

GT4 Usage Statistics

Open Discussion

Enabled by Default (NSF Requirement)

www-unix.globus.org/toolkit/docs/4.0/Usage_Stats.html



- Why are we doing this?
- The overview
- Opt-out
- · What is sent?
- How is the data sent?
- When is it sent?
- · What will the data be used for?
- Feedback

Why are we doing this?

The Globus Alliance receives support from government funding agencies. In a time of funding scarcity, these agencies must be able to demonstrate that the scientific community is benefiting from their investment. To this end, we want to provide generic usage data about such things as the following:

- how many people use GridFTP
- · how many jobs run using GRAM
- how many GT4 web services containers are running.

To this end, we have added support to the Globus Toolkit that will allow installations to send us generic usage statistics. By participating in this project, you help our funders to justify continuing their support for the software on which you rely.

How to Disable

- What is sent
 - Globus events (reasonably abstracted, not tied to user id or process names)
 - Sent to usage-stats.globus.org
- GT4 Docs describe another method
 - We found this confusing
 - Must disable for C and Java (separate steps)
 - A firewall is the simplest method (IPChains)

```
*filter
:INPUT ACCEPT [0:0]
:FORWARD DROP [0:0]
:OUTPUT ACCEPT [0:0]
# Preamble
-A FORWARD -i eth1 -o eth0 -m state --state NEW, RELATED, ESTABLISHED -j ACCEPT
-A FORWARD -i eth0 -j ACCEPT
-A INPUT -i eth0 -j ACCEPT
-A INPUT -i lo -j ACCEPT
# Disable Globus Usage Statistics Reporting from this host
-A OUTPUT -d usage-stats.globus.org -j DROP
```

Discussion

- ☐ What is the issue?
 - Reporting is on by default
 - Users must take action to turn off
- What should we do?
 - Leave on and contribute to Globus statistics, which help keep them funded!
 - Disable at all sites out of privacy concerns?
 - Allow sites to decide for themselves
- This is not an alarm, but everyone needs to know this is happening.