



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science



*Learning with Purpose*

# **PROPER**

## **Parallel Resource-Optimized Provisioning of End-to-End Requests**

### **Principal Investigator:**

Dr. Vinod Vokkarane

UMass Lowell

Vinod\_Vokkarane@uml.edu

### **External Collaborators:**

Alex Sim (LBNL)

Chin Guok (ESnet)

### **Acknowledgement:**

**DoE ASCR Grant DE-SC0012115TDD (June 2014-2017)**



Learning with Purpose

# Introduction

- “Big Data” application requirements necessitate evolution in data distribution paradigm.
  - Faster distribution of data to/from multiple storage centers.
  - Provide survivable backup paths for data transmission.
  - Split transfer of large data sets across multiple sites for quicker transmission .
- There is a discrepancy between availability and invocation of parallel technologies.
  - Data typically stored at multiple sites/repositories (Eg: Climate).
  - Networks can support guaranteed parallel transmissions via virtual circuits (OSCARS).
  - Applications/Schedulers **do not** currently take advantage of these parallelisms due to lack of storage and network resource awareness.



Learning with Purpose

## PROPER Project Objectives

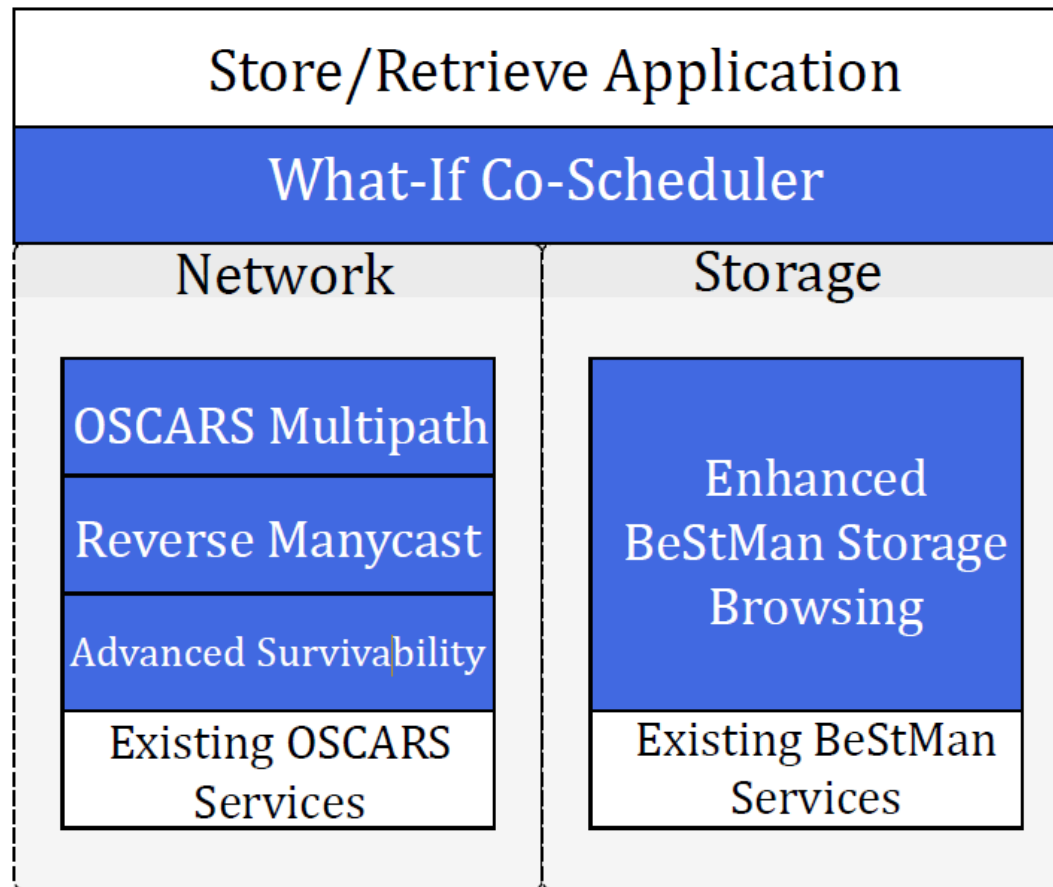
- We propose a co-scheduler that takes advantage of:
  - **Flexibility:** harness both **time** (advance reservation) and **space** (replicated data at multiple sites) domains.
  - **Parallelism:** incorporate **parallel** nature of storage and network resources into to the scheduler.
  - **Survivability:** **protection** of both network and storage resources.
  - **Negotiation:** co-scheduler will return a set of **What-If** scenarios that satisfy the user/application constraints.



Learning with Purpose

## Proposed Work

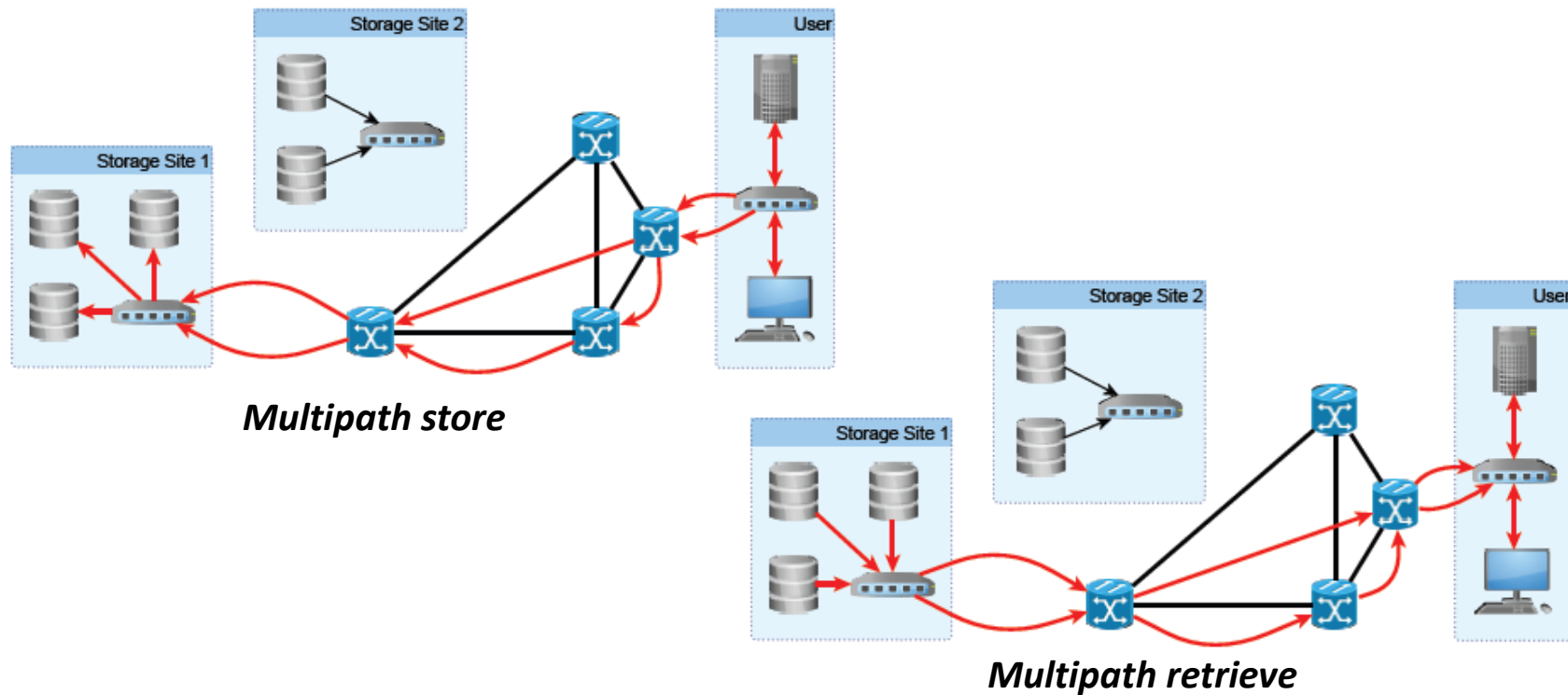
The proposed co-scheduler interfaces with the end-user application to identify areas of flexibility in both network and storage domains.



# Proposed Work

## Task 1: Parallel transfers over multipath circuits.

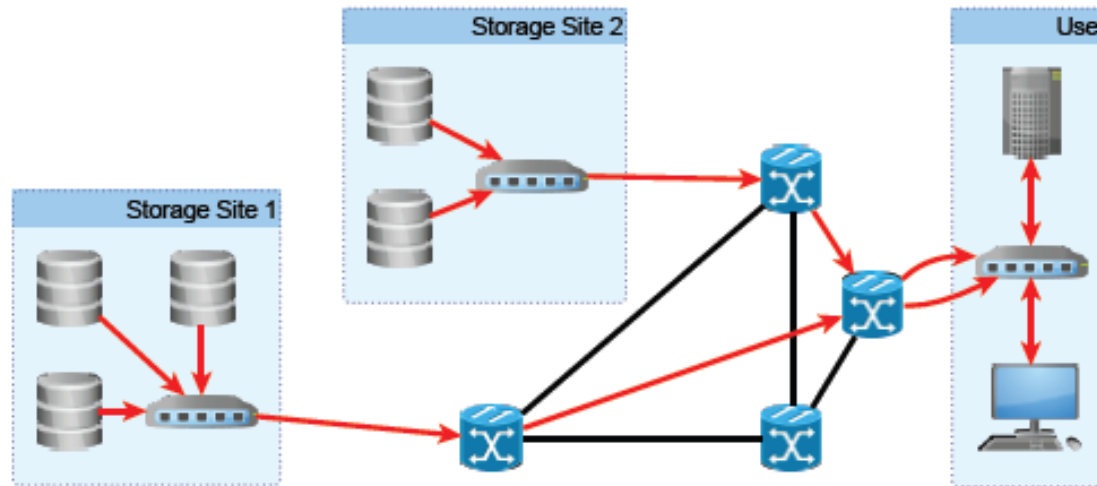
- Facilitate parallel transfer from a single sender to a single receiver.



## Proposed Work

### Task 2: “Reverse Multicast” Data Retrieval (Many to One)

- Transfer multiple chunks of a single data set from multiple repositories to a single user/datacenter.



*Reverse Multicast*

- “**Reverse Anycast**”: Transfer a single copy of replicated data from the **best** storage repository (**any to one**).



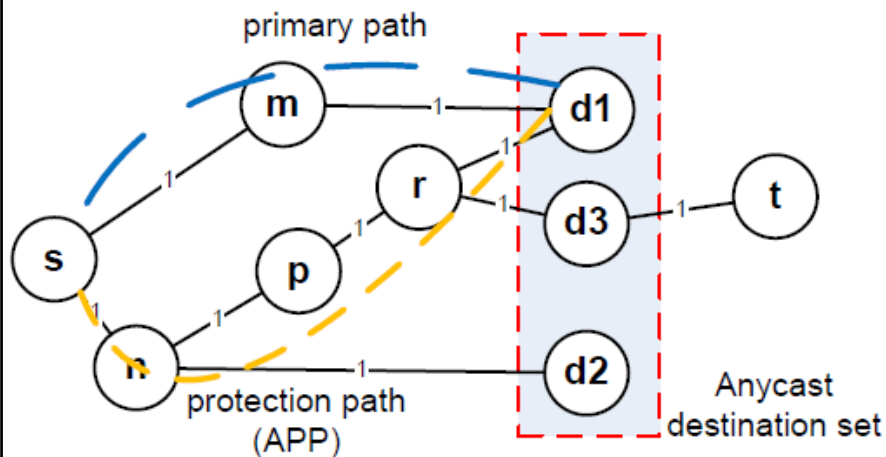
Learning with Purpose

# Proposed Work

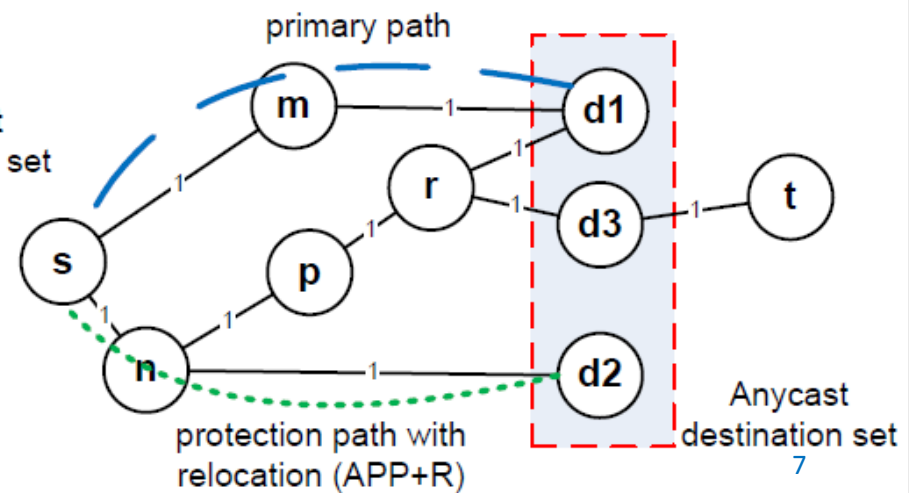
## Task 3: Survivable Anycast

- Resiliency not only against network failure, but also destination node/datacenter failure.

### *Anycast Path Protection*



### *Anycast Path Protection with Destination Relocation*



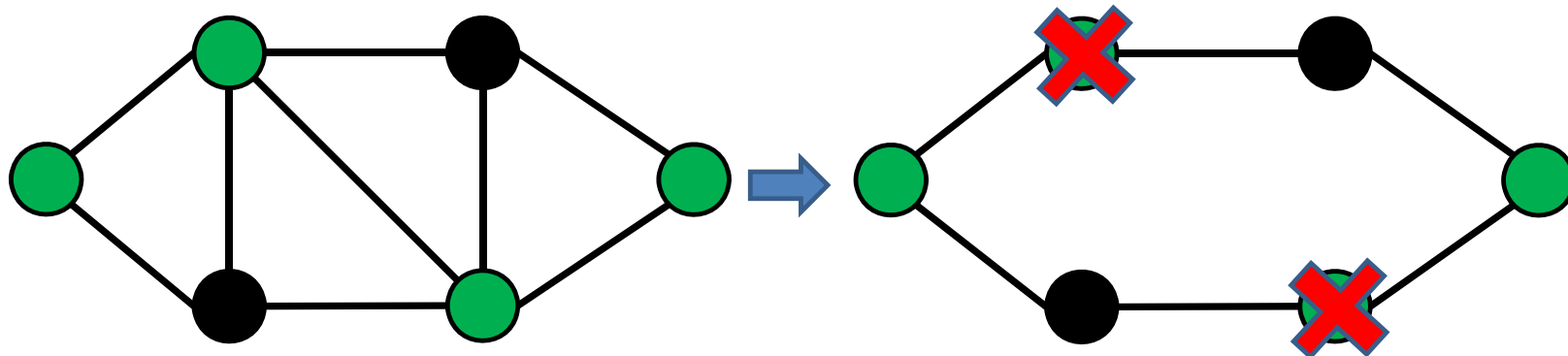


Learning with Purpose

## Proposed Work

### Task 4: Survivable Multicast (1:M and M:1)

- Survivable solutions that protect against node failure.
- Preliminary work: Network biconnectivity.
  - A biconnected graph becomes disconnected if and only if, at least two nodes are removed.
  - Objective: Create biconnected subgraphs that include all resource nodes in a set.



*Manycast Survivability using Network Biconnectivity*  
*Simple example: Ring*



# Proposed Work



Learning with Purpose

## Task 5: Berkeley Storage Manager (BeStMan) Tasks (with Alex Sim)

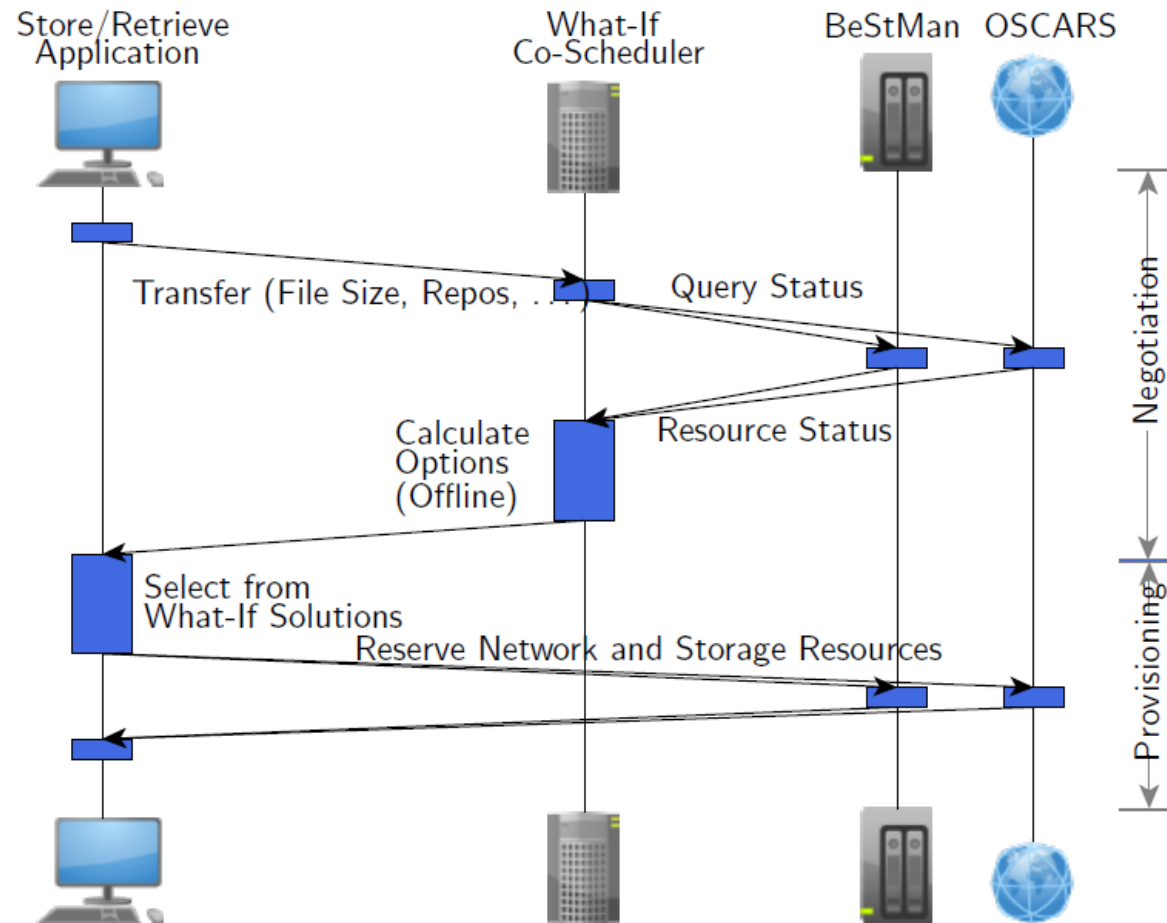
1. Extend the browsing capability of the BeStMan for its managed storage space.
2. Implement dynamic cataloging of file-system contents including database design and caching implementation.
3. Develop queuing algorithms for meta-data queries.



Learning with Purpose

# Proposed Work

## Task 6: What-If Co-scheduler



What question(s) does your research motivate you to now ask?



Learning with Purpose

3 months in, we don't have a lot of insight .....

- How do users discover these new network services?
  - Is What-If module the best place to “hide” new services like
    - anycast, anycast, multi-path, survivability?
- How do we implement this on a multi-domain network and storage?
  - ESnet now supporting NSI for multi-domain transfers.



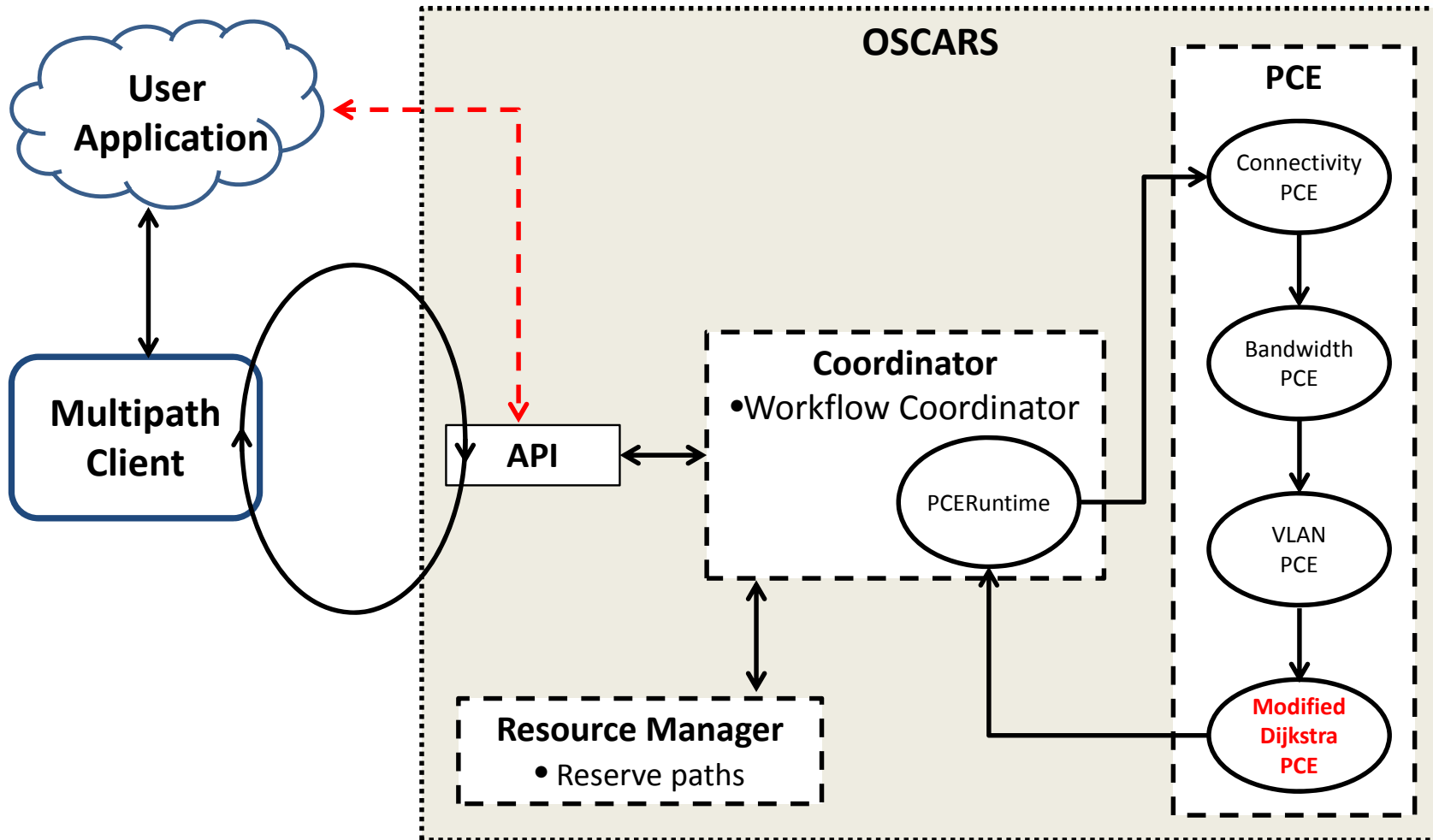
# Multipath Client for OSCARS

Demo



Learning with Purpose

# Multipath Client Design





Learning with Purpose

# Sample Multipath GUI – New Unicast

MultipathUI

**New Reservation**

Source Node: CHIC : port-5 : lin...  
 Destination Node: PNWG : port-3 : link1  
 PNWG : port-4 : link1  
 SDSC : port-1 : link1  
 STAR : port-1 : link1  
 STAR : port-2 : link1  
 STAR : port-3 : link1  
 STAR : port-4 : link1

Bandwidth (Mbps): 100  
 Number of Paths: 1

Reservation Start Time: 01 / 01 / 2014 05 : 30  
 Reservation End Time: 01 / 02 / 2014 07 : 59

**Multipath GRI**

**Unicast GRIs**  
 Show all  
 es.net-1  
 es.net-2  
 es.net-3  
 es.net-4  
 es.net-5  
 es.net-6

[ es.net-6 ]  
 RESERVED  
 Bandwidth: 100 Mbps  
 Start Time: Wed Jan 01 05:30:00 EST 2014  
 End Time: Thu Jan 02 07:59:00 EST 2014  
 Hops in Path: - CHIC --> STAR --> PNWG --> SUNN --> SDSC

UMass Lowell – ECE – Vokkarane



Learning with Purpose

# Sample Multipath GUI – New Multipath

MultipathUI

**New Reservation**

Source Node: SUNN : port-1 : li...

Destination Node: CHIC : port-1 : link1

Bandwidth (Mbps):

Number of Paths:

Reservation Start Time:  /  /   :

Reservation End Time:  /  /   :

**Multipath GRI**

MP-1

**Grouped GRIs**

Show all

es.net-9  
es.net-10  
es.net-11

---

< Multipath Group: MP-1 >

[ es.net-9 ]  
RESERVED  
Bandwidth: 100 Mbps  
Start Time: Wed Jan 01 05:30:00 EST 2014  
End Time: Sun Jan 05 11:59:00 EST 2014  
Hops in Path: - SUNN --> PNWG --> STAR --> CHIC

[ es.net-10 ]  
RESERVED

UMass Lowell - ECE - Vokkarane



Learning with Purpose

# Sample Multipath GUI – MP Group Cancel

MultipathUI

**New Reservation**

Source Node: WASH : port-1 : L...

Destination Node: ATLA : port-3 : link1  
 ATLA : port-4 : link1  
 ATLA : port-5 : link1  
 CHIC : port-1 : link1  
 CHIC : port-2 : link1  
 CHIC : port-3 : link1  
 CHIC : port-4 : link1

Bandwidth (Mbps): 25

Number of Paths: 2

Reservation Start Time: 01 / 01 / 2014 05

Reservation End Time: 01 / 05 / 2014 11

Create Reservation    Reset Fields

**Multipath GRI**

MP-1  
 MP-2  
 MP-4  
 MP-5

**Grouped GRIs**

Show all

es.net-19  
 es.net-20

to Group    Sub from Group

**Confirm Cancel?**

Are you sure you want to cancel Group 'MP-5'?

Cancel Group    Never Mind

< Multipath Group: MP-5 >  
 [ es.net-19 ]  
 RESERVED  
 Bandwidth: 25 Mbps  
 Start Time: Wed Jan 01 05:30:00 EST 2014  
 End Time: Sun Jan 05 11:59:00 EST 2014  
 Hops in Path: - WASH --> ATLA  
 [ es.net-20 ]

UMass Lowell - ECE - Vokkarane





Learning with Purpose

# Sample Multipath GUI – MP Group Cancel

MultipathUI

**New Reservation**

Source Node: WASH : port-1 : l...  
 Destination Node: ATLA : port-3 : link1  
 ATLA : port-4 : link1  
 ATLA : port-5 : link1  
 CHIC : port-1 : link1  
 CHIC : port-2 : link1  
 CHIC : port-3 : link1  
 CHIC : port-4 : link1

Bandwidth (Mbps): 25

Number of Paths: 2

Reservation Start Time: 01 / 01 / 2014 05 : 30

Reservation End Time: 01 / 05 / 2014 11 : 59

Create Reservation    Reset Fields

**Multipath GRI**

MP-1  
MP-2  
MP-4  
MP-5

**Grouped GRIs**

Show all

es.net-19  
es.net-20

Cancel Reservation    Add to Group    Sub from Group

---

< Multipath Group: MP-5 >  
 [ es.net-19 ]  
 CANCELLED  
 Bandwidth: 25 Mbps  
 Start Time: Wed Jan 01 05:30:00 EST 2014  
 End Time: Sun Jan 05 11:59:00 EST 2014  
 Hops in Path: - WASH --> ATLA  
 [ es.net-20 ]

UMass Lowell - ECE - Vokkarane



# Questions?