

iSDX: An Industrial-Scale Software-Defined IXP

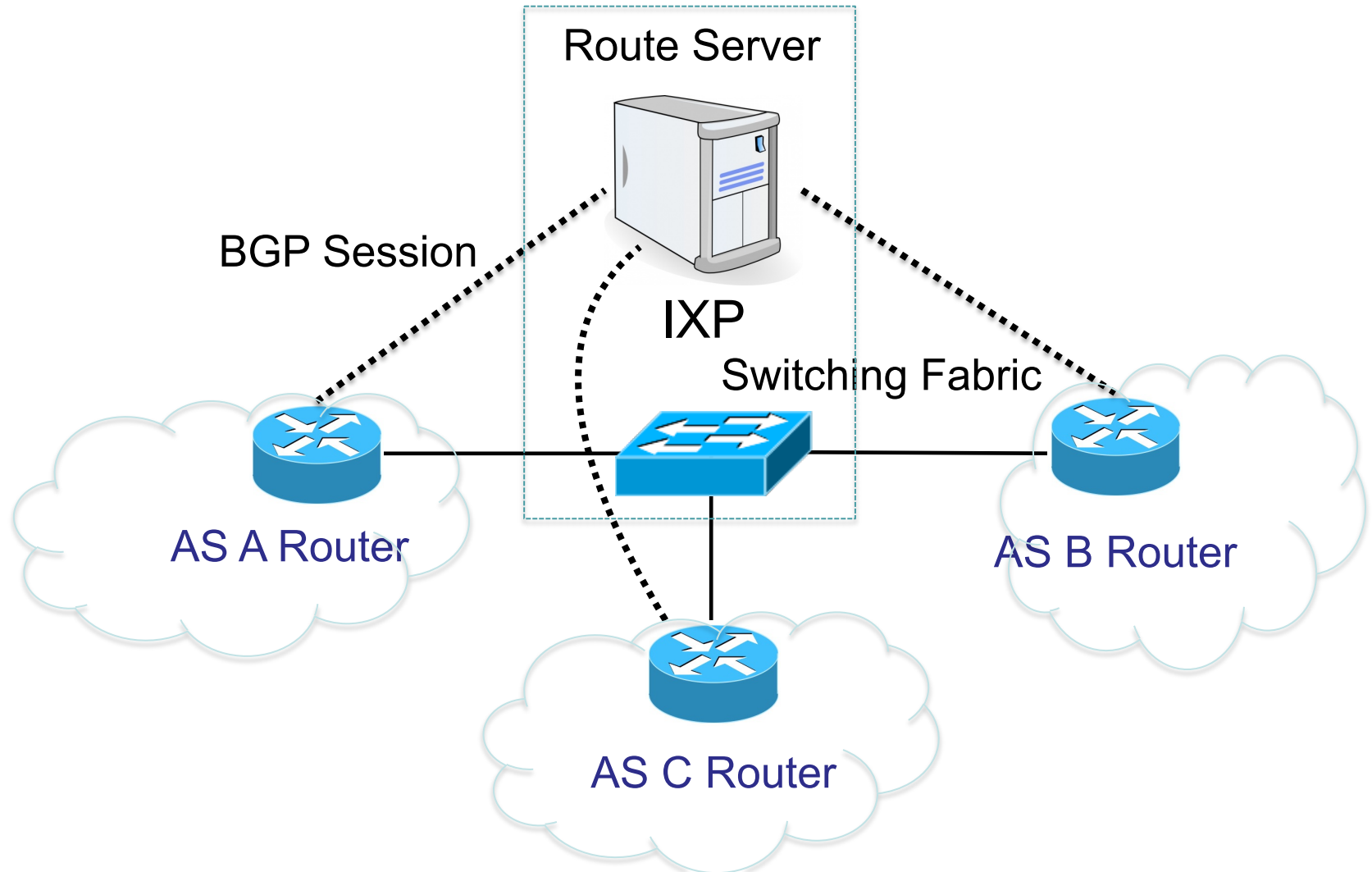
Arpit Gupta

Princeton University

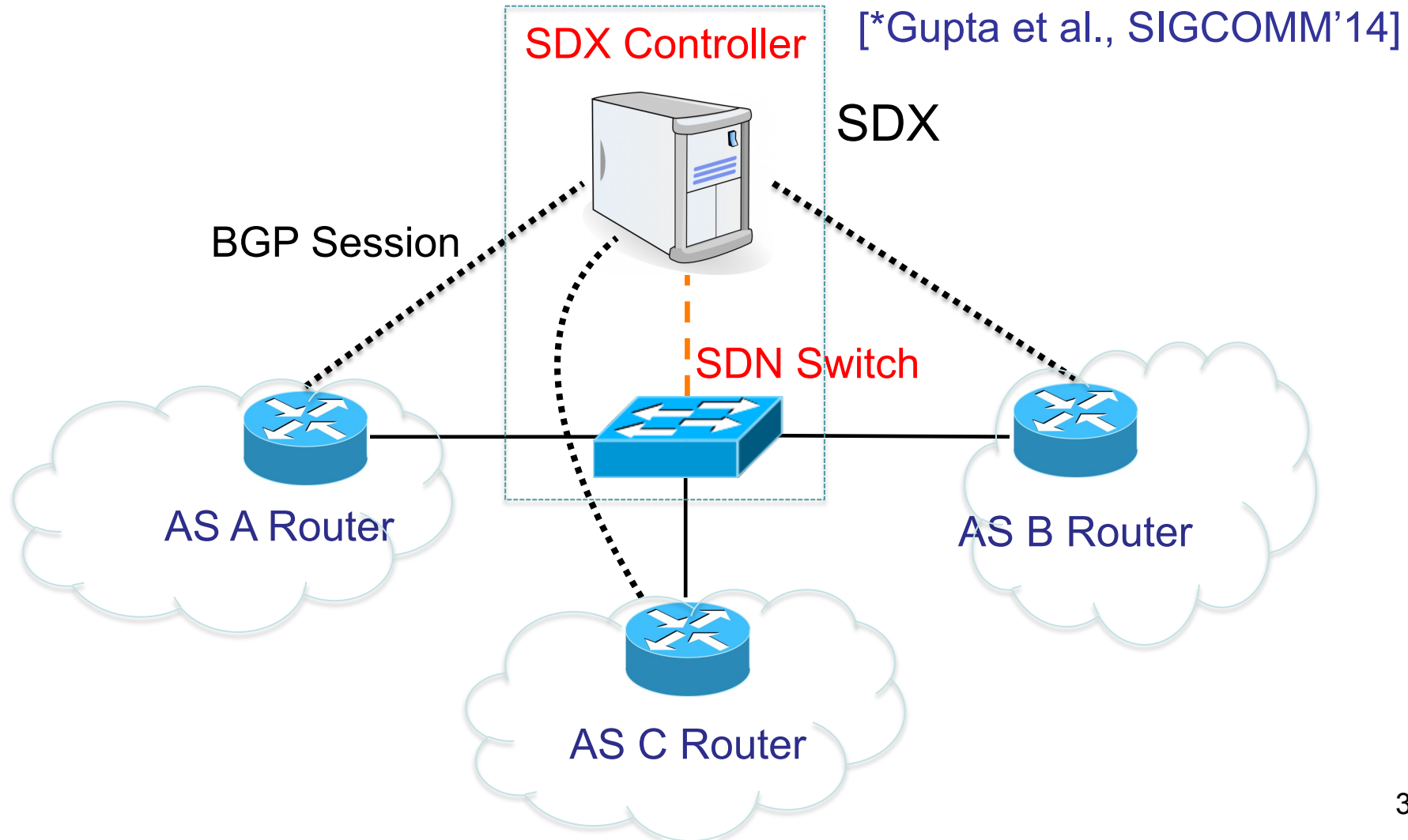
<http://sdx.cs.princeton.edu>

Robert MacDavid, Rüdiger Birkner, Marco Canini,
Nick Feamster, Jennifer Rexford, Laurent Vanbever

Internet Exchange Points (IXPs)



Software Defined IXPs (SDXs)



Deployment Ready SDX is Hard!


- **Deployment Experience:**

- Inter-Agency Exchange
- Large IXP in Europe
- Smaller IXPs in Asia



- **Challenges:**

- **Data Plane Scalability**
- ...

Data Plane Scalability Challenges

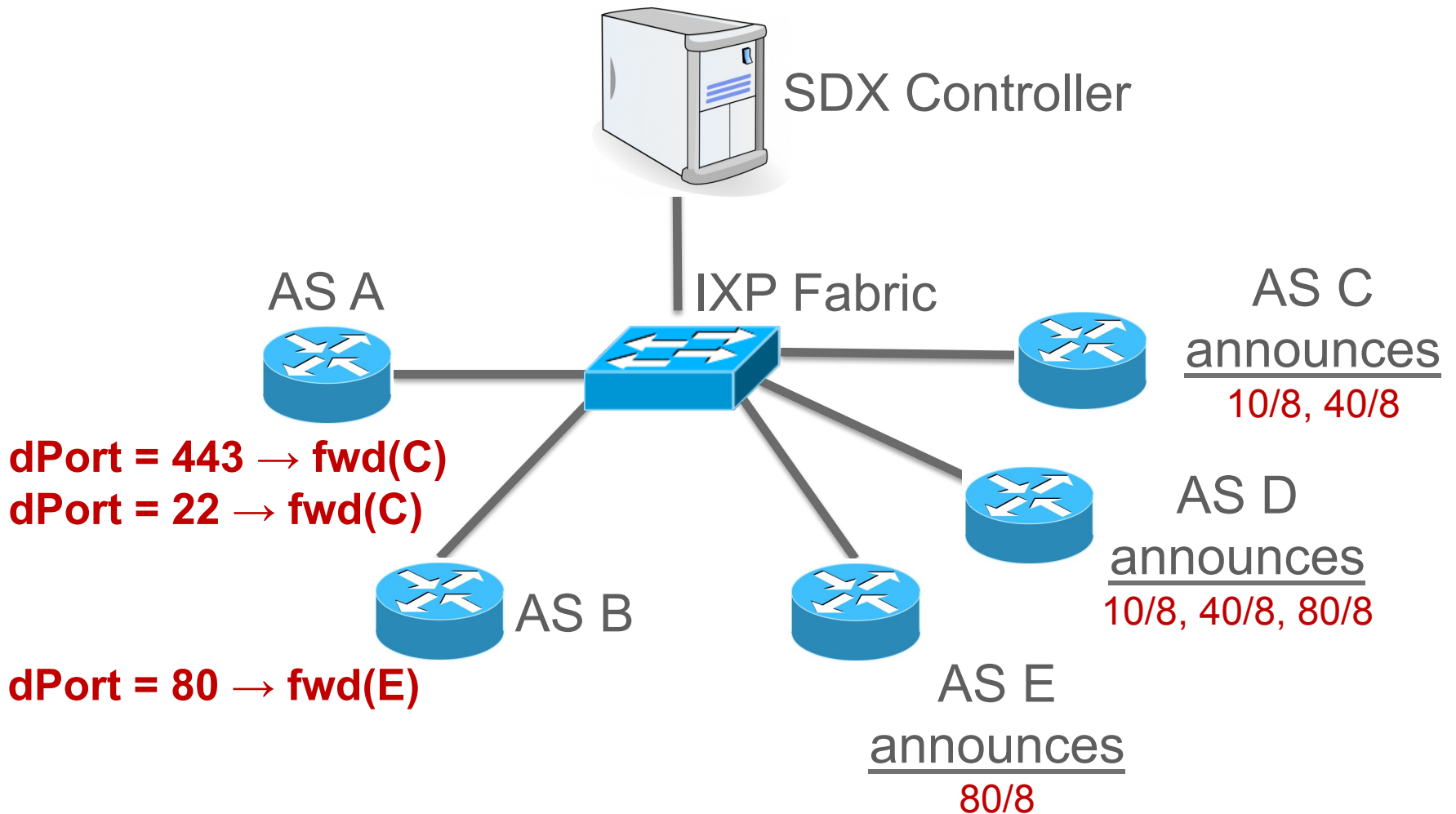
Devices	Operations	Data Plane Performance	
		State (# entries)	Update Rate (flow-mods/s)
	Match-Action on Multiple Headers	100K	2,500

Data Plane Scalability Challenges

Devices	Operations	Data Plane Performance	
		State (# entries)	Update Rate (flow-mods/s)
	Match-Action on Multiple Headers	100K	2,500
	Matches on IP Prefixes only	~1M	N/A

Problem: Optimize the usage of available devices

Simple Example



Forwarding Table Entries at SDX

SDN Policies	# Forwarding Table Entries	
dPort = 443 → fwd(C)	1	AS A
dPort = 22 → fwd(C)	1	
dPort = 80 → fwd(E)	1	AS B

Number of forwarding table entries for
A & B's Outbound SDN Policies

Goal Tracker

	Simple Example
Baseline	3

- **Large IXP Dataset:**
 - BGP RIBs & Updates from large IXP
 - 511 IXP participants
 - 96 million peering routes for 300K IP prefixes
 - 25K BGP updates for 2-hour duration

Goal Tracker

	Simple Example	Large IXP
Baseline	3	62K

- **Large IXP Dataset:**
 - BGP RIBs & Updates from large IXP
 - 511 IXP participants
 - 96 million peering routes for 300K IP prefixes
 - 25K BGP updates for 2-hour duration

Goal Tracker

	Simple Example	Large IXP
Baseline	3	62K

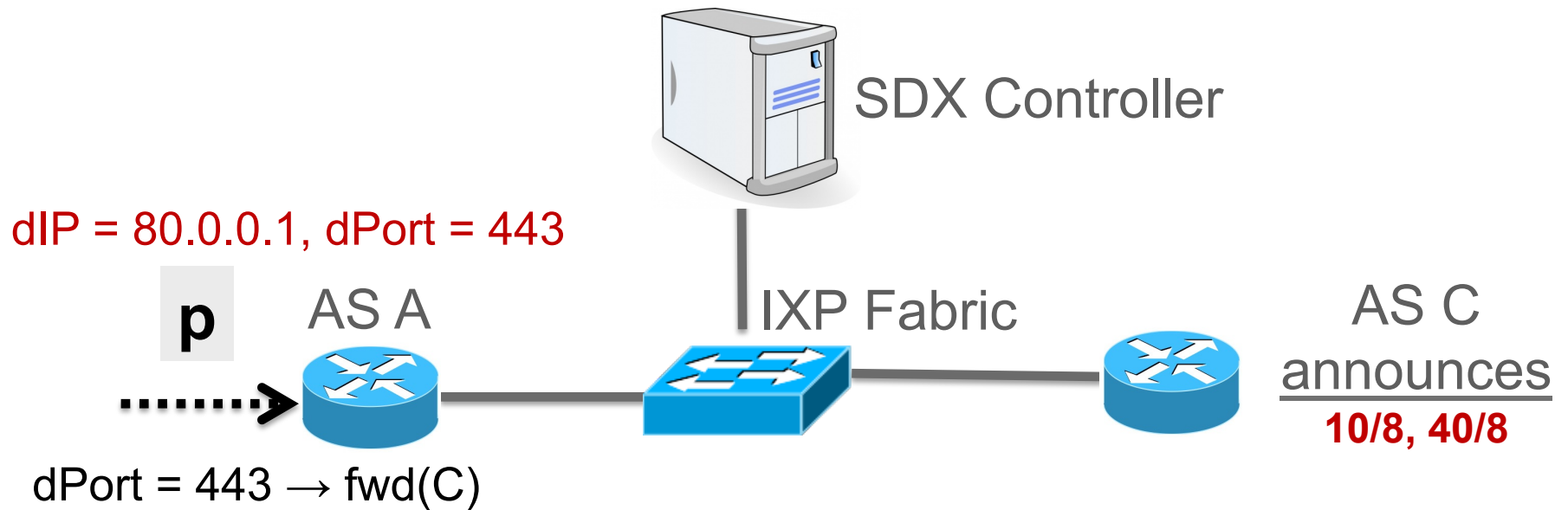
Satisfies design goals, but ...

Goal Tracker

	Simple Example	Large IXP
Baseline	3	62K

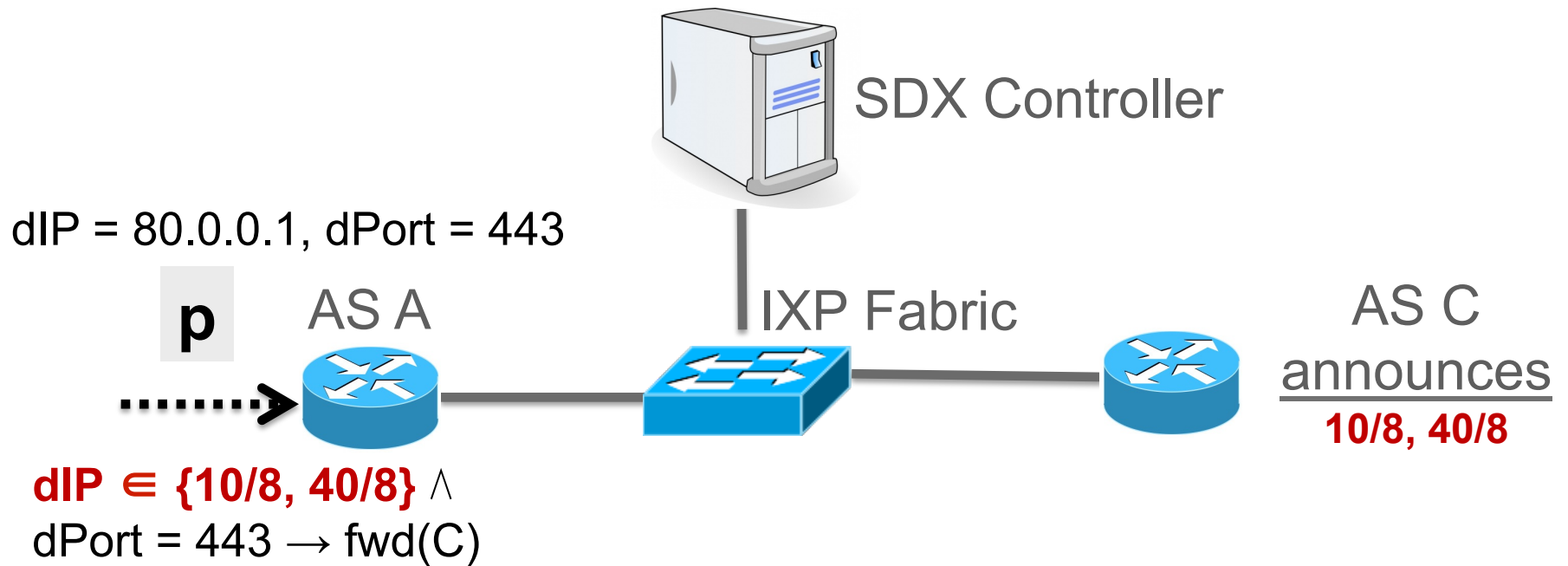
... not congruent with BGP!

Challenge: Congruence with BGP



Ensure **p** is not forwarded to C


Solution: SDN Policy Augmentation

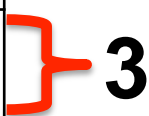


Match on prefixes advertised by C

Data Plane State Explosion!

SDN Policies	# Forwarding Table Entries		
	10/8	40/8	80/8
dPort = 443 → fwd(C)	1	1	0
dPort = 22 → fwd(C)	1	1	0
dPort = 443 → fwd(D)	1	1	1

 **4**

 **3**

SDN Policy Augmentation increases forwarding table entries

Goal Tracker

	Simple Example	Large IXP
Baseline	3	62K
Policy Augmentation	7	68M

Not possible to support these many forwarding table entries!

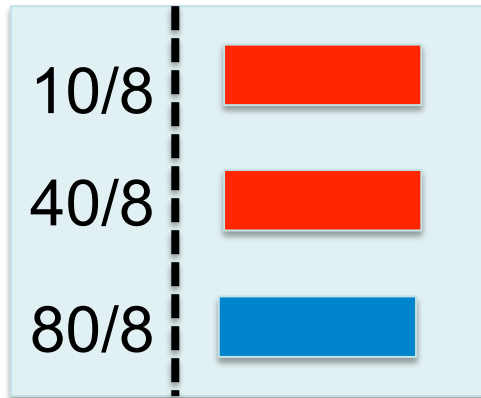
Forwarding Equivalence Classes

SDN Policies	# Forwarding Table Entries		
	10/8	40/8	80/8
dPort = 443 → fwd(C)	1	1	0
dPort = 22 → fwd(C)	1	1	0
dPort = 443 → fwd(D)	1	1	1

10/8, 40/8 exhibit similar forwarding behavior

Leveraging Forwarding Equivalence

forward to
BGP Next Hop

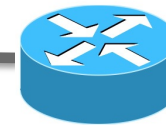


AS A

dPort = 443 → fwd(C)



IXP Fabric



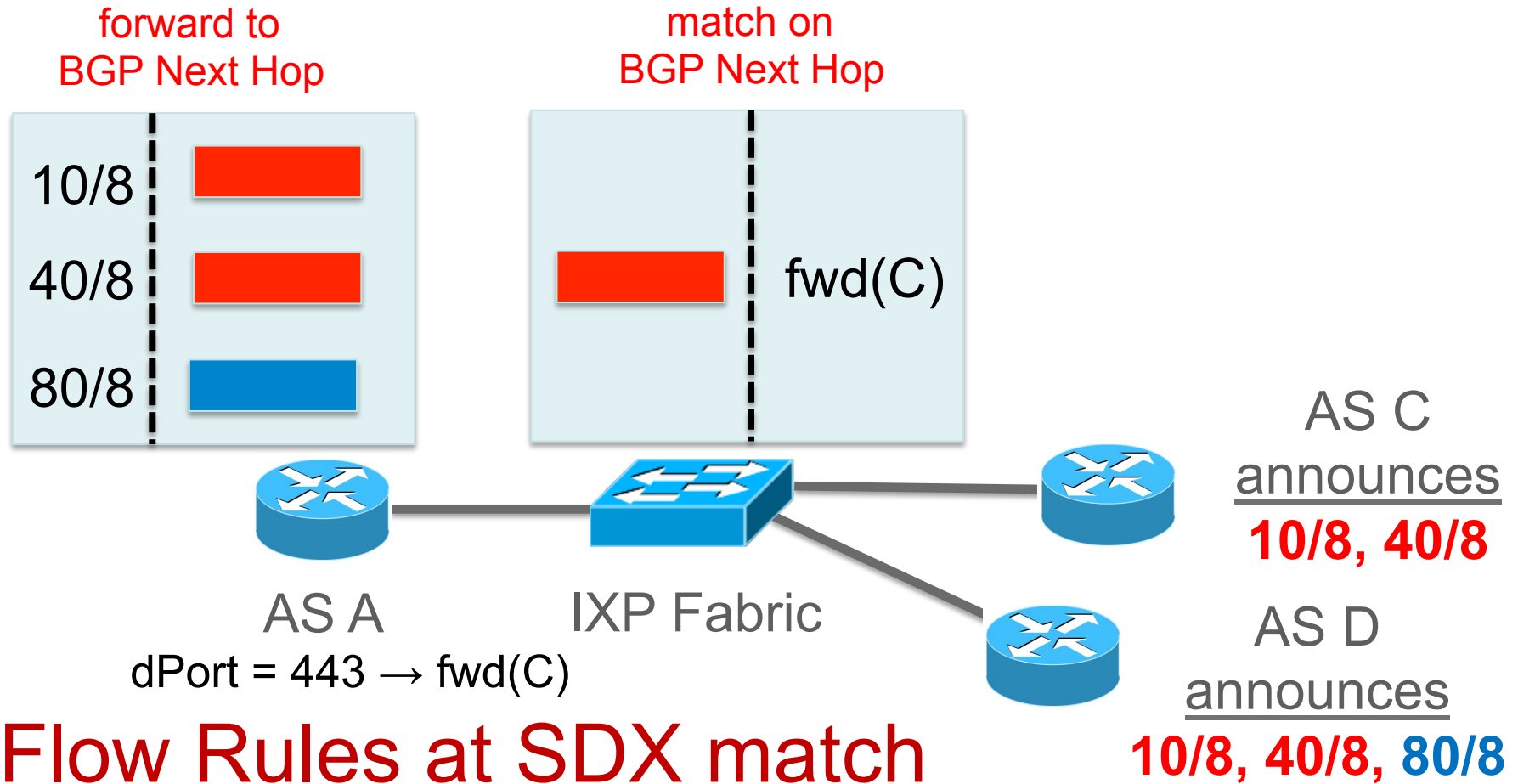
AS C
announces
10/8, 40/8



AS D
announces
10/8, 40/8, 80/8

**Single BGP Next Hop for
10/8, 40/8**

Leveraging Forwarding Equivalence



Flow Rules at SDX match on BGP Next Hops

Goal Tracker

	Simple Example	Large IXP
Baseline	3	62K
Policy Augmentation	7	68M
*FEC Computation	4	21M

[*Gupta et al., SIGCOMM'14]

Still not possible to support these many forwarding table entries!

More Efficient FEC Computation

SDN Policies	# Forwarding Table Entries	
	{10/8, 40/8}	80/8
dPort = 443 → fwd(C)	1	0
dPort = 22 → fwd(C)	1	0
dPort = 443 → fwd(D)	1	1

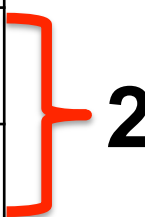
Independent FEC Computation
can be more efficient

Partitioning FEC Computation


- Large number of SDX participants
 - Many different policies on groups of prefixes
 - Leads to a large number of small FECs of prefixes
- Compute FECs independently
 - Separate computation per participant
 - Leads to small number of large FECs, and less frequent recomputation
 - Enables “scale out” of the FEC computation

FEC Computation Partitioning in Action

SDN Policies	# Forwarding Table Entries	
	{10/8, 40/8}	80/8
dPort = 443 → fwd(C)	1	0
dPort = 22 → fwd(C)	1	0

 2

dPort = 443 → fwd(D)	1
----------------------	---

 1

A & B independently compute FECs

Goal Tracker

	Simple Example	Large IXP
Baseline	3	62K
Policy Augmentation	7	68M
FEC Computation	4	21M
Independent FEC Computation	3	763K

**Also requires support for
15K flow-mods/seconds**

Undesired BGP & SDN Coupling

SDN Policies	# Forwarding Table Entries		
	10/8	40/8	80/8
dPort = 443 → fwd(C)	1	1	0
dPort = 22 → fwd(C)	1	1	0

dPort = 443 → fwd(D)	1 → 0	1	1
----------------------	-------	---	---

Incoming BGP Update:
{AS D withdraws route for prefix 10/8}

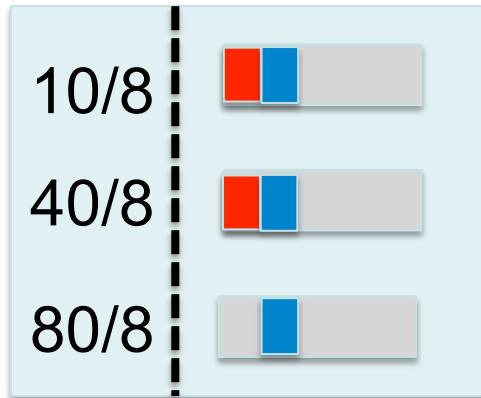
Decoupling BGP from SDN Policies

- Leverage advances in commodity hw switches
 - Support for Bitmask Matching (L2 headers)
- Extend BGP “next hop” encoding
 - So far: encode FECs (single field)
 - New idea: encode **reachability bitmask** (multi field)
- Changing only the BGP announcements
 - No need to update the SDX data plane!

Reachability Bitmask in Action

forward to
BGP Next Hop

Dedicate one bit per participant



Reachable via AS C

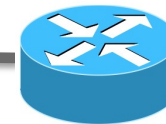


AS A

dPort = 443 → fwd(C)



IXP Fabric



AS C

announces
10/8, 40/8

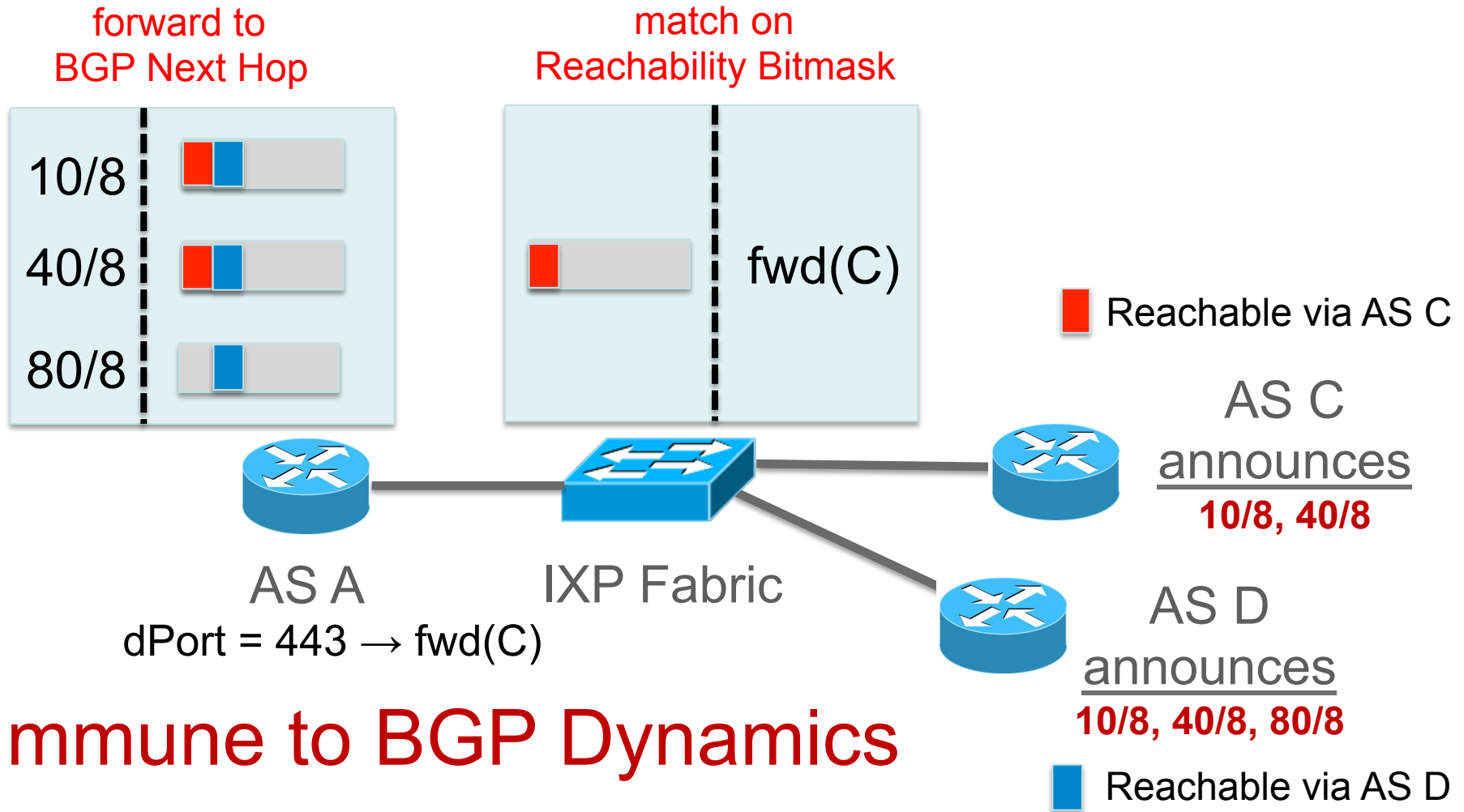


AS D

announces
10/8, 40/8, 80/8

Reachable via AS D

Reachability Bitmask in Action



Immune to BGP Dynamics

Reachability Bitmask in Action

SDN Policies	# Forwarding Table Entries
	C
dPort = 443 → fwd(C)	1
dPort = 22 → fwd(C)	1

.....

dPort = 443 → fwd(D)	1
----------------------	---

Reduces Data Plane State

Goal Tracker

	Simple Example	Large IXP
Baseline	3	62K
Policy Augmentation	7	68M
FEC Computation	4	21M
Independent FEC Computation	3	763K
Reachability Encoding	3	65K

We can now run SDX over commodity hardware switches

iSDX Evaluation Summary

- **Data Plane State:**
 - Requires **65K < 100K** forwarding table entries
- **Data Plane Update Rate:**
 - Requires **0 < 2500** flow-mods/second
- **Other Goals:**
 - Processes BGP update bursts in real time (**50 ms**)
 - Requires only **360 BGP Next Hops** compared to 25K from previous solutions

You Can Run iSDX Today!

<http://sdx.cs.princeton.edu>

- Running code
 - Vagrant & Docker based setup
 - Instructions to run with **Hardware Switches**
- ONF's Open Source SDN
 - Community: <https://community.opensourcesdn.org/wg/iSDX/dashboard>
 - Mailing List
isdx@community.OpenSourceSDN.org