



NORDUGRID-TECH-7

11/12/2008

PROTOCOLS, UNIFORM RESOURCE LOCATORS (URL) AND EXTENSIONS SUPPORTED IN ARC

Description

A.Konstantinov*

Contents

1	Protocols	2
2	Extensions	2
2.1	Options	2
2.2	Quasi-URLs/Meta-URLs	3

1 Protocols

URLs

All components of NorduGrid's ARC [1] should support following data transfer protocols and corresponding URLs:

- *ftp* - ordinary FTP without user and password fields.
- *gsiftp* - GridFTP [2], enhanced FTP protocol with security, encryption, etc. developed by The Globus Alliance [3].
- *http* - ordinary HTTP with PUT and GET methods using multiple streams.
- *httpg* - HTTP wrapped with Globus' GSI.
- *https* - HTTP wrapped with SSL.
- *file* - local file.

2 Extensions

2.1 Options

In addition to standard URL fields GM supports *options*. Options are of kind *name* or *name=value* and are inserted into URL after host and port and are separated by ','

$$protocol://host[:port][;option[;option[...]]/path$$

Following options are supported:

threads=# specify number of parallel transfers to use (currently works for GridFTP, HTTP, HTTPS and HTTPG, default is one),

blocksize=# specify size of chunks/blocks/buffers used in transactions (currently works for GridFTP, HTTP, HTTPS and HTTPG, default is protocol dependent),

cache=yes|no|renew whether file should to be cached on local side. *renew* forces a download of the file, even if the cached copy is still valid. (for transfers to local destinations, default is *yes*),

secure=yes|no whether data should be transfered using encryption (currently works for GridFTP only, default is *no* unless specified differently in configuration),

readonly=yes|no if file after transfer to local location can be stored in a way which prevents user, job, etc. from modifying it (use *readonly=no* if You plan to modify file),

checksum=cksum|md5 specifies type of checksum to be computed and provided to the indexing server,

exec=yes|no means file should be treated as executable (currently only used internally).

Example: *gsiftp://grid.domain.org:2811;threads=10;secure=yes/dir/input_12378.dat*

preserve=yes|no specify if file must be uploaded to this destination even if job processing failed (default is *no*)

pattern=pattern defines file matching pattern - currently works for file listing requests sent to *se://* endpoint (see below)

guid=yes/no make software use GUIDs instead of LFNs while communicating to indexing services - meaningful for *rls://* only (see below)

overwrite=yes/no make software try to overwrite existing file(s), i.e. before writing to destination, tools will try to remove any information/content associated with specified URL

protocol=gsi/gssapi to distinguish between two kinds of *httpg*. *gssapi* stands for implementation using only GSSAPI functions to pack and unpack data packets and *gsi* uses additional wrapping of data packets as implemented in Globus IO module (default is *gsi*)

spac token=pattern specify the space token description to be used for uploads to SRM storage elements supporting SRM version 2.2 or higher

2.2 Quasi-URLs/Meta-URLs

There are also few quasi-URLs supported. Those are URLs which are resolved to one of ordinary protocols for actual transfer of data.

- *se* - to contact “Smart” Storage Element (SSE) being developed at NorduGrid project is also supported. It’s syntax is *se://host[:port][;options]/path[?file_id]*. For more information see document “The NorduGrid “Smart” Storage Element”.
- *srm* - means Storage Resource Management (SRM) service [4]. Methods defined in versions 1.1 and 2.2 of the protocol are used. Its syntax is *se://host[:port][;options]/[service_path?SFN=]file_id*. Default *service_path* is *srm/managerv2* when the server supports v2.2, *srm/managerv1* otherwise.

Also following quasi-URLs of meta-data (indexing) servers are supported:

- *rc* - Globus Replica Catalog: *rc://[location[/location[...]]][@host[:port]/distinguished_name]/lfn*
- *rls* - Globus Replica Location Service: *rls://[url[/url[...]]@]host[:port]/lfn*
- *lfc* - LCG File Catalog: *lfc://[url[/url[...]]@]host/lfn*. Note that *lfn* must start with */*.
- *fireman* - Fireman indexing service which is part of gLite initiative [5]: *fireman://[url[/url[...]]@]host[:port]/service_path?lfn*

Here

lfn Logical File Name (LFN), that should be used for registering/querying in the Indexing Service,

distinguished_name DN of collection in LDAP server used for registering/querying,

location name of location (usually this is host[:port] of SE), each *location* can have *options* appended to it in a way *location;options*. If only *;options* part is preset, those options are treated as specified for every location.

url full or partial URL corresponding to physical location of file.

service_path endpoint path of Web service.

References

- [1] “The NorduGrid Collaboration,” Web site. [Online]. Available: <http://www.nordugrid.org>
- [2] W. Allcock *et al.*, “Data management and transfer in high-performance computational grid environments,” *Parallel Comput.*, vol. 28, no. 5, pp. 749–771, 2002.
- [3] I. Foster and C. Kesselman, “Globus: A Metacomputing Infrastructure Toolkit,” *International Journal of Super-computer Applications*, vol. 11, no. 2, pp. 115–128, 1997, available at: <http://www.globus.org>.
- [4] A. Sim, A. Shoshani and others, “The Storage Resource Manager Interface (SRM) Specification v2.2,” May 2008, GFD-R-P.129. [Online]. Available: <http://www.ggf.org/documents/GFD.129.pdf>
- [5] “gLite, Lightweight Middleware for Grid Computing,” Web site. [Online]. Available: <http://glite.web.cern.ch/glite/>