

# Seminar

Department of Computer Science

**Dr. Engin Arslan**

*Postdoctoral Research Associate*

*National Center for Supercomputing Applications*

*University of Illinois at Urbana Champaign*

## Dynamic Protocol Tuning Algorithms for High Performance Data Transfers

Date: Tuesday, February 28, 2017

Time: 11:00 AM – 12:00 Noon

Location: EGRA 219

---

### Abstract

As the data requirements of commercial and scientific applications continue to increase at an unprecedented rate, obtaining optimal end-to-end data transfer performance becomes of crucial importance for a broad range of data-intensive applications. Achieving optimal end-to-end data transfer performance requires effectively utilizing the available network bandwidth and resources, yet in practice the transfers seldom reach the levels of utilization they potentially could. Tuning protocol parameters such as pipelining, parallelism, and concurrency can significantly increase the network utilization and the transfer performance, however determining the best combination for these parameters is a challenging task since that would depend on several factors such as network characteristics, dataset characteristics and end system characteristics.

In this talk, I'll present novel algorithms for application-level tuning of protocol parameters to maximize the data transfer throughput especially in wide-area networks. The contributions of this research include: (1) analysis and prediction of optimal protocol parameter combinations based on the dataset and network characteristics using historical data as well as real-time data; (2) algorithms to cluster the datasets into comparable partitions and transfer multiple partitions concurrently for maximum transfer throughput where the transfer parameters for each partition is optimized individually; (3) dynamic monitoring of the instantaneous data transfer throughput and online tuning of the protocol parameters to detect and remedy possible transfer slowdowns.

We have developed several heuristic solutions that estimate the application-layer protocol parameters, including the number of parallel data streams per file (for large file optimization), the level of control channel pipelining (for small file optimization), and the level of concurrent file transfers to fill the long fat network pipes (for all files). The developed algorithms employ novel techniques to group and transfer set of files in order to yield the maximum possible transfer throughput. The experimental results show that our proposed heuristic algorithms outperform state-of-art solution by up to 5 times.

Although the heuristic solutions boost the data transfer throughput for networks with stable or predictable background traffic, they fall short to optimize the data transfers when network conditions change unpredictably. To address this issue, we propose predictive end-to-end data transfer optimization algorithm based on historical data analysis and real-time background traffic probing, dubbed HARP. Combining historical data analysis with real time sampling enables our algorithms to tune the application level data transfer parameters accurately and efficiently to achieve close-to-optimal end-to-end data transfer throughput with very low overhead. Our experimental analysis over a variety of network settings shows that HARP outperforms the best heuristics by up to 50% in terms of the achieved throughput.