U.S. DEPARTMENT OF ENERGY | Office of Science

UMass Dartmouth
COLLEGE OF ENGINEERING

COMMON:
Coordinated Multi-layer Multi-domain Optical Network Framework for Large-scale Science Applications (2010-2013)

PI: Dr. Vinod Vokkarane
Associate Professor, University of Massachusetts at Dartmouth
(Sabbatical: Visiting Scientist, MIT)
Contact: vvokkarane@ieee.org
Project Website: http://www.cis.umassd.edu/~vvokkarane/common/

Annual PI meeting for the ASCR Network & Middleware
Supported by DOE ASCR under grant DE-SC0004909

March 1-2, 2012

COMMON Project Team

UMass Team

- Vinod Vokkarane (PI)
- Arush Gadkar (Post-Doc)
- Joan Triay (Visiting Scholar)
- Bharath Ramaprasad (MS)
- Mark Boddie (BS-MS)
- Tim Entel (BS-MS)
- Jeremy Plante (Ph.D.)
- Thilo Schoendienst (Ph.D.)

ESnet Team

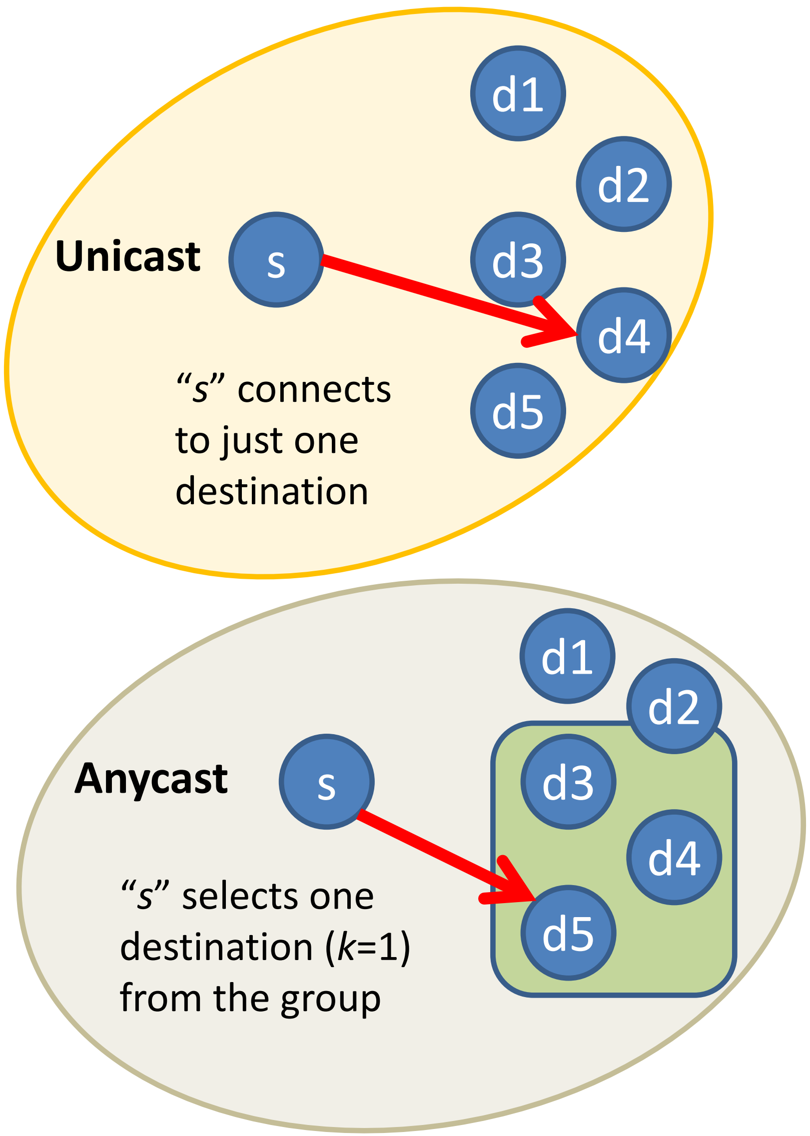
- Chin Guok
- Andrew Lake
- Eric Pouyoul
- Brian Tierney

Outline

- Introduction and project objectives
- Year 1 Project Objectives:
  - Anycast Multi-domain Service
  - Multicast-Overlay Algorithms
- Year 2 and 3 Project Objectives
  - Multi/Manycast-Overlay Deployment
  - Survivable Connections
  - QoS Support (with ARCHSTONE project)

Point-to-Point Communication Services: Unicast, Anycast

- Unicast Vs Anycast



- Unicast request: (*s*,*d*)
- Anycast request: (*s*,*{D}*) where *s* is the source and the {*D*} is the set of candidate destinations.
- Anycast: the source communicates with **any one** node from the set of candidate destinations.

Example:

Unicast: (S, d_4)

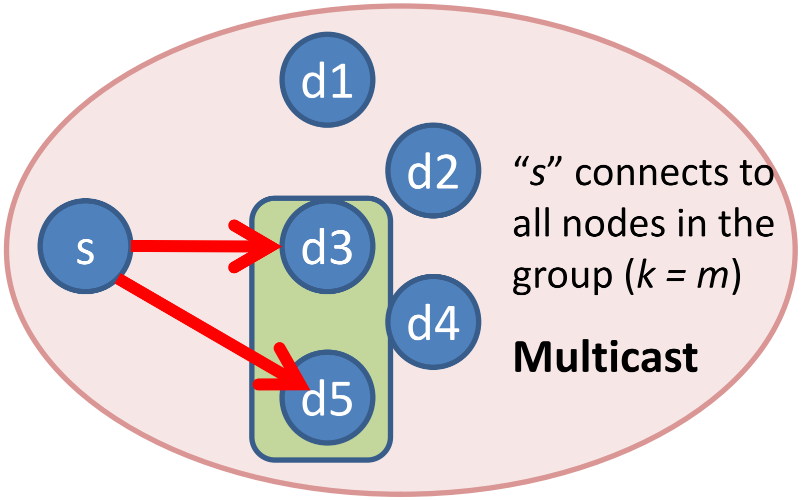
2|1 Anycast: (S,{d_3,d_4})

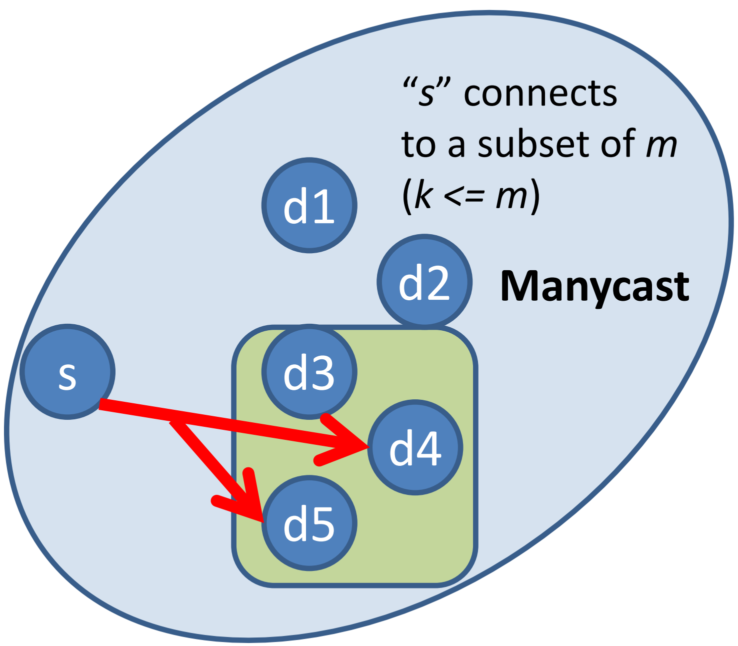
3|1 Anycast: (S,{d_3,d_4,d_5})

Note: other common request parameters such as, request duration, start time, end time, bandwidth requested are not shown on the slide.

P2MP Communication Services: Multicast/Manycast

- Multicast Vs Manycast





- Multicast request: *(s,{D})*, where s is the source node and *D* is the set of destination nodes (d_1, d_2, …,d_m}.

- In Multicast, source node communicates with **each destination** node in {*D*}.

- Manycast request: (*s,{D},k*), where *s* is the source node and the {*D*} is the set of candidate destination nodes.

- In Manycast, source node communicates with ***any k*** nodes in {*D*}.

Example:

Multicast: (S,{d_3,d_5})

Manycast: (S,{d_3,d_4,d_5},2)

(Note: other common input parameters omitted)

COMMON Project Objectives

- Design and implement new services, such as anycast, multicast, manycast, survivability, and QoS across multiple domains and multiple layers.

- Year 1:
  - Design and Deploy Anycast service on OSCARS.
  - Develop Multi/Manycast Overlay models.
- Year 2:
  - Deploy Multi/Manycast Overlay models on OSCARS.
  - Design and Deploy survivability techniques on OSCARS.
  - Design QoS mechanisms to support scientific applications on multi-domain networks.
- Year 3:
  - Extend the survivability and QoS mechanisms to multi-layer multi-domain scenarios and deploy them on OSCARS.

Year 1: Deployment of Anycast Service on OSCARS

Objectives

- Design and implement a production-ready anycast service extension to existing OSCARS framework.
- Improve connection acceptance probability and user experience for anycast-aware services.

Impact

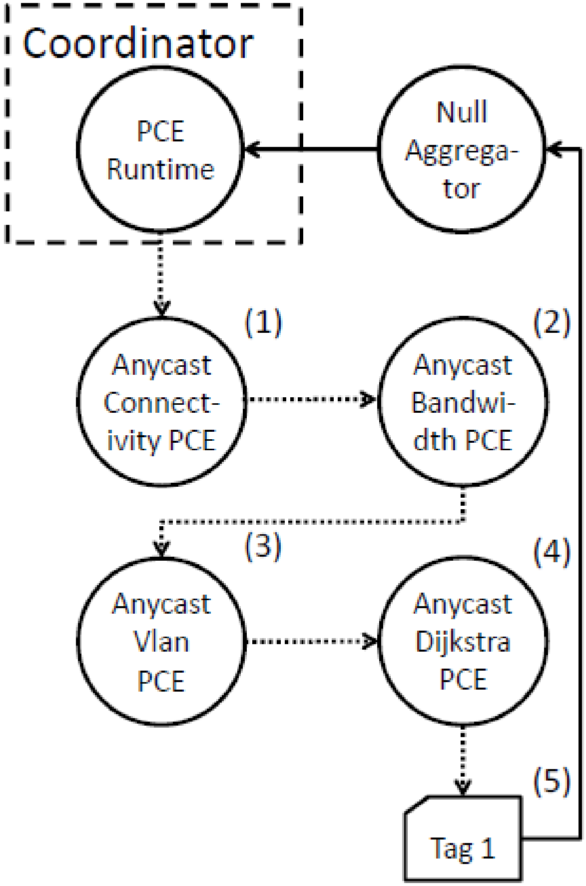
- Provide scientific community with ability to:
  - (a) Allow for destination-agnostic service hosting on large-scale networks.
  - (b) Increase service acceptance.

Design & Implementation (Complete)

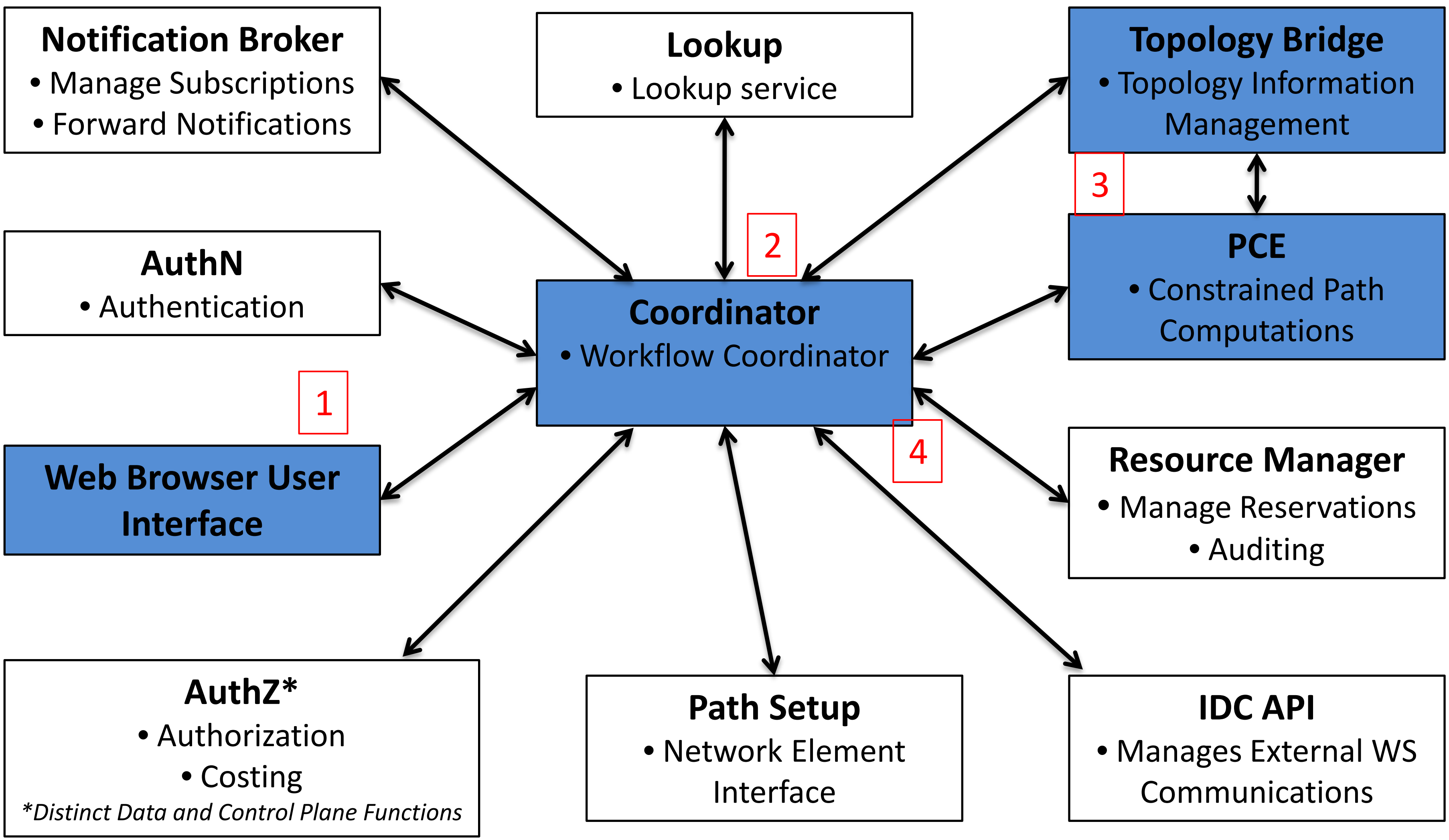
- Designed anycast service as a PCE extension.
- Implementation of the PCE modules to find anycast connectivity, remove the unavailable resources, and select the best possible destination.
- Successfully completed Stress, Regression, and Integration testing of the anycast modules on OSCARS 0.6 (Q4, 2011).
- Hot deployment ready (PnP capable) anycast version of OSCARS 0.6 available at:

https://oscars.es.net/repos/oscars/branches/common-anycast/

- Year 2: Plan to work with ESG group to attach this service to a specific application.
- Looking forward to work with other application groups.

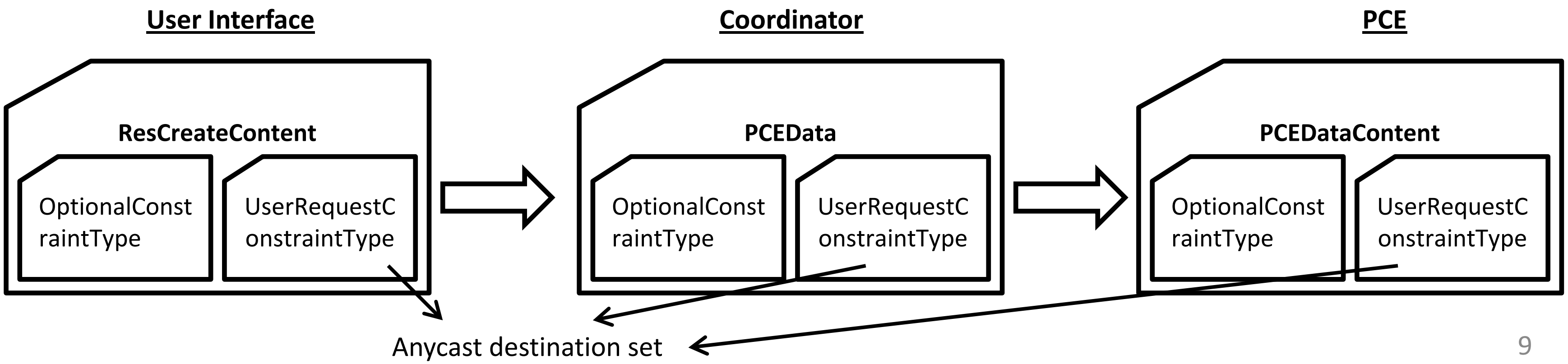


OSCARS Anycast Design

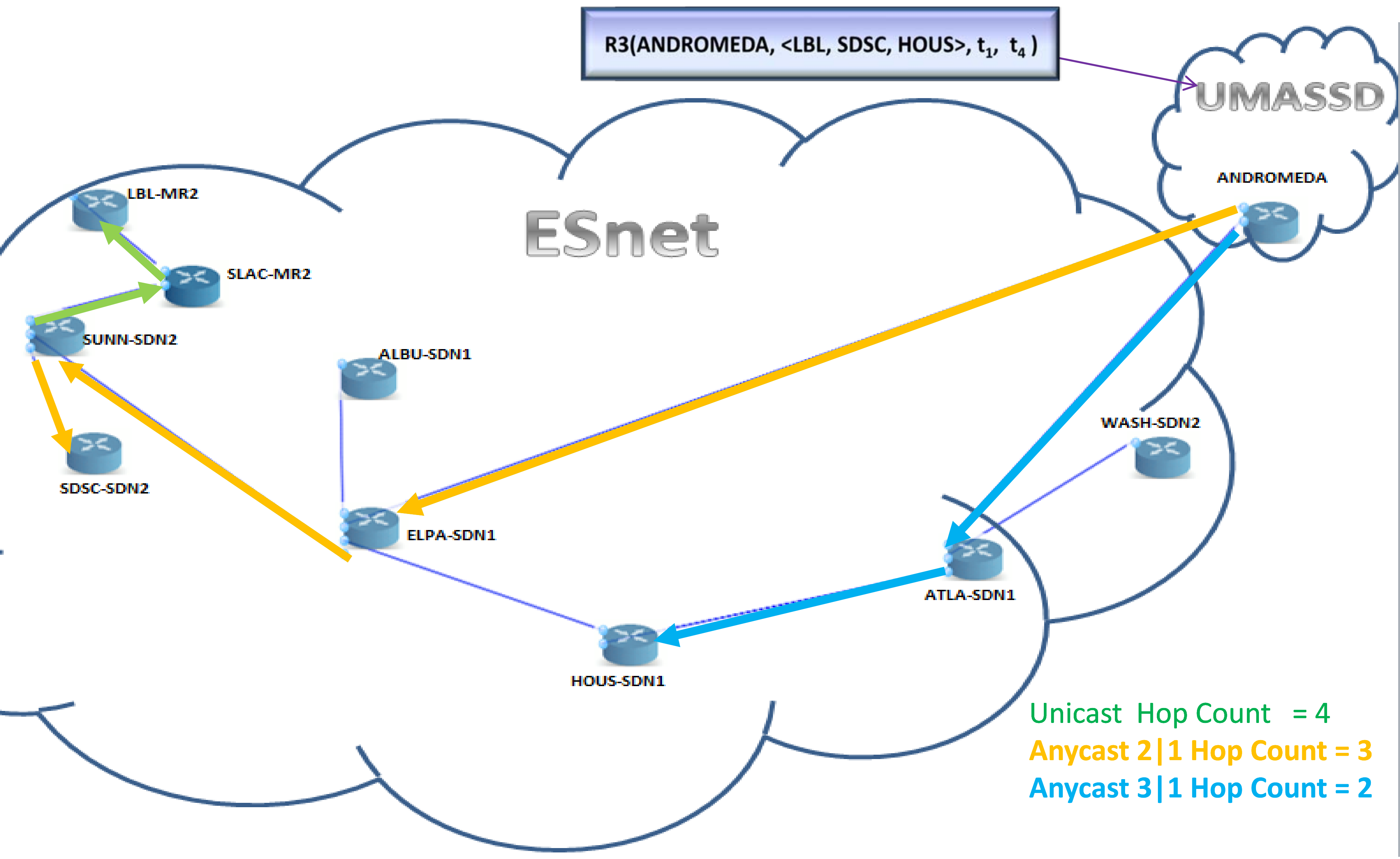


OSCARS Anycast Design

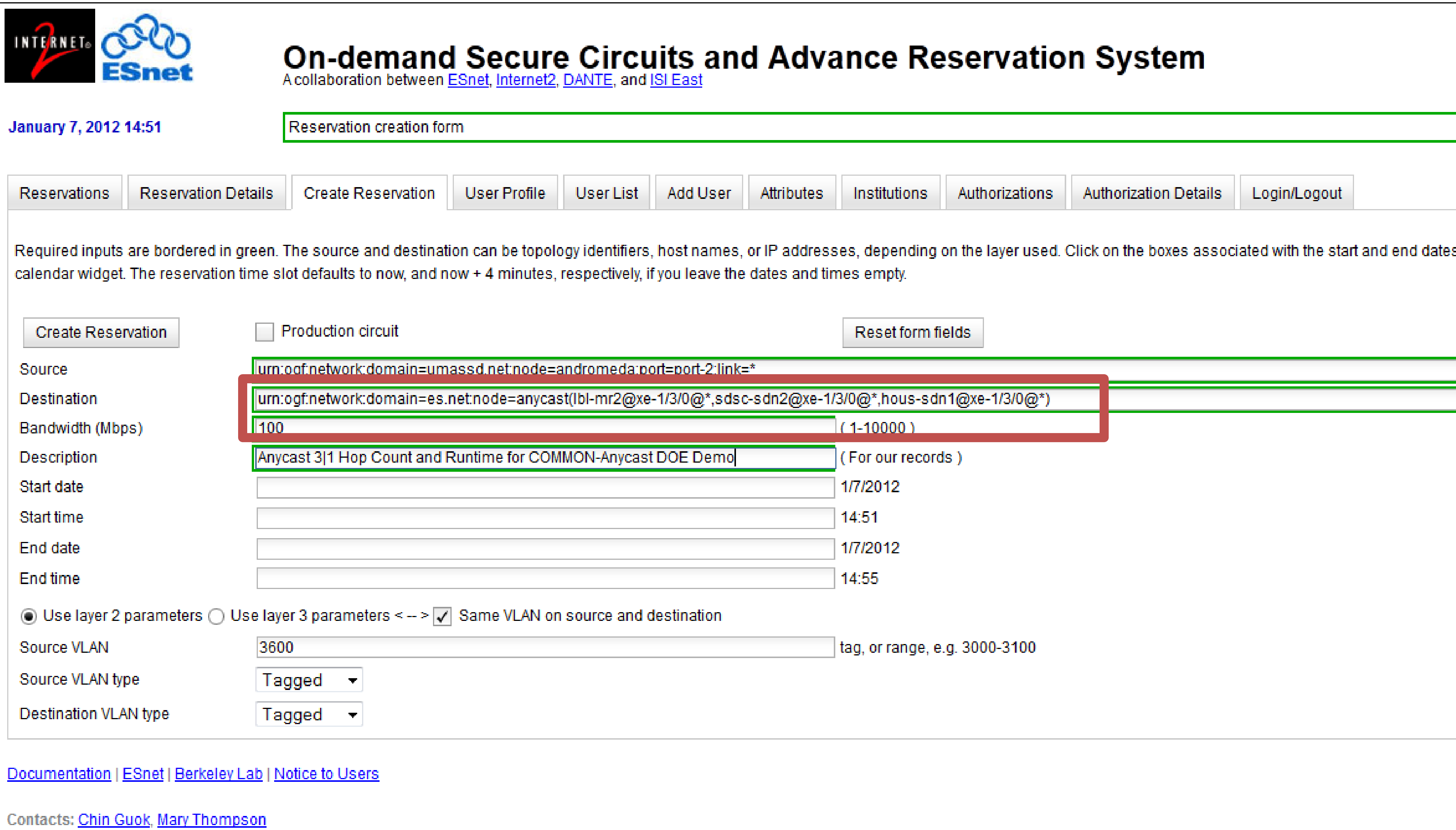
- 1 The user interface servlets would process the anycast request as a unicast request with a big exception: the destination field will be a list of destination nodes (the anycast destination set).
  - An option is to encapsulate the anycast data as an **OptionalConstraintType**, in addition to the rest of parameters mapped into a UserRequestConstraintType. Both, UserRequestConstraintType and the OptionalConstraintType will be part of the ResCreateContent.
  - The ResCreateContent will be passed to the Coordinator to further process the anycast request.
- 2 The Coordinator, through the CreateReservationRequest, will get the ResCreateContent and map the user request constraints and optional constraints into a PCEData object.
  - The PCERuntime will handle the query process to the PCE.
- 3 The PCE (using the design proposed in the following slides) will make use of the OptionalConstraintType (which carries the list of destinations).
- 4 The result of the PCE will be the path from the source node to a single destination node, so, from the path reservation and PSS modules standpoint, the rest of the flowchart will work as a unicast request.



Multi-Domain Anycast Demo



Anycast 3|1 Illustration



Anycast 3|1 Illustration

January 7, 2012 15:00

Reservation details for umassd.net-154

Reservations | Reservation Details | Create Reservation | Login/Logout

NEW GRI [ ] QUERY

REFRESH | MODIFY | CANCEL | CLONE | CREATE PATH | TEAR DOWN PATH | OVERRIDE STATUS

GRI	umassd.net-154
Status	FINISHED
User	umassd.net
Description	Anycast 3|1 Hop Count and Runtime for COMMON-Anycast DOE Demo
Start date	1/7/2012
Start time	14:52
End date	1/7/2012
End time	14:56
Created time	2012/01/07 14:52
Bandwidth (Mbps)	100
Source	urn:ogf:network:domain=umassd.net:node=andromeda:port=port-2:link=*
Destination	urn:ogf:network:domain=es.net:node=anycast(lbl-mr2@xe-1/3/0@*,sdsc-sdn2@xe-1/3/0@*,hous-sdn1@xe-1/3/0@*)

	VLAN	Hop
Local path	3600	urn:ogf:network:domain=es.net:node=atla-sdn1:port=xe-1/3/0:link=*
	0,2-4094	urn:ogf:network:domain=es.net:node=atla-sdn1:port=xe-0/1/0:link=xe-0/1/0.0
	0,2-4094	urn:ogf:network:domain=es.net:node=hous-sdn1:port=xe-7/0/0:link=xe-7/0/0.0
	3600	urn:ogf:network:domain=es.net:node=hous-sdn1:port=xe-1/3/0:link=*
Interdomain path	3600	urn:ogf:network:domain=umassd.net:node=andromeda:port=port-2:link=*
	3600	urn:ogf:network:domain=umassd.net:node=andromeda:port=port-3:link=*
	3600	urn:ogf:network:domain=es.net:node=atla-sdn1:port=xe-1/3/0:link=*
	3600	urn:ogf:network:domain=es.net:node=hous-sdn1:port=xe-1/3/0:link=*

Source VLAN	3600
Tagged	true
Destination VLAN	3600
Tagged	true

Documentation | ESnet | Berkeley Lab | Notice to Users

Contacts: Chin Guok, Evangelos Chaniotakis Andy Lake Eric Pouyoul

Benefits of Anycast over Unicast OSCARS on live deployment

In summary, from the demo we observed the following:

1. Anycast as a communication paradigm for OSCARS eliminates or reduces blocking significantly when compared to using unicast.

2. Anycast as a communication paradigm for OSCARS significantly reduces average Hop Counts required to establish circuits when compared to unicast, thereby reducing network signaling considerably as well as utilizing fewer network resources .

3. Provisioning time (run-time complexity) for Anycast M|1 for $2 \leq M \leq 4$ is comparable to that of Unicast as there is only a cumulative 2 second increase in provisioning time for an unit increase in cardinality of the Anycast set when compared to unicast .

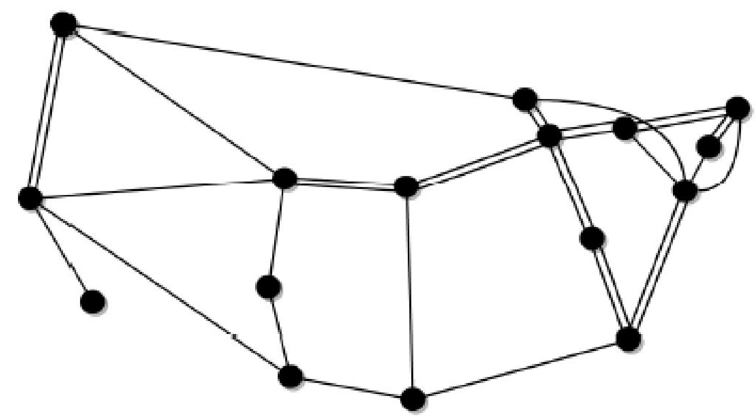
Performance of Anycast Service for OSCARS

Results for single domain

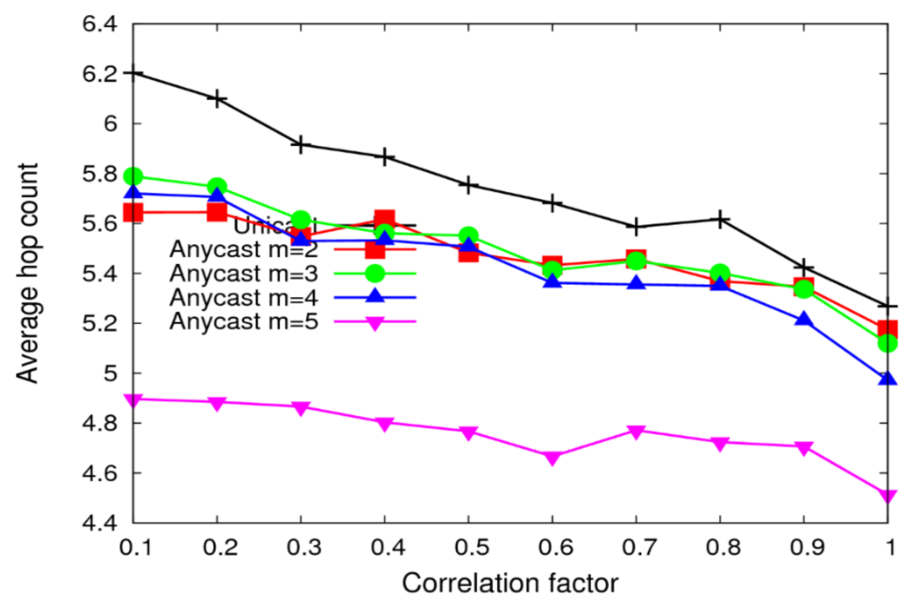
- We simulated 30 unique sets of 100 AR requests (and present the average values).
- All links are bi-directional and are assumed to have 1 Gb/s bandwidth.
- For each request, the source node and destination node(s) are uniformly distributed.
- Request bandwidth demands are uniformly distributed in the range [100 Mb/s, 500 Mb/s], in increments of 100 Mb/s.
- All requests are scheduled to reserve, transmit, and release network resources within two hours such that we stress test the network by increased traffic loads in this time frame.
- The *correlation factor* corresponds to the probability that requests overlap during that two-hour window.

Correlation	m=2	m=3	m=4	m=5
0.1	48.11	50.94	49.06	51.89
0.3	29.59	29.34	32.65	33.93
0.5	24.81	29.92	33.08	35.04
0.7	14.30	21.57	22.02	22.25
0.9	14.44	17.91	20.77	21.21
1.0	11.43	15.76	16.16	18.20

Percentage blocking reduction of anycast $m/1$ over unicast.



16-node ESnet SDN core network topology used in obtaining results.

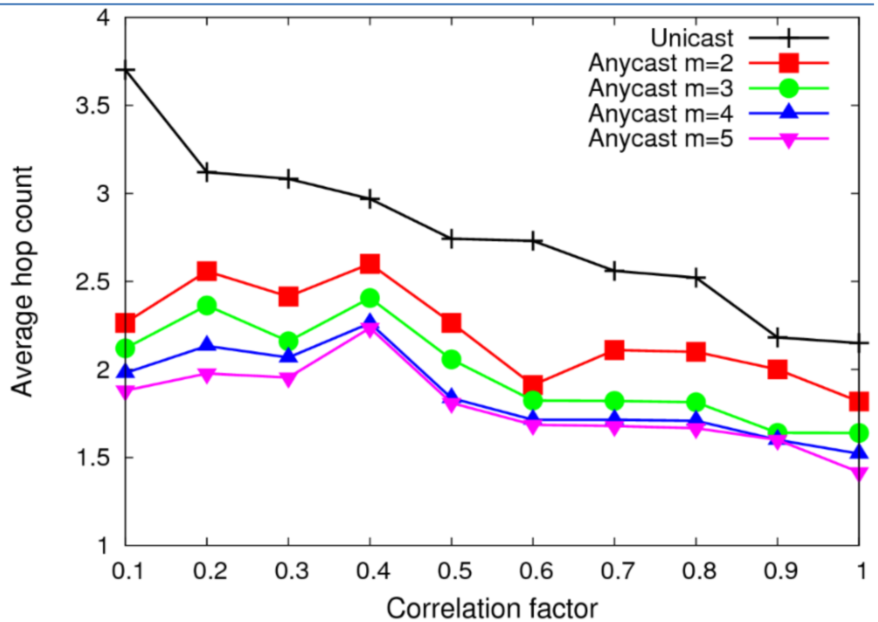


Average hop-count of successfully provisioned requests: unicast vs. anycast $m/1$.

Performance of Anycast Service for OSCARS

Results for multi-domain

- We simulated 5 unique sets of 50 AR requests (and present the average values).
- All links are bi-directional and are assumed to have 10 Gb/s bandwidth.
- Each request, has source node in ESnet and destination node(s) in GEANT.
- Request bandwidth demands are uniformly distributed in the range [1000 Mb/s, 5000 Mb/s], with step granularity of 1000 Mb/s.
- 2 inter-domain links between ESnet and GEANT.
- Remaining assumptions similar to single domain.



Average hop-count of successfully provisioned requests: unicast vs. anycast *m*/1 .

Correlation	m=2	m=3	m=4	m=5
0.1	66.67	100.0	100.0	100.0
0.3	35.71	50.00	57.14	57.14
0.5	15.79	21.05	31.58	31.58
0.7	4.00	12.00	12.00	12.00
0.9	3.57	10.71	10.71	10.71
1.0	6.67	10.00	10.00	13.33

Percentage blocking reduction of anycast *m*/1 over unicast.

Year 1-2: Deployment of Multi/Manycast Overlay on OSCARS

- Need for service to handle replicated data storage/retrieval.
- Data generated at a single site, distributed for study across multiple geographic locations.
- Fundamental obstacle: VPN (or Optical) Layer is point-point.
- Multicast and Manycast functionality must be implemented as a virtual overlay to the point-to-point VLAN (or optical layer).

Year 1-2: Deployment of Multi/Manycast Overlay on OSCARS

Objectives

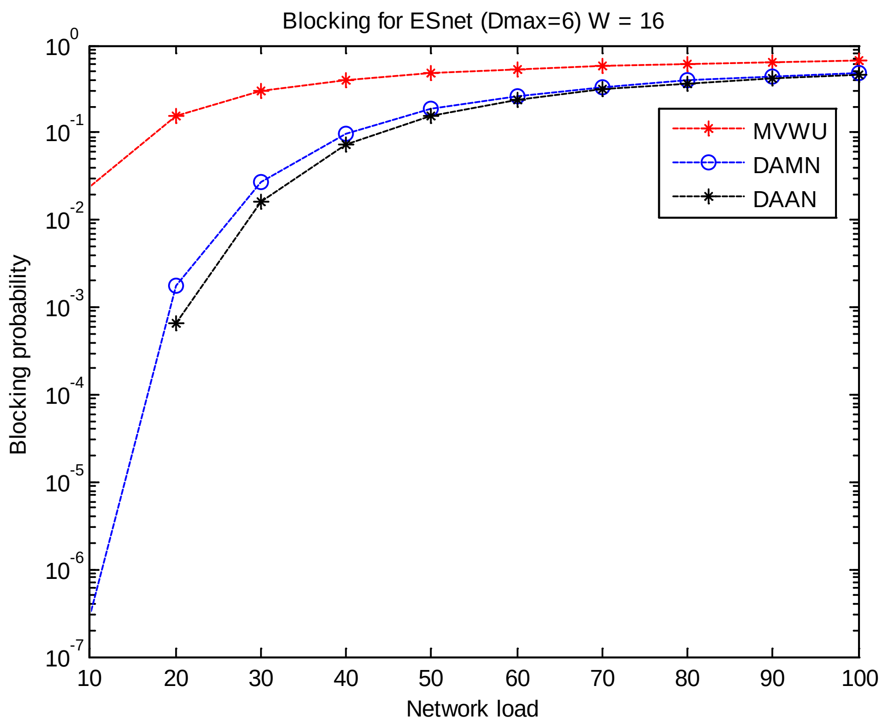
- To support point-to-multipoint connections.
- To develop an overlay model to support Multicast/Manycast communication paradigms over point-to-point unicast connections in OSCARS.

Impact

- Allow scientific community the ability to:
  - (a) Use a multicast service and increase the service acceptance.
  - (b) Provide different connection setup choices with different quality of service (QoS) to the scientists.

Progress

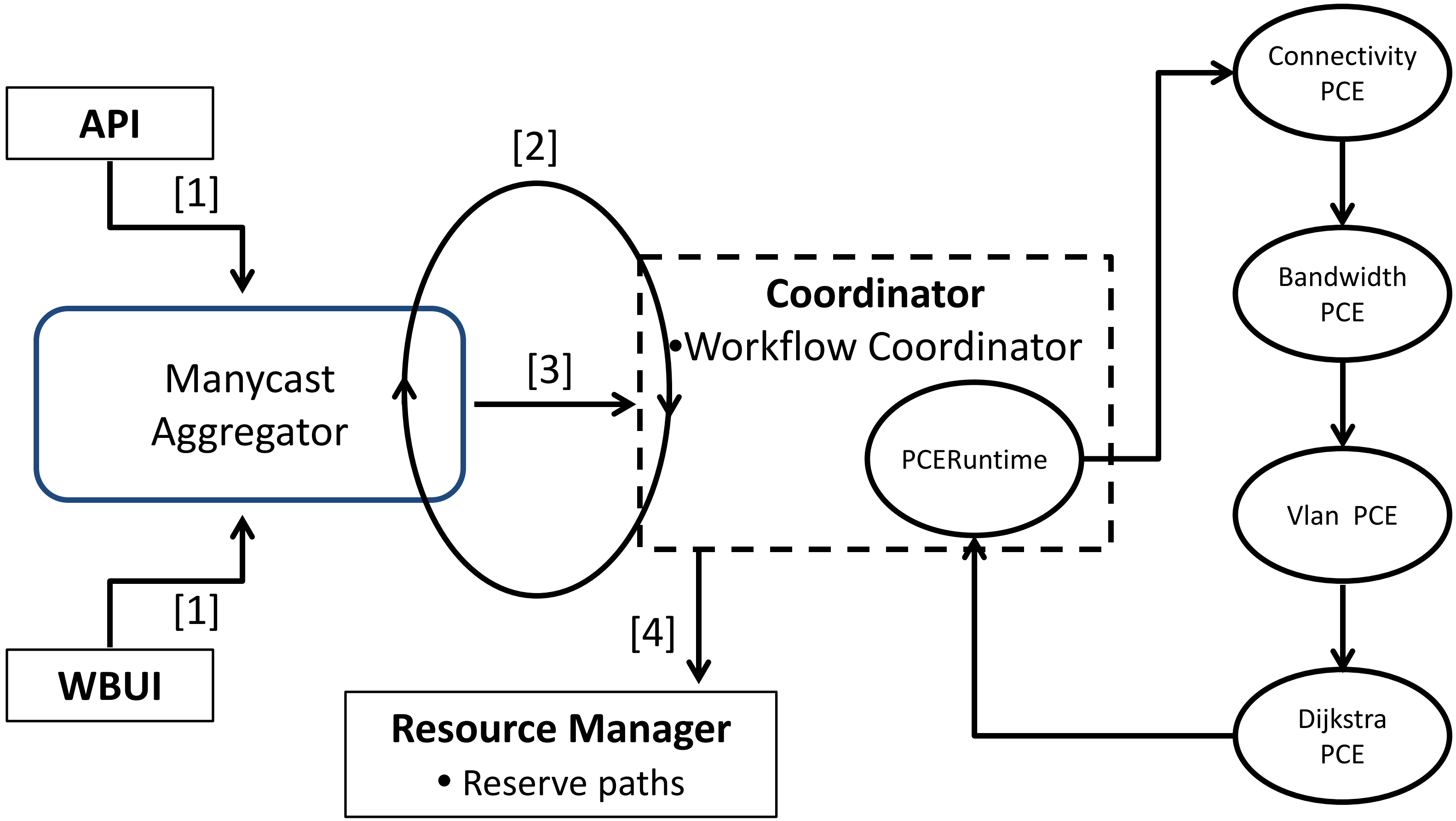
- Proposed three overlay models: Simple Overlay (MVWU), Drop at member node (DAMN) and Drop at any node (DAAN).
- Blocking performance results show significant improvement due to DAMN and DAAN algorithms
- Integrated these overlay models into the OSCARS system (Year 2-3).
  - Simple Overlay (MVWU) implementation completed yesterday!



Year 1-2: Deployment of Multi/Manycast Overlay on OSCARS

- Three Multicast/Manycast overlay approaches proposed to provide point-to-multipoint (P2MP) communication over unicast-only optical/VLAN layer.
  - **MVWU – single logical-hop overlay**
    - Source of the multicast/manycast request establishes an independent lightpath (or VC) to each destination.
    - It is possible that these lightpaths overlap, thus making inefficient use of available bandwidth.
    - This can lead to unnecessarily high connection blocking.
  - **DAMN/DAAN – multiple logical-hop overlay**
    - Source of the request establishes a lightpath (or VC) to one destination.
    - The next destination can be reached by a lightpath (or VC) from the source or from the first destination.
    - Creates a Steiner tree routing scheme wherein each lightpath (or VC) may be viewed as a hop in the logical overlay layer.
- We plan to implement both overlay mechanisms on OSCARS.

Manycast Overlay Design (Best-Effort)



[1] User requests multicast connection in WBUI/API.

- Manycast group addressing scheme which uses a single MC address to represent several destinations.
- Requests forwarded to a MC aggregator responsible for breaking down a single MC request into multiple unicast requests.
    - This aggregator is depicted as an independent module but may be more efficiently incorporated within API/WBUI modules.

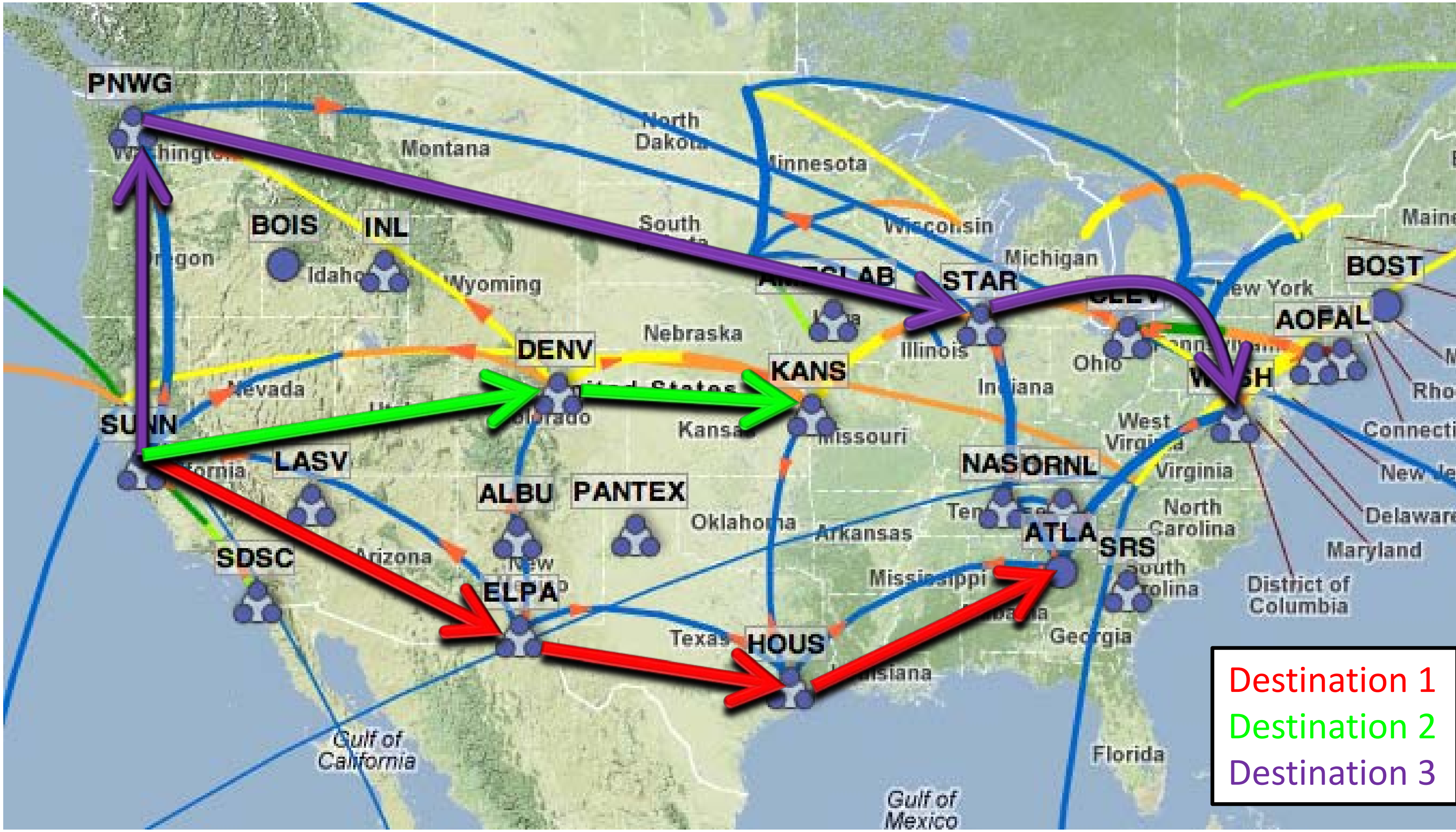
[2] MC aggregator forwards unicast requests one at a time to Coordinator.

- Coordinator forwards requests along unaltered PCE stack to identify unicast VC from source to a *single* destination.

[3]-[4] If any VC cannot be reserved, entire MC request may be blocked.

- User may be able to specify a minimum number of destinations and request may continue as Manycast.
- The current prototype bypasses these last two steps to pursue best-effort VC establishment.

Manycast Request = (SUNN, {ATLA, KANS, WASH})



Note: Links *may* be shared by paths to different destinations.
Routes use the shortest path from src to dest.

Manycast Overlay Results

Average VC Hop Count (Inter- + Intra-Nodal)

Request Set	Unicast	Multicast - 2	Multicast - 3	Multicast - 4	Multicast - 5
1	5	5.15	5.07	5.35	5.4
2	5.2	4	4.34	4.8	4.9
3	4.6	5.55	5.54	5.35	5.32
4	5	5.4	5.07	5.05	5.04
5	5.3	4.8	4.77	4.85	5.08
Average	**5.02**	**4.98**	**4.95**	**5.08**	**5.148**

- Each request set consists of 10 unique requests.
- Request bandwidth was sufficiently small, so no blocking observed.

Year 2-3: Design Survivability Techniques for use with OSCARS

Objectives

- Design and implement survivability techniques on OSCARS using path protection and destination relocation.
- Extend this feature to a multi-domain and multi-layer network.

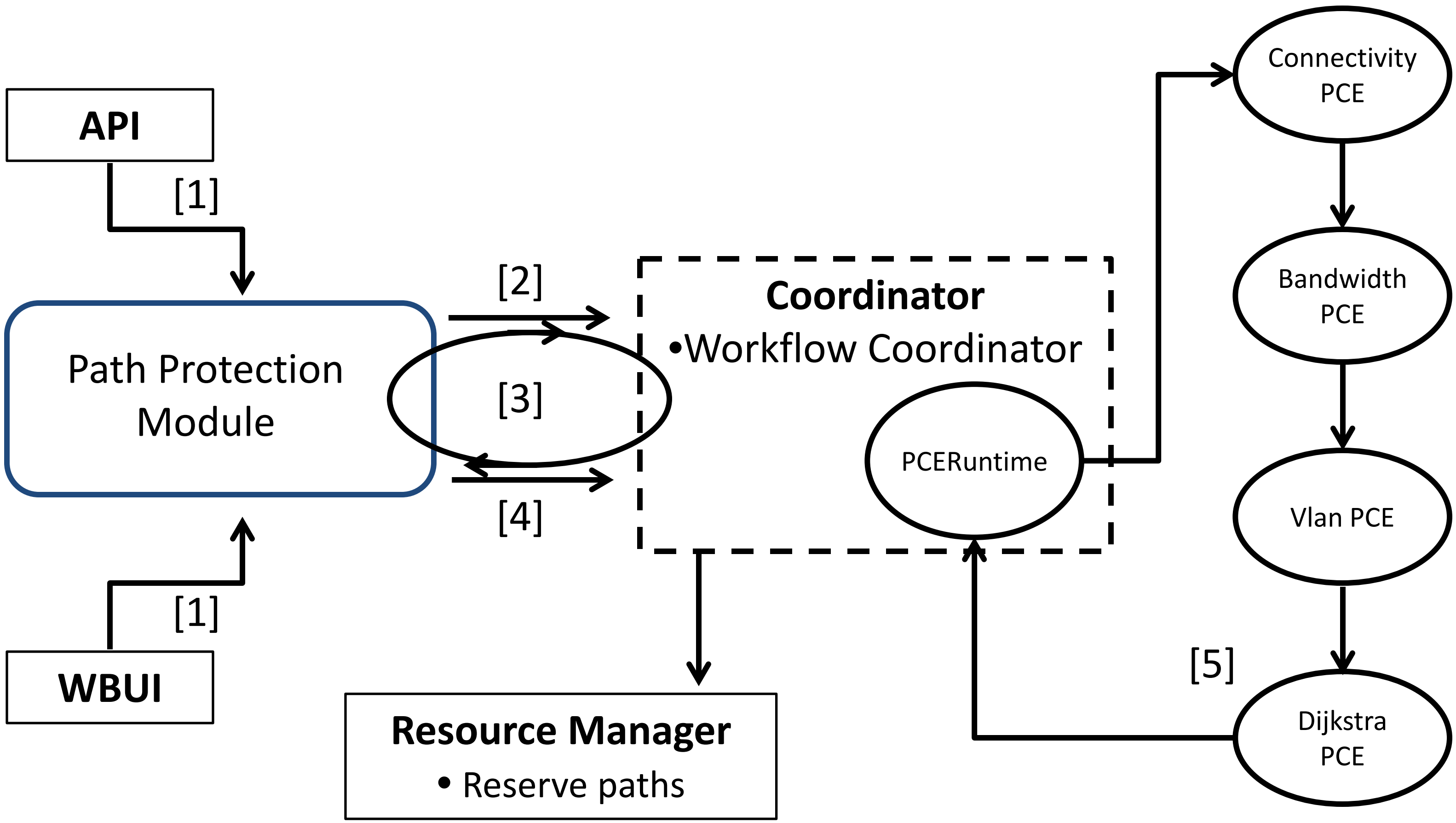
Impact

- Additional resources are reserved for each request.
- Users are protected from link failures (path protection) and destination failures (relocation).

Strategy

- Implement basic survivability techniques to protect against link and node failures.
- **Path protection** enables survivable connections by reserving a link-disjoint backup path.
  - OSCARS will provision both primary and link-disjoint backup paths for each connection request.
  - We have extended our P2MP design to implement dedicated path protection.
- The proposed design **slightly** alters existing unicast PCE stack.
  - We have adopted the pre-coordinator approach.
  - Initial implementation completed yesterday!

Survivable Request Design



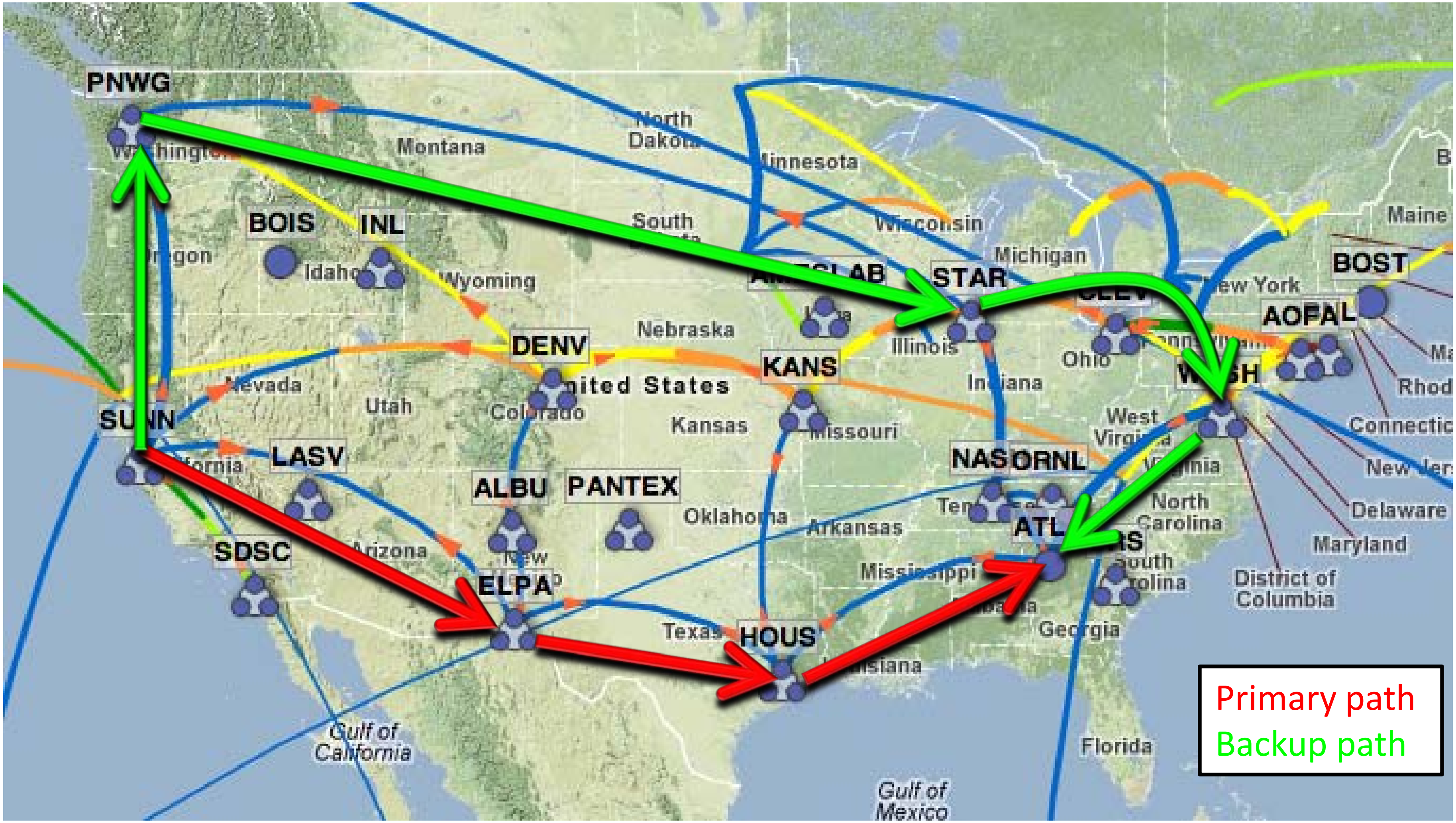
A PCE module will remove from the topology all the links traversed by the primary path, thereby guaranteeing that the backup path will be link-disjoint.

The status of the request is polled until it returns "ACTIVE", "RESERVED", or "FAILED".

[illegible] is added to the [illegible] constraints of the backup request which is then submitted.

[illegible]

Survivable Request = (SUNN : ATLA)



Path Protection: Initial Results

Average VC Hop Count (Inter- + Intra-Nodal)

Request Type	Average Hops (Primary/Backup)	Average Provisioning Time (Seconds)
Protected	7.667 / 10.333	47.87
Unprotected	7.667 / NA	33.28

Year 2-3: Design Survivability Techniques for use with OSCARS

- We plan to extend our anycast PCE design to account for single-link/node failure via anycast path protection and anycast protection using destination relocation.
- **Anycast path protection** allows the ProtectionPCE to pick the anycast destination which has the least-cost link-disjoint path pair from the source.
- **Anycast relocation** allows for the backup path to be routed to an alternate destination in the anycast set. Example: R(S, {D_1,D_2}) ➔ Primary: S-> D_1; Backup: S->D_2.

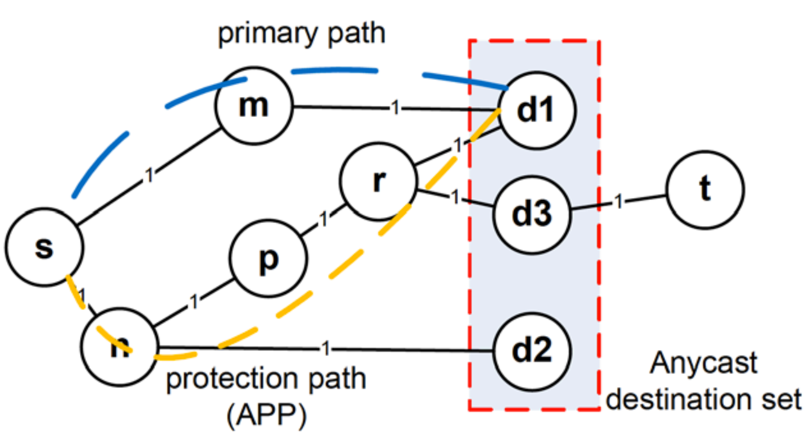


Fig A: Anycast Path Protection

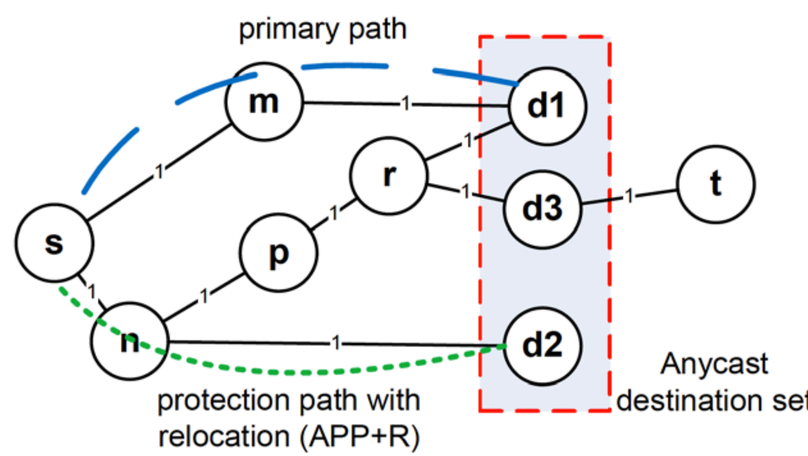


Fig B: Anycast Protection with Destination Relocation

We will look in to implementation of dynamic **restoration** mechanisms for all the above survivability techniques.

Year 2-3: Supporting Quality of Service (QoS) in OSCARS

Objectives

- To provide for user-profile based access to network resources like domains, nodes, ports, and links while provisioning connection requests.
- To support multi-layer and multi-constraint restrictions on requests based on user privileges.

Impact

- Classification of user requests based on service requirements.
- Preferential treatment for high-priority requests originating from data/resource-intensive scientific applications.

Progress

- Designing QoS user profile database extensions based on existing authentication & authorization databases.
- Research and Development on traffic policing and shaping at a request provisioning level driven by user profiles. This can be done using several metrics that balance link bandwidth and VLAN availability using policies like least-used, most-used, most recently used, least recently used, and random.

Year 2-3: What-if Driven Multi-Constrained/Layered OSCARS

Objectives

- Multi-domain offline/inline negotiation protocol for querying and/or reserving the best possible circuit by providing different viable reservation solutions which best suits the user, based on the user profile.

- Rank the various viable connection paths using key performance indicators (KPIs) to best suit the application requirements.

Impact

- Solution matches to user/application requirements as several viable reservation solutions are ranked.
- Re-attempts of reservation for failed reservation requests are minimized.

Progress (in collaboration with ARCHSTONE)

- Implementation in progress of the What-if engine and What-if driven GUI for OSCARS.
- Implementation of a separate offline and inline query + reservation workflow for what-if for multi-domain also in progress.

Research Papers and Journals relevant to COMMON progress

Year 1: Anycast and Advance Reservation

[**1**] Mark Boddie, Timothy Entel, Chin Guok, Andrew Lake, Jeremy Plante, Eric Pouyoul, Bharath H. Ramaprasad, Brian Tierney, Joan Triay, and Vinod M. Vokkarane, "On Extending ESnets OSCARS with a Multi-Domain Anycast Service", *accepted to **ONDM 2012***.

[**2**] Bharath H. Ramaprasad, Arush Gadkar, and Vinod M. Vokkarane, "Dynamic anycasting over wavelength routed networks with lightpath switching," IEEE *High Performance Switching and Routing (**HPSR 2011**},* July 2011.

[**3**] N. Charbonneau, Arush G. Gadkar, Bharath H. Ramaprasad, and Vinod M. Vokkarane, "Dynamic Circuit Provisioning in All-Optical WDM Networks Using Lightpath Switching," ***Elsevier Optical Switching and Networking****, Spl Issue on IEEE ANTS 2010*, April 2012.

[**4**] Neal Charbonneau, Chin Guok, Inder Monga, and Vinod M. Vokkarane, "Advance Reservation Frameworks in Hybrid IP-WDM Networks," **IEEE Communications Magazine**, Special Issue on Hybrid Networking: Evolution Towards Combined IP Services and Dynamic Circuit Network Capabilities, May 2011.

[**5**] N. Charbonneau and V.M. Vokkarane, "A Survey of Advance Reservation Routing and Wavelength Assignment in Wavelength-Routed WDM Networks," **IEEE Surveys and Tutorials**, 2011.

Year 2: Multicast/Manycast Overlay & QoS in OSCARS

[**6**] Arush Gadkar, Jeremy Plante, and Vinod Vokkarane, "Static Multicast Overlay in WDM Unicast Networks for Large-Scale Scientific Applications," *Proceedings, **IEEE ICCCN 2011***, Maui, Hawaii, August 2011.

[**7**] Arush Gadkar and Jeremy Plante, "Dynamic Multicasting in WDM Optical Unicast Networks for Bandwidth-Intensive Applications," *Proceedings, **IEEE Globecom 2011***, Houston, Texas, December 2011.

[**8**] Arush Gadkar, Jeremy Plante, and Vinod Vokkarane, "Manycasting: Energy-Efficient Multicasting in WDM Optical Unicast Networks," *Proceedings, **IEEE Globecom 2011***, Houston, Texas, December 2011.

[**9**] Jeremy Plante, Arush Gadkar, and Vinod Vokkarane, "Dynamic Manycasting in Optical Split-Incapable WDM Networks for Supporting High-Bandwidth Applications," *Proceedings, IEEE International Conference on Computing, Networking and Communications (**ICNC 2012**)*, Maui, Hawaii, February 2012.

[**10**] Jeremy Plante, Arush Gadkar, and Vinod Vokkarane, "Multicast Overlay for High-Bandwidth Applications", minor revision, ***IEEE Journal of Optical Communications and Networking** (JOCN)*. 2012

[**11**] J. Triay, C. Cervell´o-Pastor, and V. M. Vokkarane, "Analytical Model for Hybrid Immediate and Advance Reservation in Optical WDM Networks*," in Proc. of **IEEE GLOBECOM 2011**.*

[**12**] J. Triay, C. Cervell´o-Pastor, and V. M. Vokkarane, "Computing approximate blocking probabilities for hybrid immediate and advance reservation in optical WDM networks," *under review, **IEEE/ACM Transactions on Networking, 2011**.*