NDN-Hadoop: Exploring Applicability of NDN for Big-Data Computing

Chris Gniady and Beichuan Zhang
Department of Computer Science, University of Arizona
gniady@cs.arizona.edu, bzhang@cs.arizona.edu

Datacenter Networks Are Challenging
- Densely connected
- Lots of engineering to maximize performance before hitting limits
- Alternative protocols may simplify system and applications

Evolution of Communication Abstraction
- Telephony: name the path
- IP: name the endpoint
- Name Data Network (NDN): name the data
- Developed for the Internet
- Efficient multicast
- In-network caching
- Failure recovery

Exploring NDN in a Datacenter
- Goal: Understand applicability and benefits
- Apache Hadoop as a Big Data system

NDN Benefits for Big Data
- NDN provides network optimizations
- Simplifies development of Big Data applications
- NDN provides data multicast
- Benefits data replication in Big Data storage
- NDN provides in-network caching
- Improves performance in Big Data computing
- NDN retrieves data from alternate sources
- Transparent failure recovery in Big Data systems

Simpler Code
- 800 lines of code eliminated
- Several corner cases
- Minimal changes in current code for comparison
- Further code simplification in optimized applications

More Efficient Replication

In Network Caching

Small Scale Test System