OneDataShare: Data Transfer Optimization as a Service

PI: Tevfik Kosar, University at Buffalo; **NSF Award No:** 1724898

Overview

OneDataShare is a **universal data sharing building block** for data-intensive applications, with three major goals: (1) **optimization** of end-to-end data transfers and reduction of the time to delivery of the data; (2) **interoperation** across heterogeneous data resources and on-the-fly inter-protocol translation; and (3) **prediction** of the data delivery time to decrease the uncertainty in real-time decision-making processes.
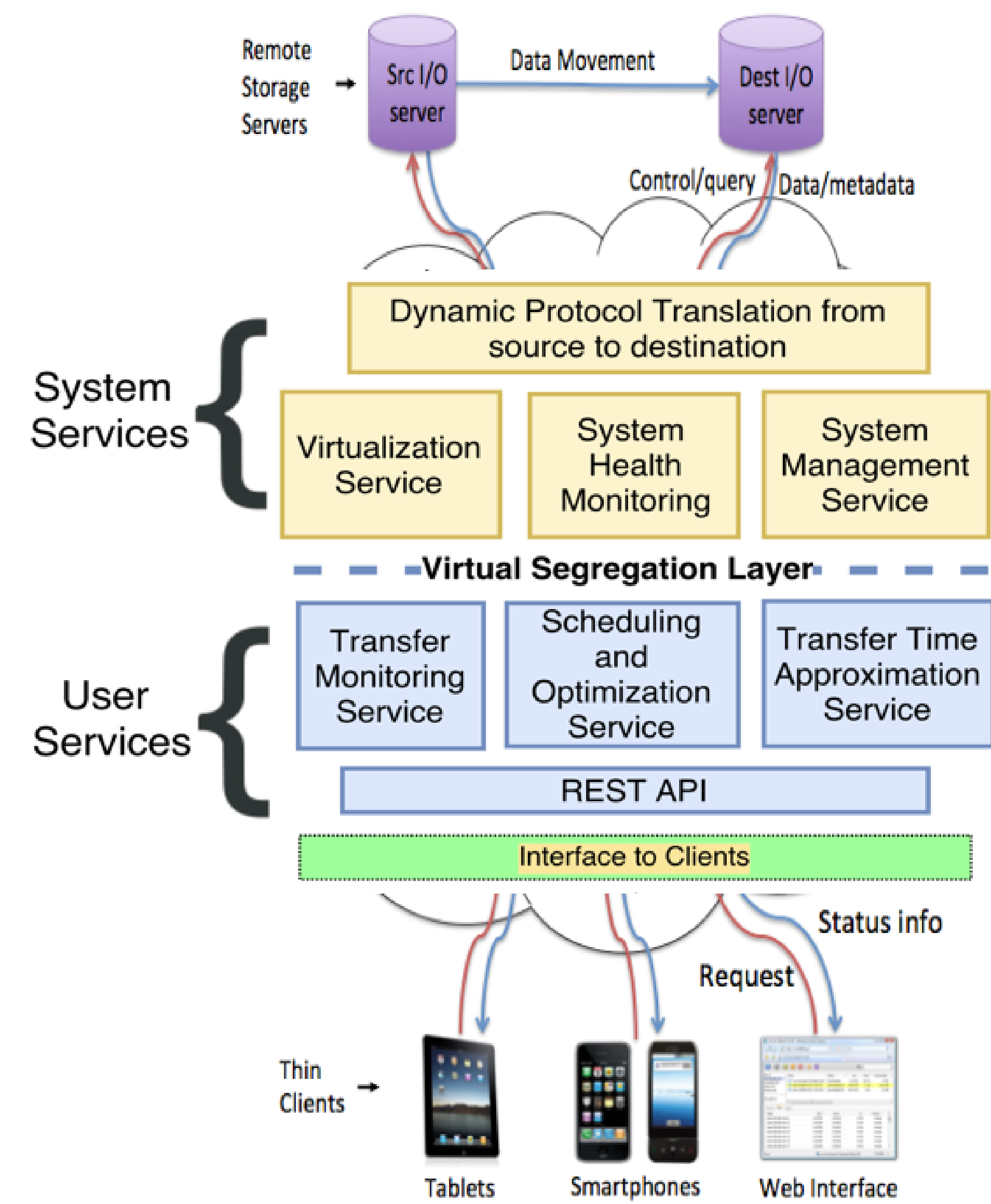

Figure 1: OneDataShare high-level overview.

Key Contributions

1. Novel data transfer optimization algorithms based on historical data analysis and real-time dynamic tuning.

2. Universal interface specification for heterogenous data storage end points.

3. Instrumentation of end-to-end data transfer time prediction capability.

Throughput Optimization

We have developed three novel end-to-end data transfer optimization algorithms based on historical log analysis and real-time dynamic tuning: HARP [1], ProMC [2], and ASM/ANN [3], which significantly outperform the state-of-the-art solutions in this area.
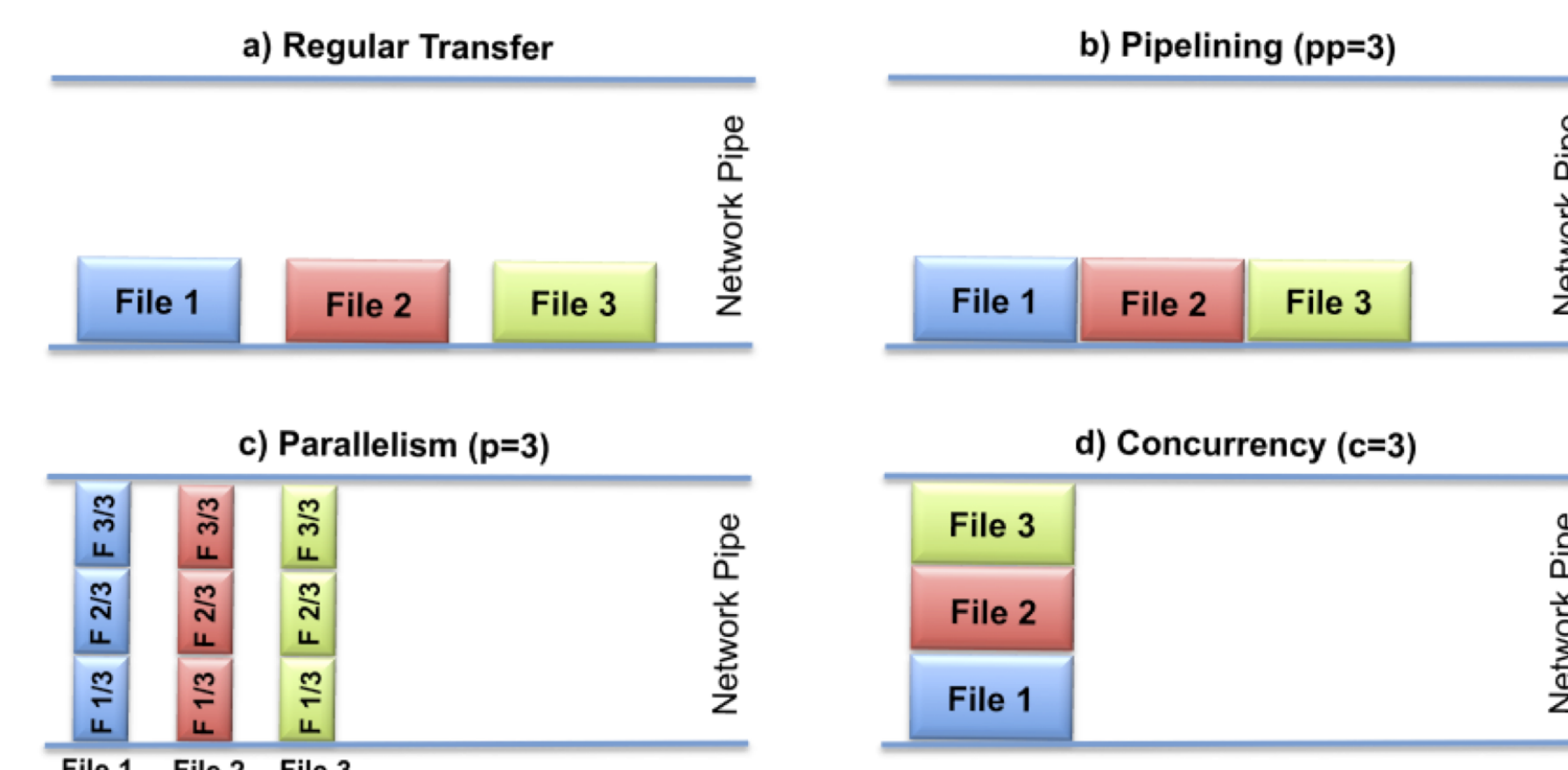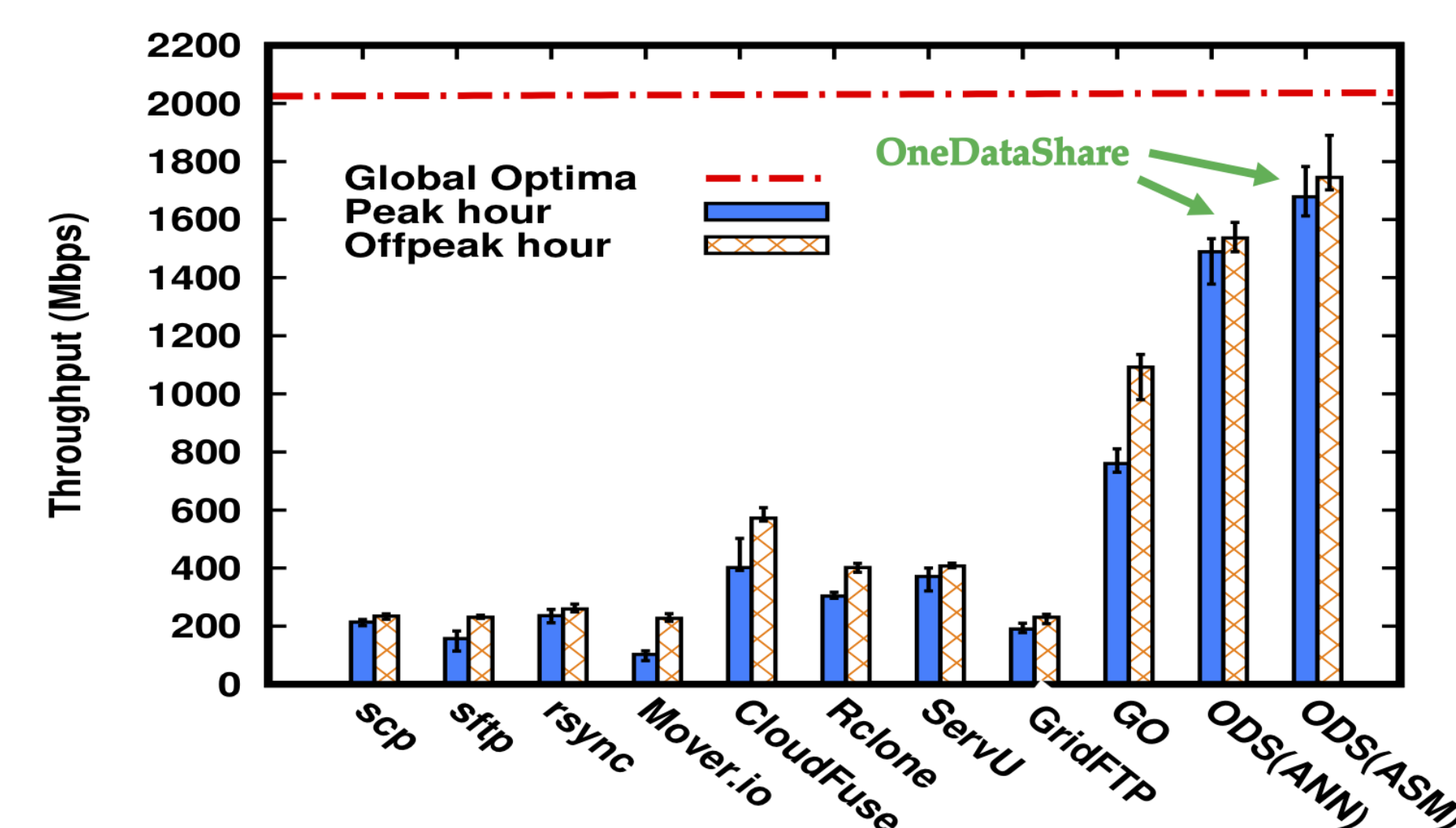

Figure 2: Protocol parameters tuned.

Figure 3: OneDataShare performance vs other MFTs on the 2Gbps IBM interdatacenter network (with SDN) between Washington, DC and San Jose, CA.
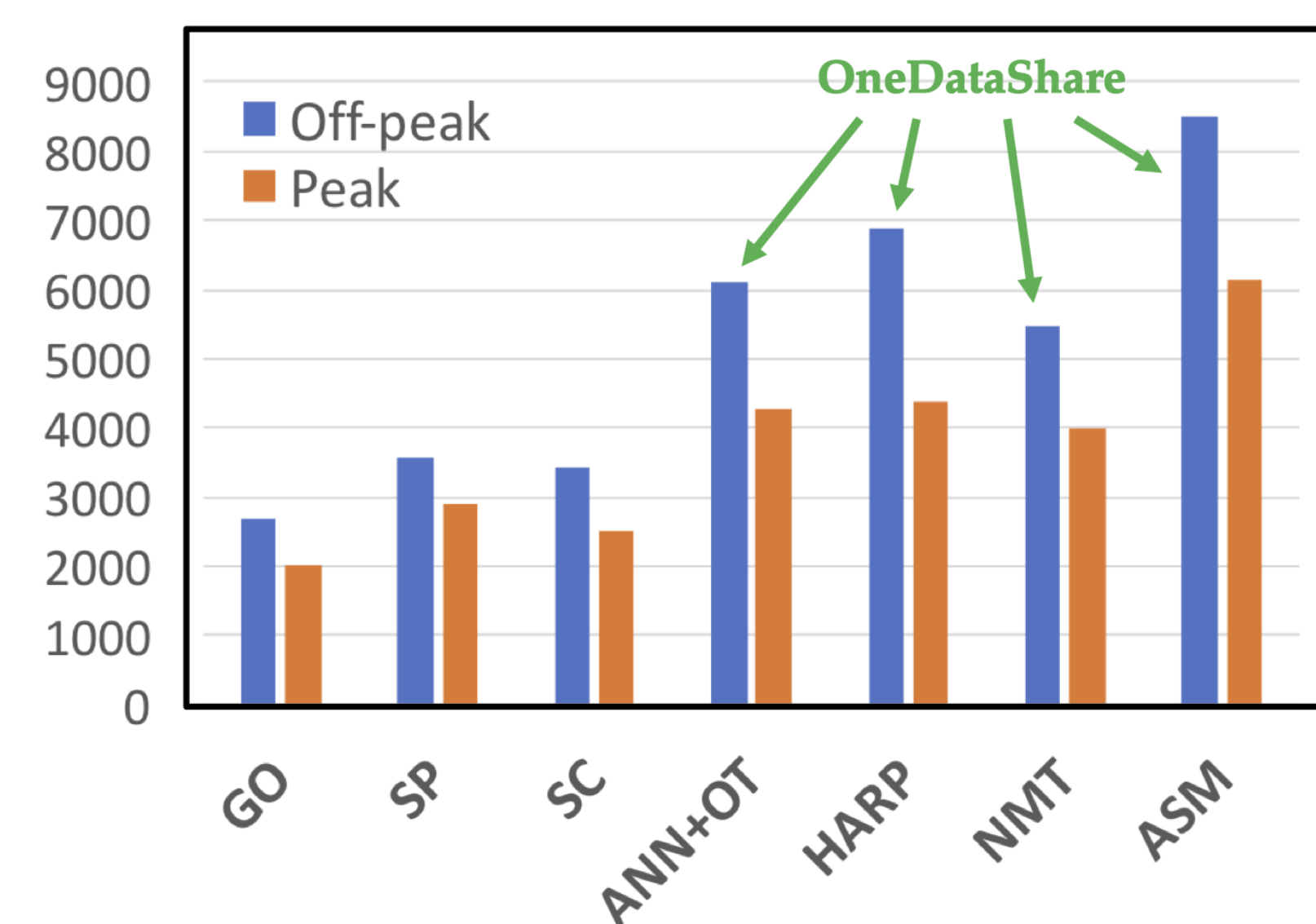

Figure 4: OneDataShare performance vs other optimization models on a 10Gbps XSEDE network between Stampede (TACC) and Gordon (SDSC).

Protocol Translation

OneDataShare [4, 5, 6] provides multi-protocol support and on-the-fly protocol translation. Currently it supports FTP, SCP, SFTP, HTTP, GridFTP, Dropbox, Google Drive, and Box.
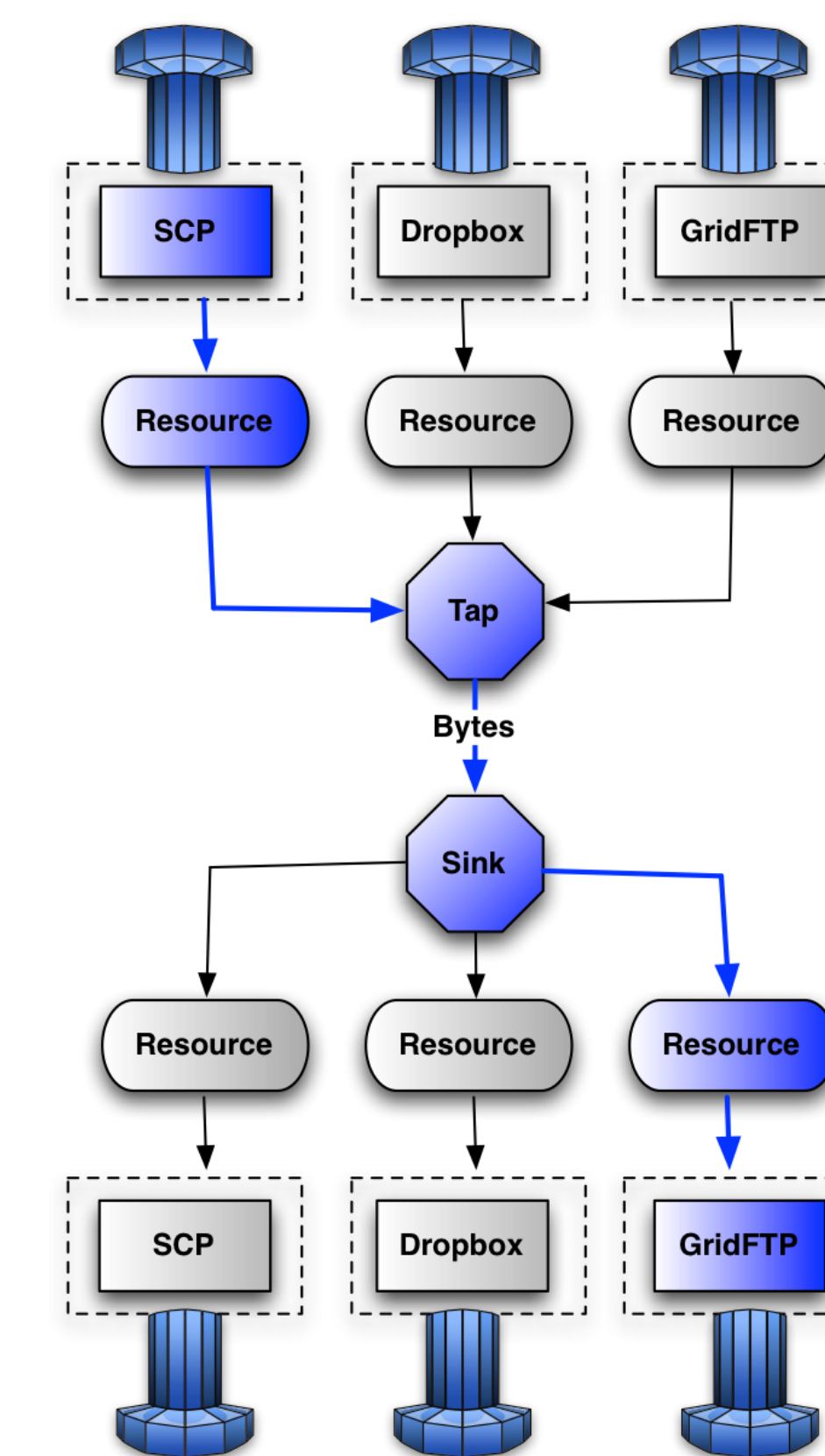

Figure 5: OneDataShare dynamic protocol translation.

Thin Client Interfaces

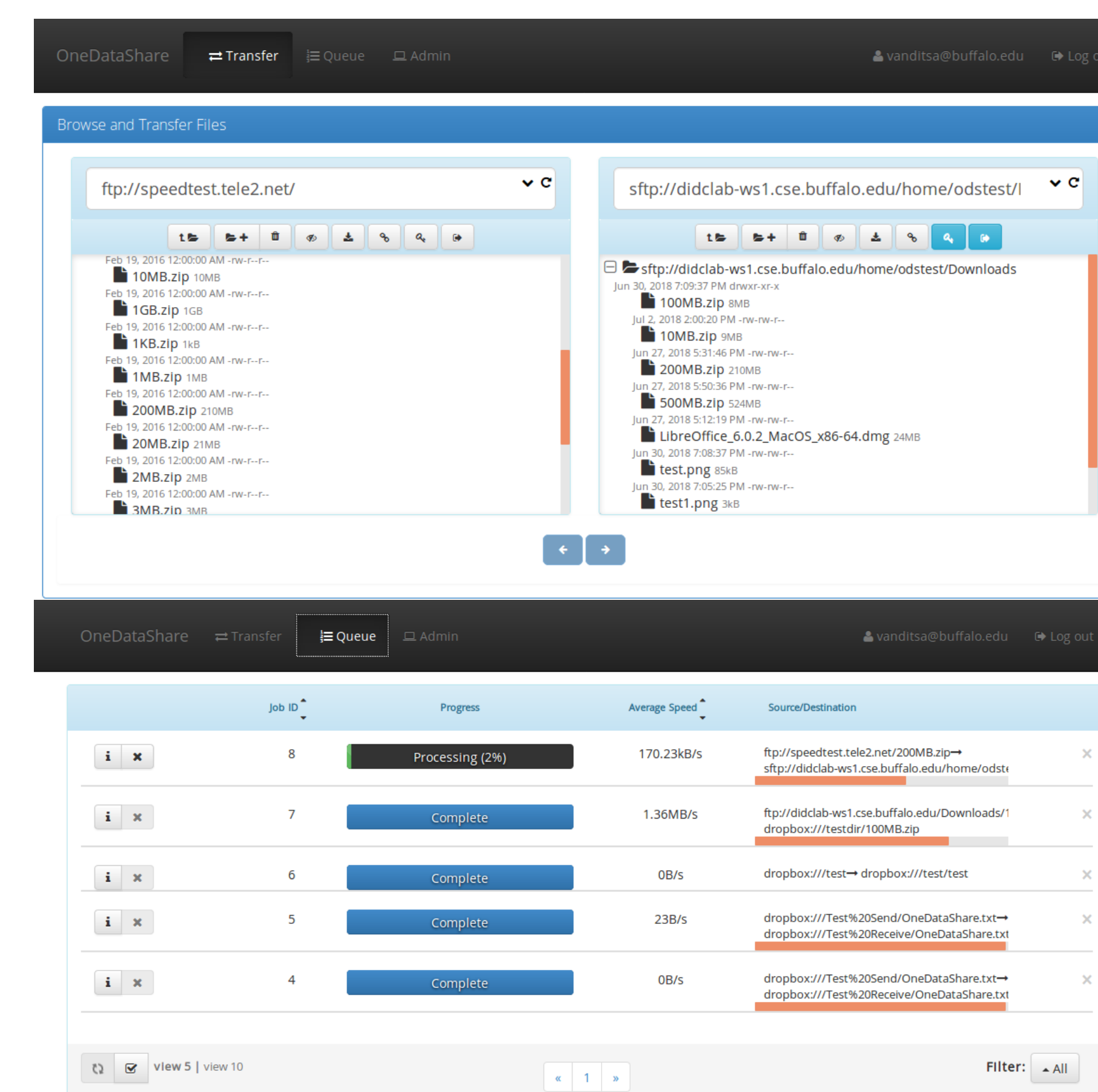Single dashboard given to the users to transfer files to any protocol via the cross-protocol architecture.

Figure 6: OneDataShare data transfer interface.

Education and Outreach

* **Training:** OneDataShare research and development team consists of 3 PhD students, 8 MS students, and 4 BS students (3 funded via NSF REU).
* **Diversity:** The team includes 3 female students and 2 african american students.
* **Internships:** We provided internship for 1 high-school student. 2 PhD students were hosted at IBM for internship via NSF INTERN supplement.
* **Outreach:** In a half-day workshop, OneDataShare technologies were introduced to a group of high school students (mostly low-income inner-city kids) interested in STEM education.
* **Release:** Available as an open-source cloud-hosted service at www.onedatashare.org.

Figure 7: OneDataShare K-12 workshop.

Selected Publications

[1] E. Arslan and T. Kosar, "High-Speed Transfer Optimization Based on Historical Analysis and Real-Time Tuning," *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, vol. 29, no. 6, pp. 1303–1316, 2018.

[2] E. Arslan, B. A. Pehlivan, and T. Kosar, "Big Data Transfer Otimization Through Adaptive Parameter Tuning," *Journal of Parallel and Distributed Computing (JPDC)*, vol. 120, pp. 89–100, 2018.

[3] M. S. Z. Nine, K. Guner, Z. Huang, X. Wang, J. Xu, and T. Kosar, "Big Data Transfer Optimization Based on Offline Knowledge Discovery and Adaptive Sampling," in *Proc. of IEEE International Conference on Big Data*, 2017.

[4] A. Imran, M. S. Nine, K. Guner, and T. Kosar, "OneDataShare: A Vision for Cloud-hosted Data Transfer Scheduling and Optimization as a Service," in *Proc. of International Conference on Cloud Computing and Services Science*, 2018.

[5] L. Di Tacchio, M. S. Z. Nine, T. Kosar, M. F. Bulut, and J. Hwang, "Cross-Layer Optimization of Big Data Transfer Throughput and Energy Consumption," in *Proc. of IEEE International Conference on Cloud Computing (CLOUD)*, 2019.

[6] B. Zhang and T. Kosar, "SMURF: Efficient and Scalable Metadata Access for Distributed Applications from Edge to the Cloud," in *Proc. of IEEE International Conference on Edge Computing (EDGE)*, 2019.