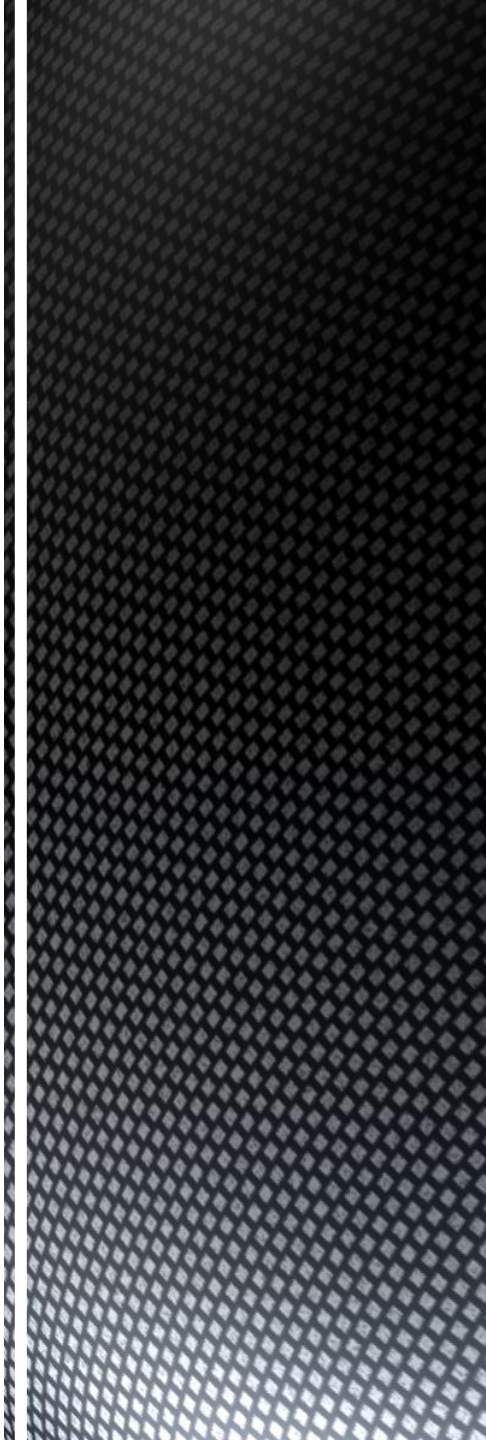
Enhancing ESnet's Unicast-Only OSCARS with a Manycast Overlay Service

Jeremy M. Plante
Vinod M. Vokkarane

Dept. of Electrical and Computer Engineering
UMass Lowell
Jeremy_Plante@student.uml.edu



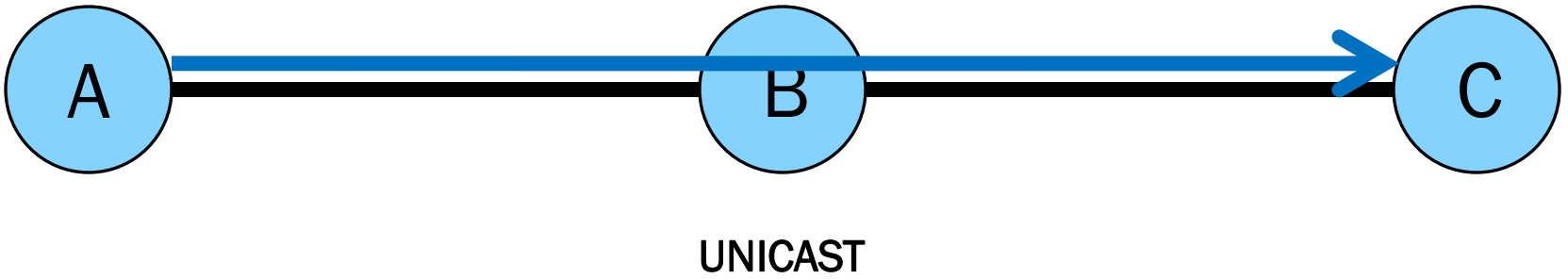
Use-Cases

- Large Hadron Collider – CMS/ATLAS
  - 2000 physicists, 170 institutions, 36 countries.
  - 300 trillion collisions per year.
  - 25 petabytes per year.
  - One copy of data maintained at CERN.
  - Second copy collectively contained by multiple “Tier 1” institutions.
  - Tier 1s send to tier 2s for science analysis.
- Climate Modeling – Earth System Grid Federation
  - Replicated data storage.
  - Distributed remote access.
- Nuclear Fusion – ITER
  - Distributed workflows
  - Coordinating and co-scheduling of compute nodes/storage/instruments.
  - Fast recalibration and experimentation.
  - Compute nodes high interdependency between tasks and tight deadline restrictions.

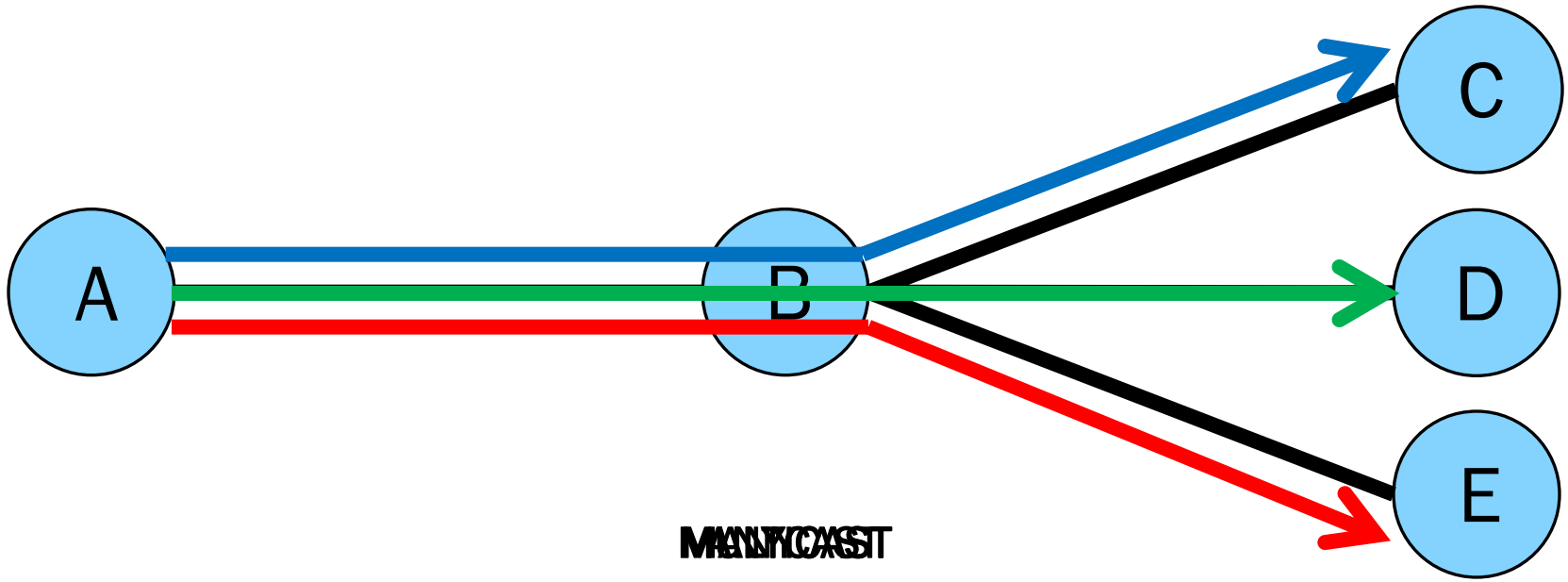
Use-Cases

- What do all these applications have in common?
  - Need to provide fast communication from a single sender to multiple receivers (hopefully in parallel).
- Need to send ALL data to ALL storage sites?
  - Costly
  - Complex
  - Inefficient
  - Unnecessary

Communication Paradigms



UNICAST



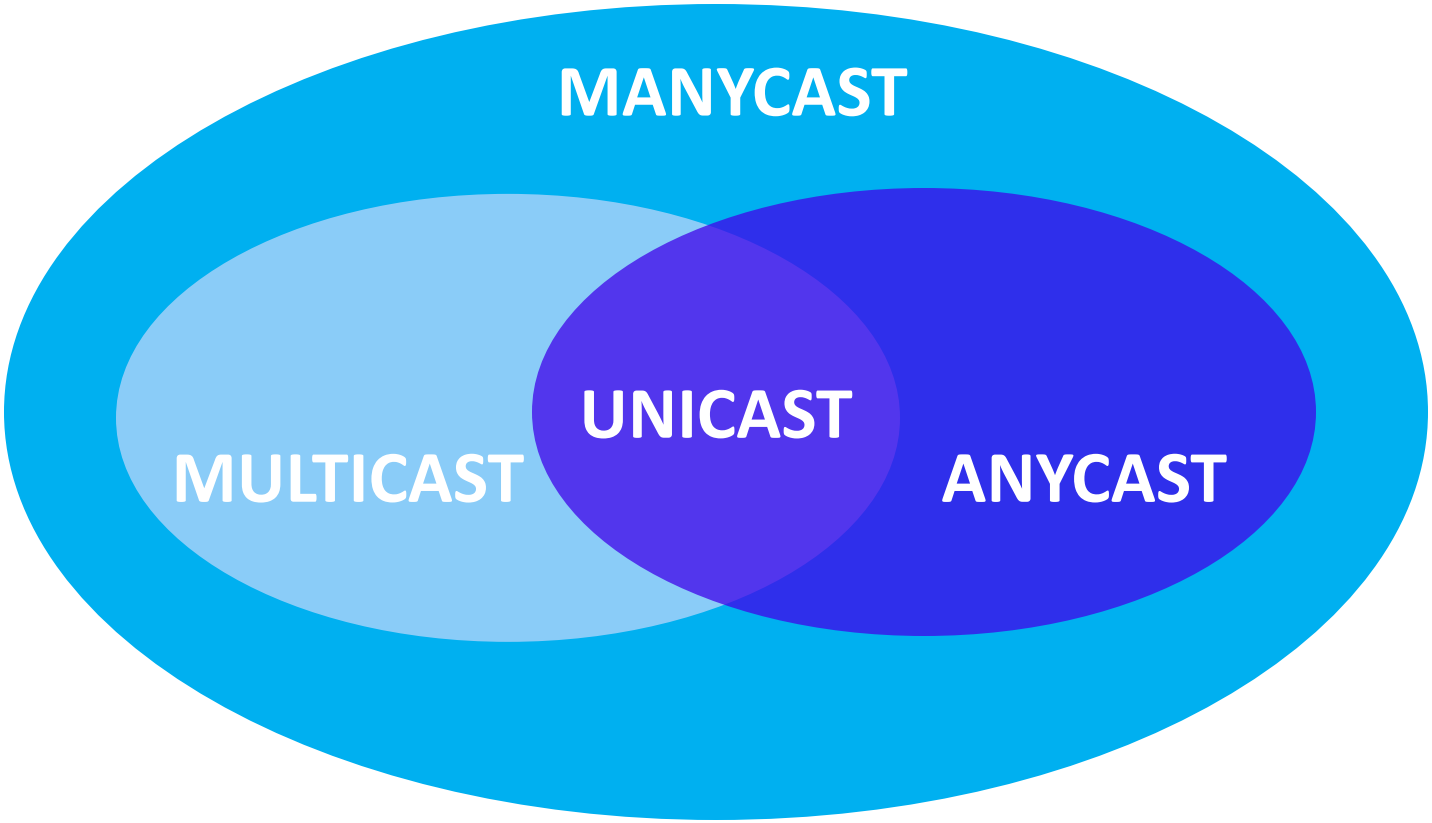
MULTICAST

Generalization

- Manycast request:
  - N nodes in network.
  - K candidate destinations.
  - K’ required destinations.

K	K’	Description
1	1	Unicast
$1 \leq K \leq N$	1	Anycast
$1 \leq K \leq N$	$K' = K$	Multicast
$1 \leq K \leq N$	$K' \leq K$	Manycast
$K = N$	$K' = K$	Broadcast

Generalization



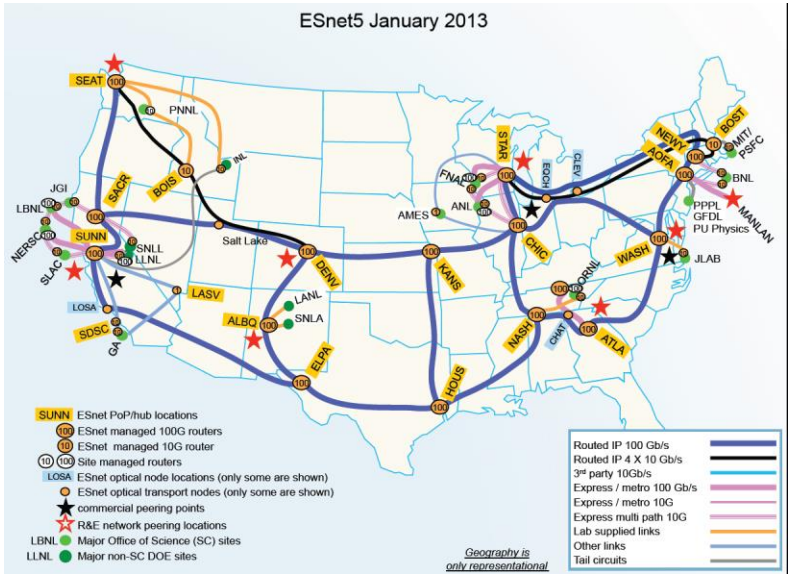
- Multicast = NP-Complete.
- Manycast = NP-Hard.

How to Select Destinations

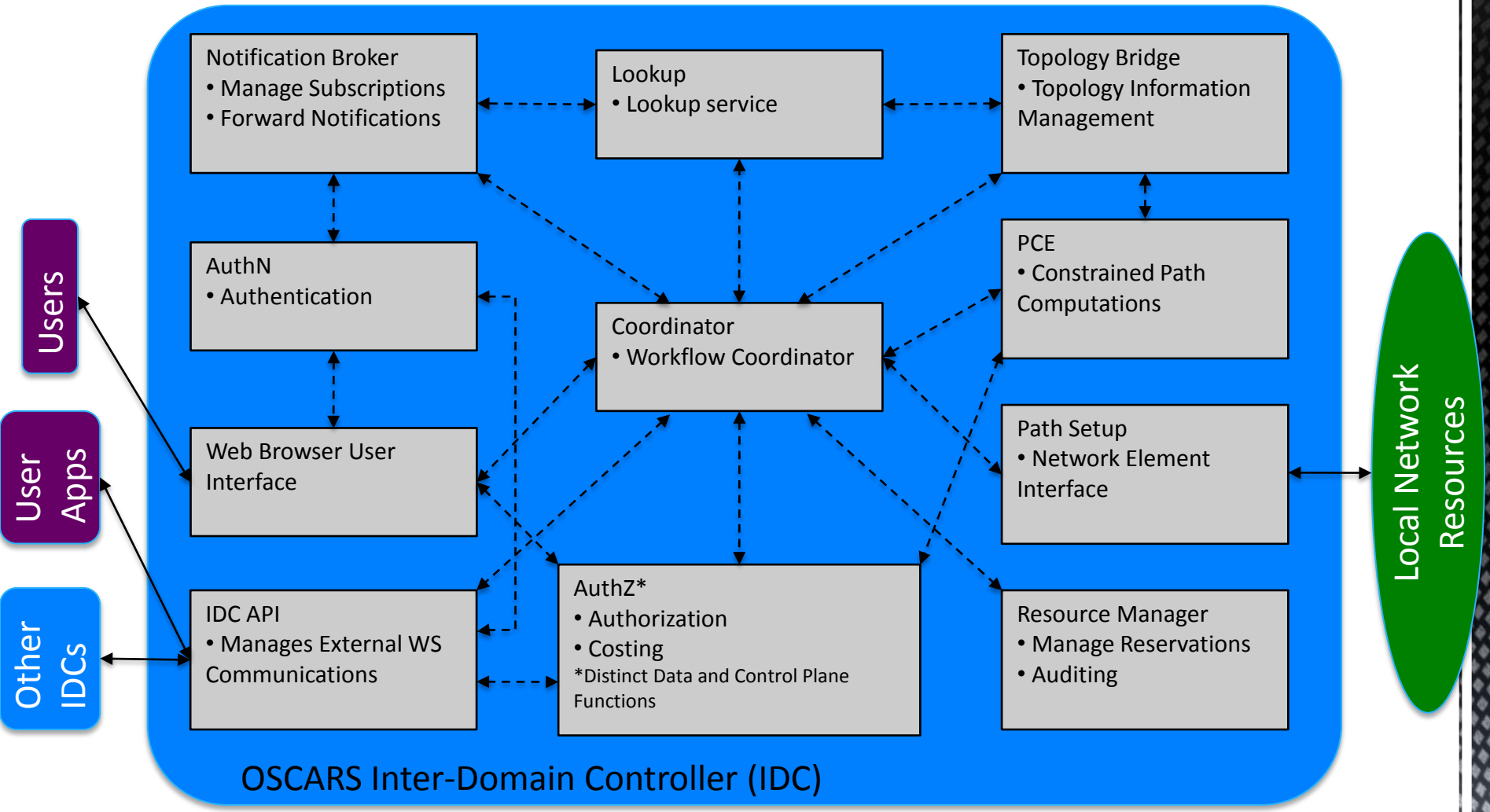
- Nearest candidate destinations.
  - Shortest paths.
  - Steiner tree.
- Least-expensive destinations.
  - Processing power.
  - Latency.
- Load-balancing.
  - Choose destinations in least-demand.
  - Choose destinations with fewest prior commitments.
- Overall Cost (non-monetary)
  - Power-efficiency.
  - Energy-efficiency.
  - GHG emissions.
- Arbitrarily
  - All destinations assumed equal
  - Reach at least K’ out of K destinations.

OSCARS

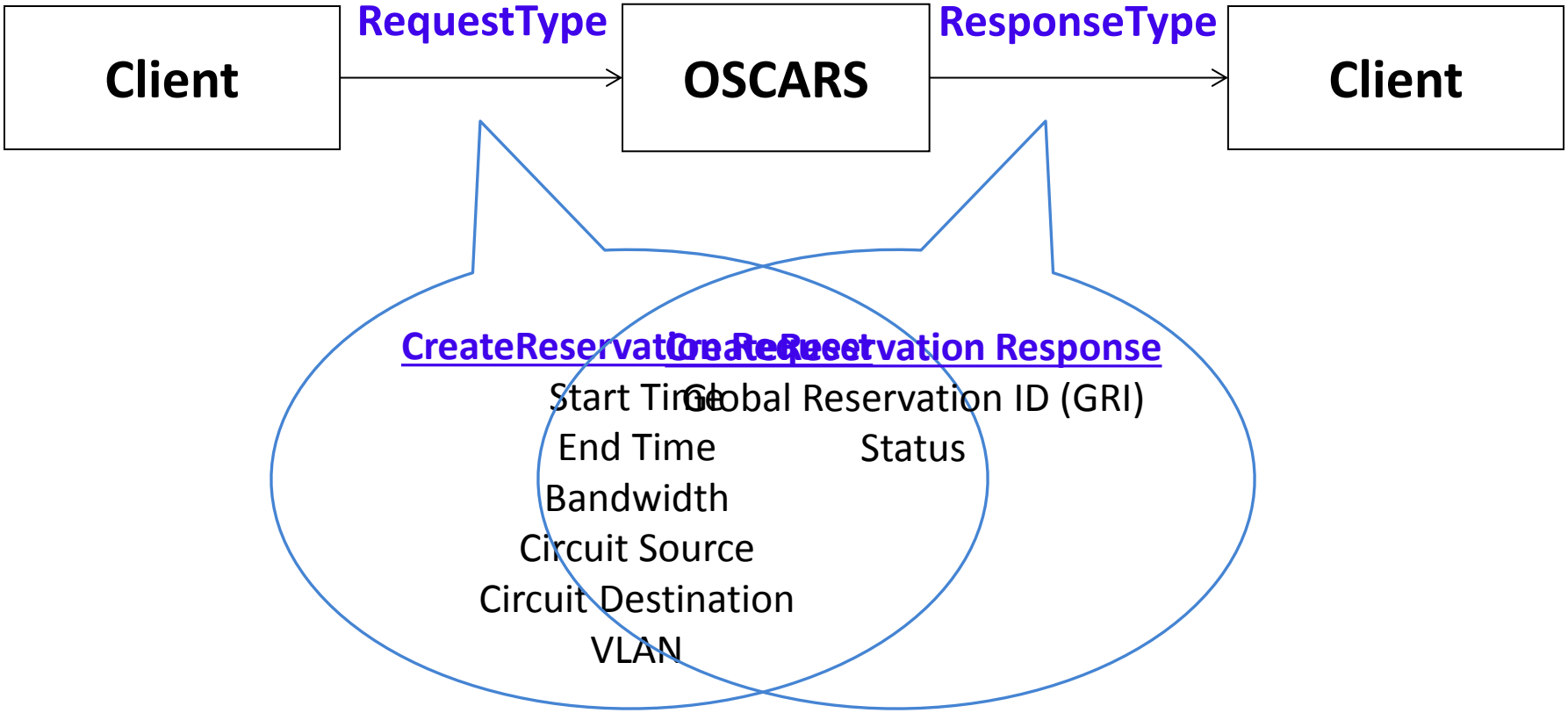
- On-demand Secure Circuits and Advance Reservation System
  - Developed by DOE's Scientific Networking Group (ESnet).
  - ESnet has built and maintains the world's fastest scientific communication network (100 Gbps).
  - Provides guaranteed performance on dedicated virtual circuits (VCs) for transmitting data.
  - Most popular circuit-provisioning software amongst networking/research communities.
  - 50% of ESnet's monthly 14 petabytes of transmitted data is carried on OSCARS VCs.



What Does OSCARS Look Like?



OSCARS/Client Interaction



OSCARS

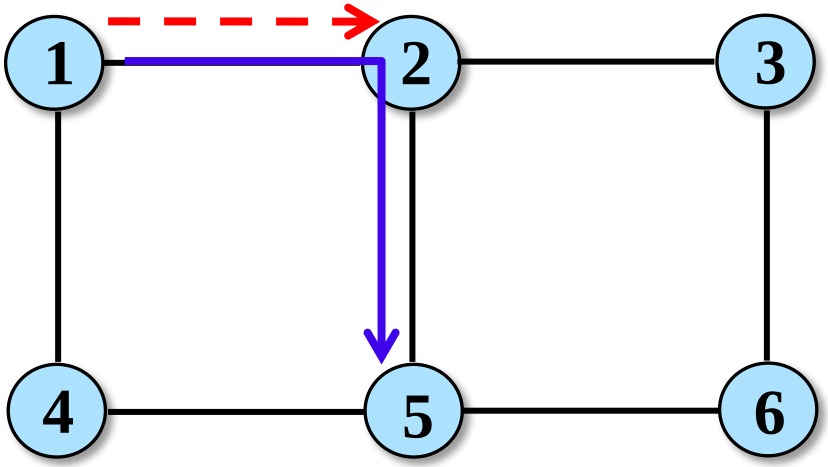
- Currently only supports point-to-point communication: Manycast/Multicast inherently not possible at the optical layer.
- Not only is OSCARS incapable of point-to-multipoint communication, but up until very recently, ESnet was limited to unicast by its hardware infrastructure.
- Provide front-end logic for grouping individual OSCARS VCs, such that their identities are transparent to the end-user.
  - This is an overlay approach to logical Manycasting!

Overlay Models

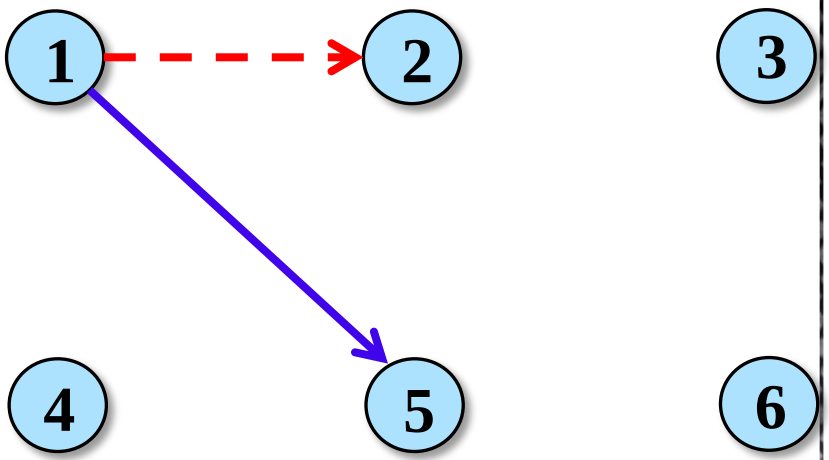
Single-hop model

- Establishes K' separate and unique end-to-end VCs from source to destinations.
- ALL lightpaths originate at the source.
- Manycast = Collection of Unicast.

Manycast Request: (1, {2, 5, 6}, 2)



Routing in Physical Topology



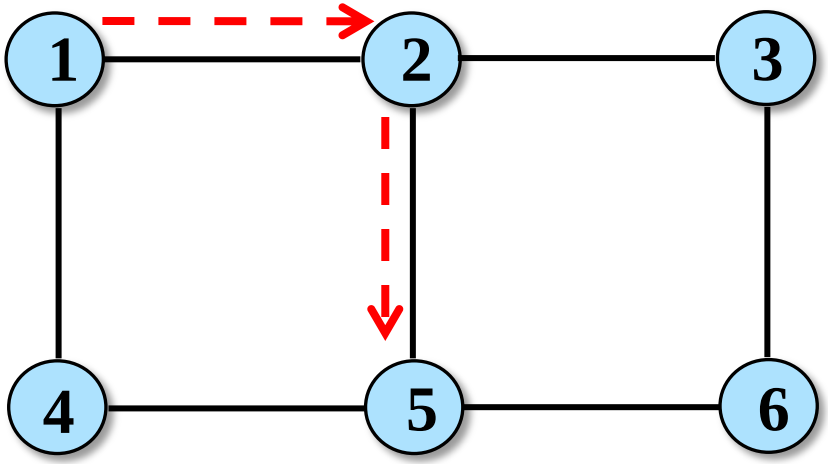
Routing in Logical Topology

Overlay Models

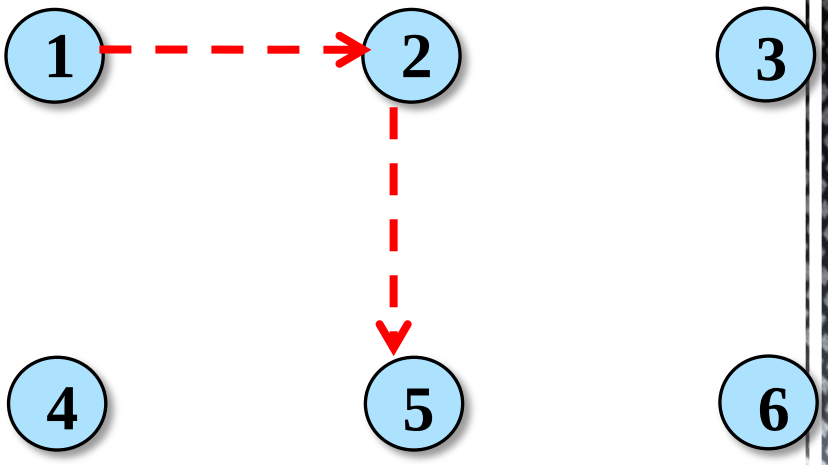
Multiple-hop model

- Establishes a logical Steiner tree, possibly consisting of multiple logical hops from the source, reaching at least K' destinations.
- VCs may originate/terminate at source OR destinations OR other network nodes, this requires temporary storage for conversion from optical signal to electronic.

Manycast Request: (1, {2, 5, 6}, 2)



Routing in Physical Topology



Routing in Logical Topology

Overlay Models

Conference Proceedings

- T. Schondienst, J. M. Plante, D. A.P. Davis, and V. M. Vokkarane, “Energy Source-Aware Manycast Overlay in WDM Networks," Proceedings, IEEE Globecom, December 2013.
- J. Plante, A. Gadkar, and V. Vokkarane, “Dynamic Manycasting in Optical Split-Incapable WDM Networks for Supporting High-Bandwidth Applications," Proceedings, ICNC February 2012.
- A. Gadkar and J. Plante, “Dynamic Multicasting in WDM Optical Unicast Networks for Bandwidth-Intensive Applications," Proceedings, IEEE Globecom, December 2011.
- A. Gadkar, J. Plante, and V. Vokkarane, “Manycasting: Energy-Efficient Multicasting in WDM Optical Unicast Networks," Proceedings, IEEE Globecom , December 2011.
- A. Gadkar, J. Plante, and V. Vokkarane, “Static Multicast Overlay in WDM Unicast Networks for Large-Scale Scientific Applications," Proceedings, IEEE ICCCN, August 2011.

Journal Publications

- A. Gadkar, J. Plante, and V. Vokkarane, “Multicast Overlay for High-Bandwidth Applications,” Journal of Optical Communications and Networking (JOCN) vol. 4, no. 8, pp. 571-585.

Overlay Models

Single-Hop

- ✓ Fast set-up.
- ✓ Entirely-front-end logic.
- ✓ No delay at “drop-nodes”.
- ✓ No network-internal storage.
- ✓ Simple.

- ✘ Bandwidth hungry.

Multiple-Hop

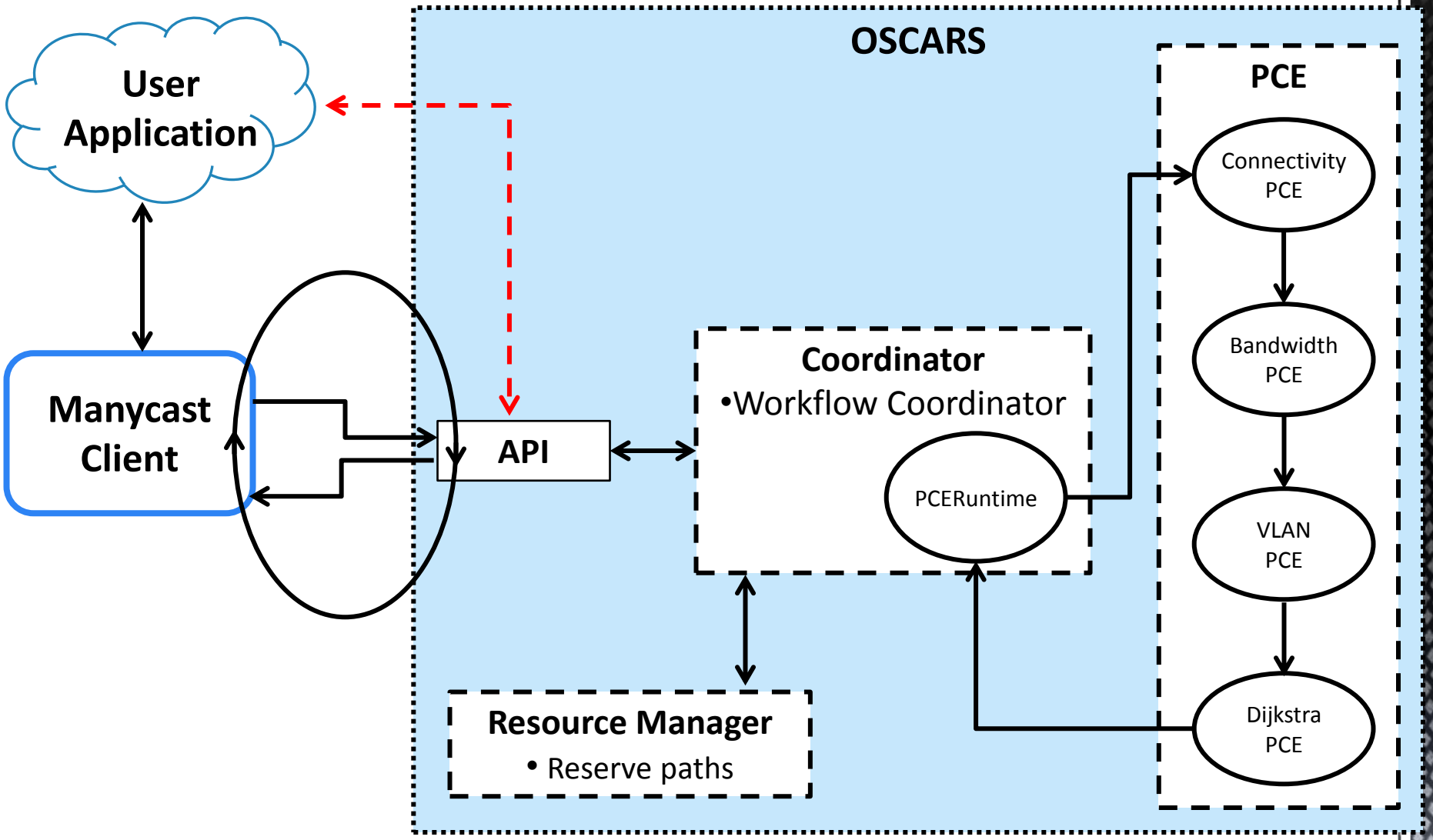
- ✓ Bandwidth-efficient.
- ✓ Better resource consumption.

- ✘ Complex.
- ✘ Requires modifications to OSCARS.
- ✘ Some delay at “drop-nodes”.
- ✘ Requires network-internal storage.

- Requirements from ESnet:
  - No modifications to OSCARS code.
  - No internal storage within the network.
  - Parallel Transfers to multiple destinations.

- Must use single-hop approach!

Manycast Client Design



Manycast Client Flexibility

- Specify a group of Manycast destinations
- Specify minimum threshold/maximum cutoff.
  - By specifying different values for threshold/cutoff, Manycast service flexibility increases:

MANYCAST OSCARS CLIENT COMMUNICATION PARADIGM

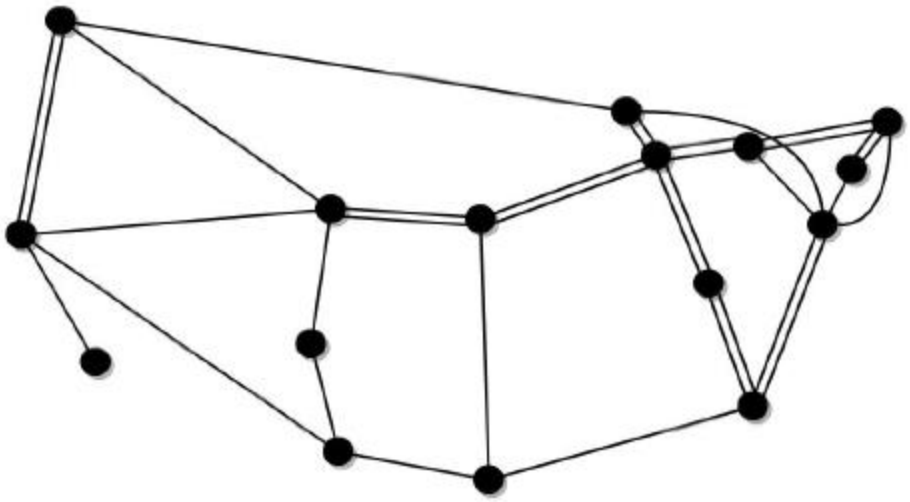
Candidate Destinations	Threshold	Cutoff	Paradigm Description
3	1	1	Anycast (3/1)
3	1	2	Best-Effort Manycast (3/2)
3	1	3	Best-Effort Multicast (3/3)
3	2	2	Manycast (3/2)
3	2	3	Bounded Best-Effort[Manycast, Multicast]
3	3	3	Multicast (3/3)

- If threshold/cutoff cannot be satisfied, extra sub-requests are cancelled (first-fit) to satisfy constraints of the Manycast request.

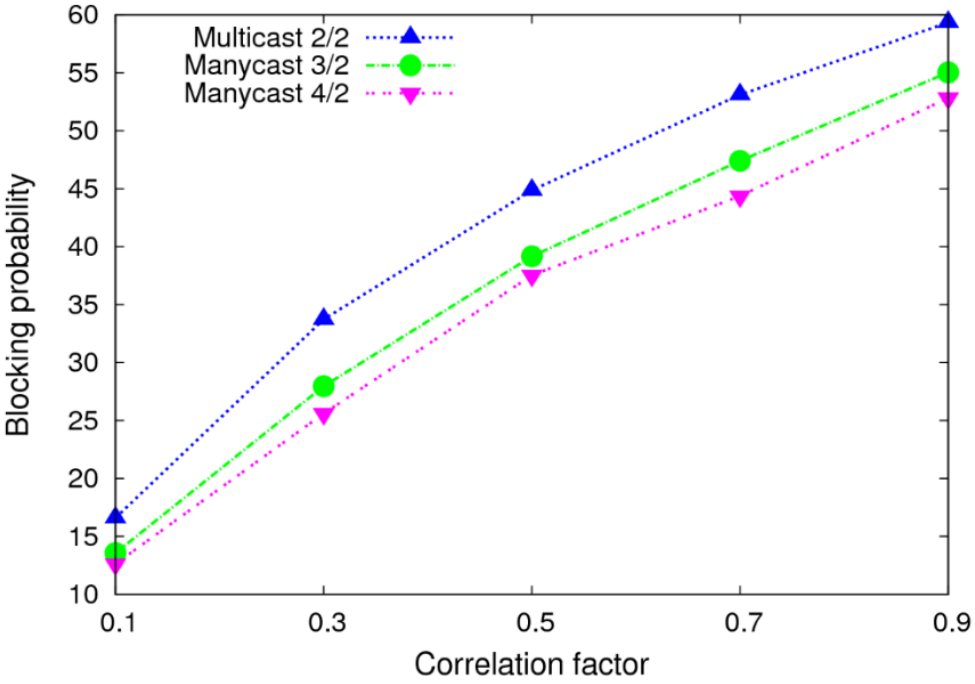
Performance Evaluation

Multicast/Manycast OSCARS comparison:

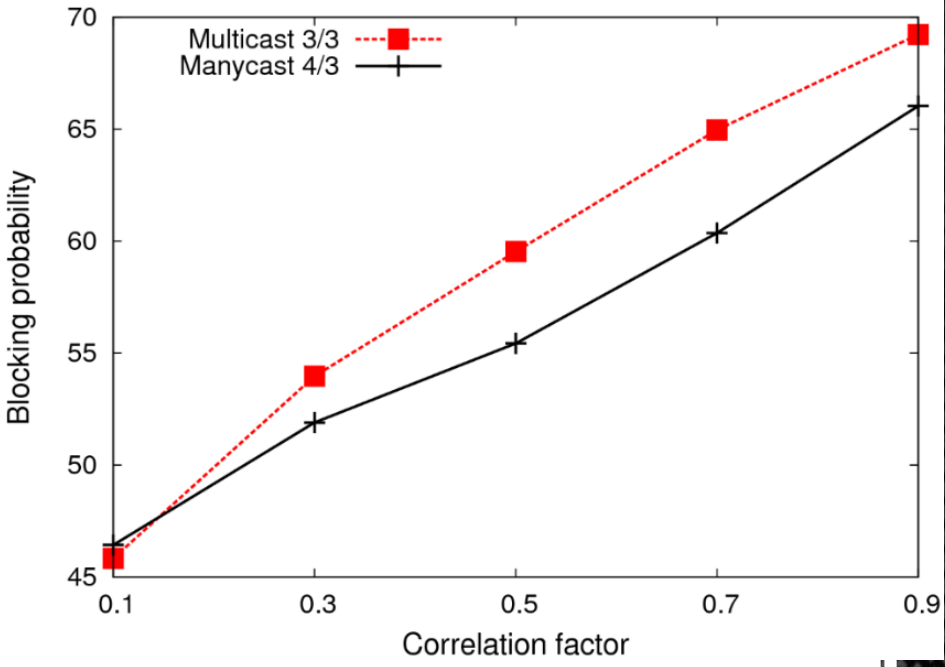
- ESnet topology.
- Advance Reservations (2-hour window)
- Correlation Factor => probability requests overlap in time.
  - 0 Correlation = time-independent set of requests.
- Request set size *R* = 100.
- Average of 10 unique request sets.



Performance Evaluation



Blocking Probability
Two destinations Required



Blocking Probability
Three destinations Required

- Manycast flexibility lowers blocking, despite same number of reached destinations.
- Particularly true at higher correlation factors.
- The relative blocking reduction due to destination flexibility is less dramatic as more candidate destinations are added.

Conclusions

- Collaboration and distributed workflows are becoming omnipresent.
- Desired parallelism must be taken into account in the network to prevent bottleneck.
  - Manycast communication.
- Many of these applications already transport data over ESnet, likely using OSCARS virtual circuits.
- Proposed Manycast client makes parallel transfers possible without *any* modification to OSCARS design.
- Current deployed system is simple, but not ideal.
  - But it's a tangible step towards a deployable Manycast overlay system!

Enhancing ESnet's Unicast-Only OSCARS with a Manycast Overlay Service

Manycast client code available
https://www.dropbox.com/sh/0jv518h9ecmz6eq/SrZSy3ug3a

Details on collaboration with ESnet
http://faculty.uml.edu/vinod_vokkarane/common/

OSCARS
http://es.net/services/oscars/

Jeremy M. Plante
Vinod M. Vokkarane

Dept. of Electrical and Computer Engineering
UMass Lowell
Jeremy_Plante@student.uml.edu