BDT Year-In-Review
Brian Bockelman, Petr Vokac, Maria Arsuaga Rios
Bulk Data Transfer (BDT) - Introduction

The BDT group was started in 2022 based on the foundation laid by the Third Party Copy (TPC) working group.

- ‘Finishing’ the transition to HTTP-TPC was a nice pivot point to re-examine the group and tweak the scope.
- Covers not only TPC aspects (there is remaining work for HTTP-TPC!) but also:
  - Migrating to new archive management protocols.
  - Transition to new authorization architectures.
  - Coordinating the wide use of network technologies such as packet marking.
- Aim for the year was to mature a slate of technologies as part of the preparation of DC24.
A critical project is to have nightly tests of token implementations amongst friendly sites. These tests:
- Cover both capability- and group-based authorization,
- Use tokens from WLCG IAM, and
- Hit a broad range of storage configurations and implementations (dCache, EOS, Echo, StoRM, XRootD).

Right now, the regular tests cover the HTTP protocol only.
- Xrootd protocol coverage just starting with manual tests.

The critical tests are OK for all supported SEs (dCache, EOS, Echo, StoRM, XRootD).
- Notably absent is DPM - see EGI GGUS migration campaign

Some VOs - notably CMS - are also working to have similar tests in their SAM framework. This will immensely help with the next tranche of sites.

All implementations pass critical tests
Year-in-Review: TPC and Tokens

It’s hard to believe the switch to HTTP-TPC was only “finalized” a year ago!

- In reality, there’s been many small details to follow-up on.
  - **Example:** After decades of experience with GridFTP, we had fine-tuned error messages to be quite useful to operators.
  - The BDT group is coordinating an [error message improvement](#) project, focusing on small quality-of-life improvements.
  - [HTTP-TPC protocol updates](#) (compatible) useful for monitoring, faster and more precise transfer failures diagnosis

We are working with experiments to prepare the documentation and guidance for sites to reconfigure their storage to support both X.509 and tokens.

- Will be in this “dual mode” for some time.
A complex topic remaining is “how should FTS handle tokens?”

- The X.509 model is “a single credential can do any read / write / delete operation anywhere”.
- The most fine-grained token approach would be a single token per transfer.
  - But this would not be achievable - token issuer would need high availability and high performance.
- A reasonable compromise would be a single VO token per-site. Limits impact of stolen credentials.

Resolving FTS issues are essential as we must work “up the stack” from storage endpoints to FTS to Rucio!

Recent scale test shows the current upper bound is ~100Hz token issuance rate for IAM. We will need to stay several integer factors under this!
Year-in-Review: Archive (tape) management

- Unify tape and transfer management under the same protocol
  - Replace complex SRM protocol (implementations with GCT dependencies) and proprietary xroot “bringonline” with simple Tape REST API
- WLCG Tape REST API task force formed at the end of last year
  - All tape storage (CTA, dCache, StoRM) and transfer service (FTS/gfal2) developers
  - May 2022 – [Tape REST API Reference document v1](#)
- HTTP Tape REST API
  - Uniform interaction despite storage target is batch or file oriented
  - Service discovery endpoint `.well-known/wlcg-tape-rest-api`
  - Per-file staging metadata can be passed to the tape system
  - Easy to introduce token support for HTTP protocol
Year-in-Review: Archive (tape) management

- **Status and tests tracked in BDT** TapeRestAPI Twiki
  - API stress tests with 1M queued files (staging, archiving, release)
    - Improvements in handling errors
  - Compliance-like tests for Tape REST functionalities used by FTS

- **Storage providers**
  - EOSCTA – current version considered production ready
  - dCache – included in 8.2 releases, finished API stress tests
  - StoRM – under development
    - part of larger refactoring (removing GCT dependencies)
    - ready during 2023

- **Client implementation status**
  - Gfal2 v2.21.0, released, deployed
  - FTS v3.12.2, released, deployed
Year-in-Review: Archive (tape) management

- Focused on deployment with X.509
  - Optimize RAL Tape access for LHCb
    - Deploy as soon as possible before the end of winter LHC stop
  - Site-by-site migration to the Tape REST API
    - Right mechanism automatically selected by FTS based on transfer protocol

- No tests with tokens yet
  - No support for `storage.stage` scope
  - More flexible timeline (no tape tests during DC24)
Year-in-Review: Network technologies

This has been a busy year for networking technologies, with ongoing work in site network monitoring, perfSONAR analytics and especially packet and flow marking.

Site Network Monitoring GOAL: Provide human readable information about a sites network (description, link information, peering information, equipment, diagrams, etc.)

- We have created a template markdown file which sites can clone and fill out following the instructions here: https://gitlab.cern.ch/wlcg-doma/site-network-information/-/tree/master/
- The template has both mandatory and optional sections. Examples are in the SitePages area in Gitlab
- The completed site specific markdown file should be converted to HTML, downloaded to a site’s web server and the URL registered in WLCG CRIC.
Packet and Flow Marking Status

**Flow Marking** (UDP firefly) implementations
- **Xrootd** 5.0+ supports UDP fireflies
  - Site admins can configure mapping of paths to experiments and user/roles to activities
  - Clients support Xroot protocol extension via `scitag` URL flag
- **dCache** PoC now ready, supports UDP fireflies
  - Testing deployment at AGLT2 (backported to 7.2) with issues reported back to Tigran

**Flow and Packet Marking**
- **Flowd** - packet and flow marking service
  - Independent service that can mark flows and packets for 3rd party services

**Collectors/Receivers**
- Initial receiver prototype developed by ESnet (available on [scitags github](https://github.com/scitags/scitags))

**Registry**
- Provides list of experiments and activities supported
  - Exposed via JSON at [api.scitags.org](http://api.scitags.org)

**Flow id propagation**
- Work needed has been agreed with Rucio and FTS (tickets were submitted to follow up)

Details in [WLCG](https://wlcg.web.cern.ch) and [Rucio](https://rucio.web.cern.ch) Presentations.
During Supercomputing 22 in Dallas, we demonstrated a number of aspects of our packet and flow marking work.

- We showed packet marking at 200 Gbps rates using flowd with both xrootd and iperf3.
- Scinet and ESnet set up packet collectors via sflow and demonstrated real-time monitoring of packets by experiment and activity.
- Demos were also run on LHCONE using equipment in the SC22 booth, KIT, UVic and CERN where packet marking for xroot transfers was monitored using a P4 programmable switch.
Looking Forward: BDT Plans in H1 2023

1. **Begin transitioning the new REST-based archive management to production use.**
   - All the pieces - FTS, storage implementations - are in place and a few scale tests have already been performed.
   - Next logical step is production systems!

2. **“Cutting metal” for FTS3 / token integration.**
   - In 2022, we specified most of the authorization flows we’d like to see for FTS3 but lacked the developer effort to implement it.
   - FTS3 team is getting additional short-term (3 yrs) effort starting early 2023 to accelerate this development.

3. **Revive Rucio-based tests for tokens.**
   - 2022 largely focused on configuring storage and nightly integration tests run by IAM team. As FTS3 becomes ready to do transfers with WLCG tokens, we will need to revive the Rucio setup done for DC21 to enable end-to-end tests.

4. **Enable packet/flow marking for DC24**
   - We would like to get larger sites to deploy flowd and get IPv6 traffic marked so we can have a significant amount of DC24 traffic identifiable by owner/activity anywhere in the network.
   - Largest site will need to deploy the site network monitoring.