

SOFTWARE-DEFINED MEASUREMENT TO SUPPORT
PROGRAMMABLE NETWORKING FOR SOYKB

Longhai Cui

Dr. Prasad Calyam, Thesis Supervisor

ABSTRACT

Campuses are increasingly adopting hybrid cloud architectures for supporting big data science applications that require "on-demand" resources, which are not always available locally on-site. Policies at the campus edge for handling multiple such applications competing for remote resources can cause bottlenecks across applications. To proactively avoid such bottlenecks, we investigate the benefits in the integration of two complementary technology paradigms of software-defined measurement and programmable networking. The integration inherently allows flexible end-to-end application performance monitoring and dynamic control of big data application flows using: (a) software-defined networking for transit selection to remote sites, and (b) pertinent selection of local or remote compute resources. Using the Soybean Knowledge Base (SoyKB) as an exemplar application, we demonstrate the benefits of software-defined measurement to support programmable networking.

As part of our study methodology, we first profiled the original data flows within SoyKB's use of iPlant public cloud resources, and identified bottleneck cases such as slow data transfer speeds, lack of performance information (e.g., such as cluster availability, job status and network health) and inflexible control of hybrid cloud resources to address application-specific needs. The profiling study motivated us to propose a new hybrid cloud architecture for SoyKB workflows that utilize end-to-end performance measurements to support a cost-optimized selection of sites for computation and effective traffic engineering at the campus-edge. We validate our approach for a SoyKB workflow use case that we setup on a wide-area overlay network testbed implementation across two geographically-distributed campuses. Our performance results show a notable performance improvement in SoyKB remote data transfer flows that utilize iRODS and TCP tuning mechanisms in the presence of cross-traffic big data flows. Additionally, we implement a SoyKB system that provides: (i) flexible workflow performance analytics at a glance to SoyKB researchers handling big data, and (ii) web service mechanisms for interfacing with popular dynamic resource management technologies such as OpenStack, HTCondor and Pegasus.