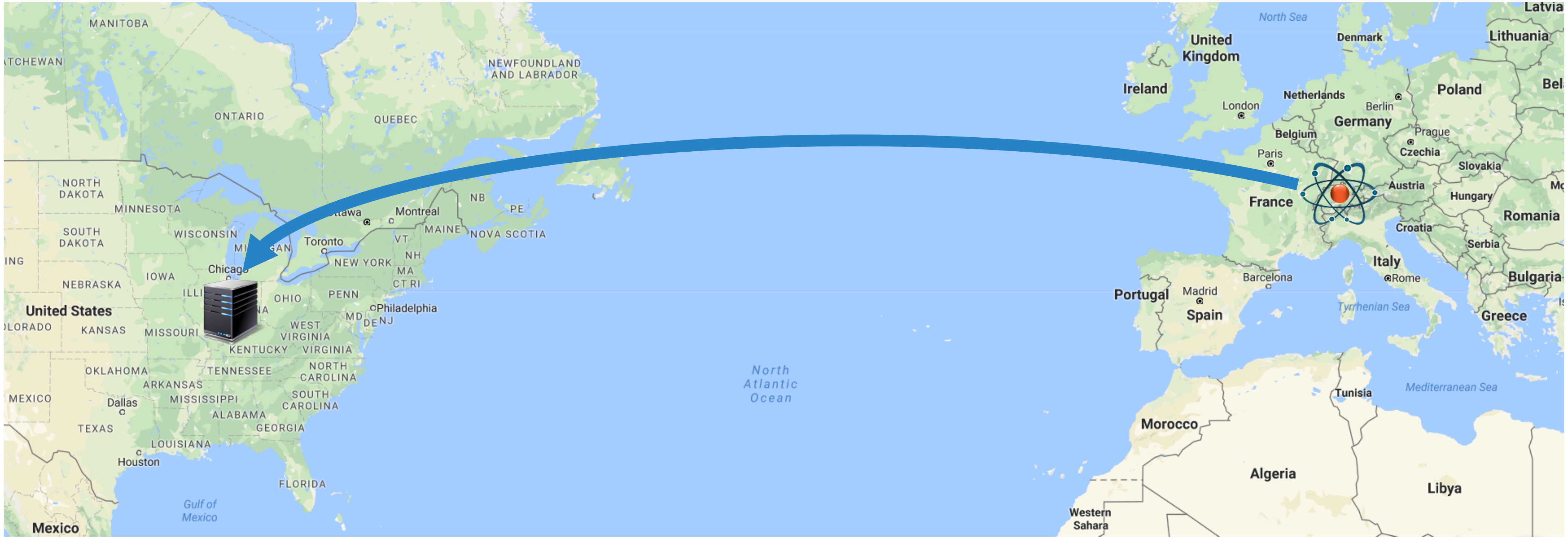
Orchestrating Intercontinental Advance Reservations with Software-Defined Exchanges

INNOVATING THE NETWORK FOR DATA INTENSIVE SCIENCE (INDIS) 2017

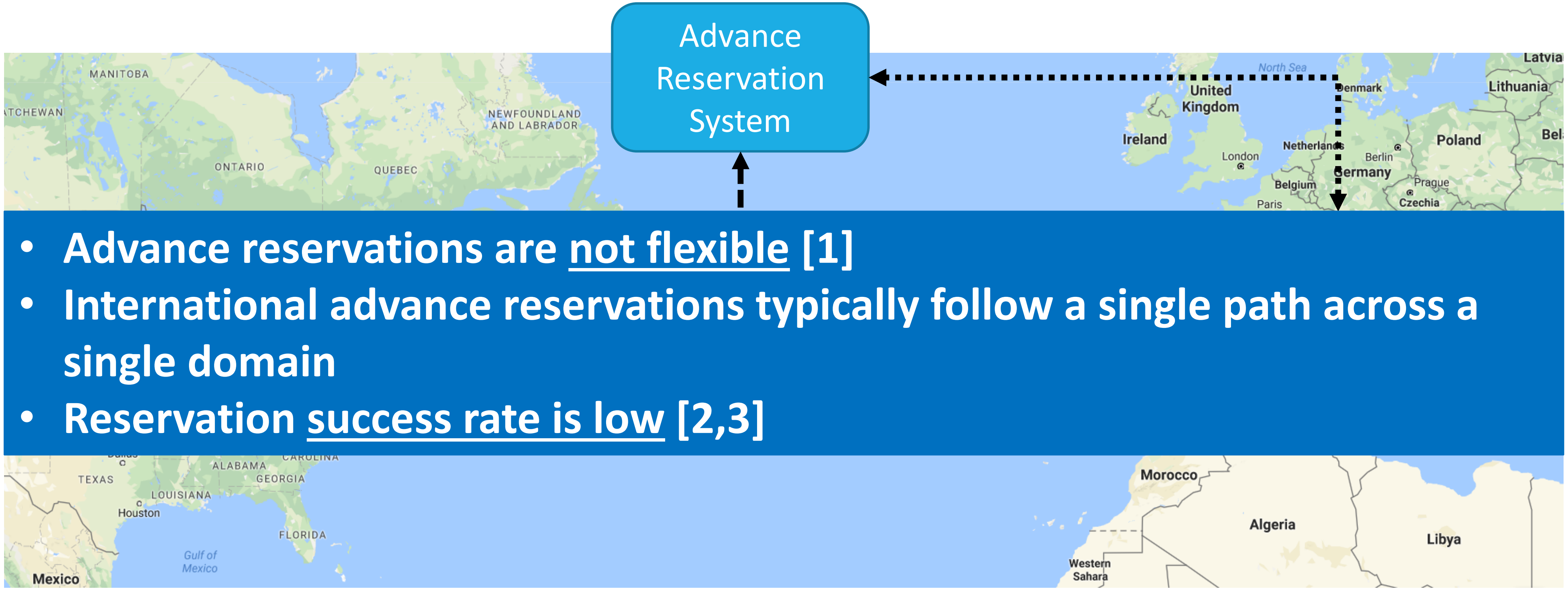
BY **JOAQUIN CHUNG,** RAJKUMAR KETTIMUTHU, NAM PHO, RUSS CLARK, HENRY OWEN

NOVEMBER 12, 2017

Motivation


MANITOBA
NEWFOUNDLAND AND LABRADOR
ONTARIO
QUEBEC
NORTH DAKOTA
MINNESOTA
SOUTH DAKOTA
WISCONSIN
Montreal
NB
PE
MAINE
NOVA SCOTIA
Toronto
VT
NH
NEW YORK
MA
CT RI
Chicago
NEBRASKA
IOWA
OHIO
PENN
Philadelphia
MD
DE NJ
United States
KANSAS
MISSOURI
WEST VIRGINIA
KENTUCKY
VIRGINIA
OKLAHOMA
TENNESSEE
NORTH CAROLINA
ARKANSAS
MEXICO
Dallas
MISSISSIPPI
SOUTH CAROLINA
ALABAMA
GEORGIA
TEXAS
LOUISIANA
Houston
FLORIDA
Gulf of Mexico
Mexico
North Atlantic Ocean
North Sea
United Kingdom
Denmark
Latvia
Lithuania
Ireland
Netherlands
London
Berlin
Poland
Germany
Belgium
Prague
Czechia
Paris
Slovakia
Austria
Hungary
France
Romania
Croatia
Serbia
Italy
Rome
Barcelona
Bulgaria
Portugal
Madrid
Spain
Tyrrhenian Sea
Greece
Tunisia
Mediterranean Sea
Morocco
Algeria
Libya
Western Sahara

Motivation

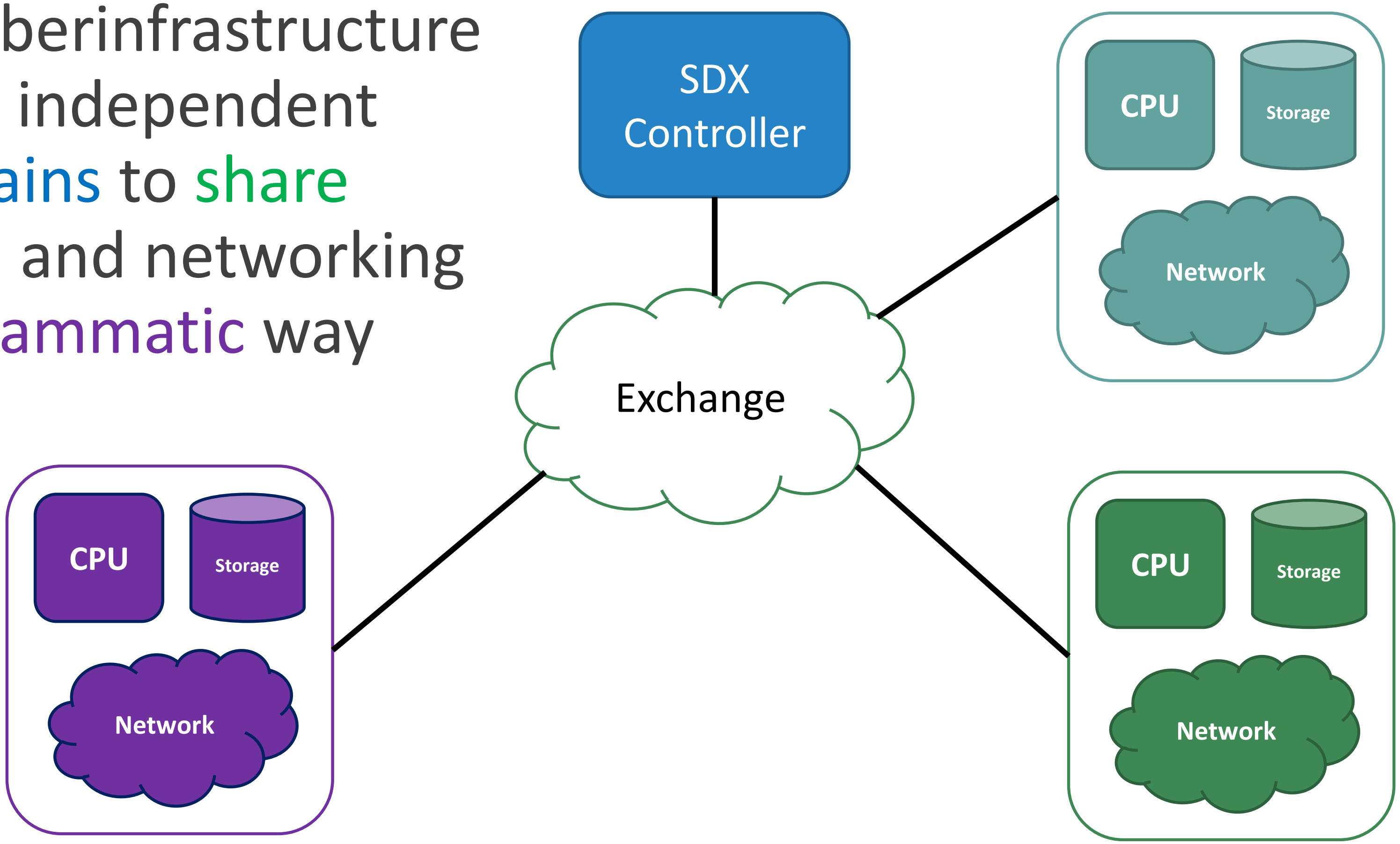


- **Advance reservations are <u>not flexible</u> [1]**
- **International advance reservations typically follow a single path across a single domain**
- **Reservation <u>success rate is low</u> [2,3]**

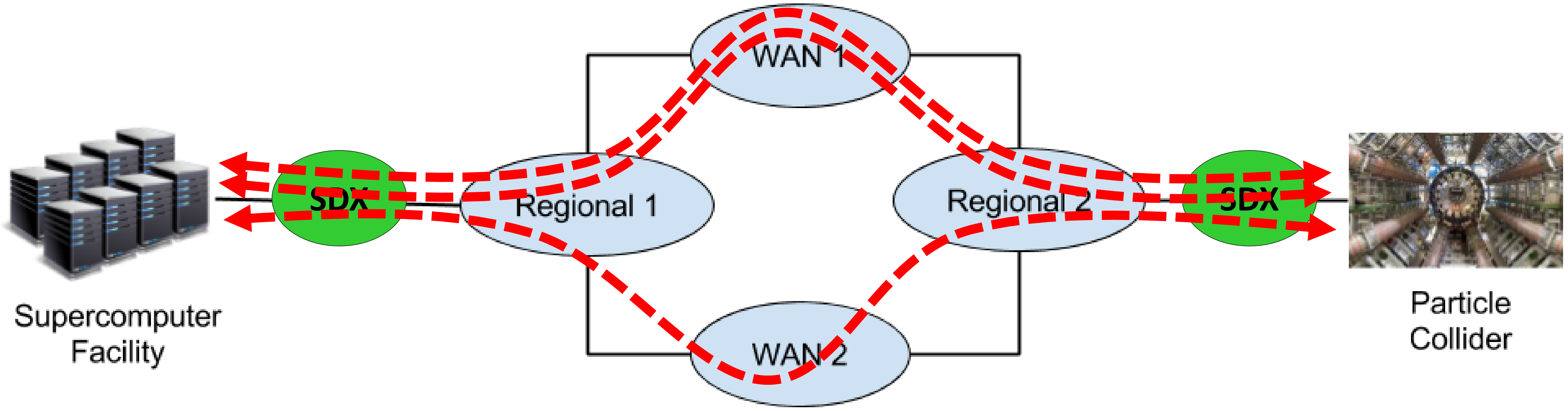
[1] M. Balman, E. Chaniotakisy, A. Shoshani, A. Sim, A flexible reservation algorithm for advance network provisioning, in: 2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis, 2010, pp. 1-11. doi:10.1109/SC.2010.4.
[2] S. Venugopal, X. Chu, R. Buyya, A negotiation mechanism for advance resource reservations using the alternate offers protocol, in: 2008 16th International Workshop on Quality of Service, 2008, pp. 40-49.
[3] P. Xiao, Z. Hu, Two-dimension relaxed reservation policy for independent tasks in grid computing, Journal of Software 6 (8) (2011) 1395-1402.
[4] INTERNET2 IP BACKBONE CAPACITY AUGMENT PRACTICE, https://www.internet2.edu/policies/ip-backbone-capacity-augment-practice/

Software-Defined Exchange (SDX)

An SDX is a novel cyberinfrastructure that allows multiple independent administrative domains to share computing, storage, and networking resources in a programmatic way



Software-Defined Exchange (SDX)

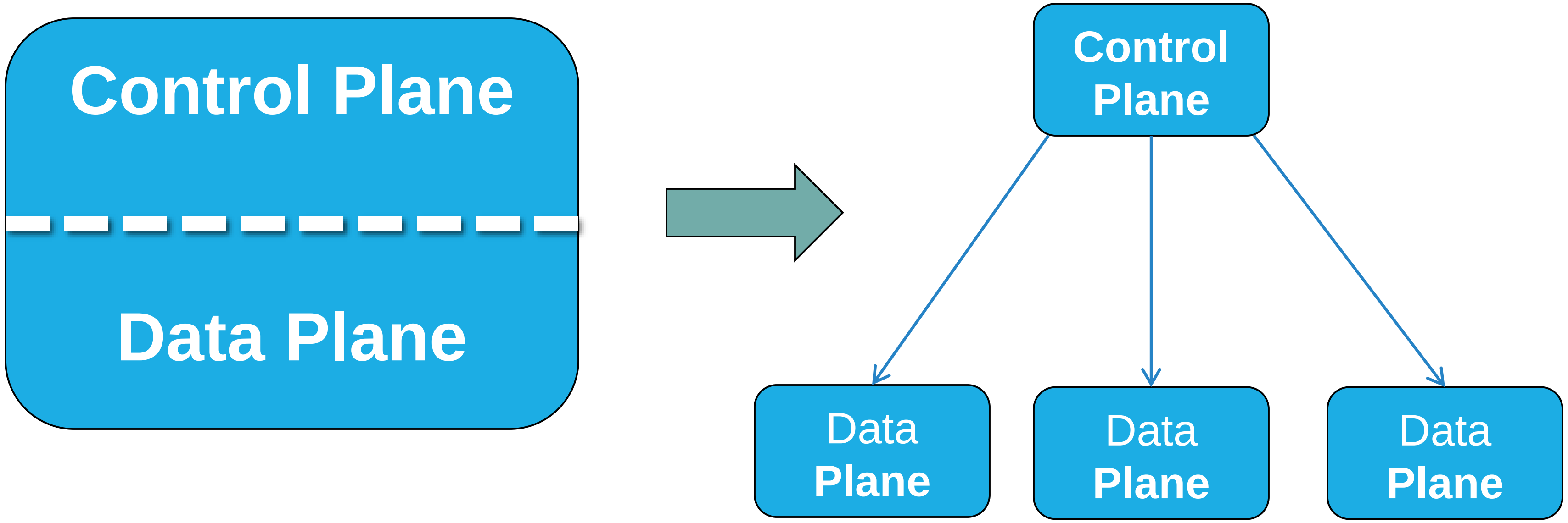


Agenda

1. Motivation
2. Background
3. Related Work
4. Architecture Overview
5. Design
6. Evaluation
7. Conclusions

What is SDN?

Software Defined Networking (SDN) separates the control plane from the data plane



Agenda

1. Motivation
2. Background
3. Related Work
4. Architecture Overview
5. Design
6. Evaluation
7. Conclusions

Related Work

Multi-domain SDN Architectures

- Multi-domain network resource management [1] → Service level specifications
- Service provider SDN (SP-SDN) [2] → Technology domains (e.g., mobile, transport, data center, etc.)

Network Resource Management

- Resource Negotiation and Pricing Protocol (RNAP) [3]
- Service Negotiation and Acquisition Protocol (SNAP) [4]

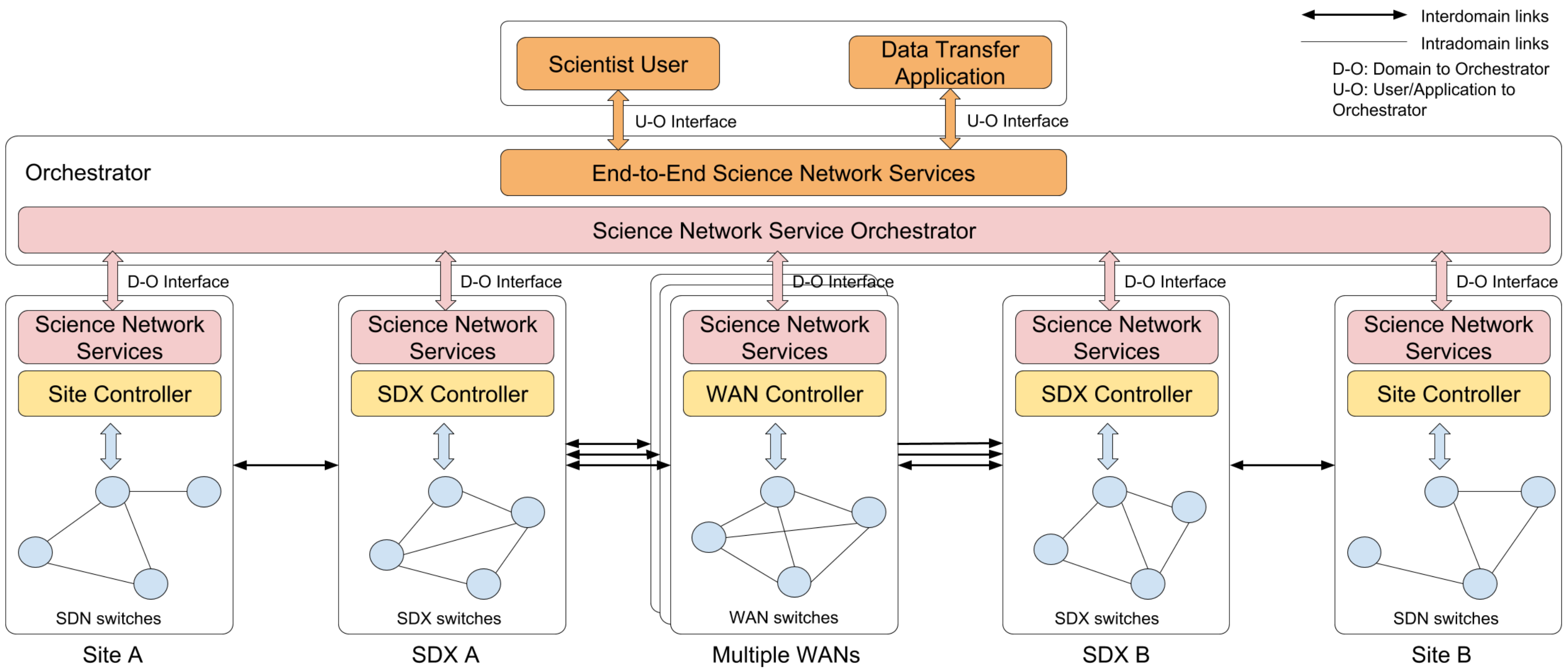
Multi-path Advance Reservations

- OpenFlow Link-layer MultiPath Switching (OLiMPS) [5]
- Multi-path extension for OSCARS client [6]

Agenda

1. Motivation
2. Background
3. Related Work
4. Architecture Overview
5. Design
6. Evaluation
7. Conclusions

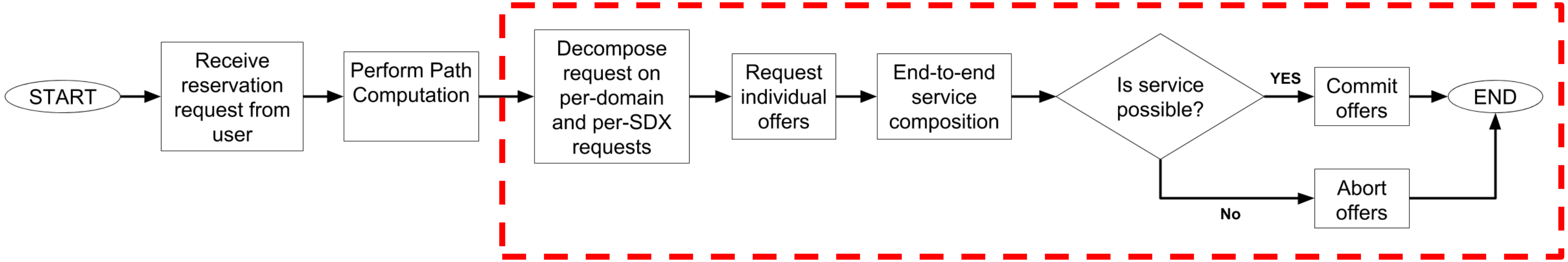
Architecture Overview



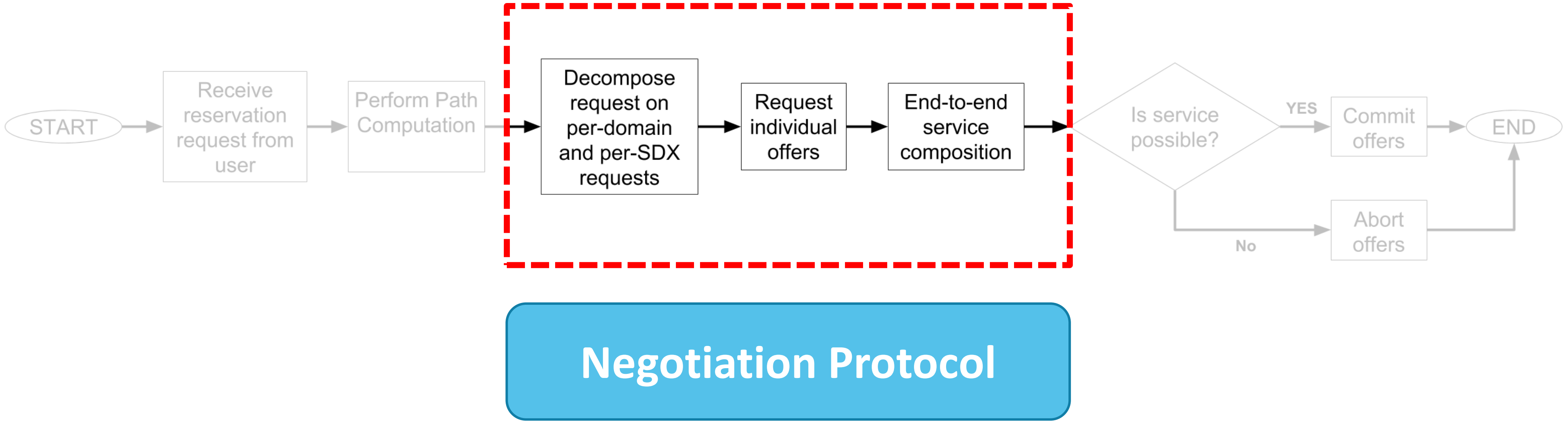
Agenda

1. Motivation
2. Background
3. Related Work
4. Architecture Overview
5. Design
6. Evaluation
7. Conclusions

Design – General Workflow



Design – Negotiation Protocol



Design – Negotiation

Types of Domains:

- **Visible domains:** provide bandwidth offers (query available bandwidth)
- **Blind domains:** cannot provide bandwidth offers (i.e., traditional advance reservation systems)

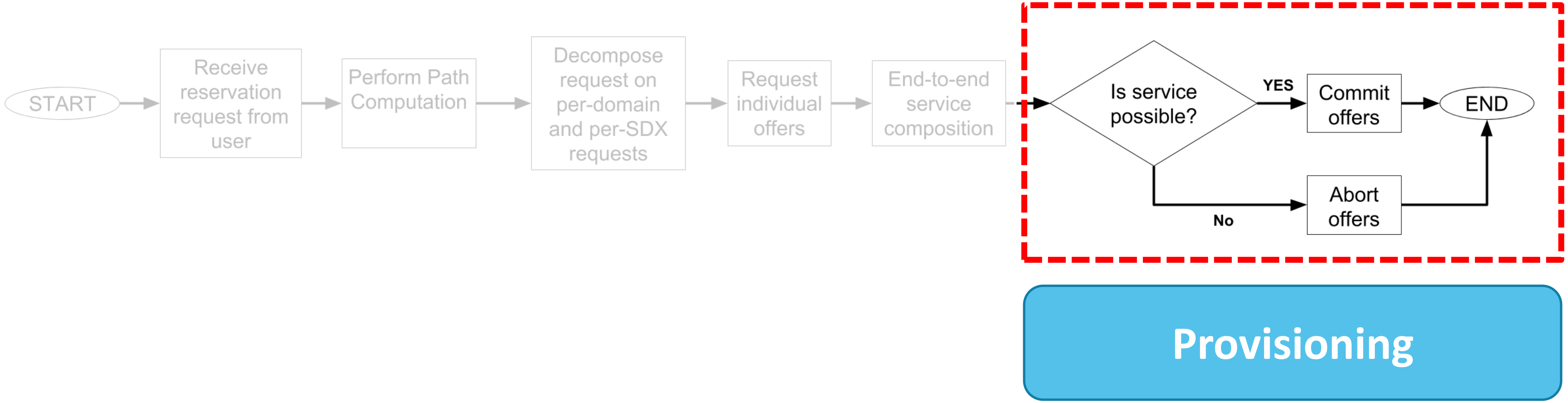
Visibility scenarios for a negotiation protocol considering *N* participant domains, with *M* visible domains and *N - M* blind domains:

1. **No visibility (M = 0):** All participant domains are blind domains
2. **Full visibility (M = N):** All participant domains are visible domains
3. **Partial visibility (M ≠ N):** blind domains and visible domains participate in the orchestration process

Negotiation Strategies

1. **Equal Splitting:** In this approach the orchestrator divides the original bandwidth request in equal parts among the participant domains
2. **Partial Offers:** In this approach the orchestrator contacts the visible domains for bandwidth offers. If the orchestrator is able to compose an end-to-end service with these offers only, the orchestrator provisions the offers. Otherwise, the orchestrator tries to request the remaining bandwidth from blind domains
3. **Full Offers:** In this approach the orchestrator contacts all participant domains for bandwidth offers. If the orchestrator is able to compose an end-to-end service with these offers, the orchestrator proceeds with provisioning, otherwise the reservation request fails

Design – Provisioning (SDX Rules)

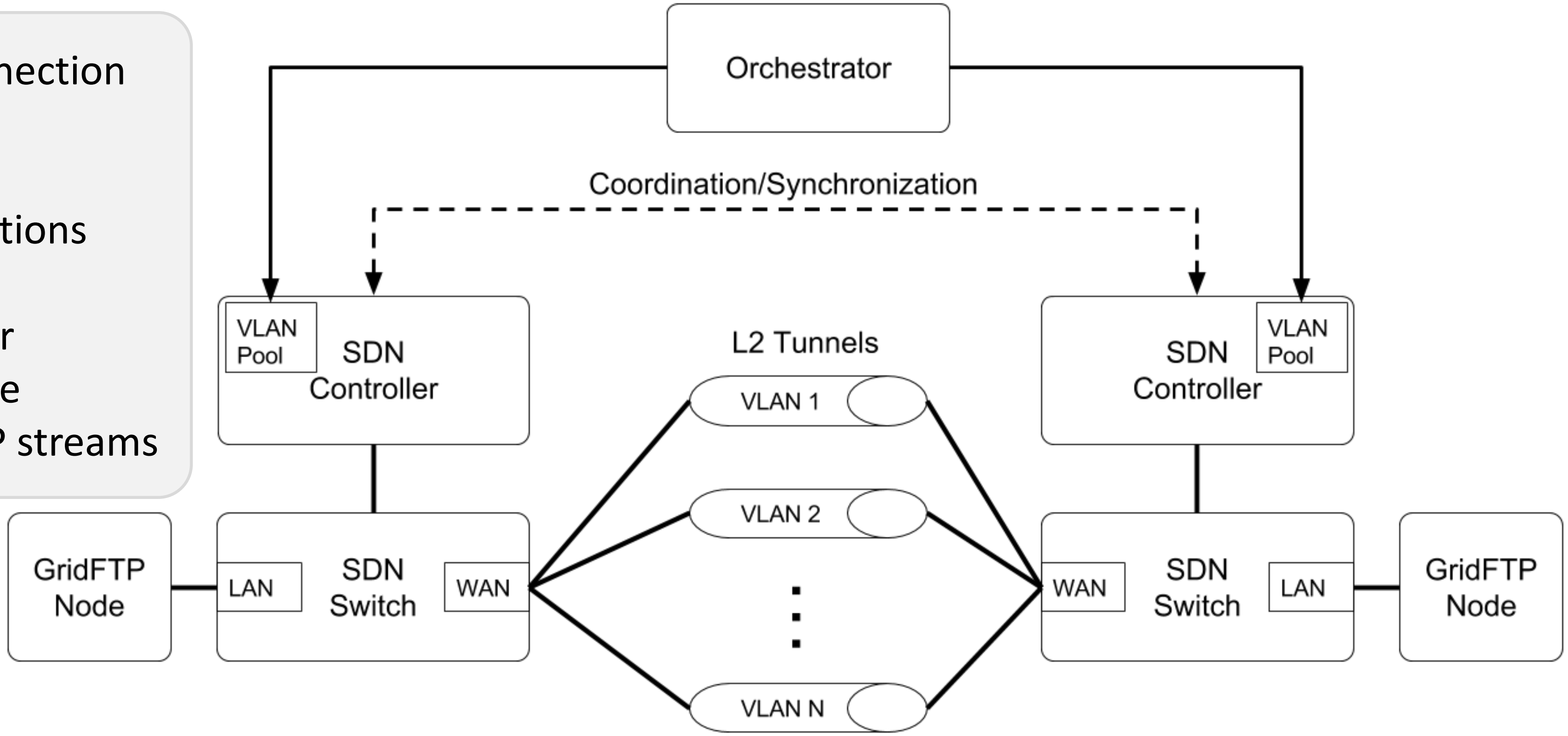


Design – SDX Rules Provisioning

SDX as interconnection points

Key insights:

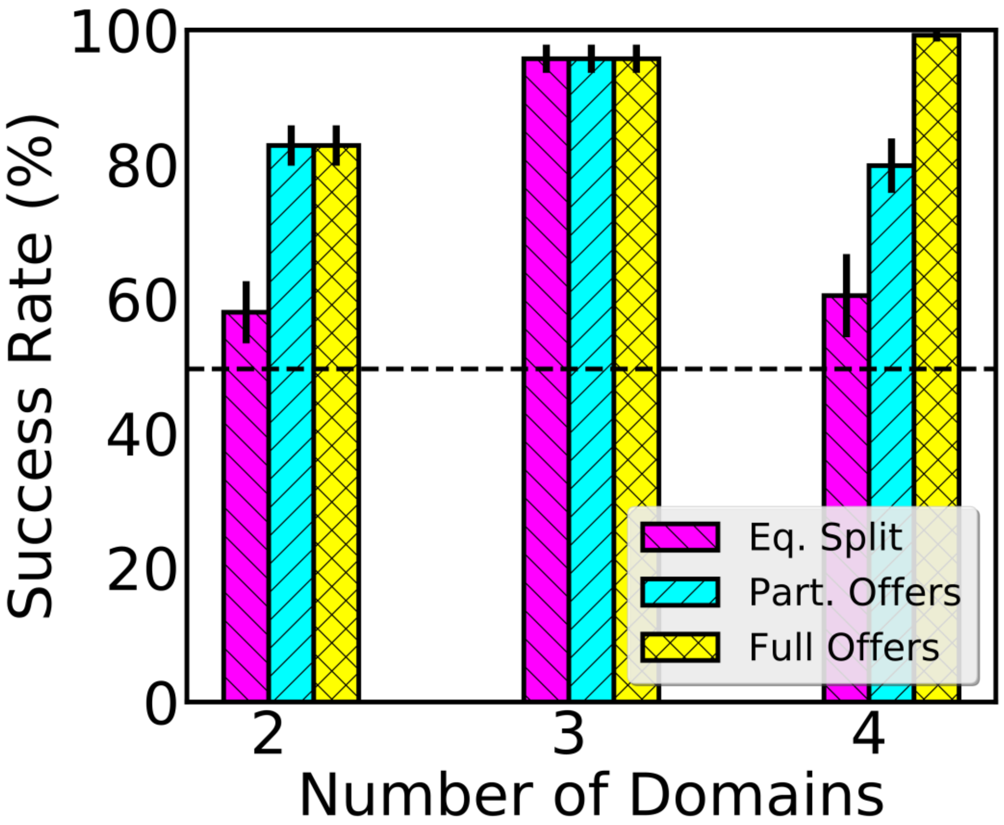
1. Adv. reservations over VLANs
2. Data transfer protocols use multiple TCP streams



Agenda

1. Motivation
2. Background
3. Related Work
4. Architecture Overview
5. Design
6. Evaluation
7. Conclusions

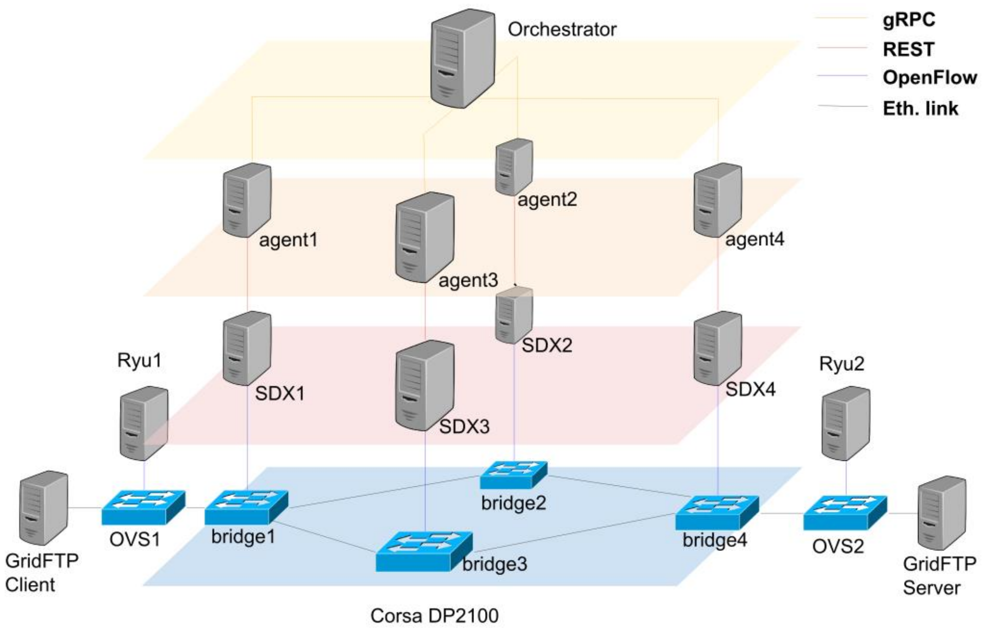
Evaluation – Negotiation Protocol



- Simulation of random user requests to an orchestrator with 2, 3, and 4 participant domains
- With 3 domains we obtained 95% success rate for any negotiation strategy
- Full offers can achieve 99% success rate with 4 domains/paths available

SDX Testbed Topology

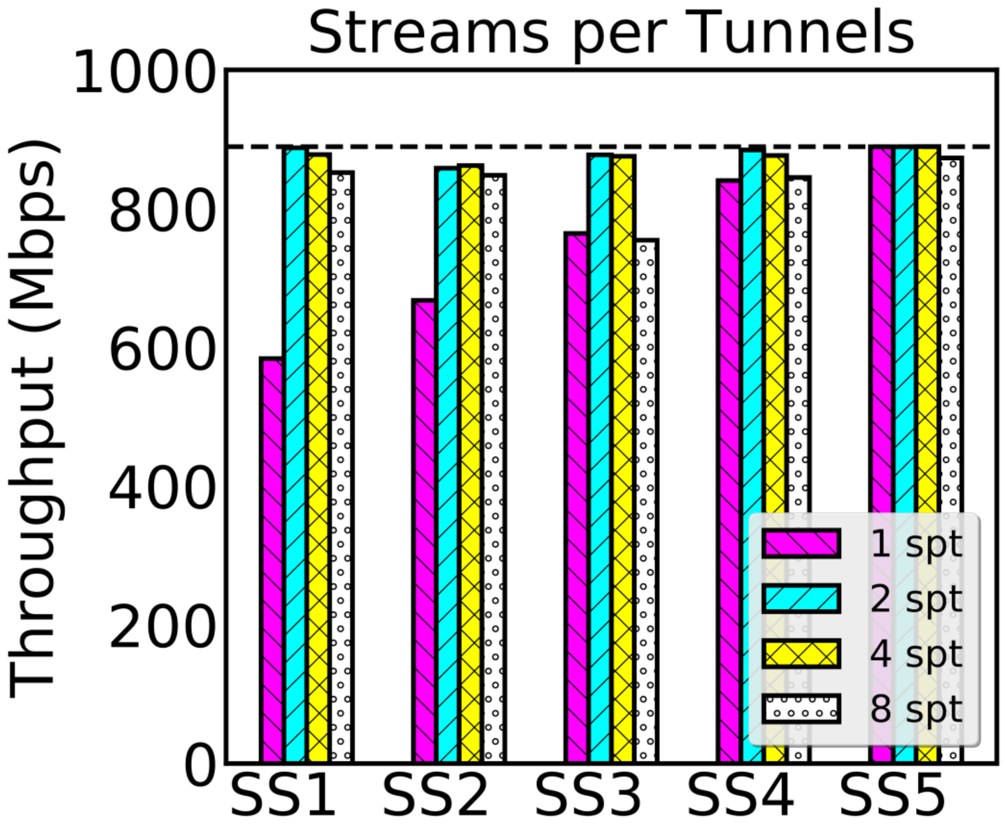
Equipment	Specifications
Corsa DP2100	OpenFlow 1.5, multiple flow tables, multi-context virtualization, 48 Gb packet buffer, 10 Gbps line-rate
Dell PowerEdge R220	Ubuntu Server 16.04, 16 GB RAM, four Intel(R) Xeon(R) CPU E3-1220 v3 @ 3.10GHz processors, four port Gigabit Ethernet card
Customized Supermicro	Ubuntu Server 16.04, 8 GB RAM, four Intel(R) Xeon(R) CPU X3430 @ 2.40GHz, two Gigabit Ethernet interfaces



90 ms RTT between endpoints

Bandwidth Splitting and TCP Streams

Code	Description
SS1	Tunnel 1: 100 Mbps, Tunnel 2: 900 Mbps
SS2	Tunnel 1: 200 Mbps, Tunnel 2: 800 Mbps
SS3	Tunnel 1: 300 Mbps, Tunnel 2: 700 Mbps
SS4	Tunnel 1: 400 Mbps, Tunnel 2: 600 Mbps
SS5	Tunnel 1: 500 Mbps, Tunnel 2: 500 Mbps



Agenda

1. Motivation
2. Background
3. Related Work
4. Architecture Overview
5. Design
6. Evaluation
7. Conclusions

Conclusions

Contributions

- An architecture for orchestrating international multi-path, multi-domain advance reservations in science networks and SDXs.
- Our orchestration architecture and negotiation protocols increases the reservation success rate from approximately 50% using single path to approximately 99% when four paths are available.
- Architectural approaches at the SDX level that enable novel science network services, while enhancing the performance of science data transfers over traditional approaches.

Future Work

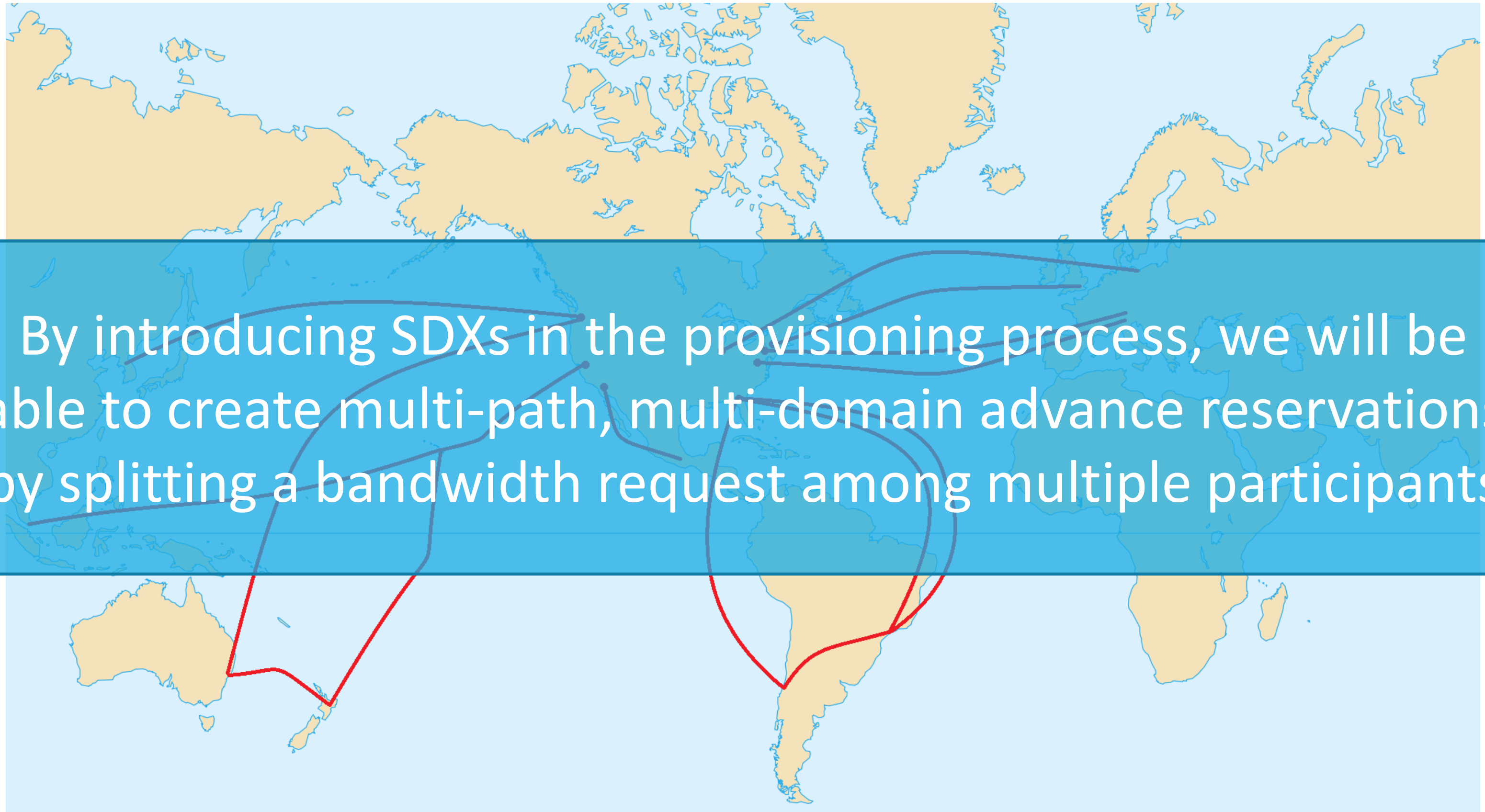
- Large scale deployments and evaluations
- Novel science network services: scheduled migrations, multipoint-to-multipoint advance reservations

References

1. S. Avallone, S. D'Antonio, M. Esposito, S. P. Romano, G. Ventre, Resource allocation in multi-domain networks based on service level specifications, Journal of Communications and Networks 8 (1) (2006) 106-115. doi:10.1109/JCN.2006.6182910.
2. J. Kempf, M. Korling, S. Baucke, S. Touati, V. McClelland, I. Mas, O. Backman, Fostering rapid, cross-domain service innovation in operator networks through service provider SDN, in: Communications (ICC), 2014 IEEE International Conference on, IEEE, 2014, pp. 3064-3069.
3. X. Wang, H. Schulzrinne, RNAP: A resource negotiation and pricing protocol, in: in International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV99), Basking, CiteSeer, 1999.
4. K. Czajkowski, I. Foster, C. Kesselman, V. Sander, S. Tuecke, SNAP: A Protocol for Negotiating Service Level Agreements and Coordinating Resource Management in Distributed Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2002, pp. 153-183. doi:10.1007/3-540-36180-4_9. URL https://doi.org/10.1007/3-540-36180-4_9
5. H. B. Newman, A. Barczyk, M. Bredel, OLiMPS. OpenFow link-layer multipath switching, Tech. rep., California Institute Technology, Pasadena, CA (United States) (2014).
6. J. M. Plante, D. A. P. Davis, V. M. Vokkarane, Parallel and survivable multipath circuit provisioning in ESNet’s OSCARS, Photonic Network Communications 30 (3) (2015) 363-375. doi:10.1007/s11107-015-0535-x. URL https://doi.org/10.1007/s11107-015-0535-x

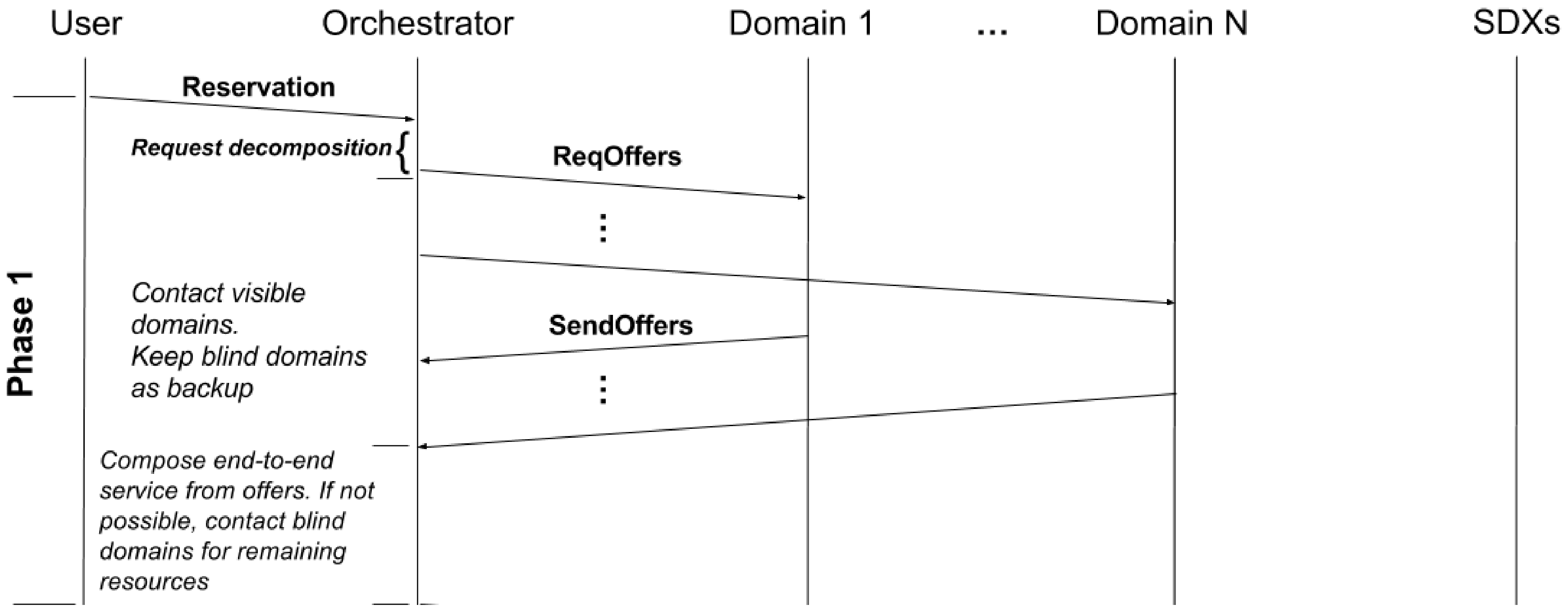
Thanks! Questions?

Bandwidth Splitting Service

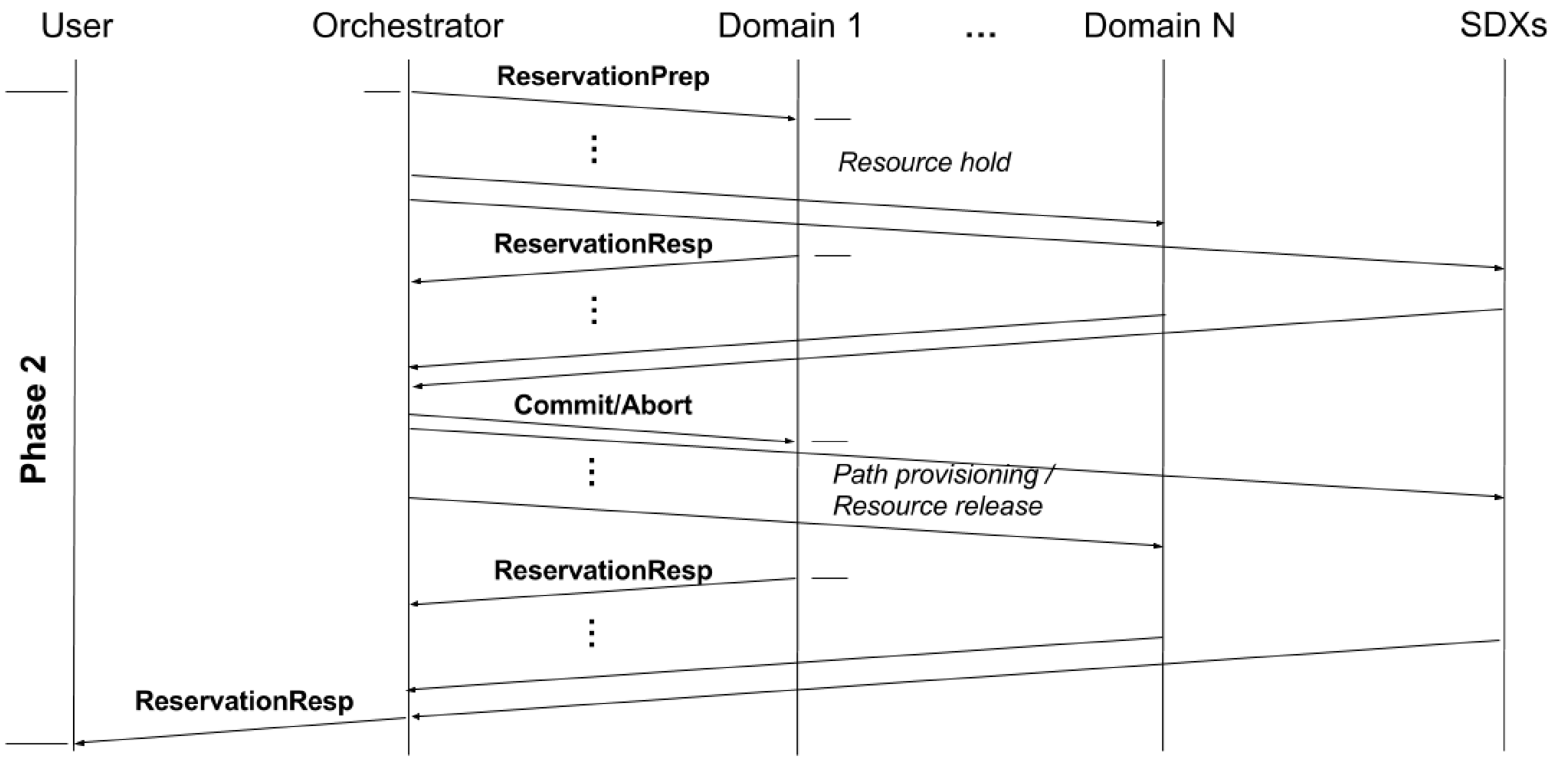


By introducing SDXs in the provisioning process, we will be able to create multi-path, multi-domain advance reservations by splitting a bandwidth request among multiple participants

Design – Negotiation Protocol



Design – Negotiation Protocol



Orchestrator Implementation

Written in Python using an agent-based approach

- We control the WAN communication channel
- Site controller can provide their own API

Orchestrator communicates with the agents using the general remote procedure call (gRPC) protocol

- HTTP/2
- Protocol buffers

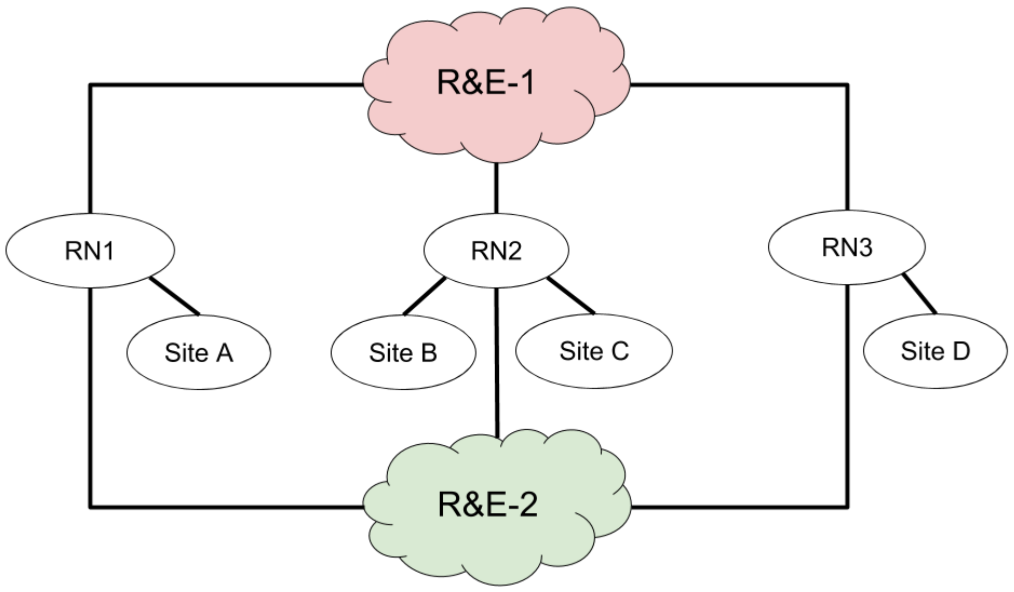
SDX Implementation

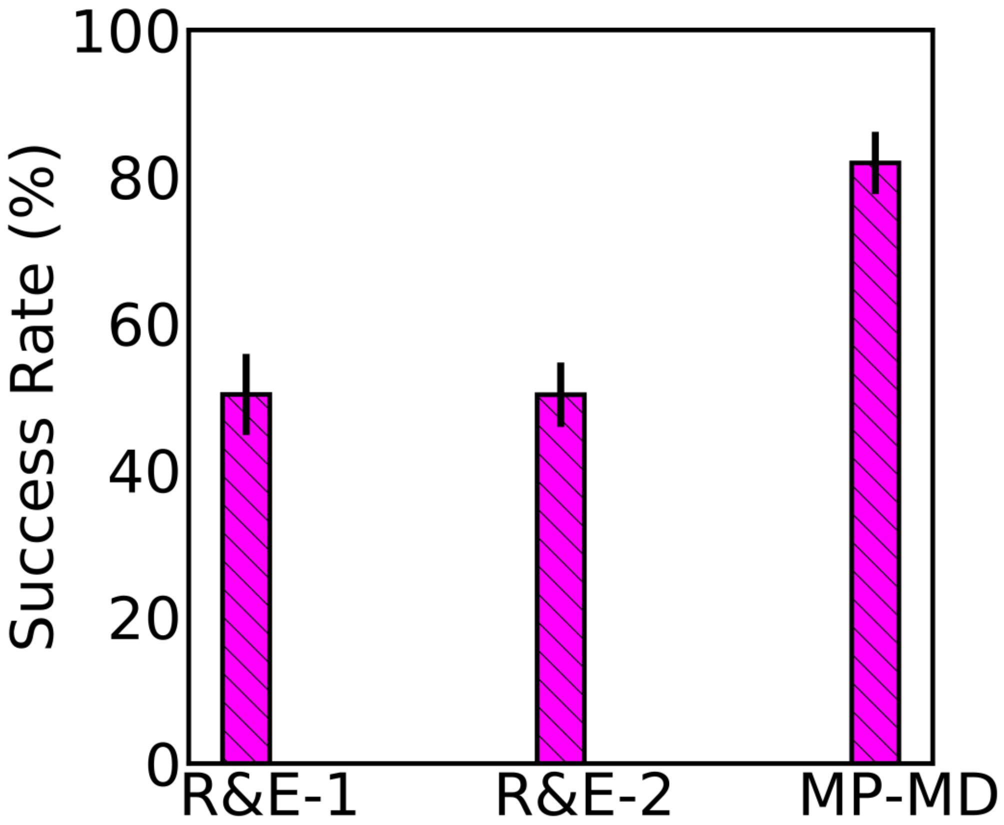
AtlanticWave/SDX controller: written in Python, using the Ryu SDN Framework, and OpenFlow

- REST API
- L2 Tunnels over VLANs
- Bandwidth offers

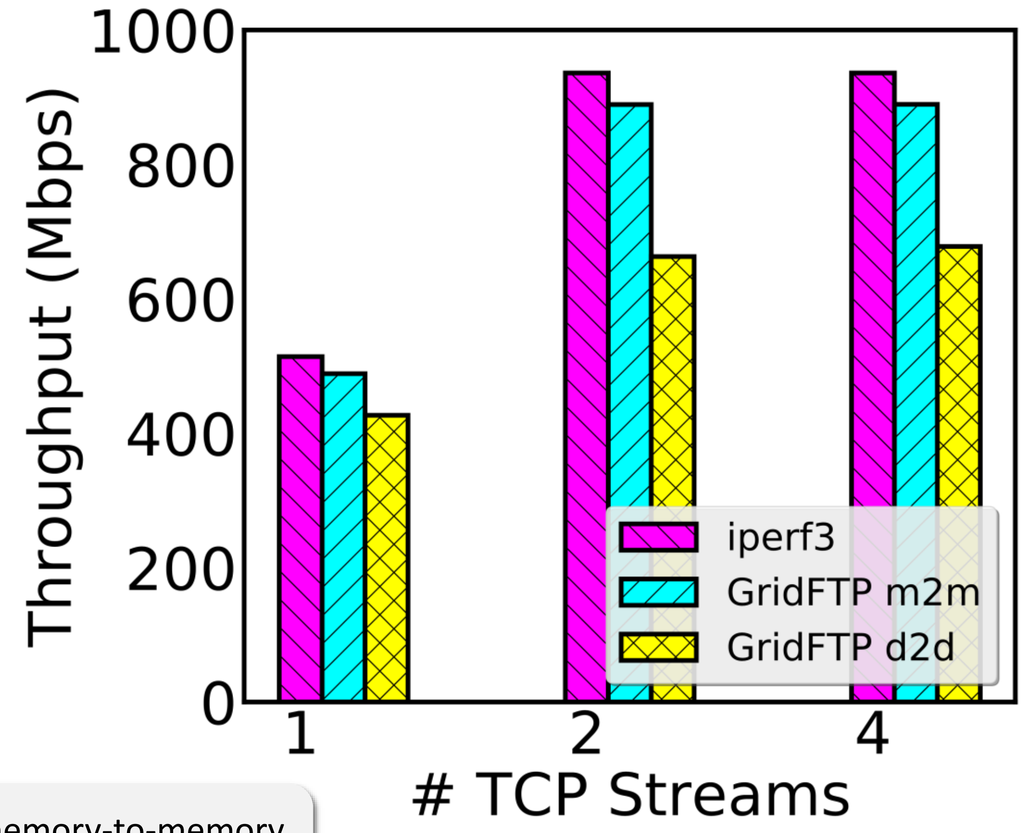
Ryu SDN controller + Open vSwitch (OVS) at each end for bandwidth splitting and aggregation

Evaluation – Negotiation Protocol

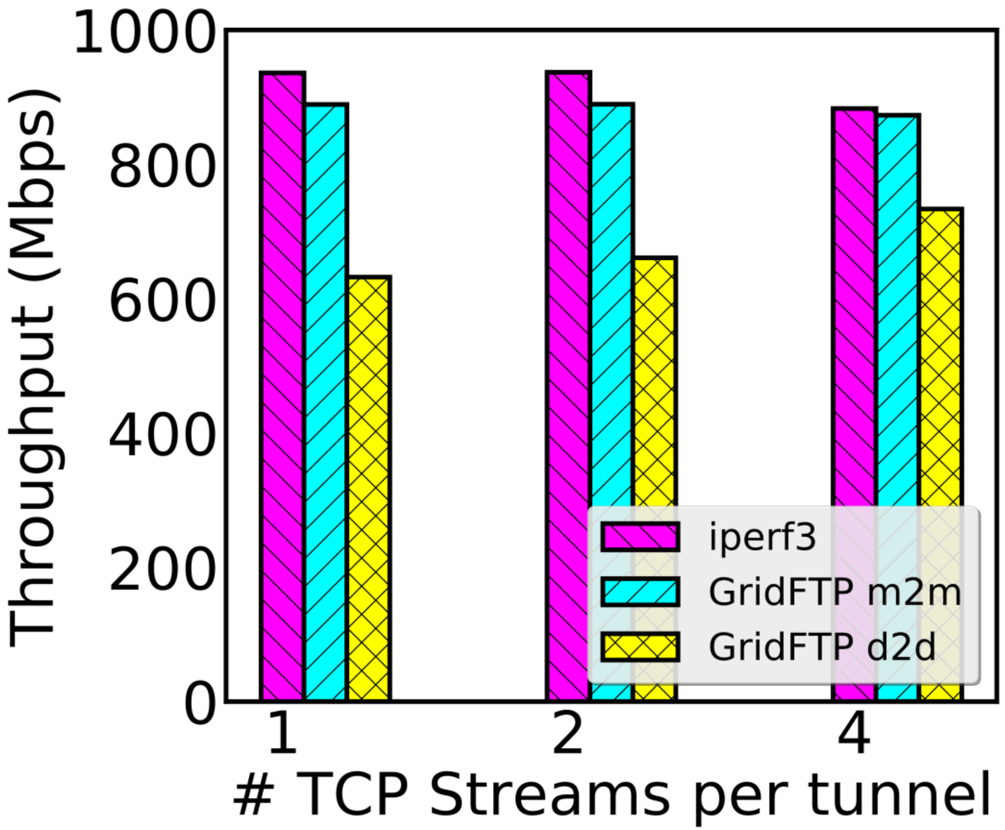




Throughput Baseline



Single L2 Tunnel @ 1 Gbps



Two L2 Tunnels @ 500 Mbps

m2m: memory-to-memory
d2d: disk-to-disk