Xrootd Thirty Party Copy for WLCG and HL-LHC

Wei Yang
on behave of the WLCG TPC Working Group
CHEP 2019
Goals:

Being able to replace GridFTP by using Xrootd, in specific:

- Enhance X509 security
  - The main task is to allow X509 proxy delegation
  - Token based authentication will wait until TLS is available in Xrootd release 5
- Support checksum verification
- Covers a broad set of the WLCG storage systems
  - And network/firewall topologies
- High performance
- Optional: containerization

Work involves both server side and client side of the SW

- C++ and Java implementations
- And coordination with gfal2/FTS
Xrootd TPC implementations

- Xrootd TPC and WLCG storages
  - C++ Xrootd and TPC implementation covers
    - SLAC Xrootd, DPM, EOS, CEPH, and posix file systems (including Storm?)
  - Java Xrootd and TPC implementation covers
    - dCache
  - Documents on implementation detail is important

- TPC and TPC Lite
  - By default TPC uses rendezvous token to for authentication
    - The token is plain text. Low overhead. Independent of security modes
    - With small security risk.
    - Will be useful along with TLS
  - With X509 proxy delegation, rendezvous token is not needed
    - TPC Lite: no such token, destination used delegated X509 proxy to fetch data from source
X509 security

● Prioritize the support X509 security
  ○ It is the current Grid computer security model.
  ○ Support of TLS require time and carefully planned/tested implementation

● Initial ideal of using robot certificates met resistance from sites
  ○ many of them support multiple experiments and VOs
  ○ So in TPC R&D phase 1, a quick switch to:

● Implement X509 proxy delegation
  ○ Delegate client’s X509 proxy to the destination server
  ○ Destination uses a X509 credential (delegated from client) to “pull” data from source
  ○ No need to maintain a robot certificate
    ■ Simplify deployment task
    ■ Support multiple VOs in one instance
Enhance X509 implementation for proxy delegation

**Enhancement:**
- Sign Diffie-Hellman parameters
  - DH key exchange establishes a (much faster) symmetric encryption key
  - Server signs the DH parameters using its host certificate (private key)
  - Client verifies the signature (using server’s public key)
    - Prevent Man-in-the-Middle attack
- Support RFC 2818 - Subject Alternative Name in host certificates
  - To prevent DNS spoofing
  - Xrootd supports RFC2818
    - Enforce RFC2818 when delegation is required

**Document the C++ implementation**
- Effort starts in C++ implementation, we want to have the same in Java
- Thanks for the dCache team for the implementation.
Stability, scalability and performance

- In C++, Xrootd X509 and VOMSxrd used OpenSSL
  - VOMSxrd plugin extracts client’s VOMS info for authorization
  - Worked fine under RHEL5 but showed memory leaks in RHEL6/7
    - Switch to used new OpenSSL API in Xrootd X509 and VOMSxrd
- Xrootd clustering mechanism works with TPC
  - To scale up
  - Support checksumming by individual server
- Improve multiple TCP streams performance in Xrootd client
  - Async IO is well supported at server side in C++ implementation
  - Measure and tune XrdCl internal parameters for optimal performance
  - XrdCl with multiple TCP streams can now match the performance of bbcp.
Other things

- **TPC in Xrootd proxy mode**
  - Xrootd proxy mode provides a gateway when the storage is inaccessible from outside.
  - TPC works in Xrootd proxy mode, as a DTN
    - Require a simple shell script

- **Checksum**
  - All WLCG storages, including Xrootd proxy mode support checksuming in Xroot protocol
  - Some may require a recent version of the storage software

- **Performance mark**
  - Xrdcp and its java counterpart support progress mark
  - FTS uses performance mark to check the health of the transfer

- **Object path in CEPH**
  - CEPH Xrootd plugin now support double slashes in object path.
  - Making it identical to other storage systems

- **Support of ALICE token forwarding**
Other things: cont’d

● **Containerization**
  ○ Running DTN in Singularity container at SLAC.
  ○ Should work with Docker as well

● **Limitation**
  ○ No suid, use ACL in Xrootd authorization DB file to control access
  ○ Work with the WLCG distribution data model - VO owns the data
  ○ Not working well when requiring individual users to “own” data

● **Tests organized by WLCG TPC WG**
  ○ Both smoke test (functional test) and stress test
<table>
<thead>
<tr>
<th>SCORE</th>
<th>ENDPT</th>
<th>TYPE</th>
<th>SRC</th>
<th>DST</th>
<th>DN</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>BRUSSELS</td>
<td>dCache</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>20</td>
<td>CERN-EOS</td>
<td>EOS</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>20</td>
<td>CERN-TRUNK</td>
<td>DPM</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>20</td>
<td>FNAL</td>
<td>dCache</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>20</td>
<td>OU</td>
<td>XrootD</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>20</td>
<td>PRAGUE</td>
<td>DPM</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>20</td>
<td>UKI-LANC</td>
<td>DPM</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>20</td>
<td>UKI-MAN1</td>
<td>DPM</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>20</td>
<td>UKI-MAN2</td>
<td>DPM</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>20</td>
<td>UNI-BONN</td>
<td>CephFS</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>19</td>
<td>DESY-PROM</td>
<td>dCache</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>19</td>
<td>UKI-BRUNEL</td>
<td>DPM</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>18</td>
<td>SLAC</td>
<td>XrootD</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
</tbody>
</table>

**PROBLEMATIC ENDPOINTS**

<table>
<thead>
<tr>
<th>SCORE</th>
<th>ENDPT</th>
<th>TYPE</th>
<th>SRC</th>
<th>DST</th>
<th>DN</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>CERN-RC</td>
<td>DPM</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>9</td>
<td>CALTECH</td>
<td>HDF5</td>
<td>P</td>
<td>F</td>
<td>P</td>
</tr>
<tr>
<td>0</td>
<td>BNL</td>
<td>dCache</td>
<td>F</td>
<td>-</td>
<td>F</td>
</tr>
</tbody>
</table>

**ERROR DETAILS**

1. **CERN-RC**
   - TPC_DST_D (round-trip-8-tpc-dst-d): 000000 : [ERROR] Server responded with an error: [3011] No such file or directory

2. **CALTECH**

3. **BNL**
   - UPLOAD (round-trip-17-upload): 000438 : Run: [ERROR] CheckSum error
   - TPC_DST_D (round-trip-17-tpc-dst-d): 000519 : Run: [ERROR] CheckSum error
   - DOWNLOAD (round-trip-17-download): 000383 : Run: [ERROR] CheckSum error
Moving forward

- Sites deployment, reliability, scalability
  - Now that technical issues with EOS, dCache, ECHO are mostly addressed.
  - Pay a bit more attention to sites and help them

- TLS support in Xrootd 5
  - Required to support WLCG token based AAI
  - Will also evaluate how this will change/optimize the way Xrootd TPC work

- Xroot and HTTP(s) sharing one instance
  - The immediate thing to look for VOMSxrd and XrdHTTPVOMS
    - For VOMS info extraction
  - Can one of them work with both Xroot and HTTP(s) protocol?

- New TPC mechanism
  - “Pull” and “Push”. “Push” mode is desired in some cases to workaround constraints.
Summary

● Has been a long way and we now have
  ○ All the pieces we need with major bug fixed
  ○ Have workable setups for all major WLCG storage systems
    ■ including dCache, DPM, EOS, Xrootd, ECHO, Posix
  ○ Various checks and tests to help us diagnosis problem

● Next focus
  ○ TLS/token bases AAI
  ○ Sites deployment