Automated XRootD Integration Testing using Kubernetes in GridPP

Robert Currie[1], *Wenlong* Yuan[1]

[1]University of Edinburgh, Edinburgh, UK

Abstract. XRootD is a robust and scalable service that enables distributed data management for scientific communities, including Astronomy and High-Energy Physics (HEP). Within GridPP, XRootD is crucial for managing over 100PB of storage. However, due to diverse site configurations, integration testing has been historically challenging. This paper presents "XKIT" a Kubernetes-based system for automated XRootD testing, reducing the burden on administrators while ensuring reliability and reproducibility. We discuss the implementation, testing strategies, and benefits for the broader GridPP community.

1 Introduction

XRootD[1] is a widely adopted, scalable solution for distributed data access across the WLCG[2] community. GridPP[3] serves as the UK's distributed computing infrastructure for particle physics, providing computational resources and data storage capabilities essential for analyzing the massive datasets generated by experiments at CERN and other facilities. This mission requires highly reliable, high-throughput data access to support thousands of concurrent analysis jobs across geographically distributed sites.

Within GridPP, multiple XRootD instances deployed across different Tier-2 centers form a federated storage system that enables seamless data access regardless of the physical location of files. These distributed XRootD deployments allow researchers to access petabytes of experimental data efficiently, supporting GridPP's role in enabling cutting-edge particle physics research. However, the diversity of site configurations across these XRootD instances poses significant challenges for integration testing. Historically, testing new releases required extensive manual effort and coordination between multiple sites. Our work introduces XKIT, an automated testing platform using Kubernetes[4] to deploy virtual grid sites and execute high-level integration tests.

2 Motivation

Integration testing is essential due to the complexity of XRootD deployments within GridPP. Key challenges include:

- XRootD is highly configurable with diverse plugins and versions.
- There are a large number of possible configurations across sites.
- Need to ensure backward compatibility and seamless upgrades to support scientific code.

As its adoption has increased, ensuring stable and well-tested deployments has become a critical concern. Traditional testing methods required manual validation at grid sites, engaging multiple experts, and consuming valuable resources. Furthermore, enabling various XRootD extensions in production without proper testing has historically led to unexpected behaviors, including performance degradation, misconfigured authentication mechanisms, and interoperability issues with different storage systems. These challenges are exacerbated by the lack of a standardized testing environment, making it difficult to predict how a particular site configuration will behave when new features are introduced. Addressing these issues requires a comprehensive integration testing framework that systematically evaluates different configurations before they are deployed in a production setting.

With a reduced number of core storage sites, the importance of ensuring high availability and reliability for each individual site has increased. Failures or misconfigurations in a single storage site can have a greater impact on the overall GridPP infrastructure, making rigorous testing and validation essential to maintaining stable operations. Ensuring that all extensions and configurations are tested in controlled environments before deployment helps mitigate risks associated with unexpected behavior in production. Automating this process reduces effort and enhances reproducibility.

3 XKIT Design and Implementation

Several key features in the design of XKIT test framework include:

- Containerized XRootD deployments.
- Kubernetes for managing service orchestration.
- Automated execution of client-server test cases.
- Logging and result collection in a structured format.

This setup is designed to support automated testing of XRootD configurations reflecting different site setups.

Using multiple XRootD servers deployed in Kubernetes clusters with predefined configurations gives us the flexibility to test various scenarios not captured in throughput or isolated unit-testing. Coupled with reproducible client tests we aim to be able to run automated production-like workflows. Figure 1 shows the XKIT testing strategy. The methodology is divided into three stages: Client-Side Testing, Server-Side Testing, and End-to-End Validation. A design choice within this strategy is to prioritize using well-defined APIs to reduce boilerplate code.

Credential Management: In order to perform full-stack integration testing, XKIT utilizes secure tokens and X.509 user-proxy certificates. These are manually generated locally by the user and are then securely passed into the containerized test client. This approach ensures that real grid credentials are properly isolated within the testing environment whilst maintaining the security standards required for GridPP operations. Host-based x509 certificates for testing have also been issued by the GridPP CA following best practices. These certificates are mounted into server-side containers on a per-server basis to support a full grid security stack.

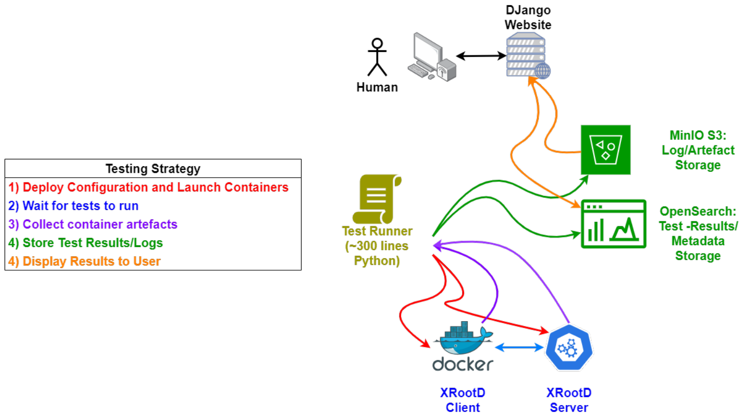

Figure 1. This figure shows the XKIT Testing Strategy. The methodology is divided into three stages: Client-Side Testing, Server-Side Testing, and End-to-End Validation. A design choice within this strategy is to prioritize using well-defined APIs to reduce boilerplate code.

4 Testing Methodology

Testing within XKIT is separated into 3 different categories of tests:

1. **Client-Side Testing**: Running command-line utilities (e.g., `xrdcp`, `xrdfs`) and API calls to validate expected behavior.

2. **Server-Side Testing**: Verifying read/write transfers, plugin compatibility, and system logs.

3. **End-to-End Validation**: Ensuring seamless operation in realistic site-like environments.

Our testing approach is explicitly designed to support GridPP sites by replicating real-world deployments across major UK Tier-2 centers. Since site configurations are diverse, covering different versions of XRootD, authentication mechanisms, and third-party integration requirements, it is essential to have a comprehensive testing framework that can accommodate these variations. Each test suite is structured to validate compatibility with past, present, and upcoming XRootD releases, aiming to allow for seamless upgrades for GridPP sites.

Key considerations we've taken into account in our testing strategy include:

- Emulating grid site configurations through containerized test environments.
- Supporting backward compatibility checks to assist sites running older versions.
- Automating regression testing to detect configuration issues before deployment.
- Centralizing test management and results collection to provide a reference for GridPP administrators.

Service-Oriented Design: Unlike monolithic test frameworks requiring extensive boilerplate, XKIT leverages Kubernetes to create ephemeral, site-like environments, allowing for automated, scalable, and reusable test cases. Helm charts provide structured configuration management, while integration with CI/CD pipelines ensures that test execution is both automated and reproducible. This API-driven approach reduces maintenance overhead, enhances reusability, and facilitates rapid adaptation to evolving GridPP requirements.

Integration Testing at Scale: Unlike unit testing, which focuses on isolated code correctness, integration testing ensures that multiple components function together as expected. The challenge in large-scale integration testing for GridPP is the diversity of site-specific configurations. Capturing and documenting individual site setups is crucial but complex, as configurations differ in authentication methods, network topology, and plugin usage. XKIT addresses this by providing parameterized, reproducible test environments that can adapt to site variations, reducing manual intervention while ensuring comprehensive validation.

Reference Configurations for Production Storage: One of the key outcomes of XKIT testing is the development of a set of reference configurations that can be used for production storage setups. By systematically testing different XRootD deployments, authentication mechanisms, and network topologies, XKIT enables the identification of best-practice configurations suited to various GridPP site requirements. These reference configurations serve as a guide for site administrators, helping them deploy XRootD in a manner that is both optimized and fully validated against known working setups. This reduces configuration drift across sites and ensures consistency in production deployments.

By addressing these concerns, XKIT aims to reduce the reliance on manual validation by individual sites, allowing GridPP members to focus on optimizing their storage infrastructure while benefiting from consistent, high-quality testing outcomes.

5 Conclusion

We are in a position where we are growing the amount of different testing environments which we're testing XRootD with. Our aim is to build a robust and reliable testing framework with XKIT, enabling automated, reproducible testing of XRootD configurations, reducing administrative overhead and improving reliability for GridPP. By using containerized deployments and Kubernetes, we can validate processes more effectively, which benefits the wider scientific community.

Acknowledgments

We acknowledge the support of GridPP, the University of Edinburgh, and the CHEP 2024 conference organizers.

References

[1] A. Hanushevsky et al, XROOTD - A highly scalable architecture for data access. WSEAS Transactions on Computers. *Volume 4*. pages 348-353. (2005).
[2] Jamie Shiers, The Worldwide LHC Computing Grid (worldwide LCG), Computer Physics Communications, *Volume 177*, Issues 1–2, (2007).
[3] Faulkner, P. J. W. and others, GridPP: Development of the UK computing Grid for particle physics, J. Phys. G, *Volume 32*, pages N1–N20, (2012).
[4] B. Burns et al., *Kubernetes: Up and Running*, O'Reilly Media, (2017).