

NDN for Scientific Computing

Susmit Shannigrahi, CSU

Chengyu Fan, CSU

Christos Papadopoulos, CSU

Items

- Deadline based data transfers
- NDN-Genomics work with Clemson

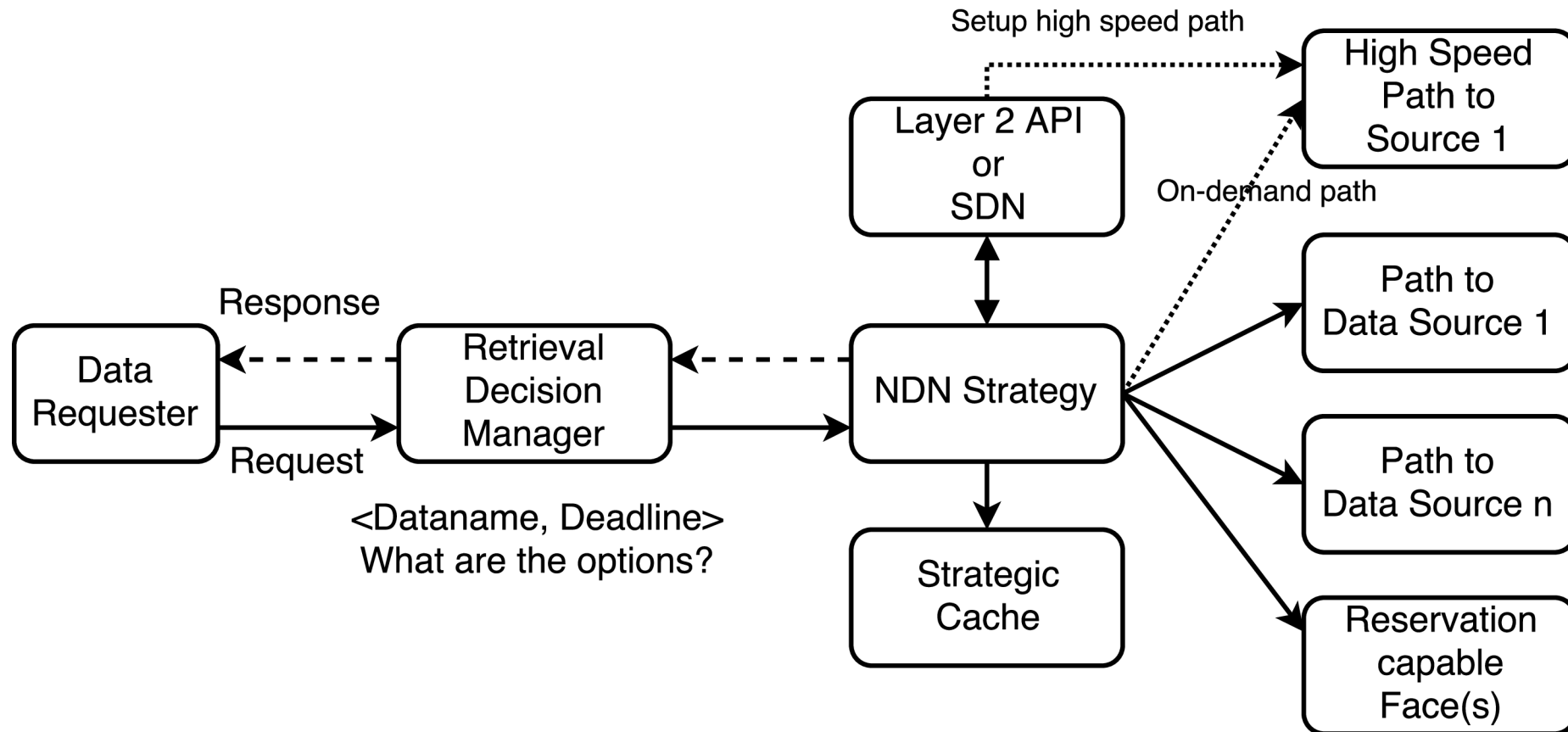
Concept

- Scientific communities uses batch transfers
- Clients submit requests and deadlines to a system
 - NDN network can find out the means
- Aggregate requests, strategically store content in the network

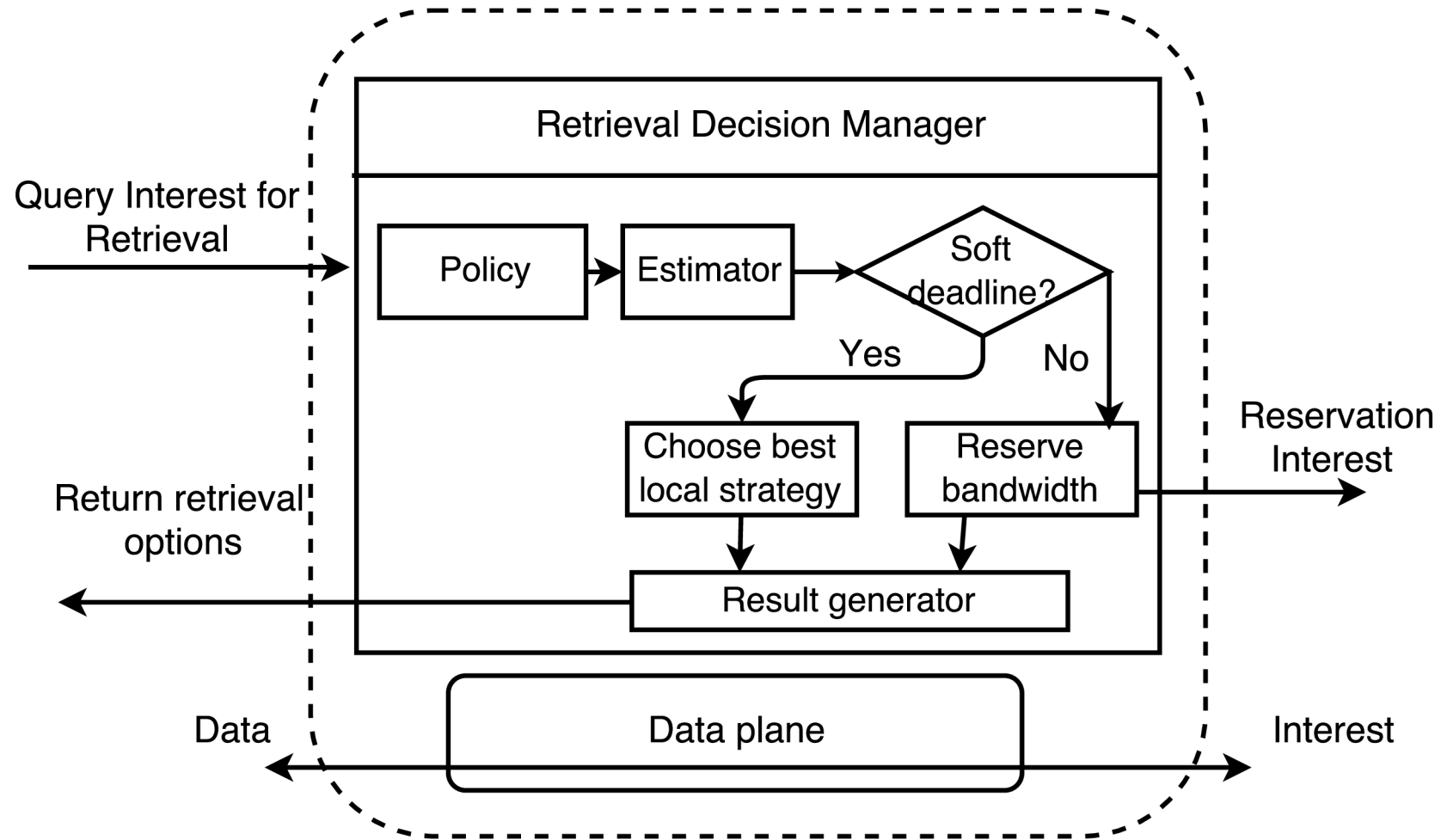
IP vs NDN

- End-to-end reserved channels
- Per-user/per-application flows, no reusability
- NDN can do better
 - Reuse reserved resources for serving requests that falls within a deadline
 - Trade in-network storage for bandwidth
 - Aggregated requests use less resources

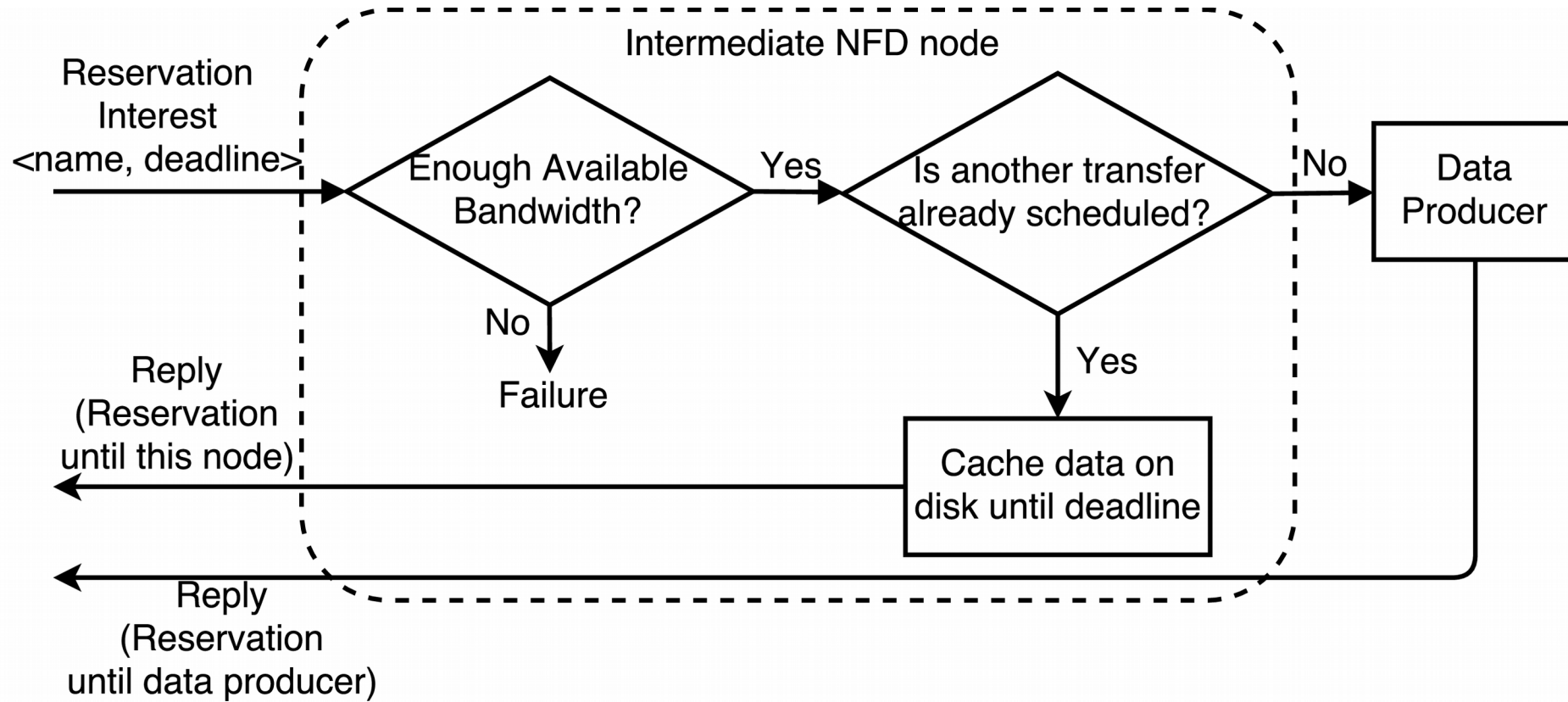
Overview



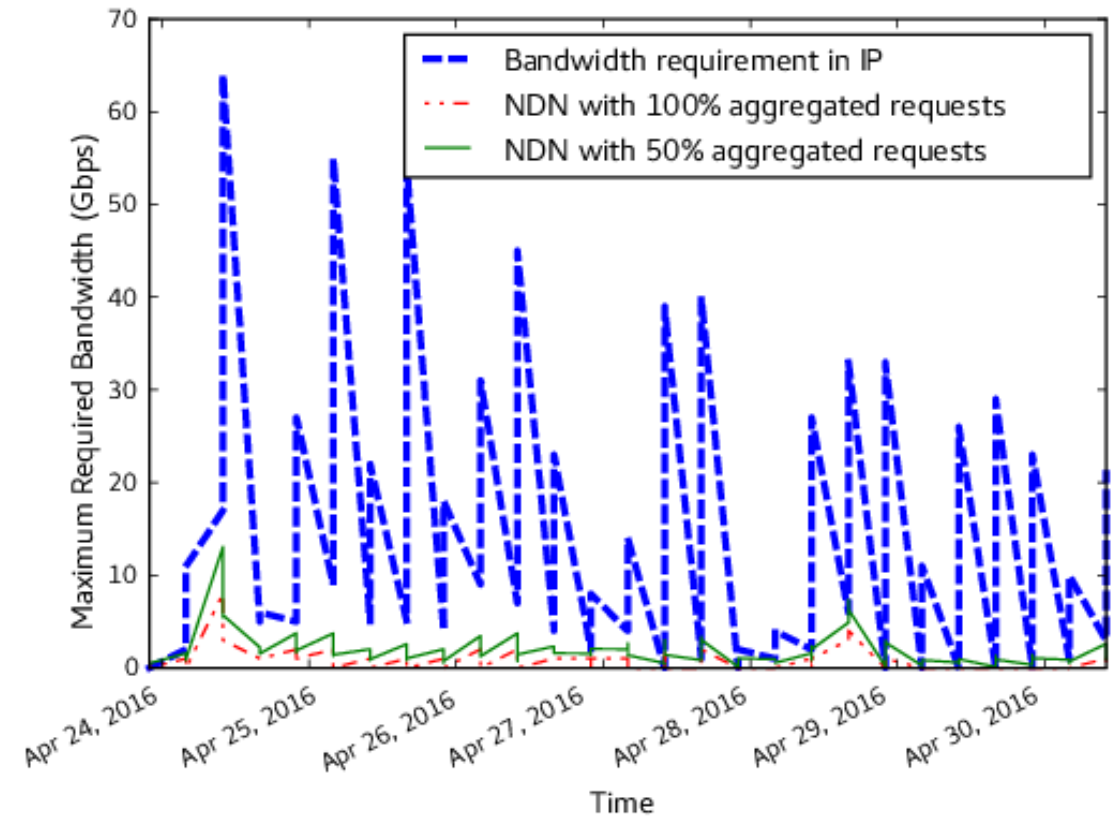
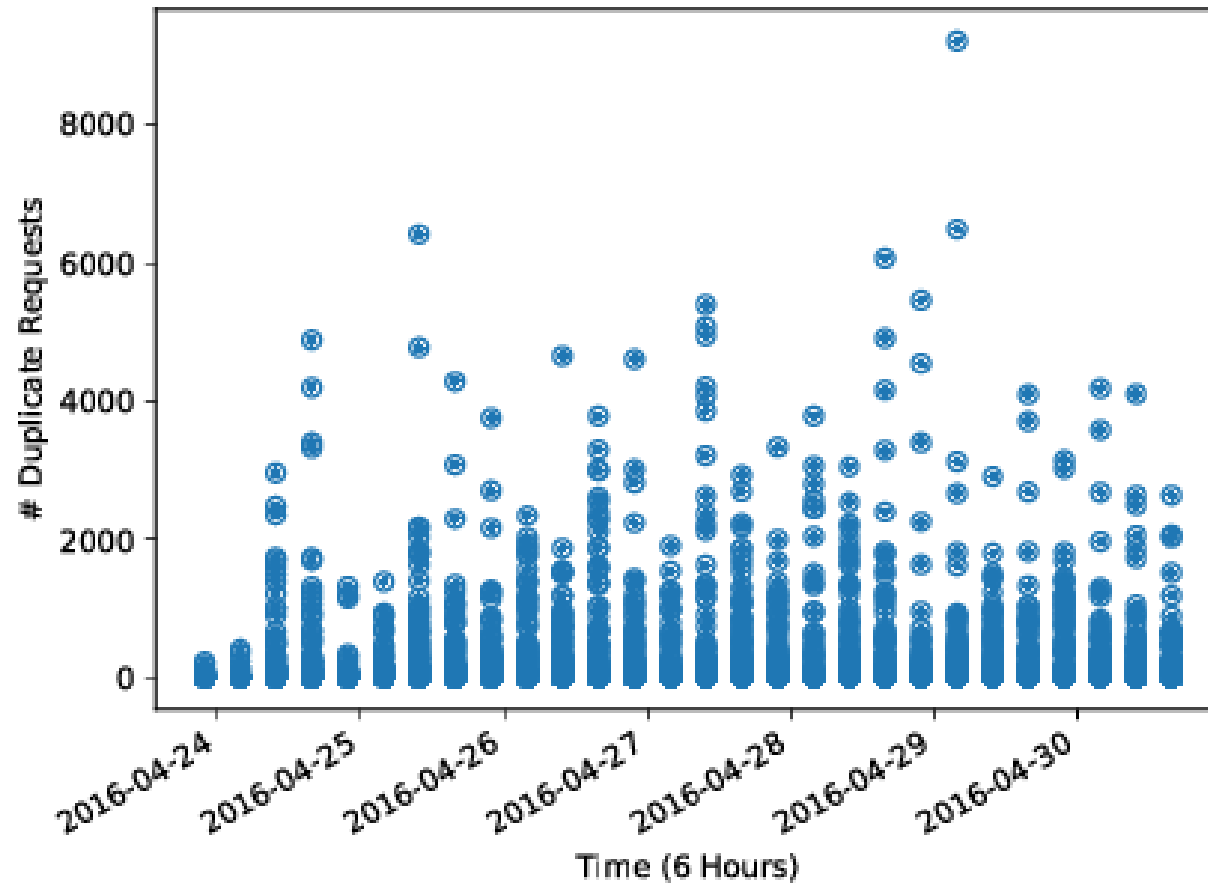
Retrieval Manager



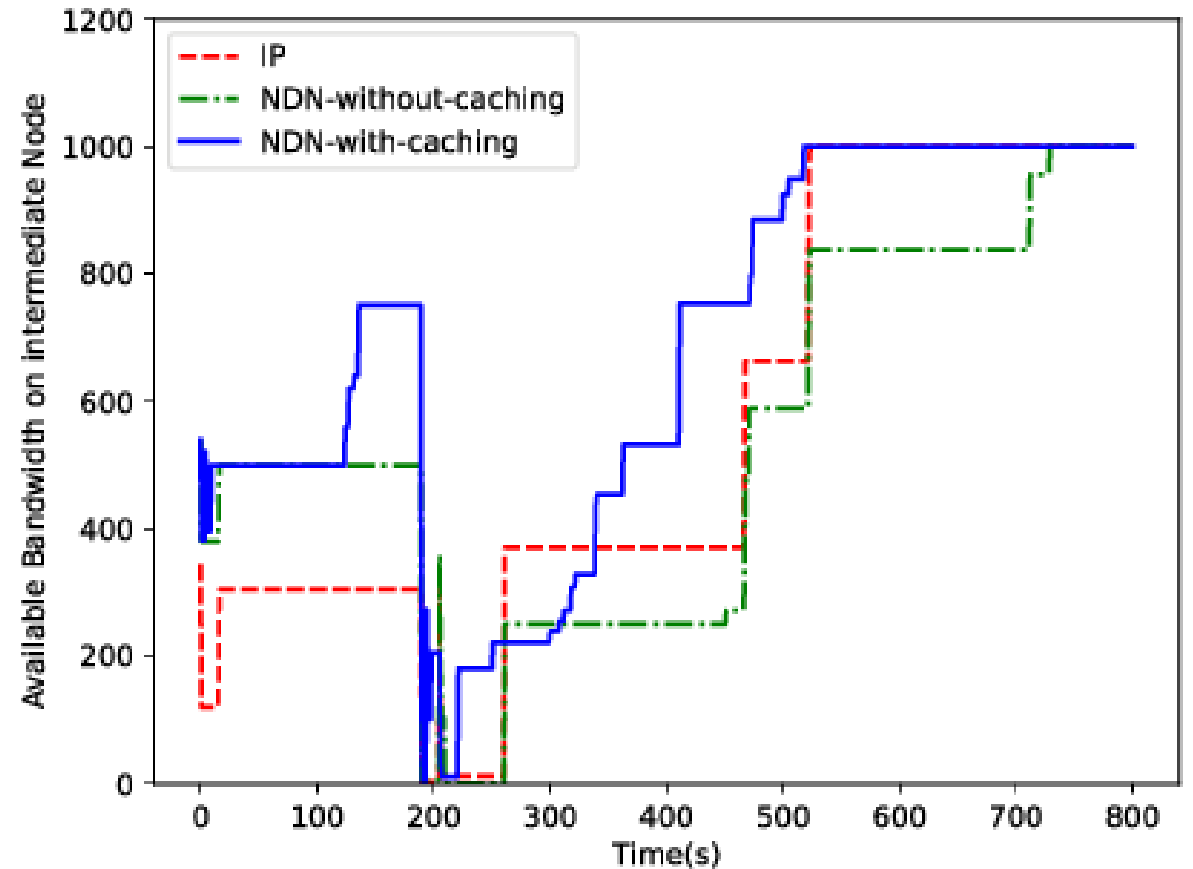
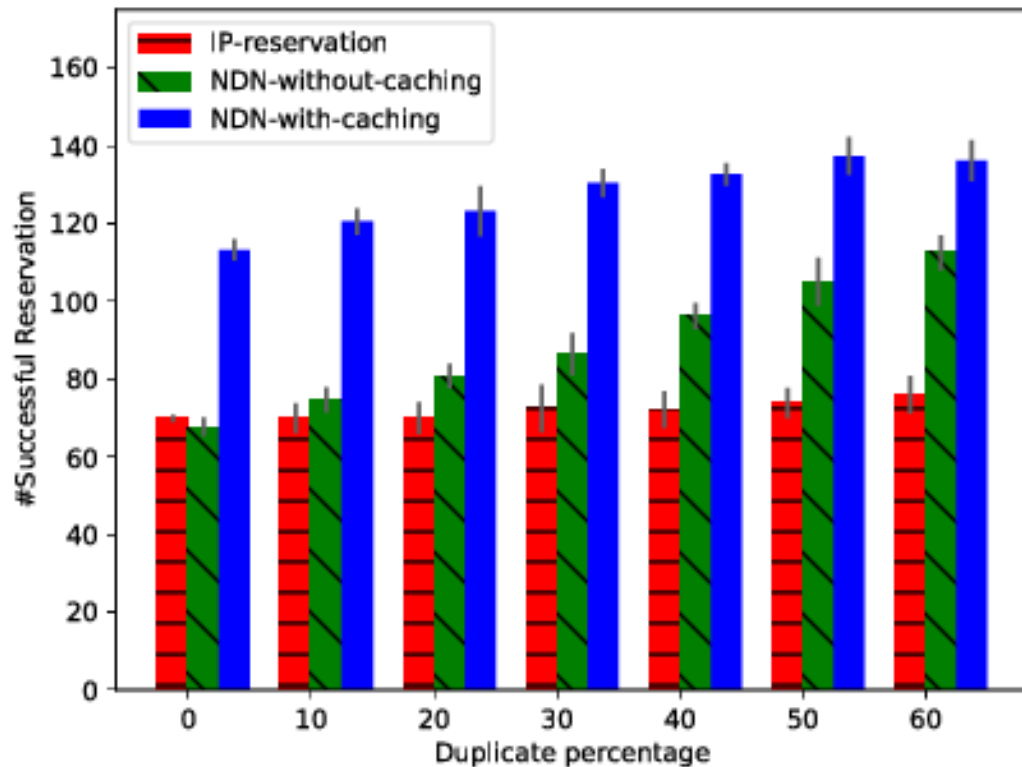
Reservation and Strategic Caching



Analytical evaluation



A reservation protocol (SCARI) and ndnSIM evaluation



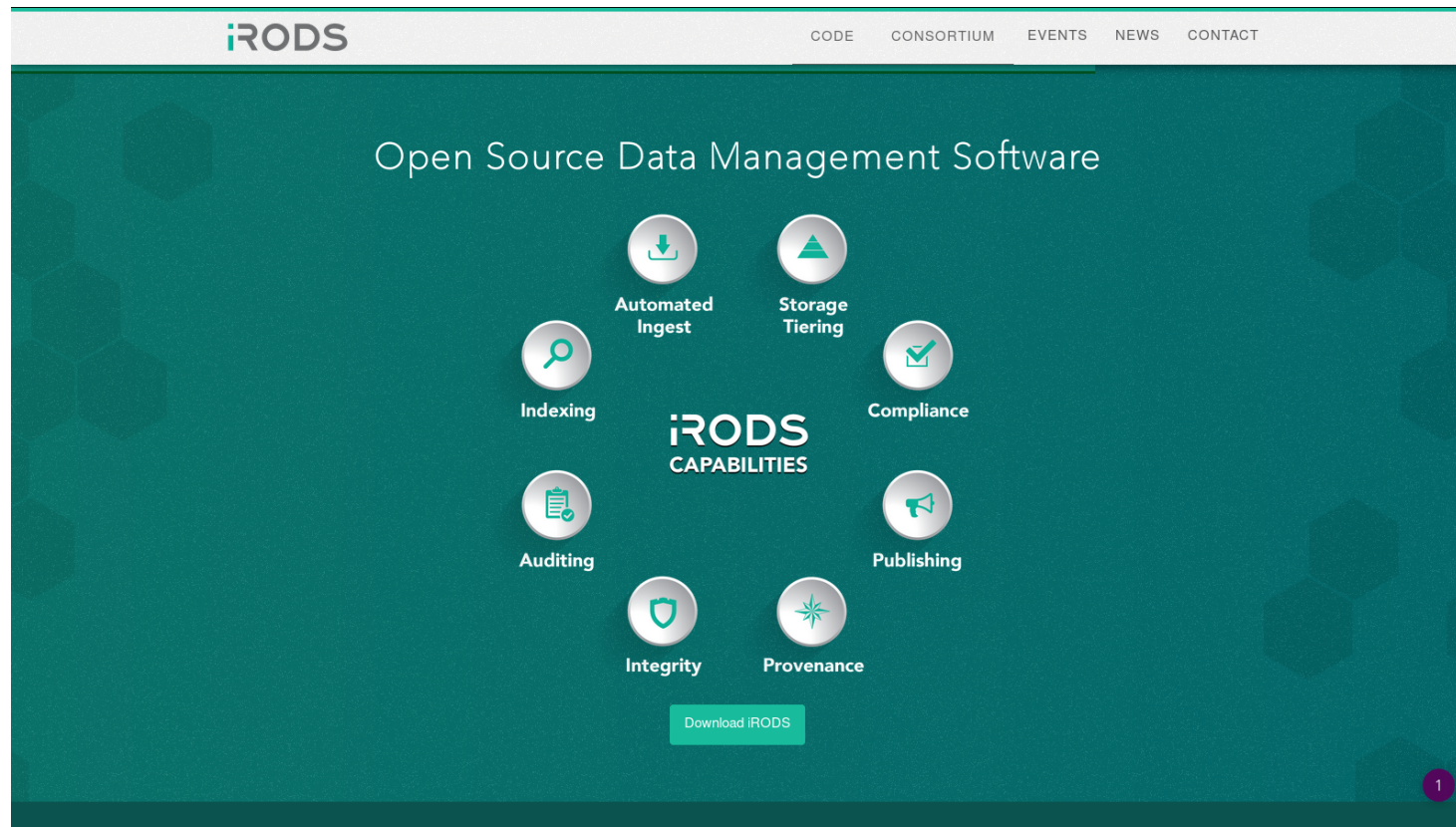
NDN for HEP - SANDIE

- Ran's talk tomorrow

NDN-Genomics pilot

- NSF CC*
- National Cyberinfrastructure for Scientific Data Analysis at Scale (SciDAS)
 - Clemson
 - Washington State University
- Federate access to multiple national various infrastructure
 - Find data
 - design and execute necessary computations on found data

IRODS



Use case

- Large amount of data
 - NCBI in Maryland has 14.4 petabytes of high-throughput DNA sequence data
- Thousands of researchers access the same data
 - Some genomes are popular
- Both static and dynamic data
- Currently hard to find, individual ftp/http servers
 -

How NDN can help

- NDN is the unifying entity among these various infrastructure
 - Name based data discovery
 - Name based data retrieval
 - Name based data and computation staging
 - Caching popular datasets

Progress

- Names are already hierarchical and easily imported into NDN
 - [genus]_[species]{_[infraspecific name]}-[assembly_name].{hisat2 extension}
 - We translated them into NDN names
- Use NDN-SCI for providing publication, search, retrieval and staging
- Use NDN in the background
 - Users request data the same way, requests go through NDN core
 - Provides NDN benefits to iRODS users
 - We should have real data request patterns for NDN research

Questions and Comments?