

# xrootd Update

OSG All Hands Meeting

University of Nebraska

March 19-23, 2012

Andrew Hanushevsky, SLAC

---

<http://xrootd.org>

# Outline

---

- # Additions & Changes (available now in 3.1.1)
- # On the horizon (anticipated for 3.2 release 3Q12)
- # Near future (anticipated for 3.3 release)
- # The xrootd collaboration
- # Conclusion
- # Acknowledgements

# Additions & Changes (3.1)

---

- # Extended Attribute Framework
- # Integrated checksums
- # Shared-Everything File System Support
- # Static ENOENT redirection
  - Enhances creation of new topologies
- # Caching proxy server
- # Federated site shares
- # Monitoring Extended
- # Read-fork-read client

# Agnostic Extended Attributes

## # FS-Independent Extended Attribute Framework

- Used to save file-specific information
  - File system must support extended attributes
- Current Attributes
  - XrdCks.xxx xxx (e.g. md5) checksum
  - XrdFrm.yyy **F**ile **R**esidency **M**anager information
- Future Attributes
  - Original file creator
  - Extended Access Information
    - Number of parallel readers, % read, access time
      - Can be used for better migration/purging

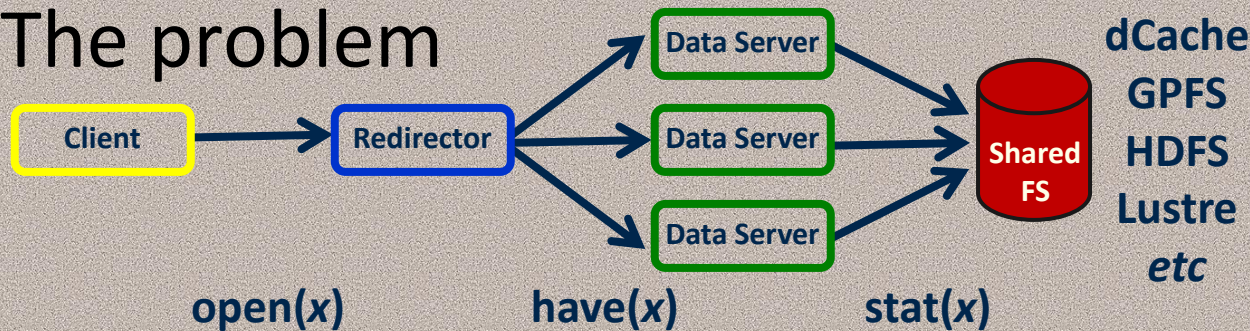
# Integrated Checksums

- # Originally, checksums supported via callout
- # Now, xrootd internally handles checksums
  - **xrootd.chksum** [*max num*] {adler32|crc32|md5}
    - Add new ones via **ofs.ckslib** directive and plug-in
  - Checksum saved in extended attributes
    - Returned on query
    - Automatically recomputed when file changes
      - Can be disabled
  - External program not needed; but still supported



# Shared FS Support

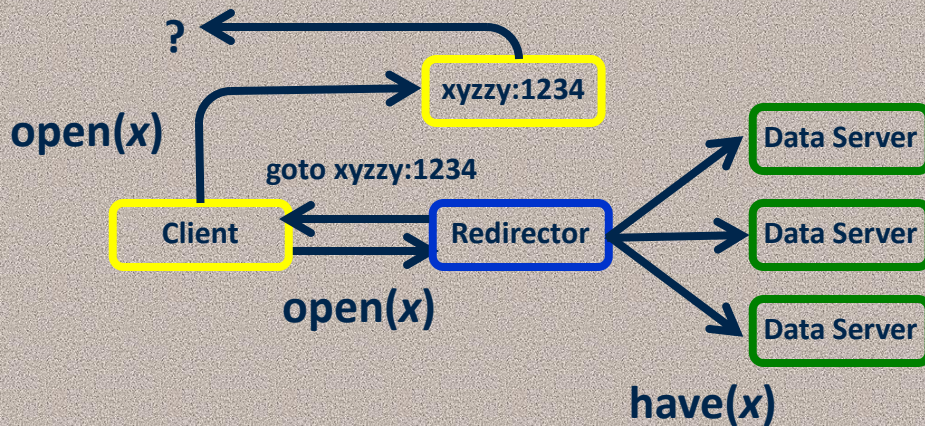
## # The problem



## # The solution

- # Use **cms.dfs** directive to eliminate duplicate hits
- # Many tuning options available
  - # Lookup at redirector, limits, caching, etc
  - # See [http://xrootd.org/doc/prod/cms\\_config.htm](http://xrootd.org/doc/prod/cms_config.htm)

# Static ENOENT Redirection

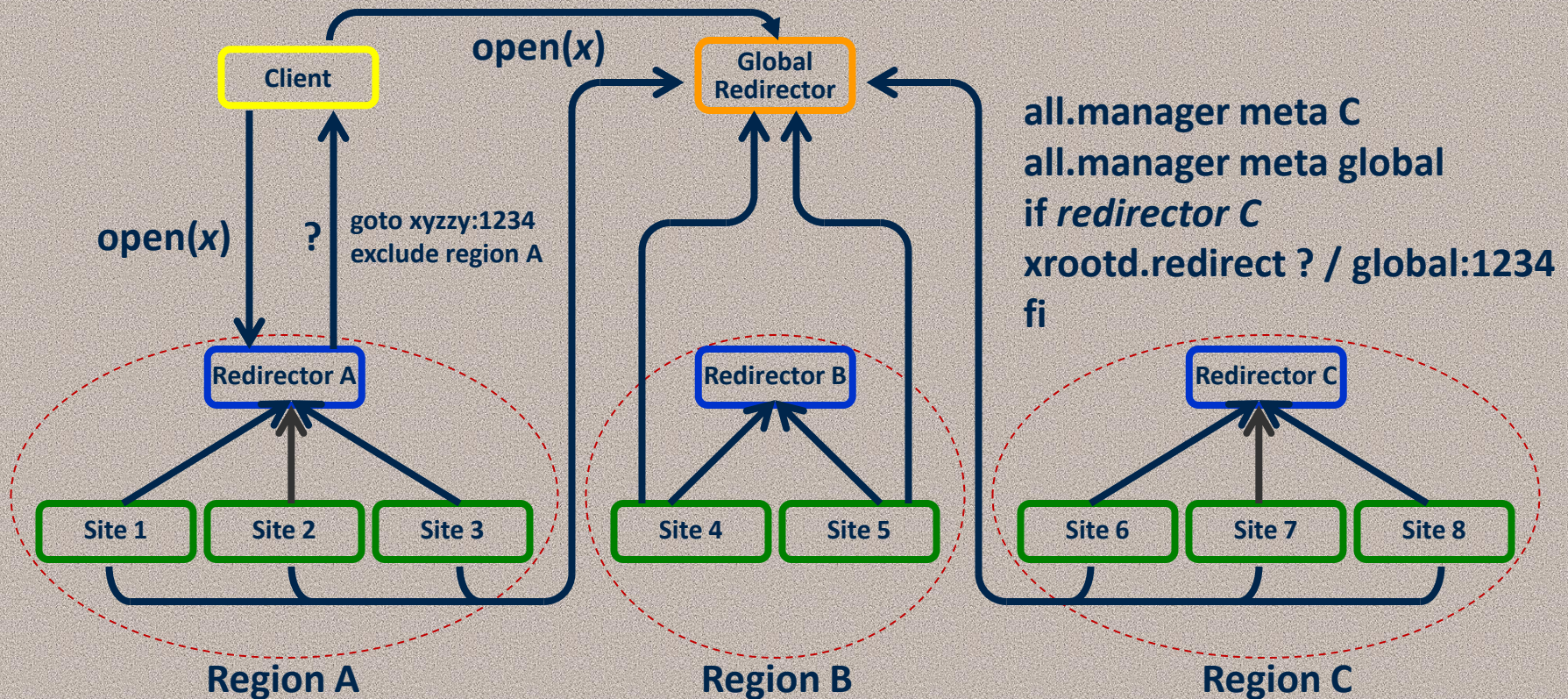


```
if redirector_host  
xrootd.redirect ? / xyzzy:1234  
fi
```

## # What we envisioned. . .

- # xyzzy is a caching proxy server or proxy cluster
- # Provides high performance WAN access
- # Client accesses data via WAN in this case

# What Others Envisioned!



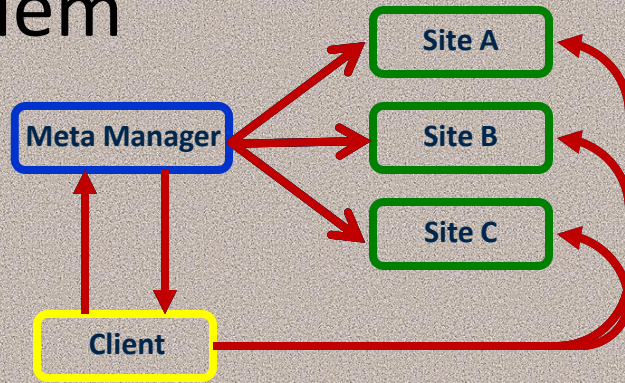


# Caching Proxy Server

- # The proxy server plug-in has in-memory cache
  - Must be enabled via **pss.memcache** directive
- # Many tuning options; some of which are . . .
  - Cache and page size
  - Read ahead size
  - Maximum block size to cache
  - Root file access optimization
- # See [http://xrootd.org/doc/prod/ofs\\_config.htm](http://xrootd.org/doc/prod/ofs_config.htm)

# Federated Site Shares

## # The problem



**Each site gets an equal number of requests all things being equal**

## # What to do if sites do not want to be equal?

- # Use **cms.sched** directive to establish site share

- # Use the **gshr** and **gsdfit** options

- # See [http://xrootd.org/doc/prod/cms\\_config.htm](http://xrootd.org/doc/prod/cms_config.htm)

# Monitoring

- # Readv requests fully monitored
  - By default, only a readv request summary provided
    - Can request a full unwind of a readv request
- # Per client I/O monitoring now flushable
  - Before, I/O statistics flushed when buffer full
  - Can specify a flush window
    - Based on code provided by Matevz Tadel, CMS
- # Authentication can now be fully monitored

# Read-Fork-Read Client

- # The current client allows forking
  - This allows sharing data between processes
    - Read conditions data
    - Fork  $n$  times for parallel processing
    - Read event data using pre-read conditions data
- # Extensively used by CMS
  - Substantially reduces memory load
    - Critical for large multi-core worker nodes



# git & cmake & EPEL

# Now using git repository for source code

- git clone <http://xrootd.org//repo/xrootd.git>

# Standardized on cmake for builds

- See README file in the top level source directory

# Adhering to EPEL guidelines

- All '.a' files are replaced by '.so' files
  - Caused increase in number of installed .so files
    - We consolidated libraries; but more than we would like
    - Unfortunately, changing things afterwards is difficult in EPEL



# On the horizon (3.2)

---

- # New fully asynchronous client
  - Guards against low performing servers
    - To be added in subsequent release
- # Extended monitoring
  - Redirect information
  - Authentication summary information
- # Integrated 3<sup>rd</sup> party copy
  - Allows client directed server-to-server copies
- # Dropping RH4 support

# Things Within A Year (3.3+)

---

- # Disk Caching Proxy Server
  - Extension of memory caching (UCSD effort)
- # Automatic checksum validation
- # Integrated alerts
- # New more effective async I/O model
- # IPV6

# Xrootd Collaboration

- # Mutually interested institutions contributing effort for development and maintenance
  - SLAC (founder)
  - CERN (2010)
  - Duke (spring 2011)
  - JINR (fall 2011)
  - UCSD (winter 2011)
    - Newest member!

# Conclusion

# **xrootd** is under active development

- Always looking for new ideas
  - Feel free to suggest them
- Be a contributor
  - You too can contribute to the code base
- Consider joining the **xrootd** collaboration
  - It costs no money to join

# See more at <http://xrootd.org/>

# Acknowledgements

## # Current Software Contributors

- ATLAS: Doug Benjamin, Patrick McGuigan, Danila Oleynik, Artem Petrosyan
- CERN: Fabrizio Furano, Lukasz Janyst, Andreas Peters, David Smith
- CMS: Brian Bockelman (unl), Matevz Tadel (ucsd)
- Duke: Douglas Benjamin
- Fermi/GLAST: Tony Johnson
- LBNL: Alex Sim, Junmin Gu, Vijaya Natarajan (BeStMan team)
- Root: Gerri Ganis, Beterand Bellenet, Fons Rademakers
- OSG: Tim Cartwright, Tanya Levshina
- SLAC: Andrew Hanushevsky, Wilko Kroeger, Daniel Wang, Wei Yang

## # Operational Collaborators

- ANL, BNL, CERN, FZK, IN2P3, SLAC, UCSD, UTA, UoC, UNL, UVIC, UWisc

## # US Department of Energy

- Contract DE-AC02-76SF00515 with Stanford University