Building User Friendly Data Transfers Nodes
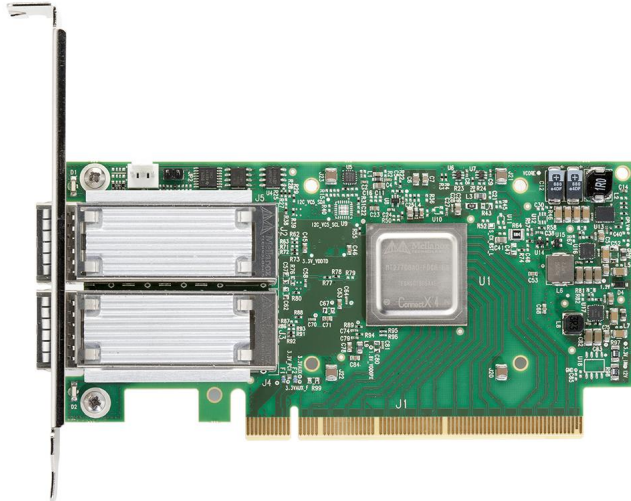
Joseph Hill
3 December 2018

Data Transfer Node (DTN)

- A specialized system to conduct high performance data transfers
- Can be used in a variety of work flows
- Many possible configurations depending on the workflow

User to User Transfer

- Sender and receiver are a long distance apart
- Each has access to a nearby DTN
- DTNs have a high speed link between them
- User driven workflow

```
  Sender ----- DTN ========== DTN ----- Receiver
```
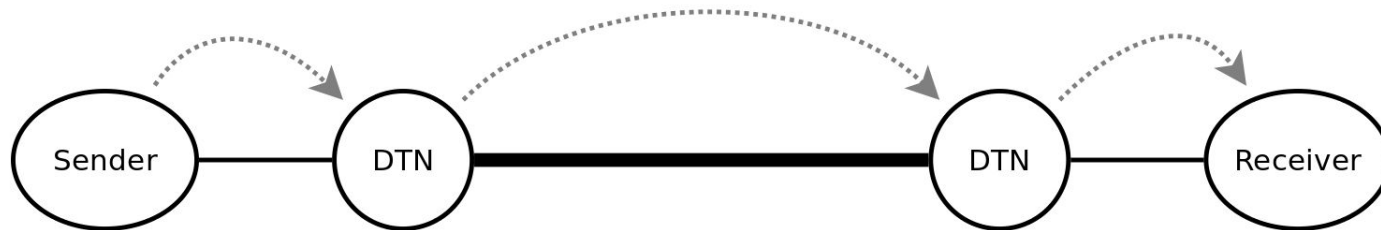
DTN Demonstration

- Workstations not optimized for long distance transfers can benefit by using an optimized path with DTNs

- Scenario shows a file transfer from a workstation in the US to a workstation in Europe

- Compares a file transfer over an optimized path with DTNs versus path over the internet

- Indirect path utilizing DTNs achieves substantially better performance then a file transfer directly between the workstations

- Just one example of how DTNs can act as an interface to a path optimized for a specific data transfer
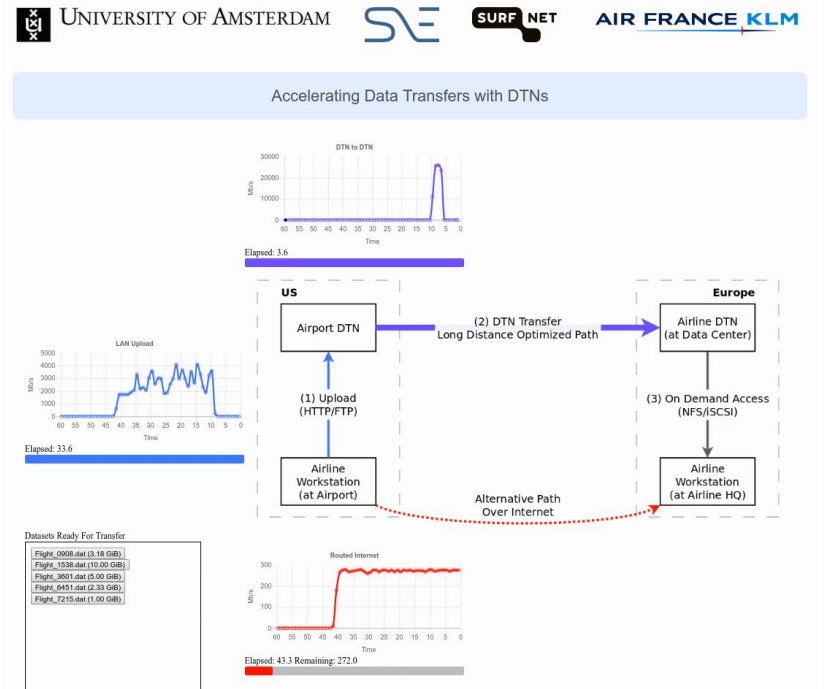
Motivation

- Ideally users could send directly to each other utilizing all available bandwidth
  - Administrative barriers
  - Requires common protocol
- Path may supply greater bandwidth
- High speed link can be faster even with multiple file copy steps
- Goal: Provide a means to use high speed links while minimizing changes for researchers

FileSender

- Sender uploads to the site, recipient downloads from the site
- Uses HTML5 for transfers
  - Multiple TCP streams
  - Supports large files
- Only one user needs an account
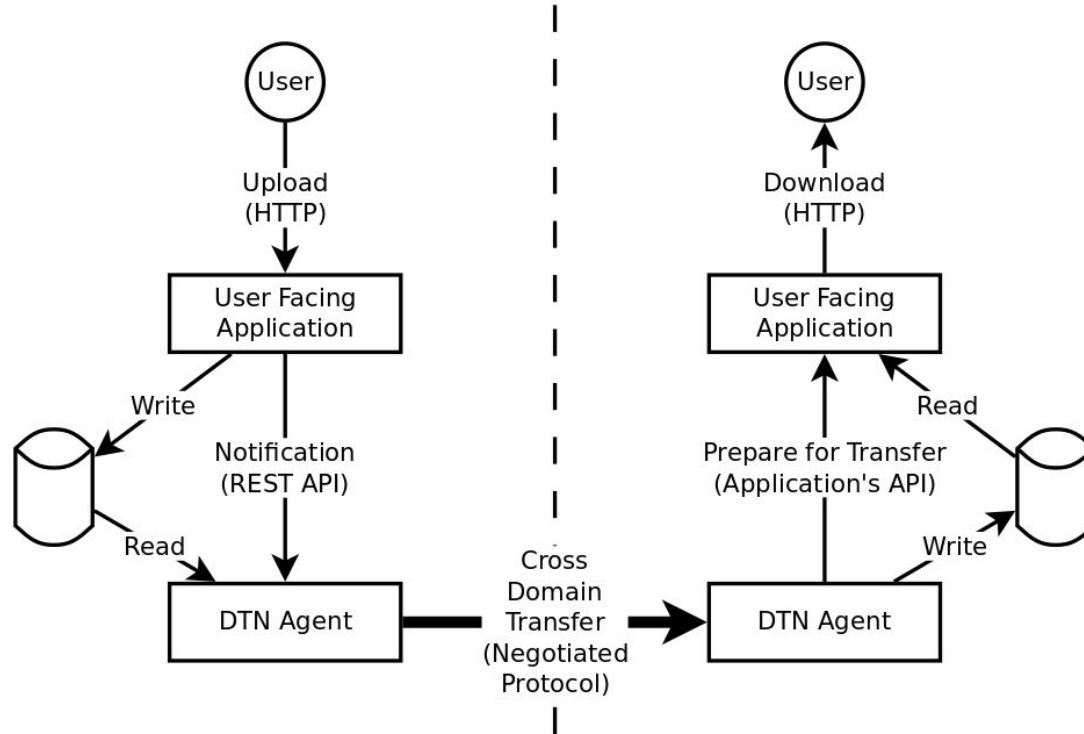 - Only requires email address of other user

ownCloud

- Data lives within the system
- Support for sharing between users
- Support for federation
- Change to user interface required to support DTN transfers
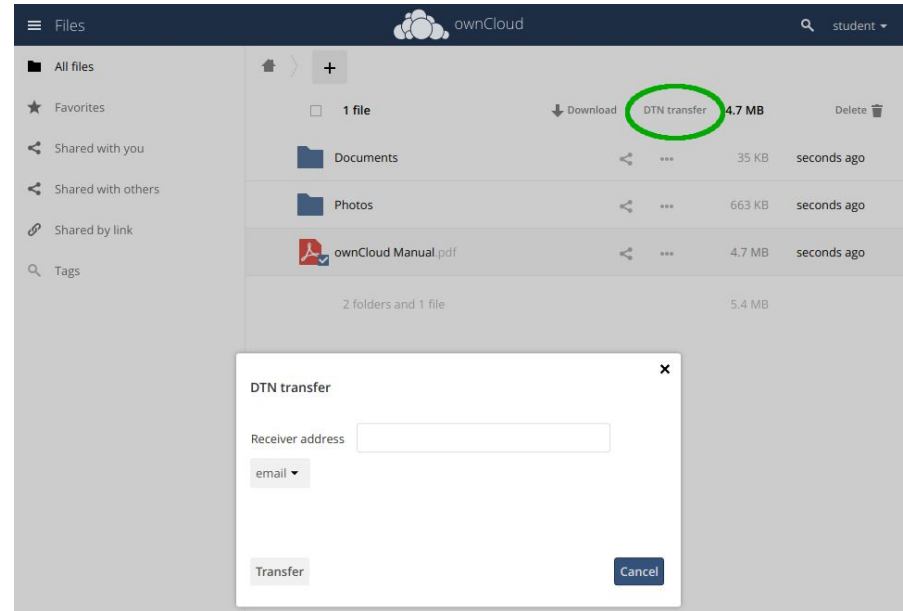
ownCloud changes done by Antoon Prins of SURFsara

DTN Assisted Transfer

DTN Agent

- Provides the user facing application with an interface to a high speed link
- Applications do not need to be the same across domains
- Agent has a module for each application that it supports
- Application's existing API is reused where possible
- Agent handles discovery of the peer DTN

DTN Discovery

- What do we want the sender to have to know about the recipient
- DNS service records can be used similarly to MX records
- Users may want to set a different DTN for themselves
- Alternatively a DTN registry could be setup using LDAP
  - A trusted root would need to be established

```
_dtn._tcp.uva.nl   IN  SRV 0 0 3001 dtn.uva.nl.
_dtn._tcp.surf.nl  IN  SRV 0 0 3001 dtn.surf.nl.
```

Root

uva.nl
10.0.0.1

joe.hill@uva.nl
172.16.0.1

surf.nl
172.16.0.1
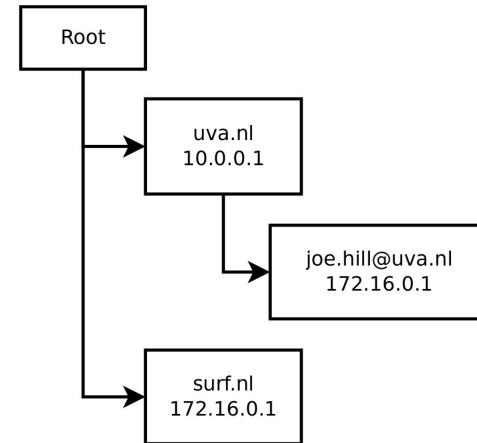
Data Transfer Protocols

- Negotiated by agents
- GridFTP, FDT, mdtmFTP, SCP
- Security typically relies on a known user
- Each transfer application implemented as a module

Future Work

- Use AutoGOLE to dynamically setup high speed links
- Agent to Agent Protocol not finalized, many security questions remain
- Add support for one time tokens to data transfer applications
- Protocol negotiation based on dataset
- Determine a protocol of last resort
- Transfer status and error handling

Questions?