverage Standard Unix Tools**

Raj Kettimuthu

Argonne National Laboratory and The University of Chicago



# Motivations

- **GridFTP uses GSI for security**
  - Provides strong security on both control and data channel
  - Delegation – allows 3<sup>rd</sup> party to authenticate on user's behalf
- **Ease of use**
  - Obtaining and installing a X509 certificate is a complex process
  - Managing trust roots, keeping CRLs up-to-date



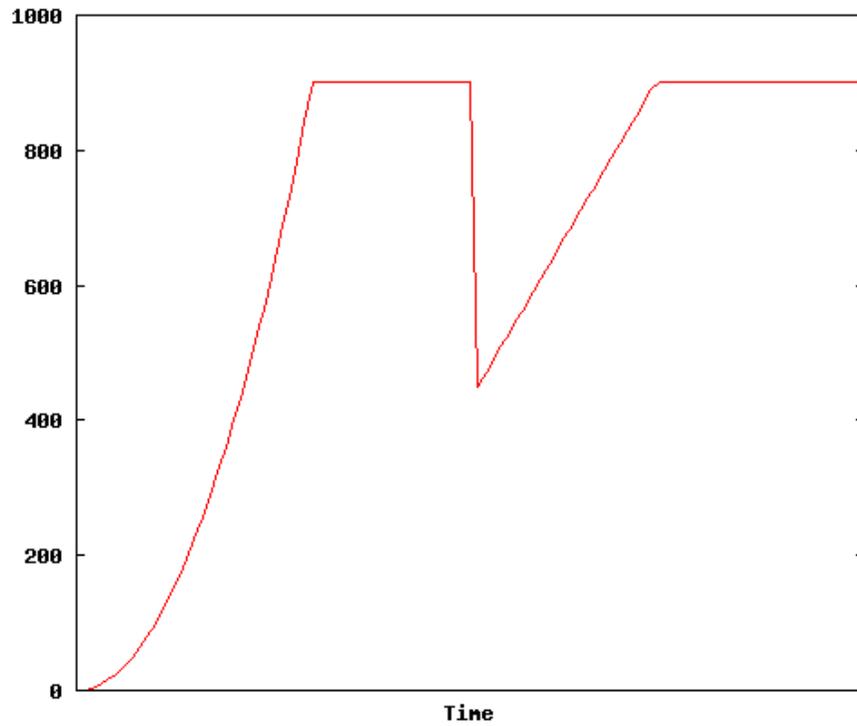
# Alternate security mechanism

- **How about SSH?**
  - Ubiquitously available
- **How to we leverage it?**
- **Can we use Unix pipes to leverage it?**
  - Construct powerful Unix command lines by combining Unix commands
  - Data is processed by each command and passed on to the next command
- **Globus XIO Popen driver allow clients to combine GridFTP with other Unix tools**

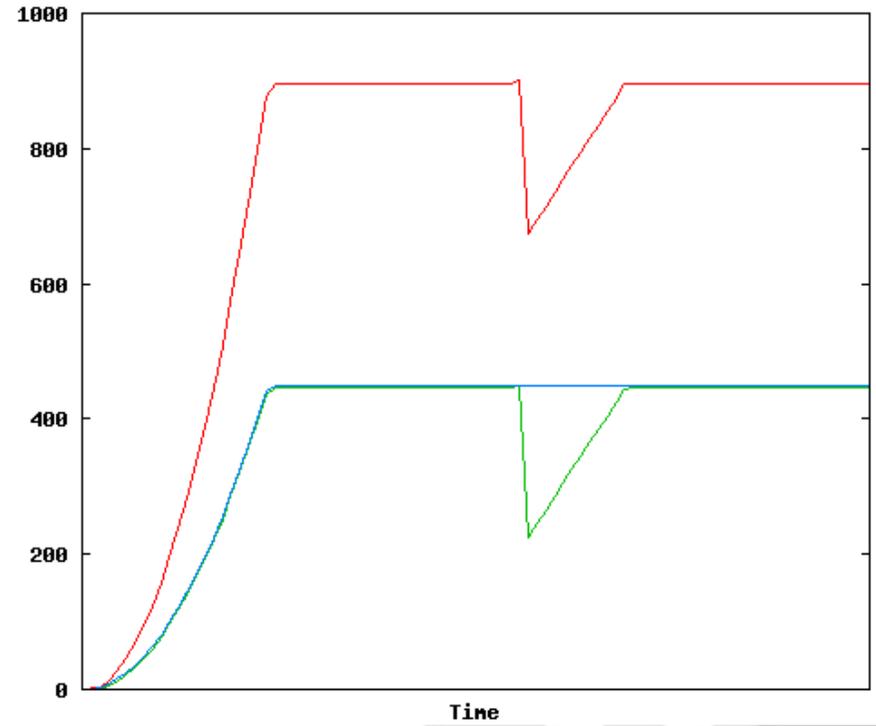


# Parallel Streams

One Stream



Two Streams





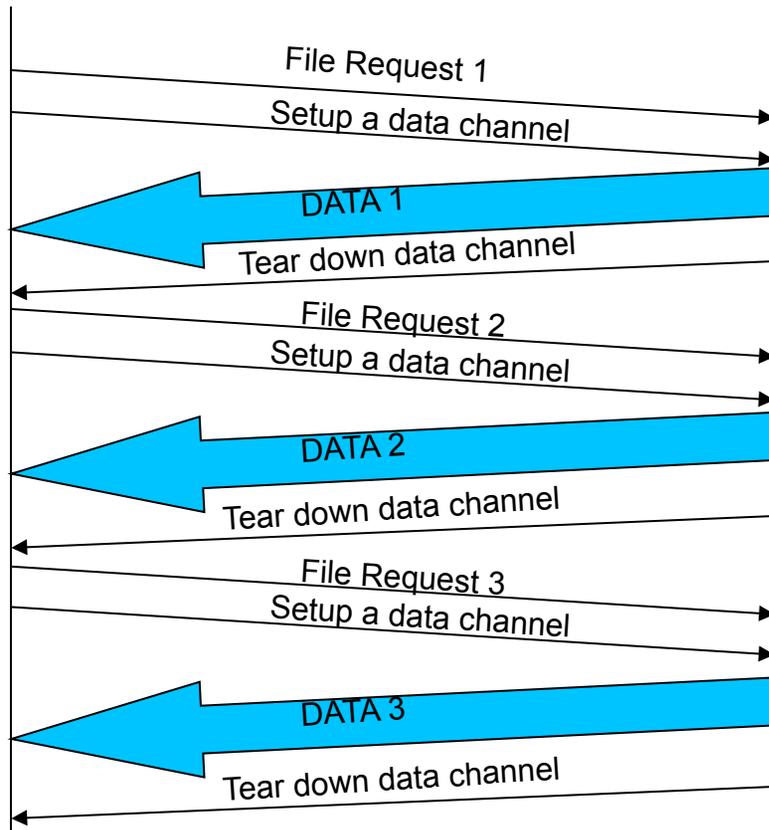
## Lots of small files

- **Parallel streams benefit large files**
- **Often times, the entire dataset is large (terabytes) but is partitioned into a large numbers (millions) of small files**
- **Parallel streams do not help improve the performance for these data sets**
  - Might even hurt the performance
- **We have implemented some solutions to address this in the past**

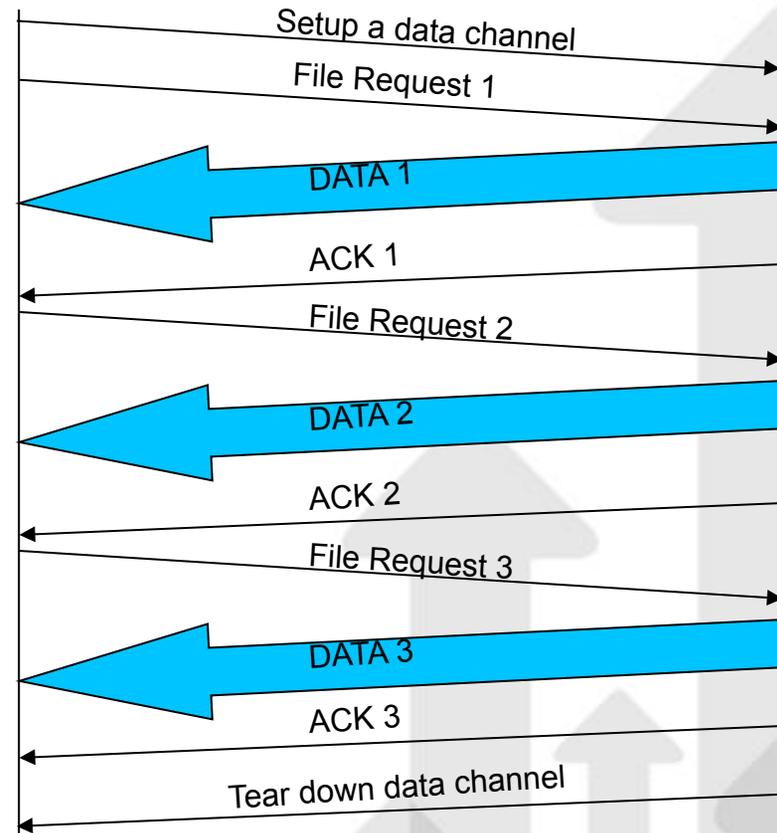


# Data Channel Caching

Standard FTP



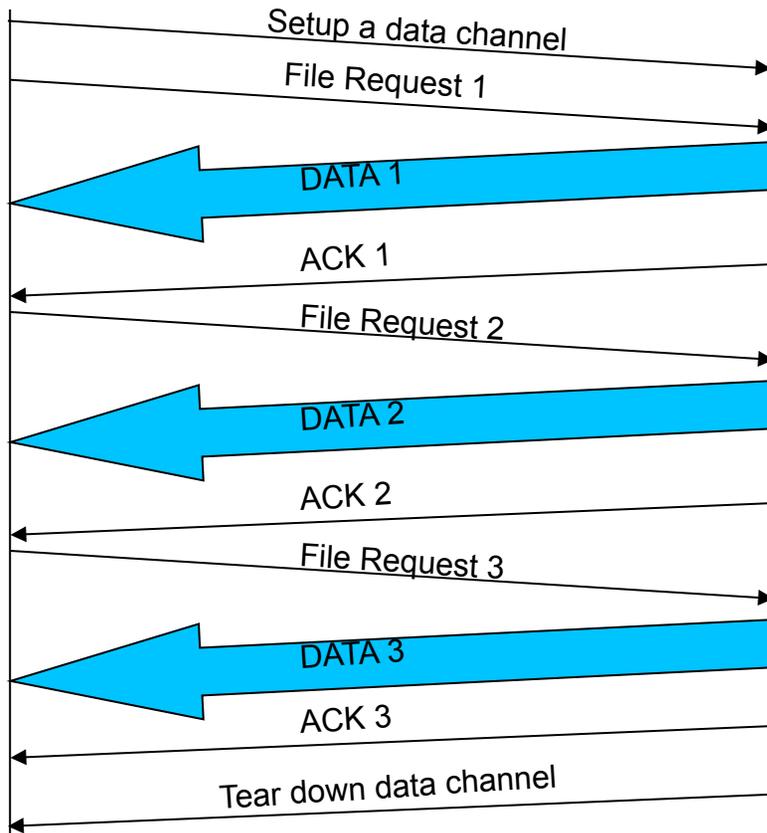
GridFTP



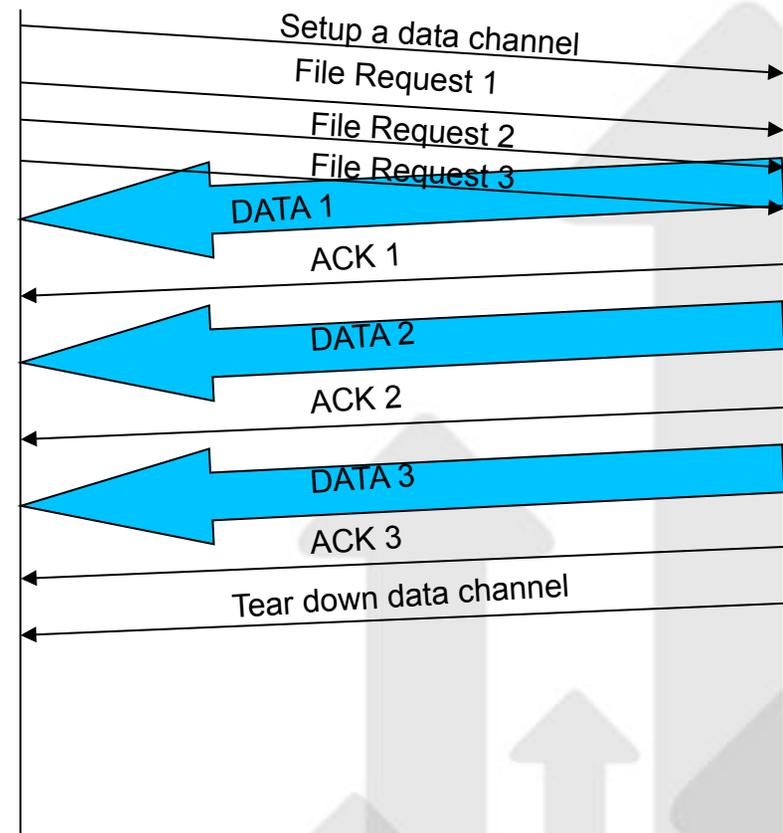


# Pipelining

GridFTP

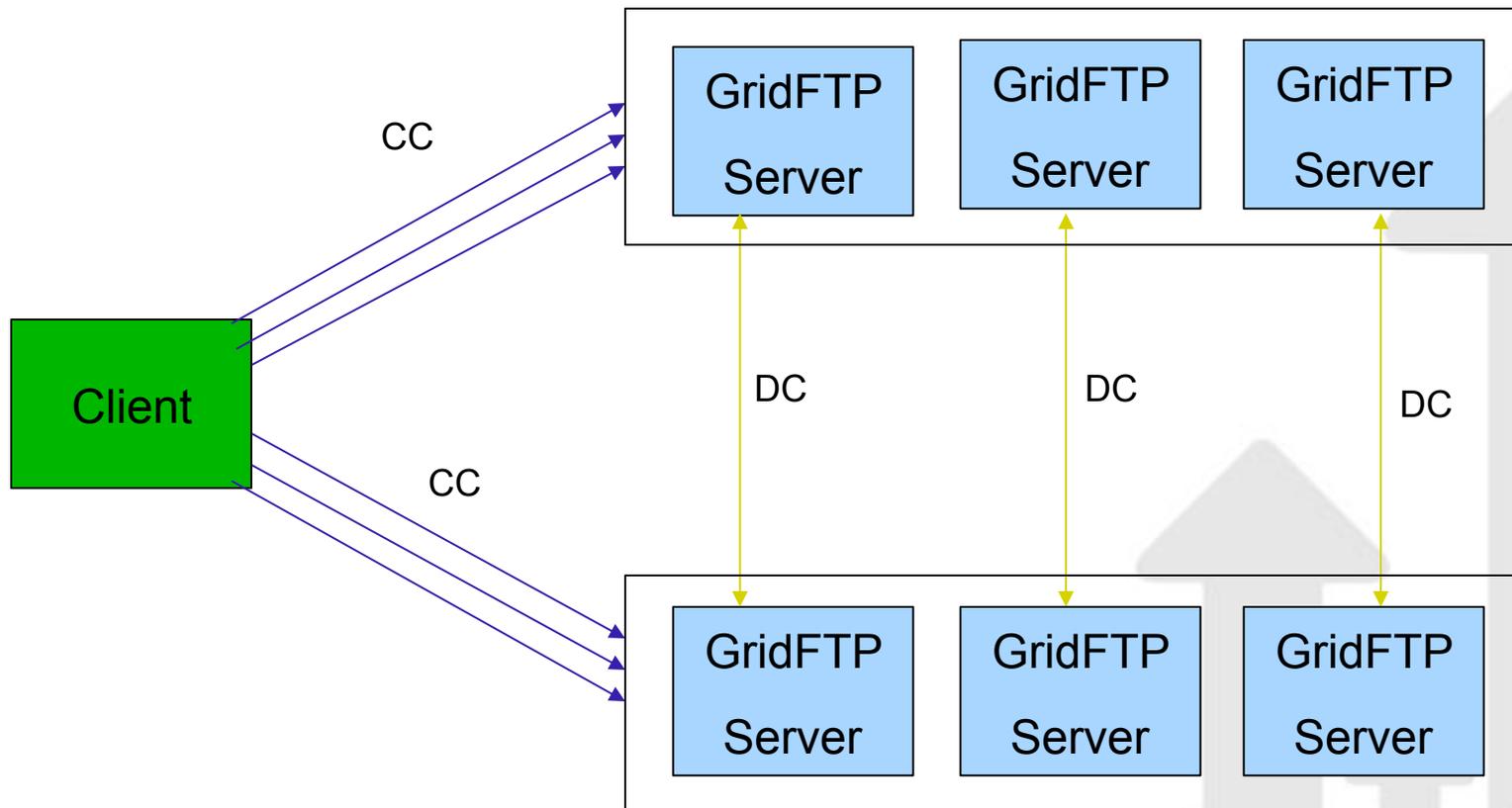


GridFTP with Pipelining





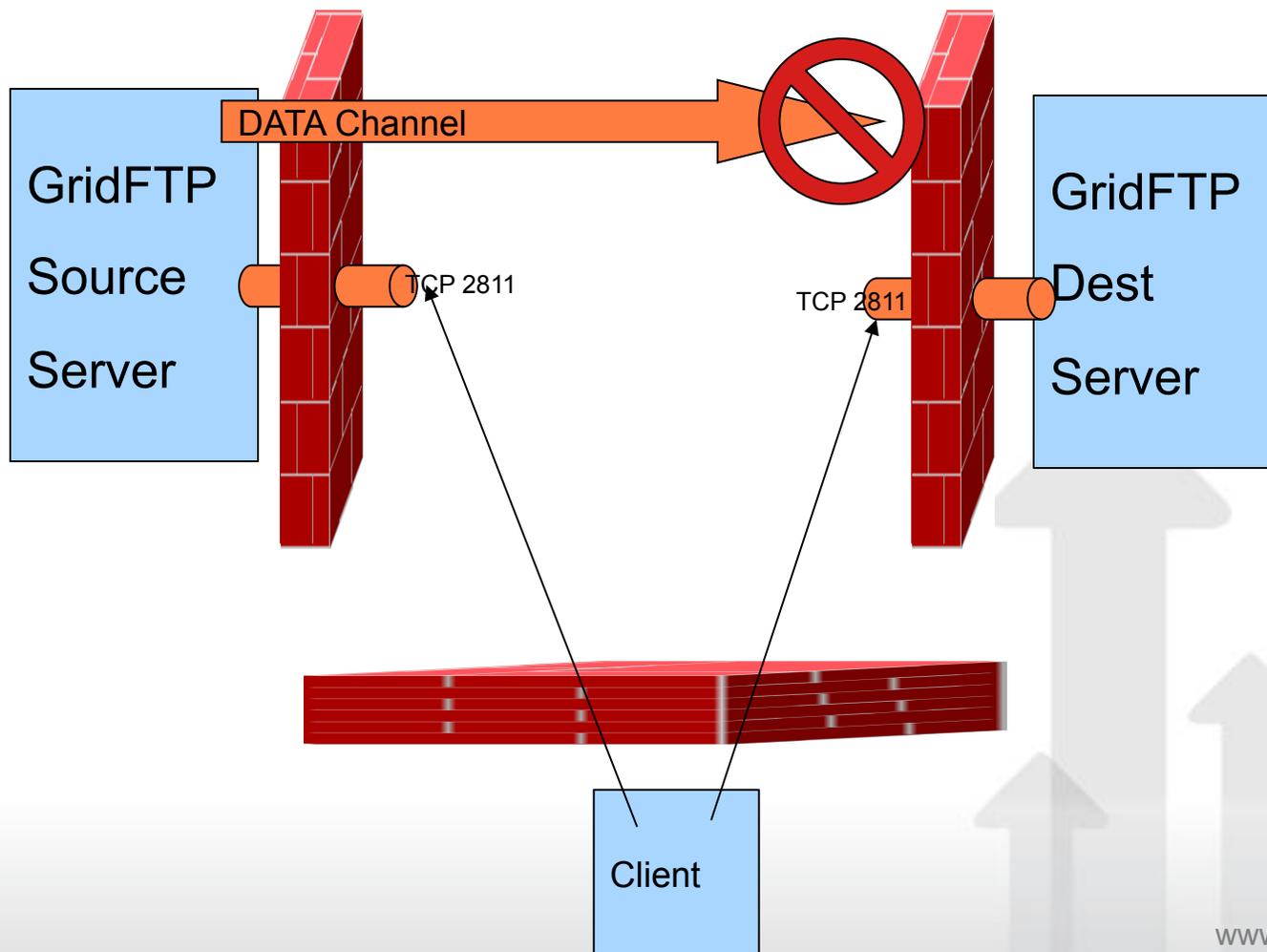
# Concurrency





# Firewall

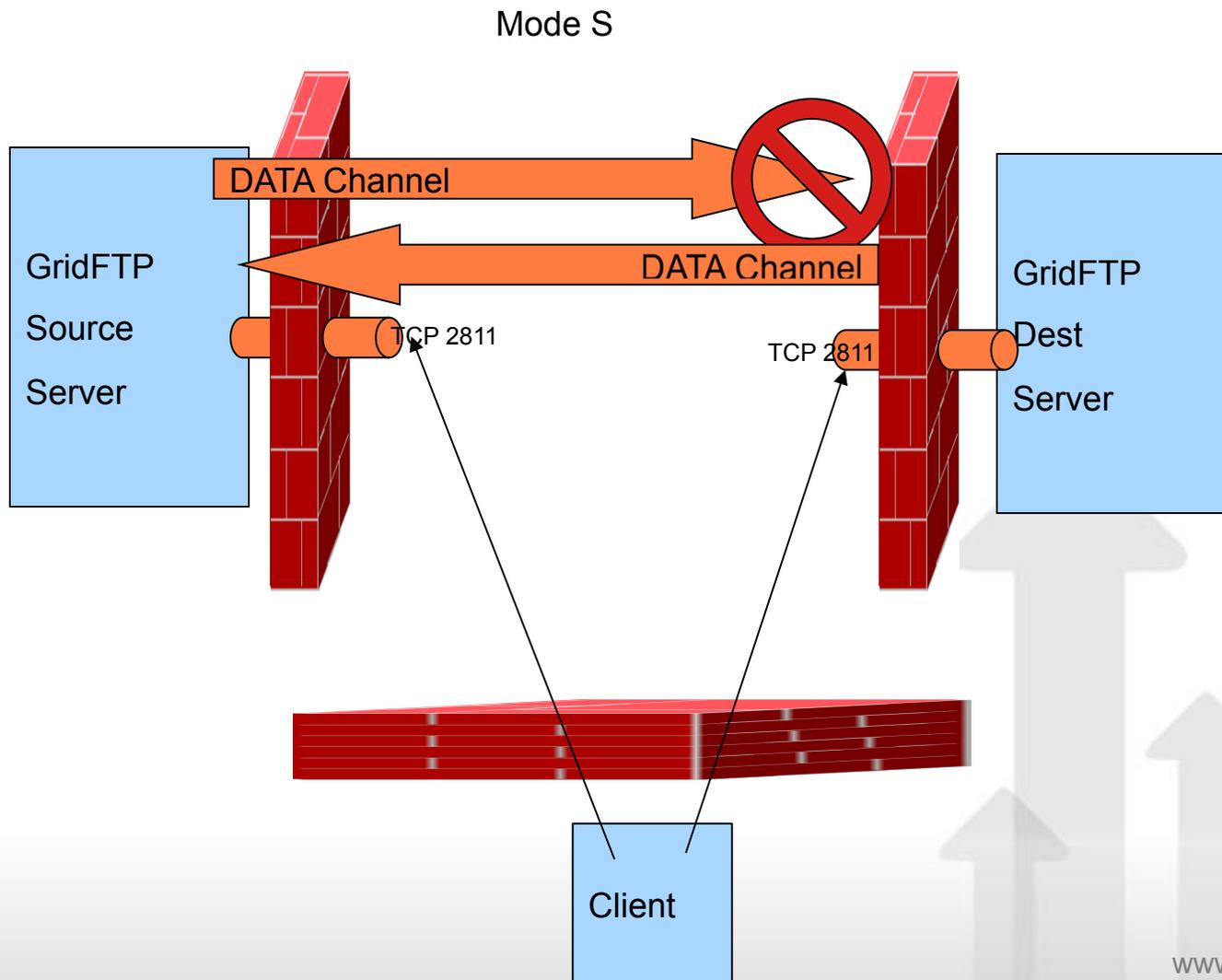
- Outgoing allowed at sender, incoming blocked at receiver





# Firewall

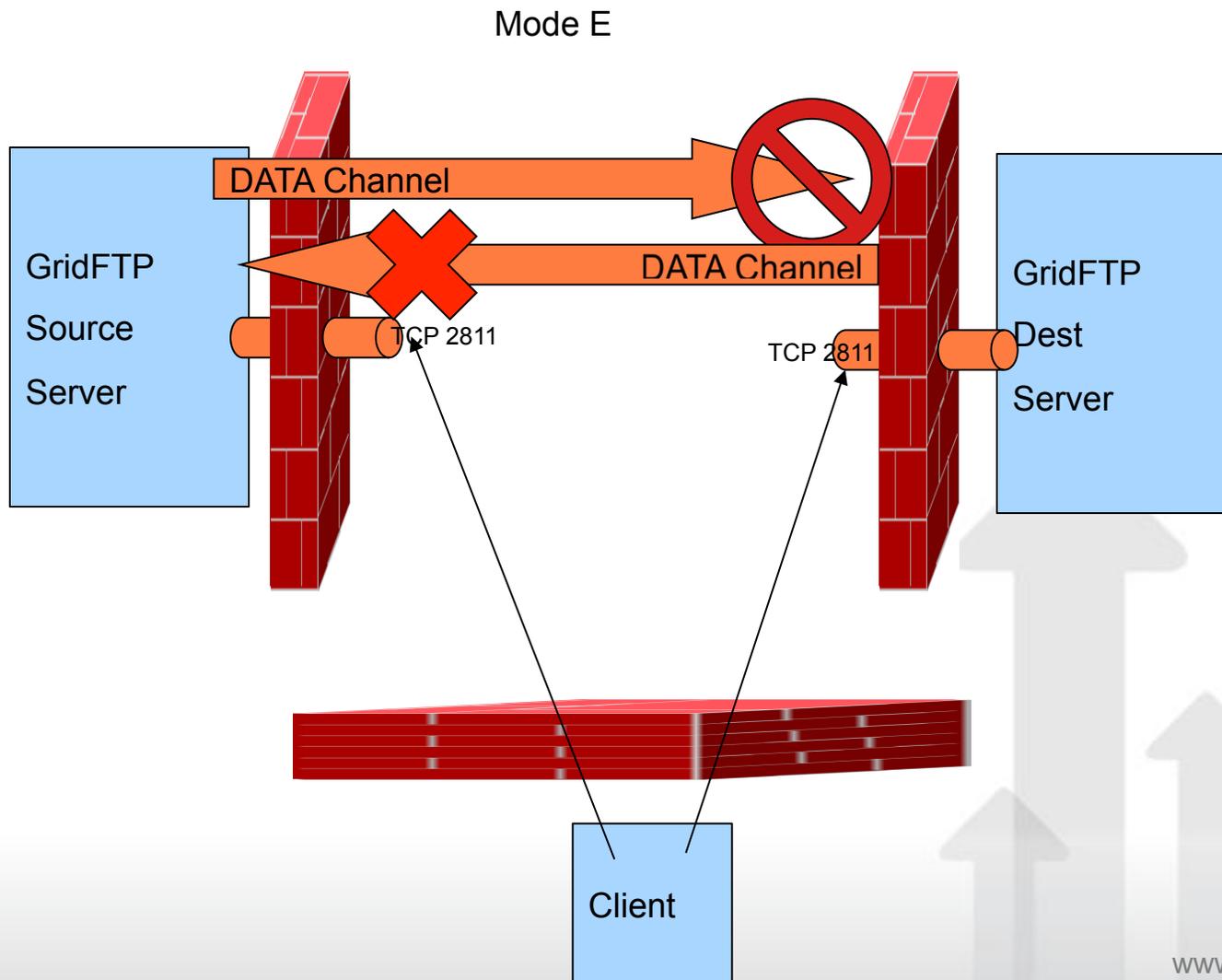
- Outgoing allowed at sender, incoming blocked at receiver





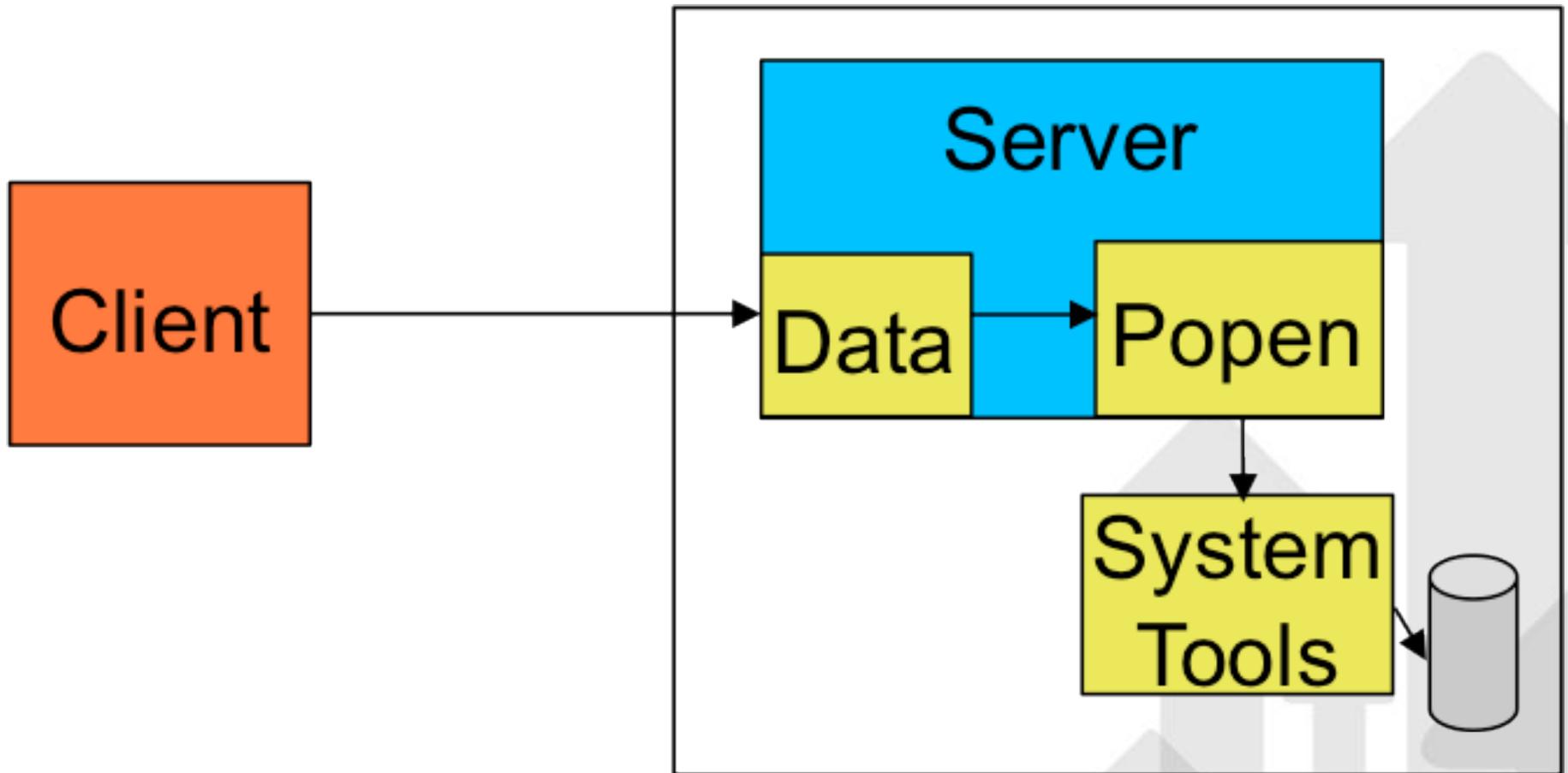
# Firewall

- Outgoing allowed at sender, incoming blocked at receiver





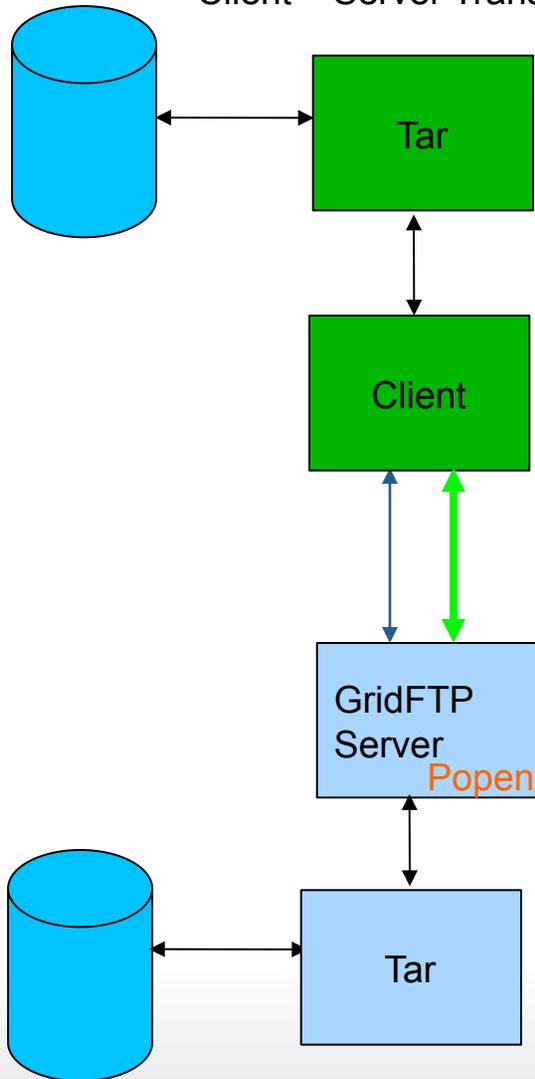
# Globus XIO Pipe Open



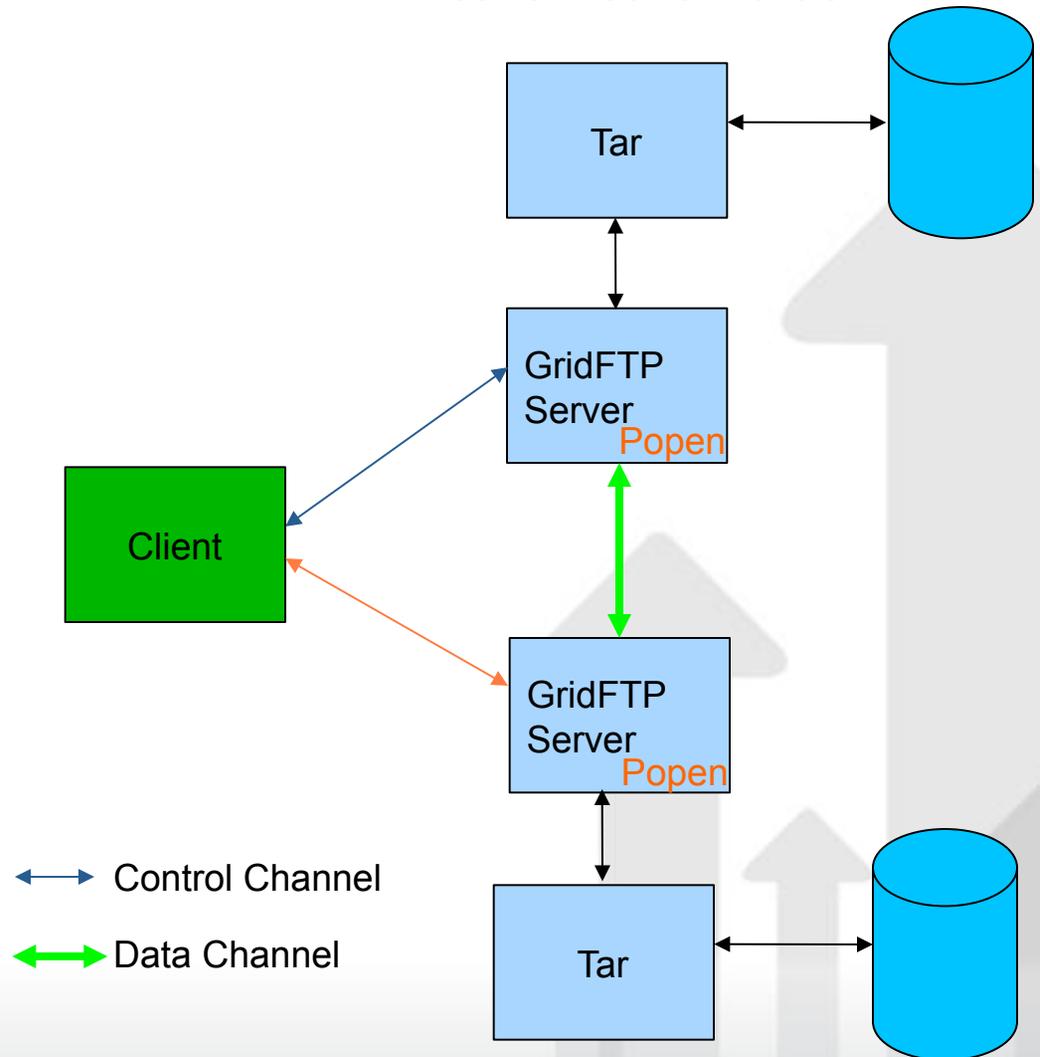


# On-the-fly tar

Client – Server Transfer



Server – Server Transfer

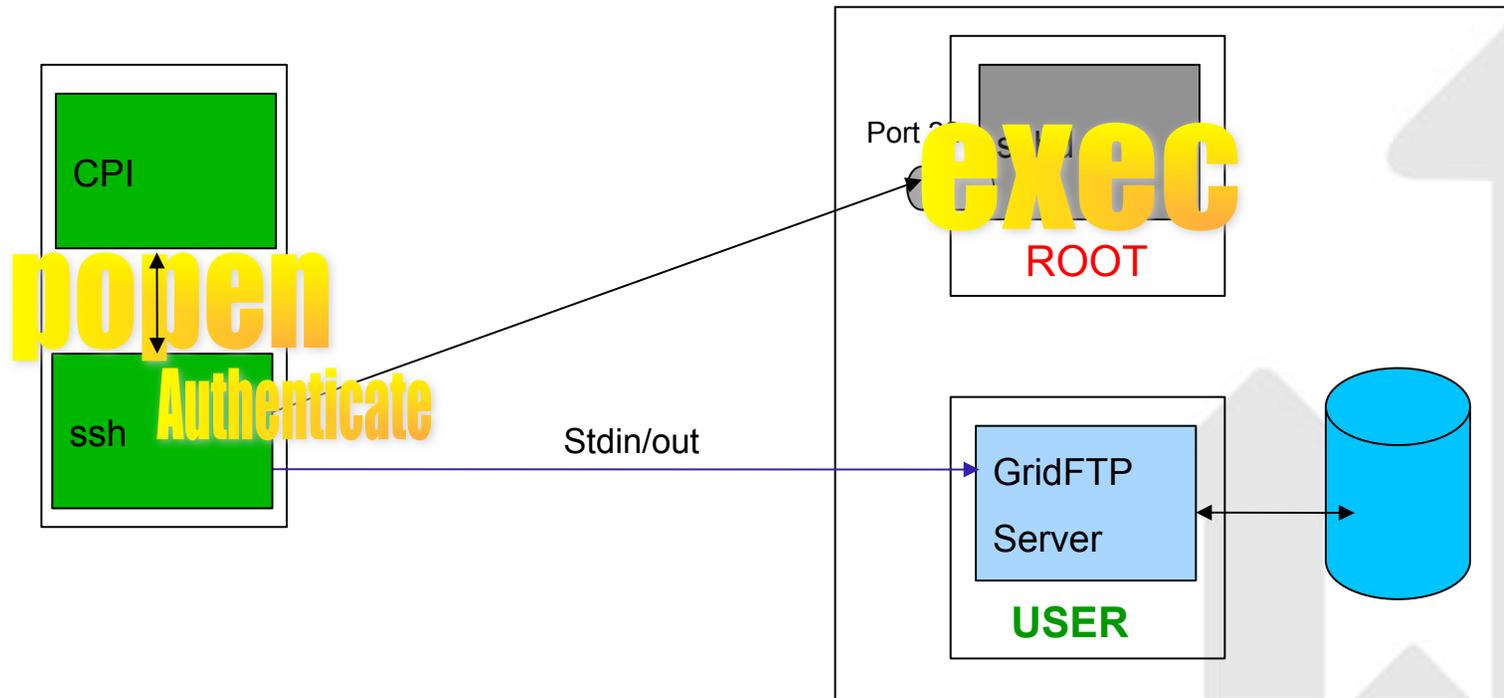


↔ Control Channel

↔ Data Channel



# sshftp:// Interactions





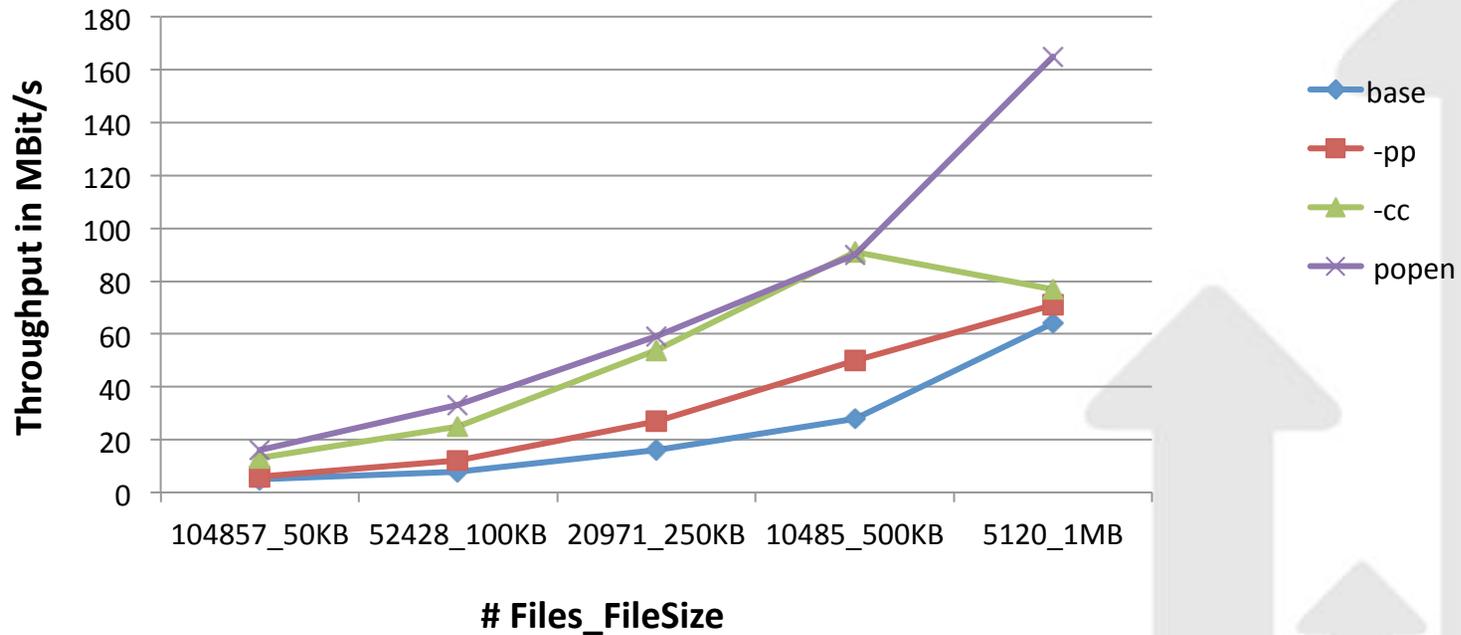
## Security risks with Popen

- **Execution of arbitrary programs as part of a data transfer opens a potential security risk**
- **For this behavior to be allowed on the server, admins have to explicitly enable it**
  - All programs that the user might execute using the Popen driver, as well as the Popen driver itself must be explicitly whitelist'd at the time the GridFTP server is run.
- **globus-gridftp-server**
- **-fs-whitelist popen,file,ordering**
- **-popen-whitelist tar:/bin/tar**



# On-the-fly tar performance

Comparison of Data Transfer between PSC Pople and Purdue Steele. RTT = 20ms





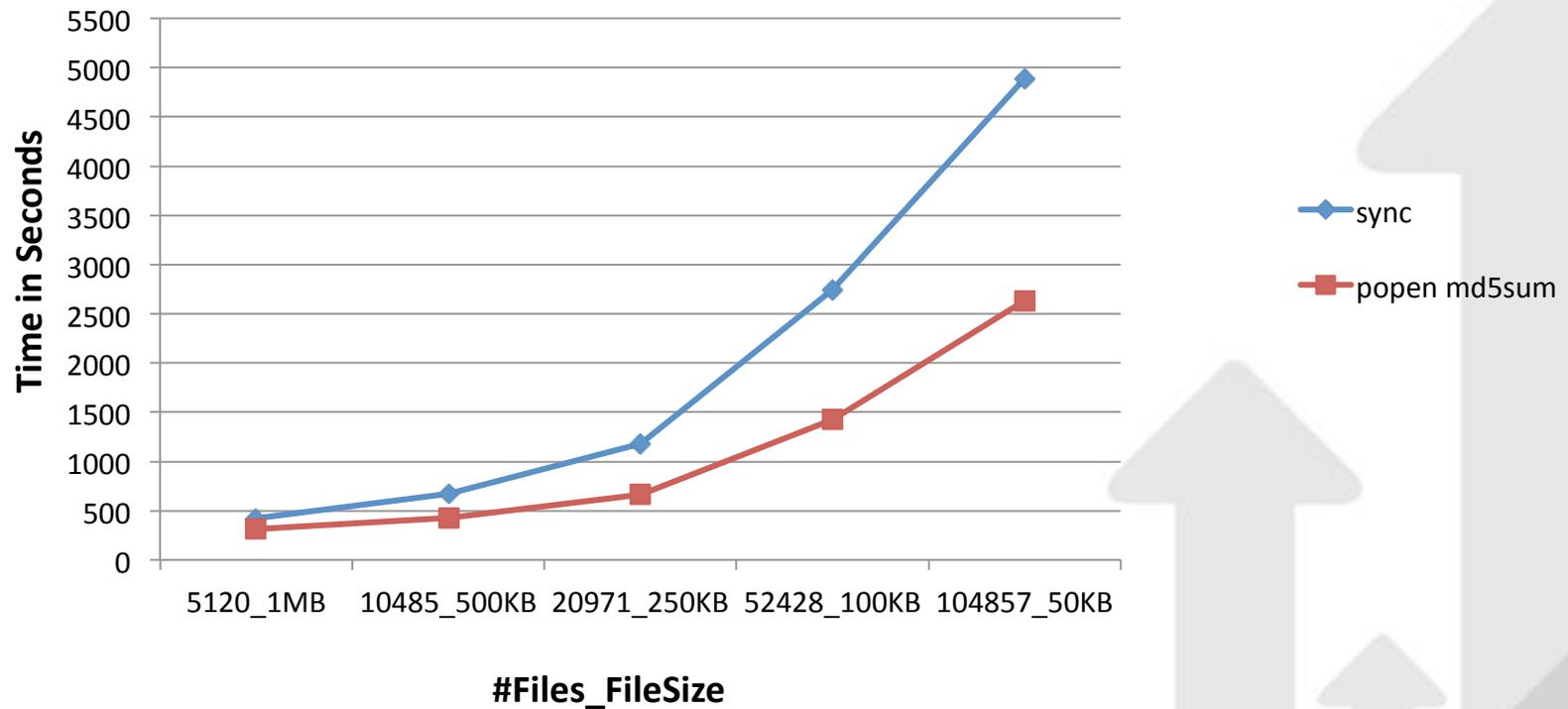
## Checksum via Popen

- **GridFTP protocol has a CKSM command to checksum a file**
- **To checksum a directory containing large number of files**
  - Traverse the remote directory
  - Perform checksum on files, one at a time with the CKSM command
  - Time-consuming.
- **Popen driver allows us to pipe commands through md5sum and complete a checksum entire directory**



# Checksum via Popen

Comparison Checksumming between PSC Pople and Purdue Steele. RTT = 20ms





**Questions?**

