Gridftp Data Integrity Verification

Draft proposal, v1.1

Timur Perelmutov, FNAL

This document addresses the issues of the Data Integrity Verification as it was proposed in the GridFTP Protocol Improvements document

(https://forge.gridforum.org/projects/gridftp-wg/document/GridFTP_v1.0_Improvements).

I propose to add 2 new commands "SETCKSM" and "CKSM" which would allow a checksum to be sent/received for STO/ESTO/RETR/ERET transfers.

The proposal to extend the Extended Block Header format and add a recovery from corruption mechanism was included into the "Modification of Extended Block Mode Protocol" proposal by Igor Mandrichenko, so it is removed from this document.

SETCKSM command

Command syntax is
SETCKSM <SP> <algorithm> <SP> <checksum value as a hexadecimal number> <CRLF>

The gridftp specification will define a list of string constants corresponding to possible supported checksum algorithms. This list should include, but should not be limited to "CRC", "ADLER32", "MD5".

The purpose of the SETCKSM command is to communicate the checksum of the data transferred by the first STOR / ESTO commands following issue of SETCKSM command. The transfer to the gridftp server can be in either stream or extended block mode and can be either full or partial. The server can use the information transmitted by the SETCKSM command for: immediate data integrity verification, storage in transferred file metadata, etc. In case of success the server responses with the return code 200.

CKSM command

Command syntax is
CKSM <SP> <algorithm> [<SP> <offset> [<SP> <length>]] <SP> <pathname> <CRLF>

Note: offset and length fields are optional. If offset is omitted the checksum for the entire file is returned, and if the length is either -1 or omitted, the length is assumed to be the length of the entire file.

The purpose of the CKSM command is to allow the client to discover the checksum of the full / partial file data prior to receiving the this data via RETR / ERET commands and then use this information for performing the verification of the received data integrity. The successful response to the command is the return code 200 followed by checksum as a hexadecimal number.

Feature Set Negotiation

If server implements the CKSM and SETCKSM commands then the following feature names will be included in the response of the FEAT command:

CKSM <algorithm1> <algorithm2> ... <algorithmN>: server supports CKSM command with the algorithms algorithm1, algorithm2 , ... , algorithmN.

SETCKSM <algorithm1> <algorithm2> ... <algorithmN>: server supports SETCKSM command with the algorithms algorithm1, algorithm2 , ... , algorithmN.