XRootD Packet Marking

Presentation at RNTWG
June 7, 2022

Andrew Hanushevsky, SLAC
http://xrootd.org

XRootD Packet Marking

- Two aspects
  - Firefly UDP Packet generation
  - IPv6 Flow Label setting
- Common information between the two
  - Experiment as a 9 bit number
  - Activity as a 6 bit number
  - Scitag a URL cgi element that encodes above
    - scitag.flow=(expcode<<6) | actcode

Determining Experiment & Activity

- The overly optimistic Scitag
  - URL providers were to tag URL's with Scitag
    - As they should actually know the Exp-Act
  - It's been way too slow in coming
    - I don't know of any provider that is planning it
- **XRootD** uses a practical temporary bypass
  - Data Access Context Resolution (DACR)
    - Mechanism widely used for authorization
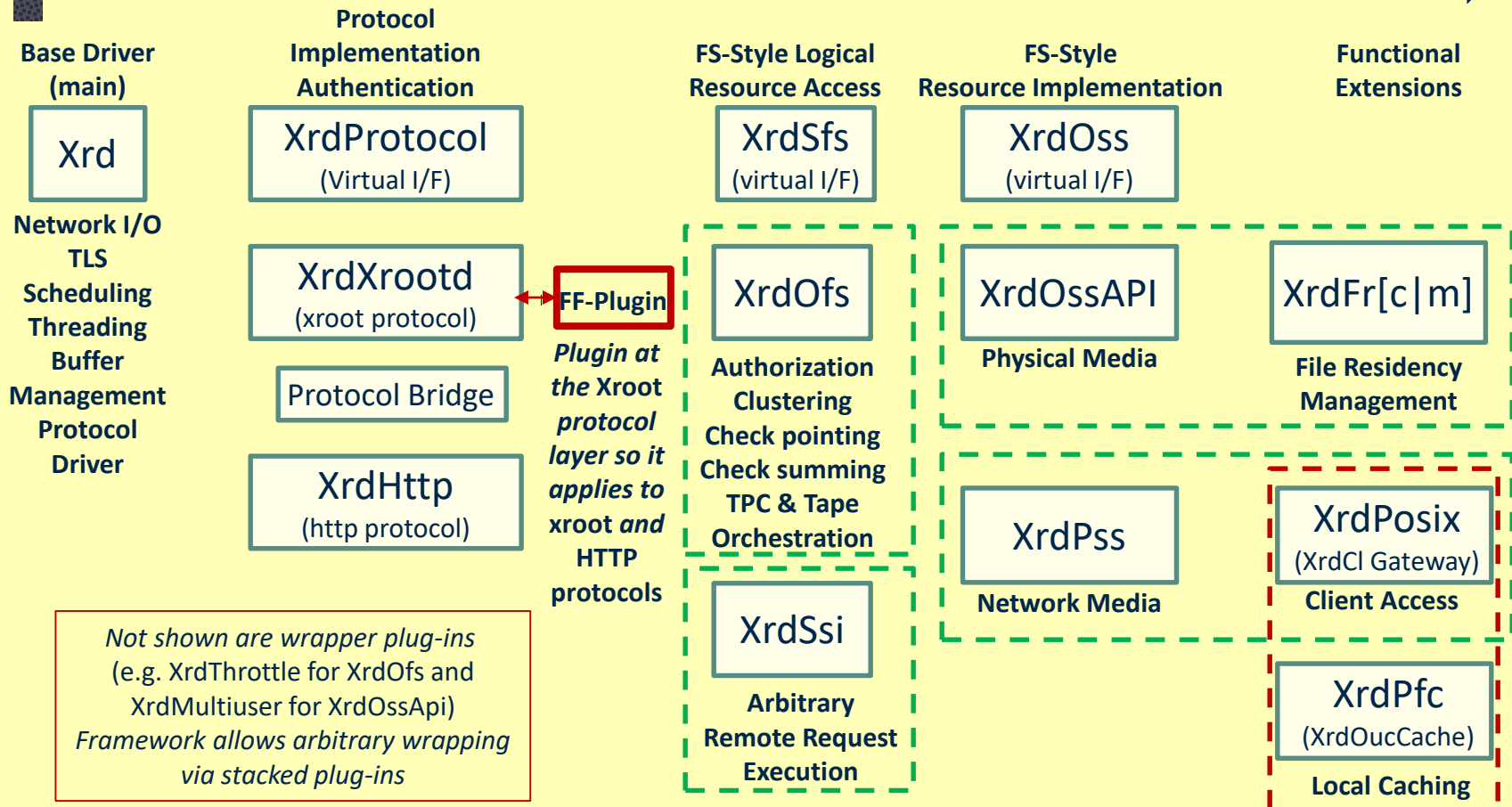   - Big issue for multi-discipline sites

XRootD DACR

- If Scitag exists it determines Exp-Act O/W
- Experiment determined by configured
  - Path being opened or
  - Virtual organization membership (x509)
  - Site default if neither applies
- Activity determined by configured
  - Client's username or
  - Client's role (x509)
  - Site default if neither applies

XRootD Firefly Implementation

- Exists as an internal plug-in
  - Loaded if Firefly configured
    - Cloned for each connection as Exp-Act may differ
- On first file open…
  - DACR applied to determine Exp-Act
  - Firefly packet emitted
    - Currently, no periodic echo to avoid UDP storm
- On connection termination
  - Ending Firefly packet emitted

Firefly Plug-in Positioning

Base Driver (main)	Protocol Implementation Authentication		FS-Style Logical Resource Access	FS-Style Resource Implementation	Functional Extensions
Xrd	**XrdProtocol** (Virtual I/F)		**XrdSfs** (virtual I/F)	**XrdOss** (virtual I/F)	

Network I/O
TLS
Scheduling
Threading
Buffer
Management
Protocol
Driver

XrdXrootd (xroot protocol) ←→ **FF-Plugin**

Plugin at the Xroot protocol layer so it applies to xroot and HTTP protocols

Protocol Bridge

XrdHttp (http protocol)

XrdOfs

Authorization
Clustering
Check pointing
Check summing
TPC & Tape
Orchestration

XrdSsi

Arbitrary
Remote Request
Execution

XrdOssAPI

Physical Media

XrdPss

Network Media

XrdFr[c|m]

File Residency
Management

XrdPosix (XrdCl Gateway)

Client Access

XrdPfc (XrdOucCache)

Local Caching

Not shown are wrapper plug-ins (e.g. XrdThrottle for XrdOfs and XrdMultiuser for XrdOssApi) Framework allows arbitrary wrapping via stacked plug-ins

Some Firefly Caveats

- ⊞ All flow tagged with Exp-Act of 1ˢᵗ open
  - Technically violates JWT implementation
    - With JWT's Exp-Act can change for each open
      - This is a nightmare scenario that I am ignoring
        - Frankly, we cannot reasonably implement this
          - Probability of ever being used is practically zero
- ⊞ Periodic Firefly packets not being sent
  - This is to avoid UDP storms and packet loss
    - Servers typically have thousands of connections
      - Still researching how this could be effectively done

SLAC
NATIONAL ACCELERATOR LABORATORY

XRootD IPv6 Flow Labels

- Implementation on hold
  - Curious oddities make IPv6 FL problematic
    - Label can only be set by connection initiator
      - That means every client must set correct flow label
        - Server must specify label is to be reflected
          - Not yet implemented on all platforms

- These are roadblocks
  - No obvious good path forward

XRootD IPv6 FL Roadblocks I

- Determining correct experiment-activity
  - No reasonable way other than scitag
    - Recall that scitags are themselves problematic
- Each worker node needs proper FL config
  - May be hard to come by based on IPv6 issues
- Deployment of IPv6 FL capable clients
  - Typically a multi-year effort
    - For any kind of significant penetration

XRootD **IPv6 FL Roadblocks II**

- Clients deployed via k8s cannot set FL
  - This will be an ongoing problem
    - If k8s becomes the path for workload deployment
- Does not address non-xroot clients
  - Curl, wget, aria, gfal, any 3rd party client

Third Party Clients?

- The **XRootD** framework supports HTTP
  - Accessed only via 3rd party clients
- HTTP traffic is approaching 50% usage
  - Due to switch to using HTTP Third Party Copy
- Retrofitting 3rd party clients is a no-go
  - Except in certain limited scenarios
    - E.G. **XRootD** curl based HTTP TPC
    - Likely not possible for other platforms
- This is a real problem

Don't take this as being negative!

- Anything can be overcome but …
  - It's a question of time and effort
  - Team already considering alternatives
    - Seems that Flow Labels in general under thought
    - Net community is scrambling to make sense of it
      - Current approach is to only use random flow labels
        - Latest Linux release makes this the default
          - This actually doesn't make for a happy situation
  - Looking forward to what comes next!

SLAC
NATIONAL ACCELERATOR LABORATORY

Conclusion

- Firefly implementation is solid
  - Available since release 5.4.0
    - https://xrootd.slac.stanford.edu/doc/dev55/xrd_config.htm#_Toc88514010
- Flow Label implementation is on hold
  - We need more clarification & commitments
    - Scitag commitment or reasonable alternative
  - Perhaps a totally different approach?
    - A server-side approach would be far better