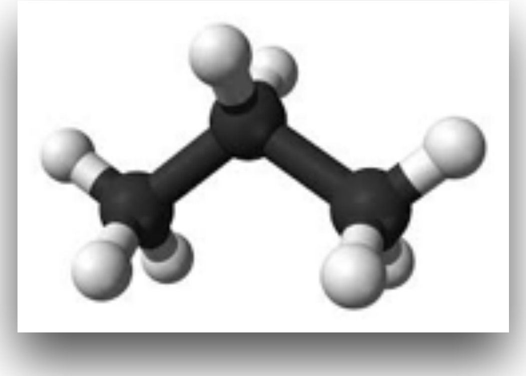
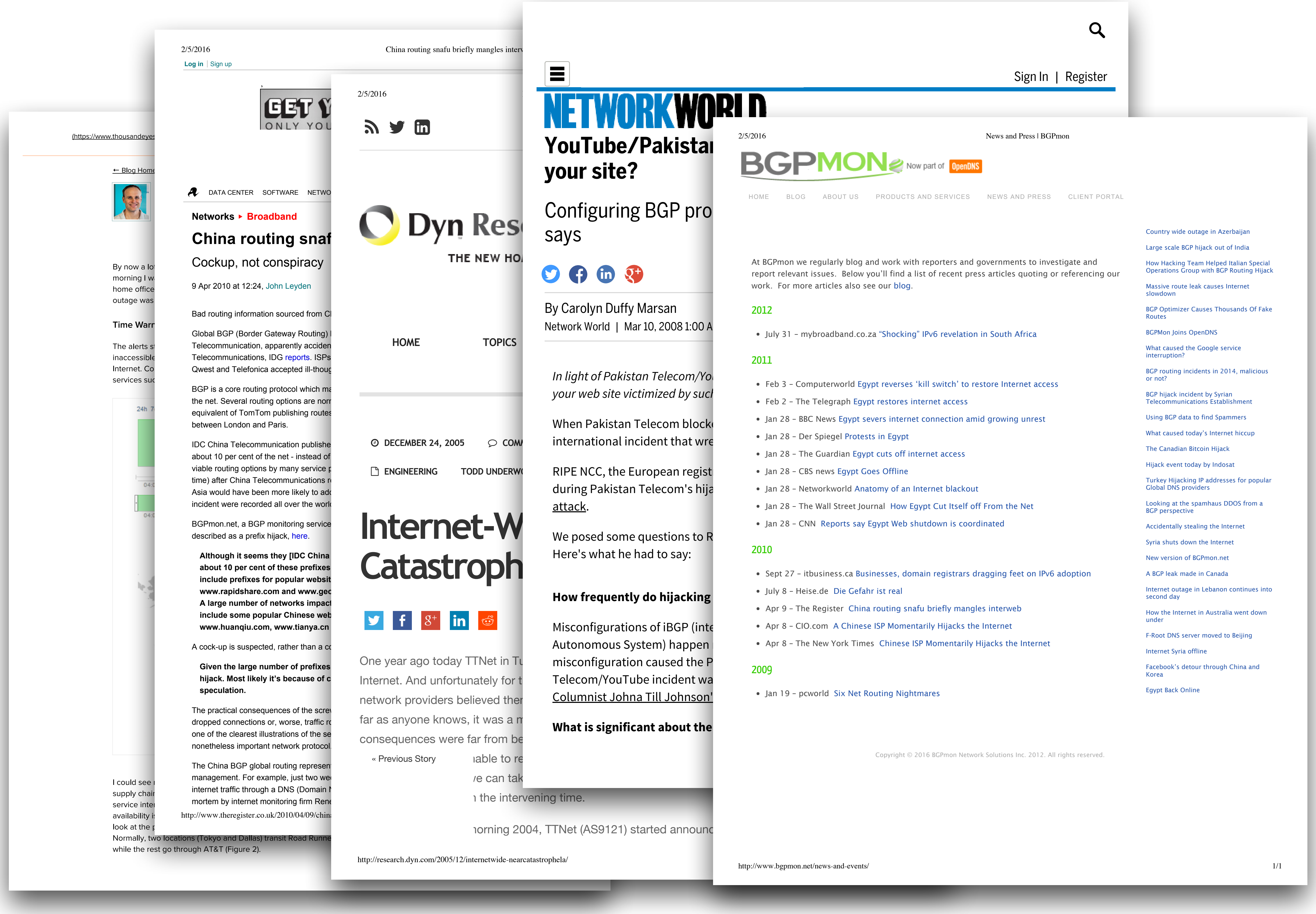


Propane

Programming Distributed Control Planes

Ryan Beckett (Princeton & MSR)
Ratul Mahajan (MSR)
Todd Millstein (UCLA)
Jitu Padhye (MSR)
David Walker (Princeton)

Configuring Networks is Error-Prone



(https://www.thousandeyes

← Blog Home

By now a lot
morning I w
home office
outage was

Time Warr

The alerts s
inaccessible
Internet. Co
services suc

I could see
supply chain
service inter
availability is
look at the p
Normally, two locations (Tokyo and Dallas) transit Road Runne
while the rest go through AT&T (Figure 2).

2/5/2016 China routing snafu briefly mangles interw

Log in | Sign up

GET Y
ONLY YOU

DATA CENTER SOFTWARE NETWO

Networks ▸ Broadband

China routing snaf

Cockup, not conspiracy

9 Apr 2010 at 12:24, John Leyden

Bad routing information sourced from Cl

Global BGP (Border Gateway Routing) I
Telecommunication, apparently acciden
Telecommunications, IDG reports. ISPs
Qwest and Telefonica accepted ill-thoug

BGP is a core routing protocol which ma
the net. Several routing options are norr
equivalent of TomTom publishing routes
between London and Paris.

IDC China Telecommunication publishe
about 10 per cent of the net - instead of
viable routing options by many service p
time) after China Telecommunications r
Asia would have been more likely to add
incident were recorded all over the world

BGPmon.net, a BGP monitoring service
described as a prefix hijack, here.

**Although it seems they [IDC China
about 10 per cent of these prefixes
include prefixes for popular websit
www.rapidshare.com and www.geo
A large number of networks impact
include some popular Chinese web
www.huanqiu.com, www.tianya.cn**

A cock-up is suspected, rather than a co

**Given the large number of prefixes
hijack. Most likely it's because of c
speculation.**

The practical consequences of the screw
dropped connections or, worse, traffic ro
one of the clearest illustrations of the se
nonetheless important network protocol.

The China BGP global routing represent
management. For example, just two wee
internet traffic through a DNS (Domain N
mortem by internet monitoring firm Rene

http://www.theregister.co.uk/2010/04/09/china

2/5/2016

Dyn Rese

THE NEW HO

HOME TOPICS

DECEMBER 24, 2005 COMM

ENGINEERING TODD UNDERWO

Internet-W
Catastroph

One year ago today TTNet in Tu
Internet. And unfortunately for t
network providers believed ther
far as anyone knows, it was a m
consequences were far from be
« Previous Story able to re
we can tak
n the intervening time.

norning 2004, TTNet (AS9121) started announc

http://research.dyn.com/2005/12/internetwide-nearcatastrophela/

Sign In | Register

NETWORKWORLD

YouTube/Pakistan
your site?

Configuring BGP pro
says

By Carolyn Duffy Marsan
Network World | Mar 10, 2008 1:00 A

*In light of Pakistan Telecom/Yo
your web site victimized by such*

When Pakistan Telecom blocke
international incident that wre

RIPE NCC, the European regist
during Pakistan Telecom's hija
attack.

We posed some questions to R
Here's what he had to say:

How frequently do hijacking

Misconfigurations of iBGP (inte
Autonomous System) happen
misconfiguration caused the P
Telecom/YouTube incident wa
Columnist Johna Till Johnson'

What is significant about the

2/5/2016 News and Press | BGPmon

BGPMON Now part of OpenDNS

HOME BLOG ABOUT US PRODUCTS AND SERVICES NEWS AND PRESS CLIENT PORTAL

At BGPmon we regularly blog and work with reporters and governments to investigate and report relevant issues. Below you'll find a list of recent press articles quoting or referencing our work. For more articles also see our blog.

2012

- July 31 – mybroadband.co.za "Shocking" IPv6 revelation in South Africa

2011

- Feb 3 – Computerworld Egypt reverses 'kill switch' to restore Internet access
- Feb 2 – The Telegraph Egypt restores internet access
- Jan 28 – BBC News Egypt severs internet connection amid growing unrest
- Jan 28 – Der Spiegel Protests in Egypt
- Jan 28 – The Guardian Egypt cuts off internet access
- Jan 28 – CBS news Egypt Goes Offline
- Jan 28 – Networkworld Anatomy of an Internet blackout
- Jan 28 – The Wall Street Journal How Egypt Cut Itself off From the Net
- Jan 28 – CNN Reports say Egypt Web shutdown is coordinated

2010

- Sept 27 – itbusiness.ca Businesses, domain registrars dragging feet on IPv6 adoption
- July 8 – Heise.de Die Gefahr ist real
- Apr 9 – The Register China routing snafu briefly mangles interweb
- Apr 8 – CIO.com A Chinese ISP Momentarily Hijacks the Internet
- Apr 8 – The New York Times Chinese ISP Momentarily Hijacks the Internet

2009

- Jan 19 – pcworld Six Net Routing Nightmares

Country wide outage in Azerbaijan

Large scale BGP hijack out of India

How Hacking Team Helped Italian Special Operations Group with BGP Routing Hijack

Massive route leak causes Internet slowdown

BGP Optimizer Causes Thousands Of Fake Routes

BGPMon Joins OpenDNS

What caused the Google service interruption?

BGP routing incidents in 2014, malicious or not?

BGP hijack incident by Syrian Telecommunications Establishment

Using BGP data to find Spammers

What caused today's Internet hiccup

The Canadian Bitcoin Hijack

Hijack event today by Indosat

Turkey Hijacking IP addresses for popular Global DNS providers

Looking at the spamhaus DDOS from a BGP perspective

Accidentally stealing the Internet

Syria shuts down the Internet

New version of BGPmon.net

A BGP leak made in Canada

Internet outage in Lebanon continues into second day

How the Internet in Australia went down under

F-Root DNS server moved to Beijing

Internet Syria offline

Facebook's detour through China and Korea

Egypt Back Online

Copyright © 2016 BGPmon Network Solutions Inc. 2012. All rights reserved.

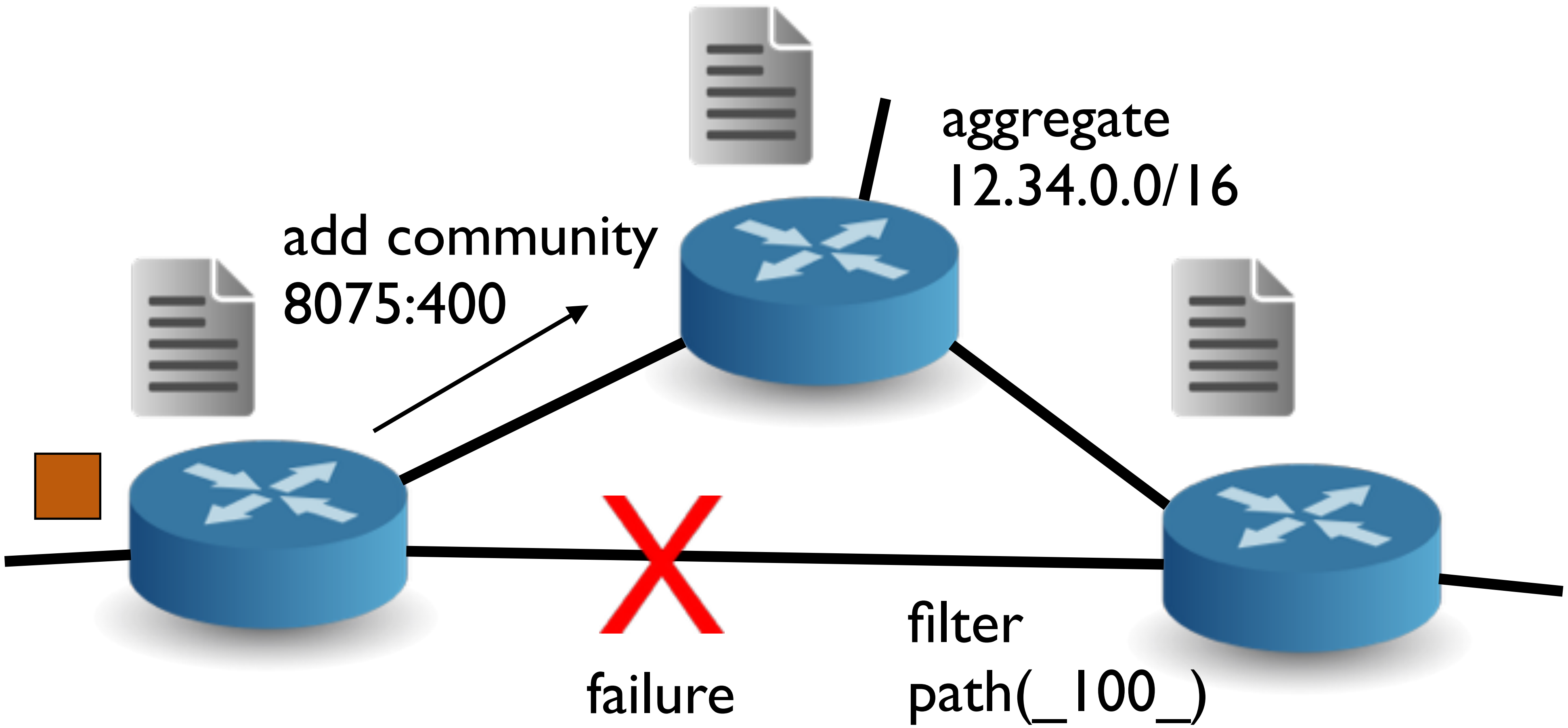
http://www.bgpmon.net/news-and-events/ 1/1

Objectives ⟵ GAP ⟶ Mechanisms

Objectives: Network-wide

- Prefer traffic to send traffic through customers over providers
- Don’t use our network as transit between A and B
- Traffic must stay within national boundaries

Mechanisms: Device-by-Device



aggregate
12.34.0.0/16

add community
8075:400

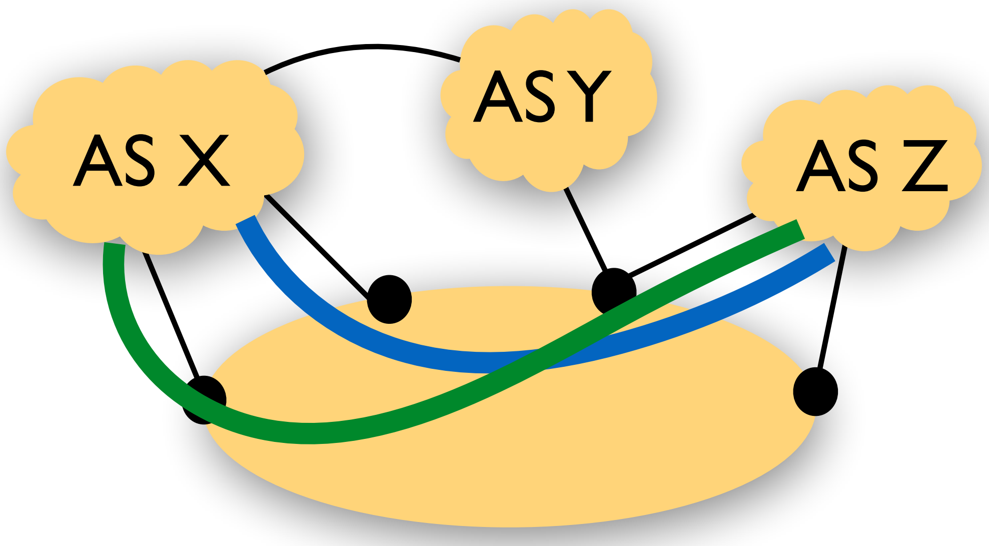
failure

filter
path(_100_)

Propane:
Programming a Distributed Control Plane

1) Language for expressing high-level objectives with:

- Path constraints and relative preferences with fall-backs in case of failures
- Uniform abstractions for intra- and inter-domain routing



2) Compiler that generates low-level, distributed configs:

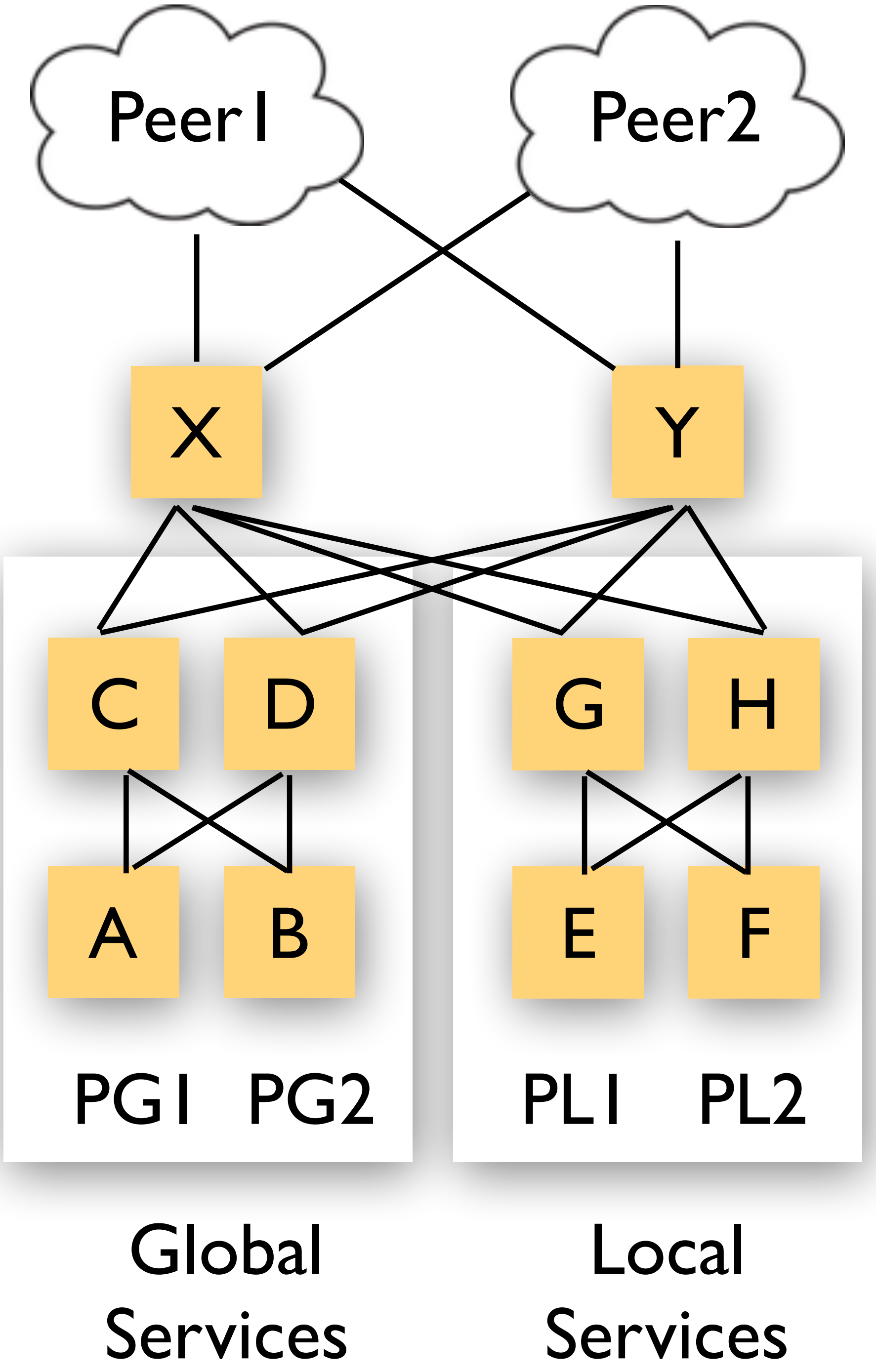
- Efficient algorithms to synthesize a set of policy-compliant BGP configs
- Failure analysis guarantees *policy compliance* under all failures

Example: A Data Center Network

A Data Center Network

Goals:

- P1: Announce global services externally as the aggregate PG
- P2: Do not announce local services externally
- P3: Prefer Backbone1 to Backbone2



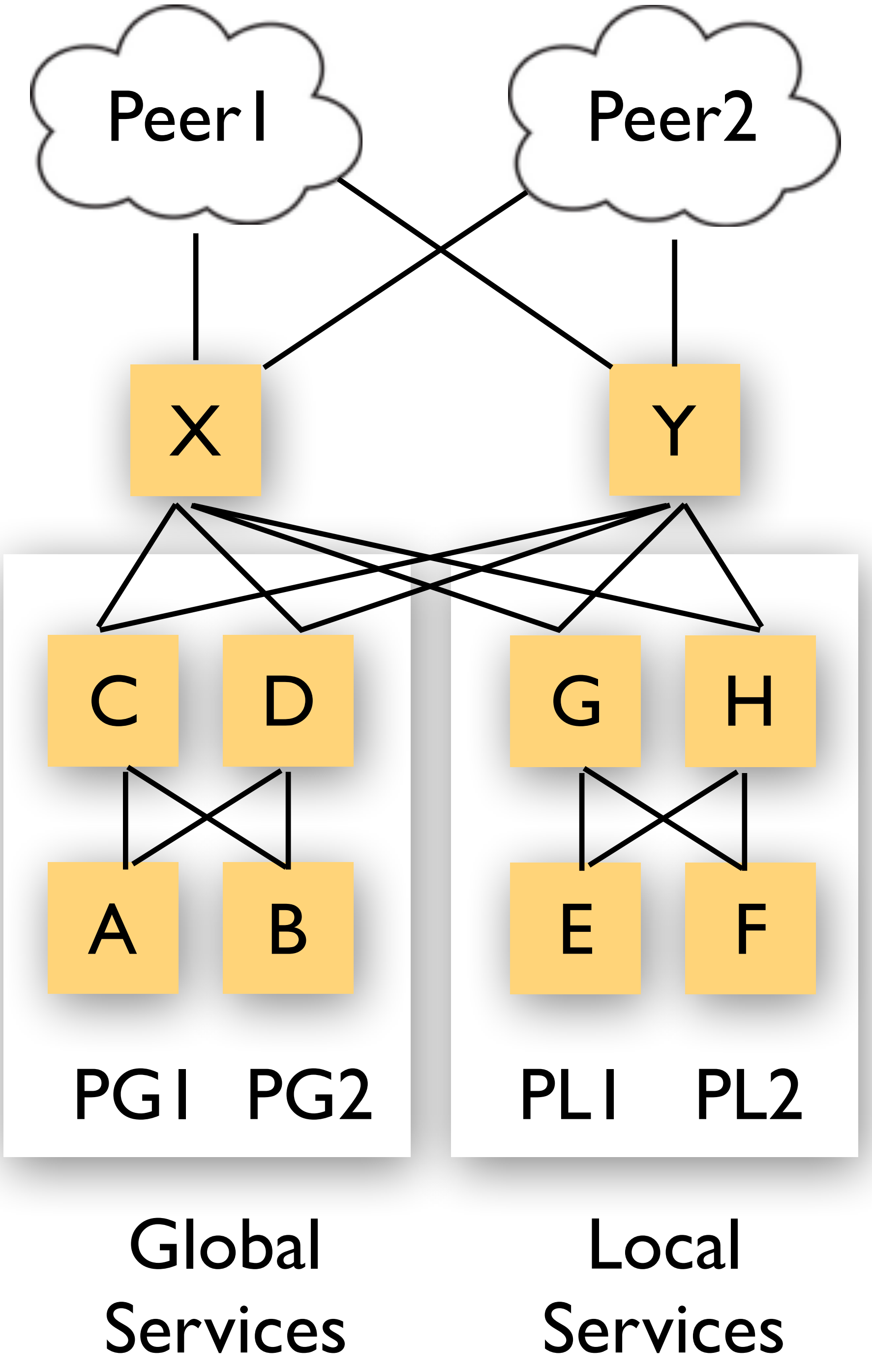
A Data Center Network

Goals:

- P1: Announce global services externally as the aggregate PG
- P2: Do not announce local services externally
- P3: Prefer Backbone1 to Backbone2

Implementation Techniques for X, Y:

- do export announce's from C, D outside
- do *not* export announce's from G,H outside
  - appeal: X,Y do not need to know which prefixes are local vs global
- aggregate to PG if announce is subset of PG



A Data Center Network

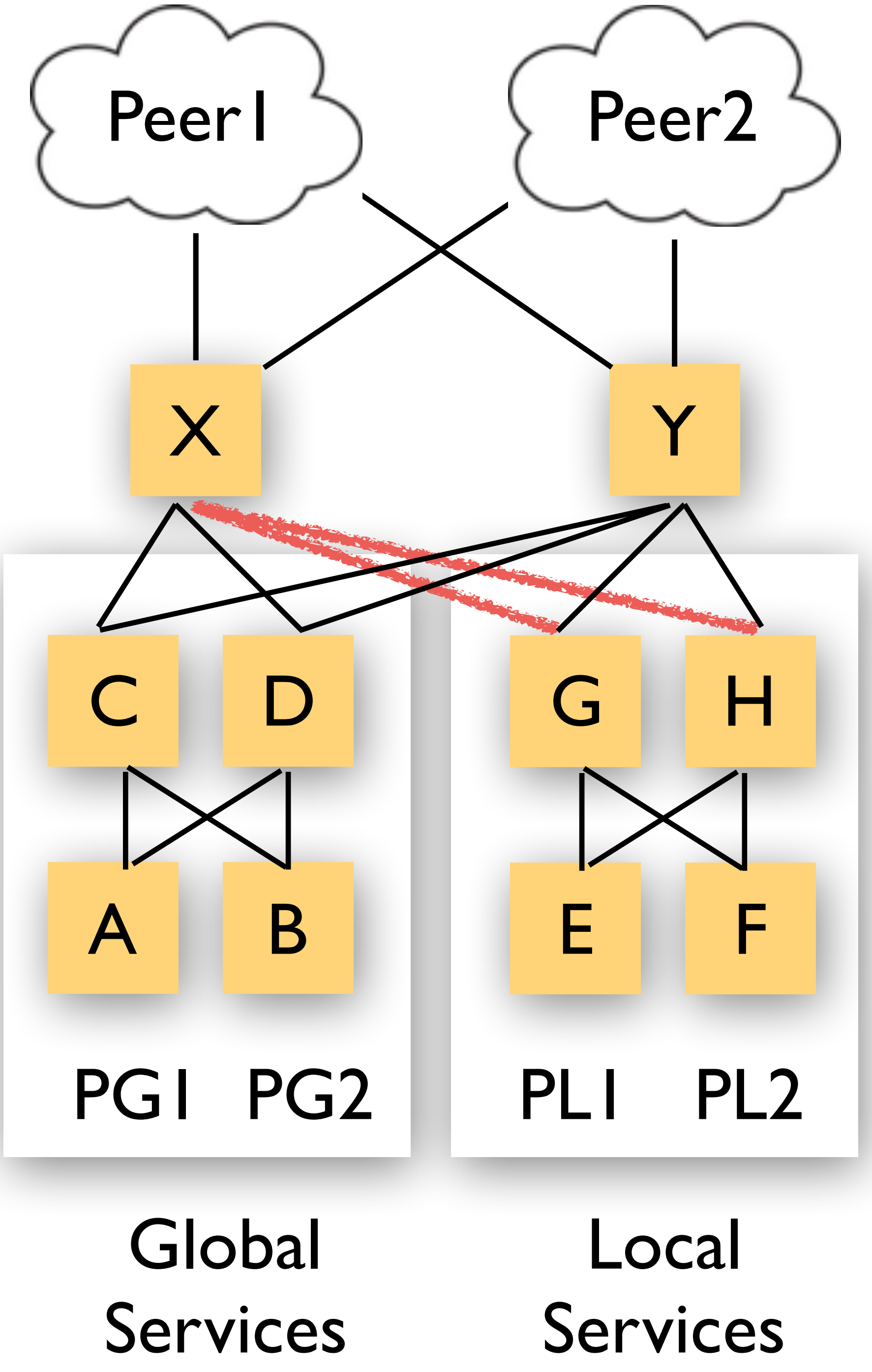
Goals:

- P1: Announce global services externally as the aggregate PG
- P2: Do not announce local services externally
- P3: Prefer Backbone1 to Backbone2

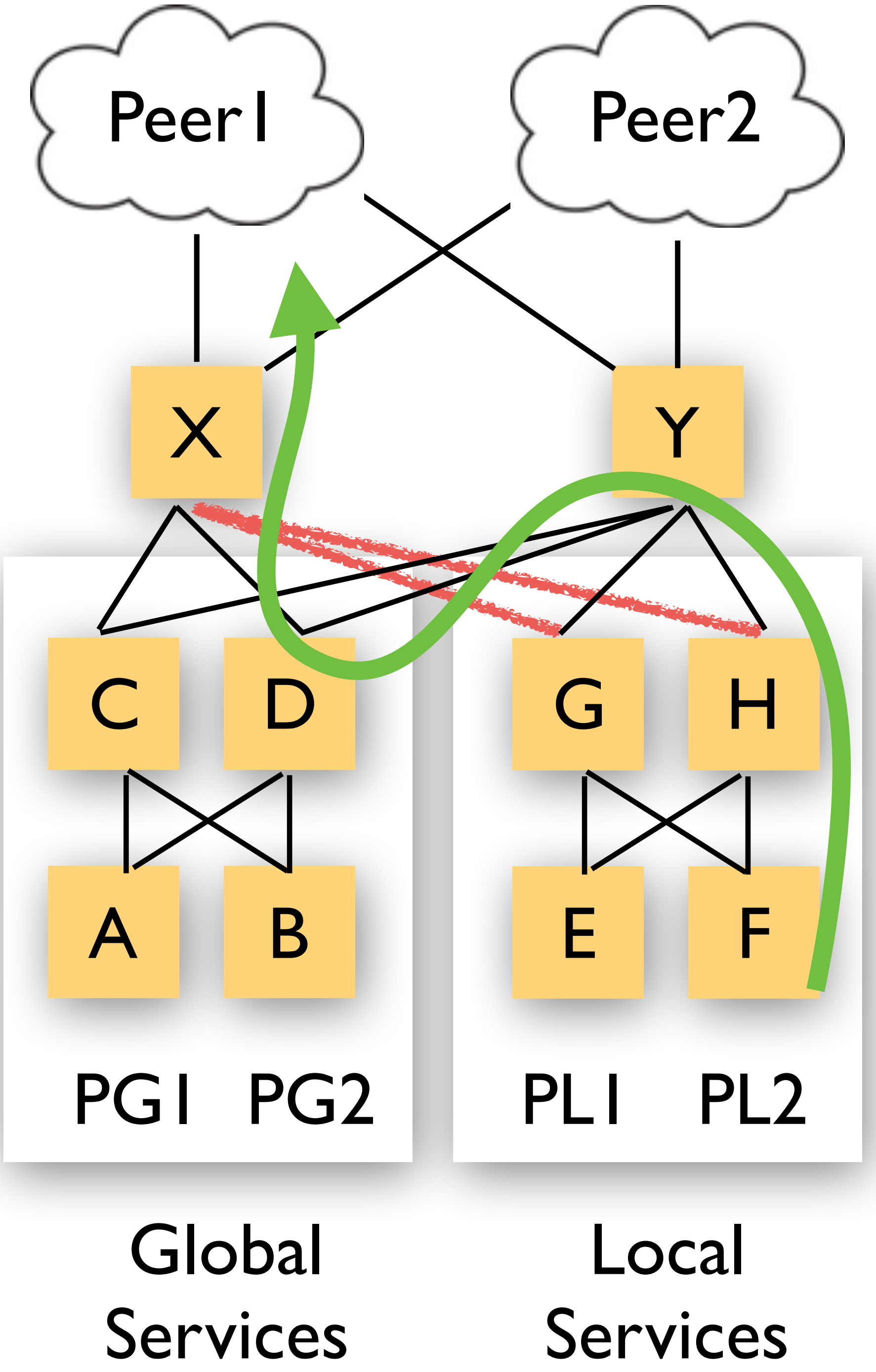
Implementation Techniques for X, Y:

- do export announce's from C, D outside
- do *not* export announce's from G,H outside
  - appeal: X,Y do not need to know which prefixes are local vs global
- aggregate to PG if announce is subset of PG

Consider X-G, X-H Failure:



A Data Center Network



Goals:

- P1: Announce global services externally as the aggregate PG
- P2: Do not announce local services externally
- P3: Prefer Backbone1 to Backbone2

Implementation Techniques for X, Y:

- do export announce's from C, D outside
- do *not* export announce's from G,H outside
  - appeal: X, Y do not need to know which prefixes are local vs global
- aggregate to PG if announce is subset of PG

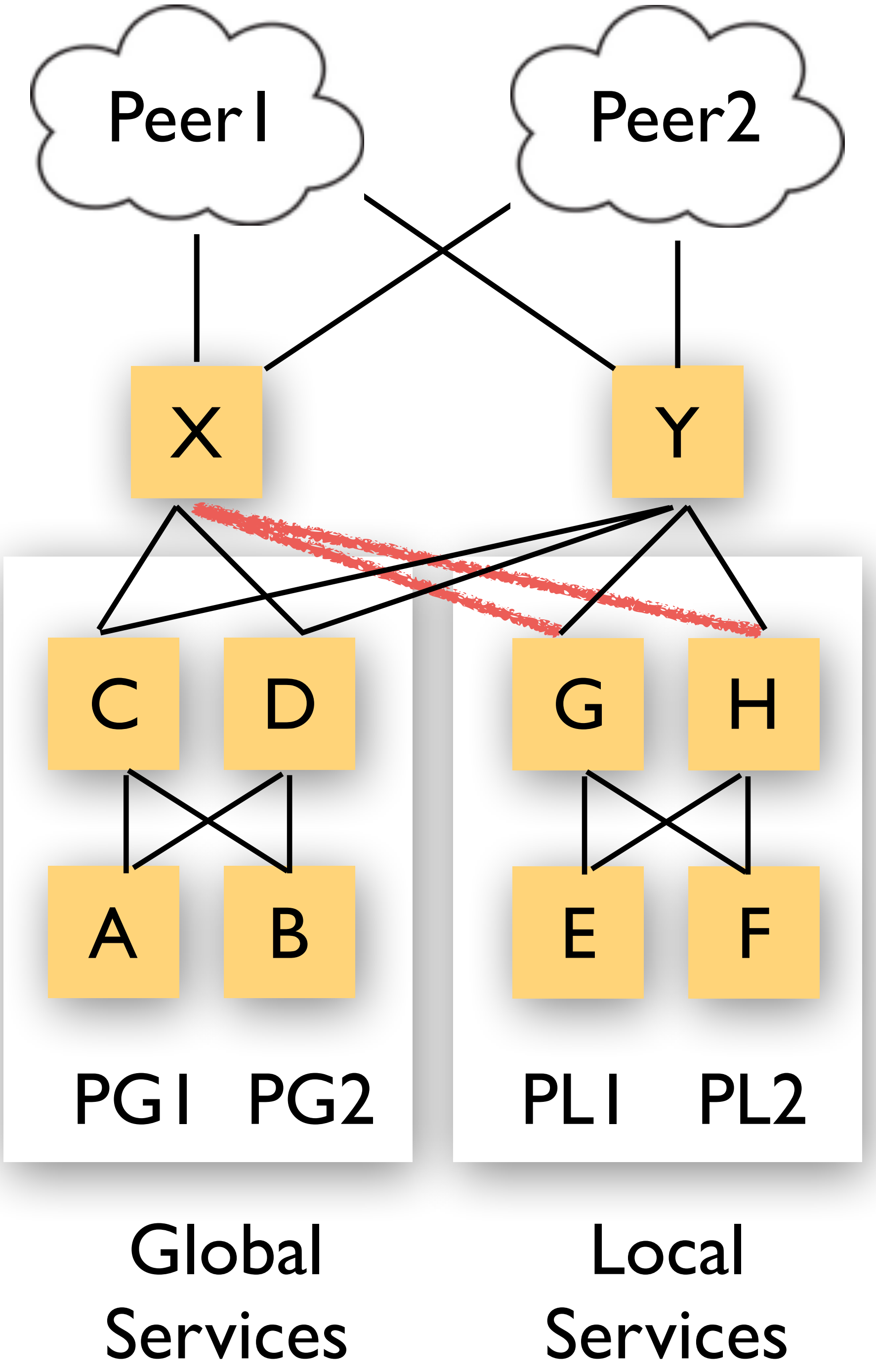
Consider X-G, X-H Failure:

- PL* announcements travel H-Y-D-X
- PL* announcements are then leaked

A Data Center Network

Implementation Techniques for X, Y:

- do export announce's from C, D outside
- do *not* export announce's from G,H outside
    - appeal: X,Y do not need to know which prefixes are local vs global
- aggregate to PG if announce is subset of PG
- **disallow "valley" paths**

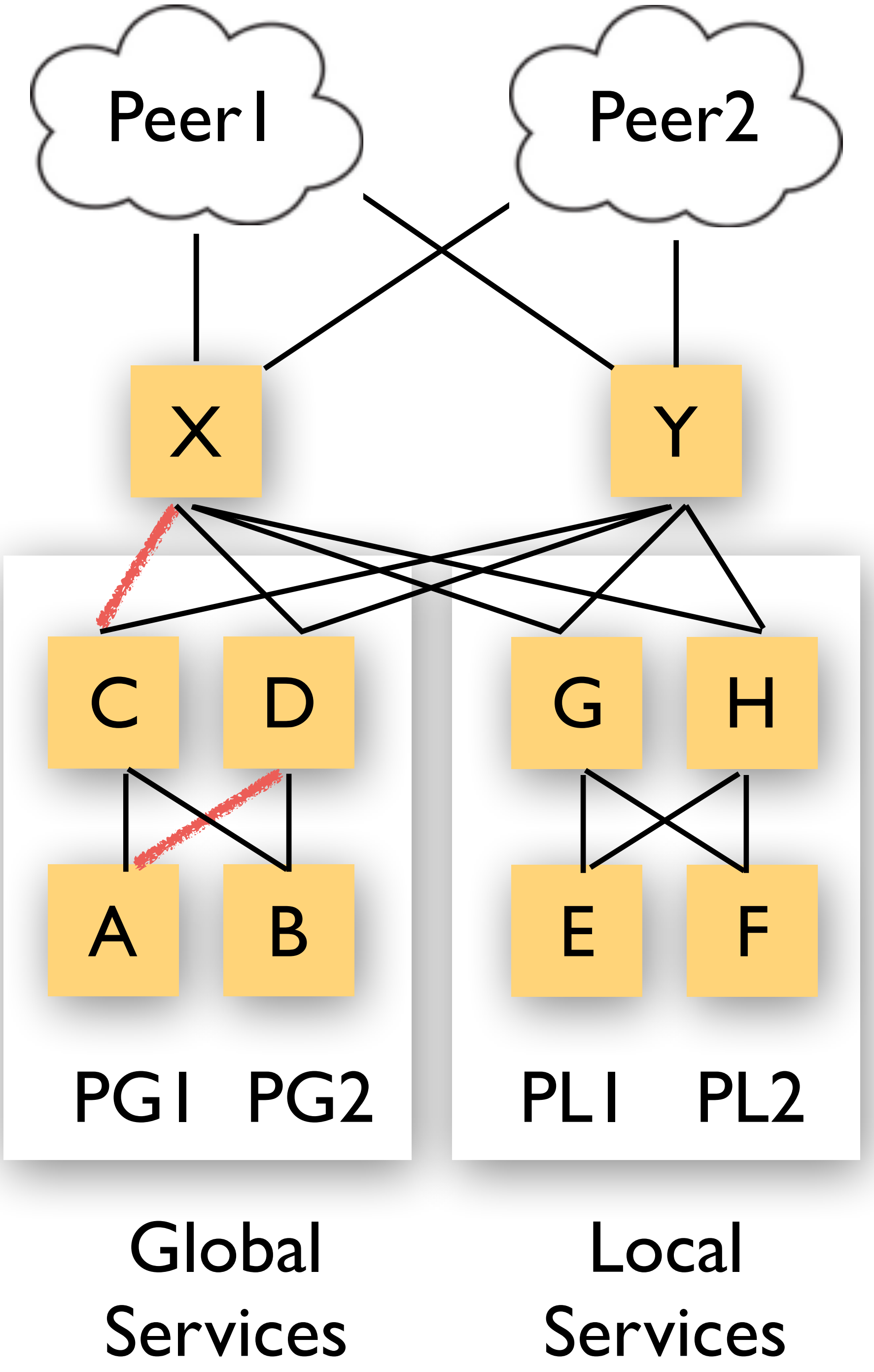


A Data Center Network

Implementation Techniques for X, Y:

- do export announce's from C, D outside
- do *not* export announce's from G,H outside
  - appeal: X,Y do not need to know which prefixes are local vs global
- aggregate to PG if announce is subset of PG
- **disallow "valley" paths**

Consider D-A, X-C Failure:



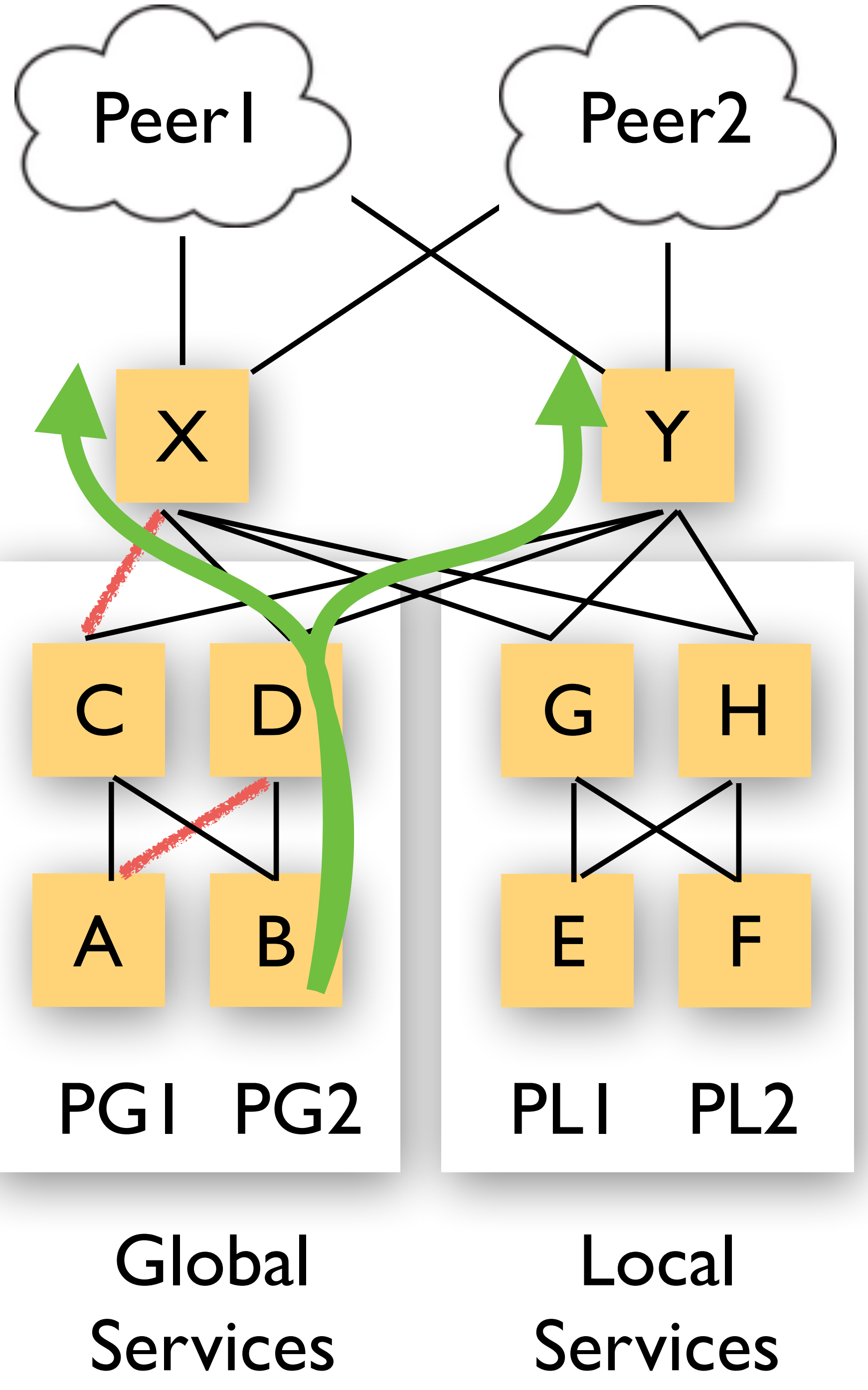
A Data Center Network

Implementation Techniques for X, Y:

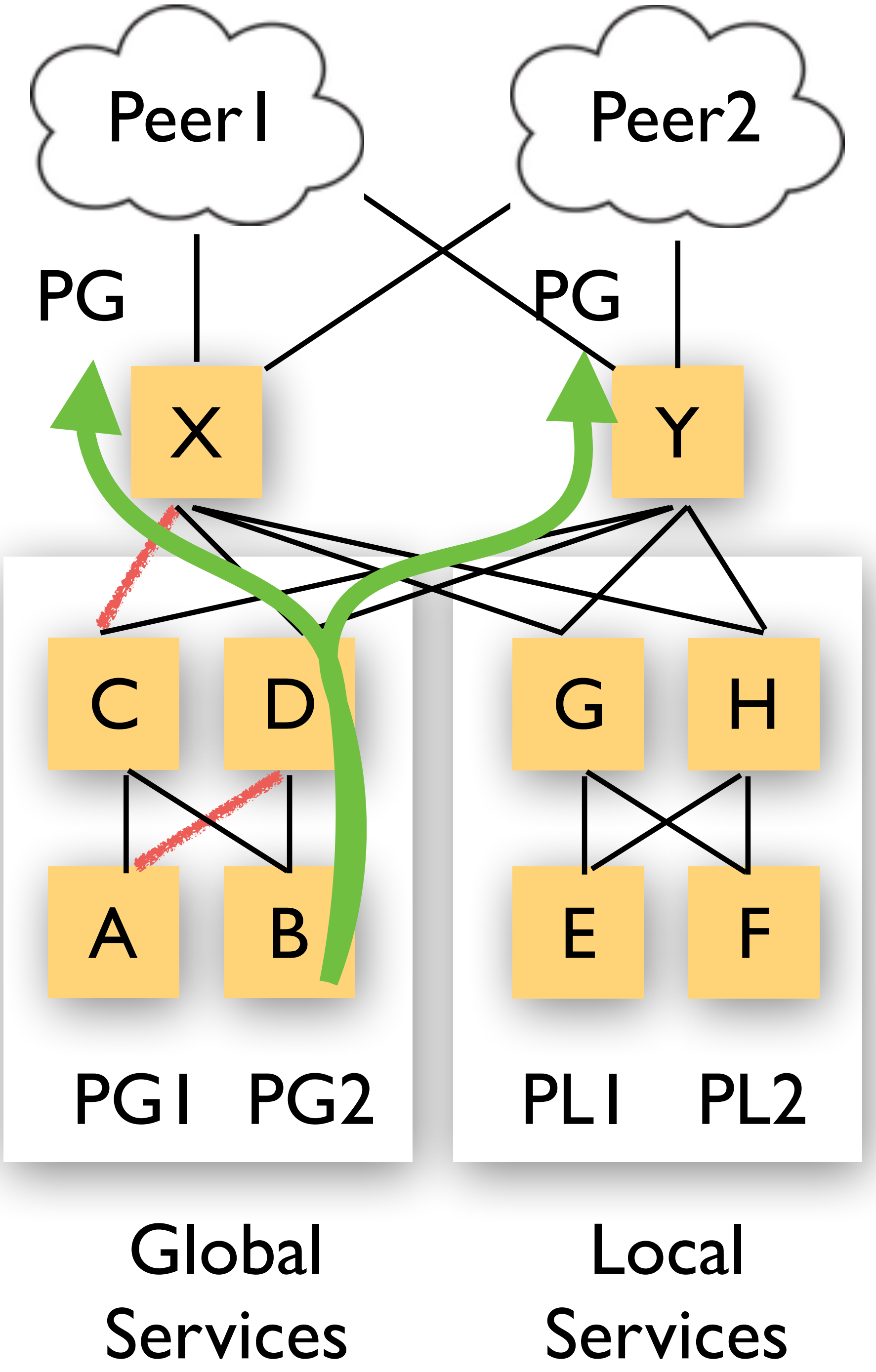
- do export announce's from C, D outside
- do *not* export announce's from G,H outside
  - appeal: X,Y do not need to know which prefixes are local vs global
- aggregate to PG if announce is subset of PG
- disallow "valley" paths

Consider D-A, X-C Failure:

- X and Y will hear PG2



A Data Center Network



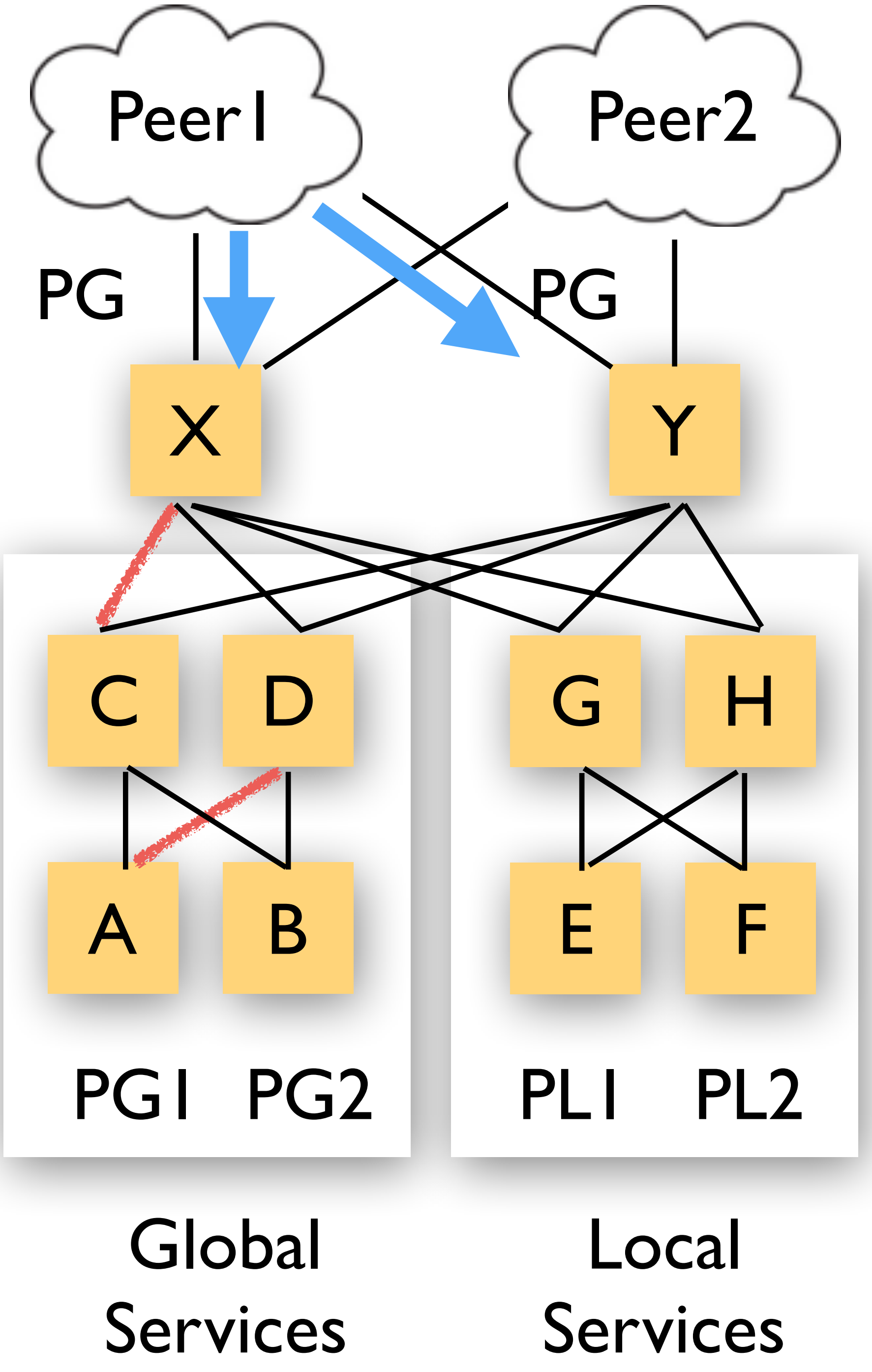
Implementation Techniques for X, Y:

- do export announce's from C, D outside
- do *not* export announce's from G,H outside
  - appeal: X,Y do not need to know which prefixes are local vs global
- aggregate to PG if announce is subset of PG
- **disallow "valley" paths**

Consider D-A, X-C Failure:

- X and Y will hear PG2
- X and Y will announce aggregate PG

A Data Center Network



Implementation Techniques for X, Y:

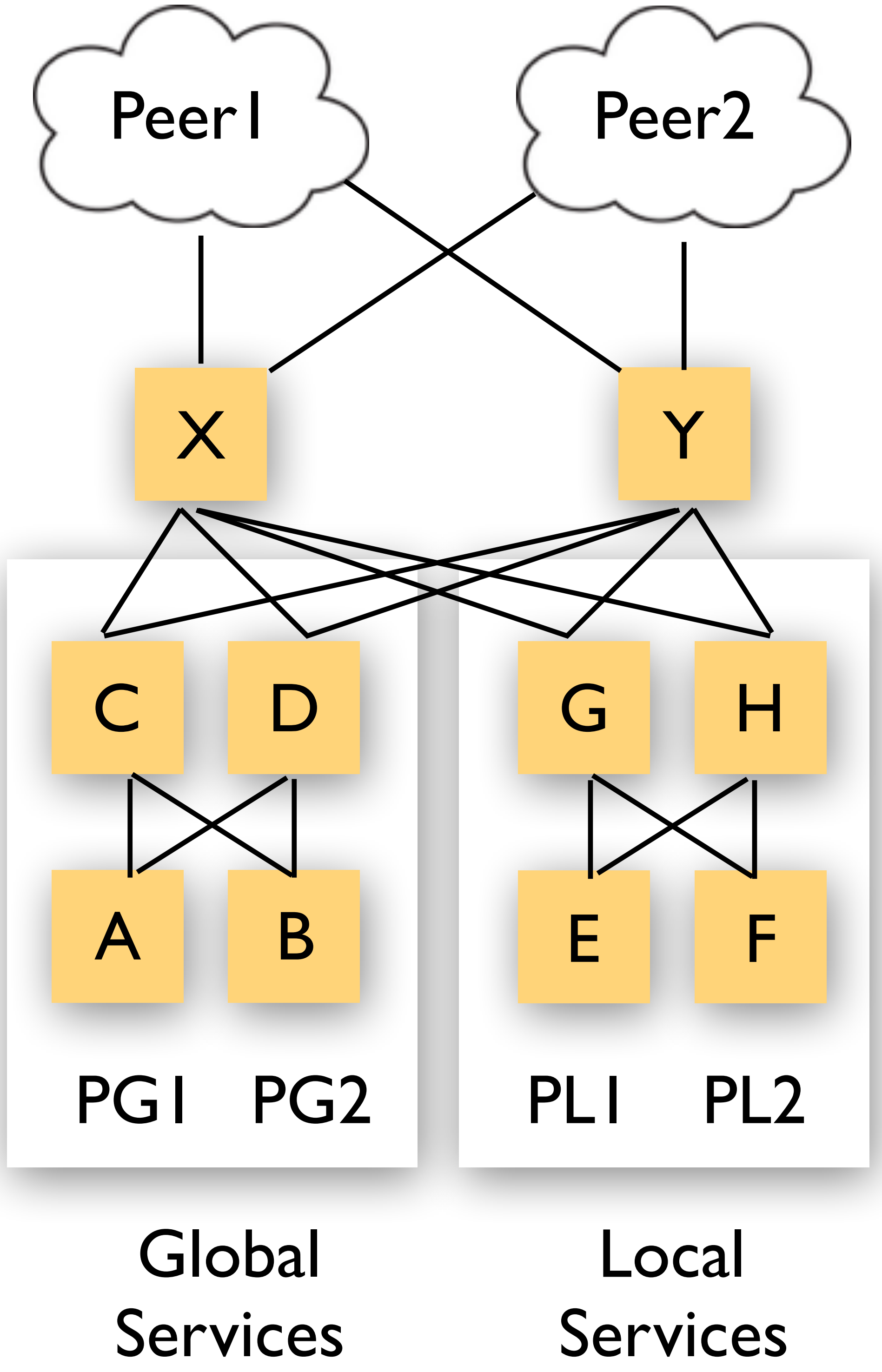
- do export announce's from C, D outside
- do *not* export announce's from G,H outside
  - appeal: X,Y do not need to know which prefixes are local vs global
- aggregate to PG if announce is subset of PG
- **disallow "valley" paths**

Consider D-A, X-C Failure:

- X and Y will hear PG2
- X and Y will announce aggregate PG
- But PG1 is inaccessible through X because there is no valley routing
- An aggregation-induced black hole is created [See Le et al, CoNext '11]

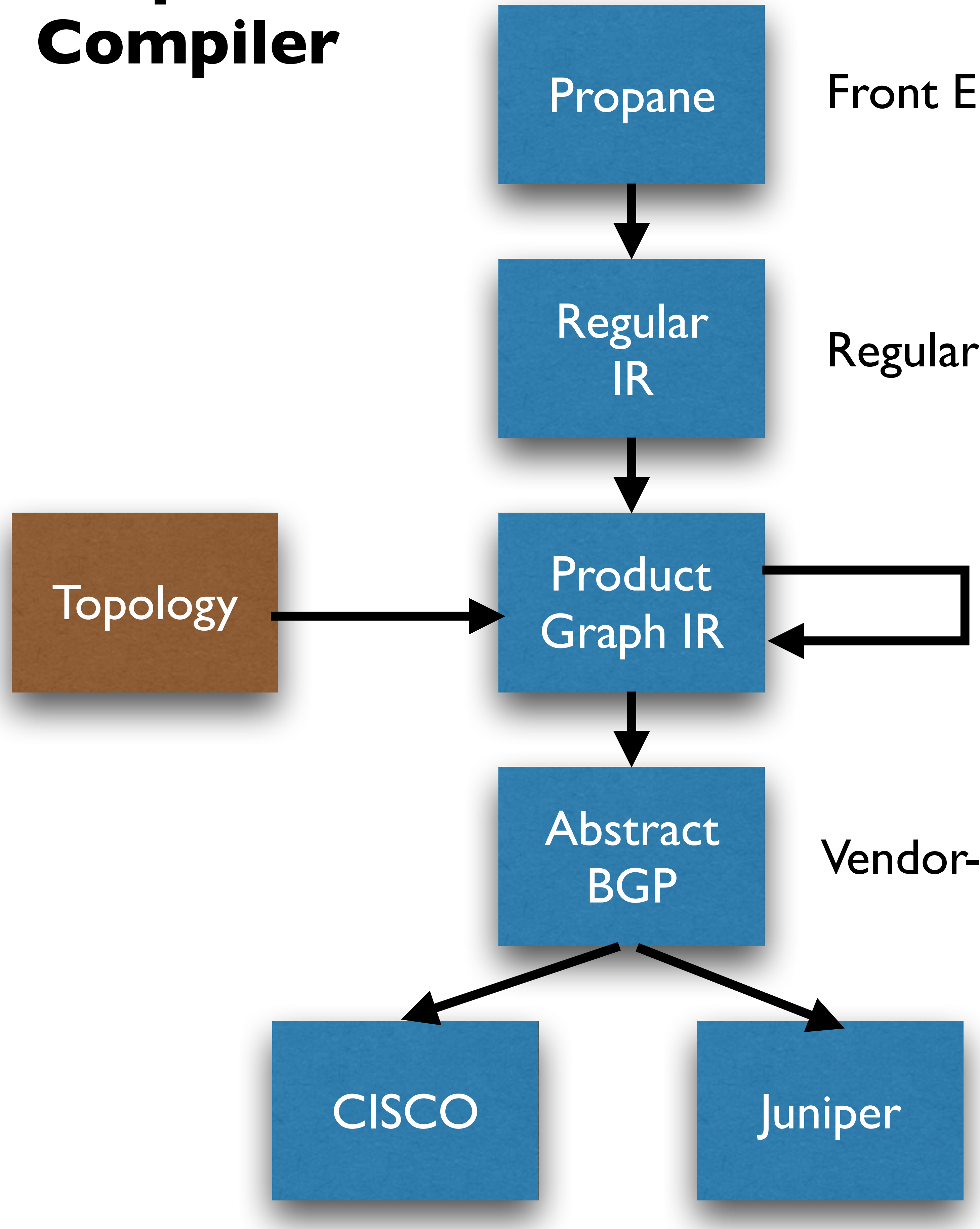
Quick Demo

- Originate prefixes for each TOR router
- Do not announce local services externally
- Aggregation on global prefixes
- Prefer Peer1 over Peer2
- Prevent transit between Back1 and Back2



Compiling Propane

Propane Compiler



Propane

Front End Constraint Language

Regular IR

Regular Expression-based IR

Topology

Product Graph IR

Failure Analyses

Abstract BGP

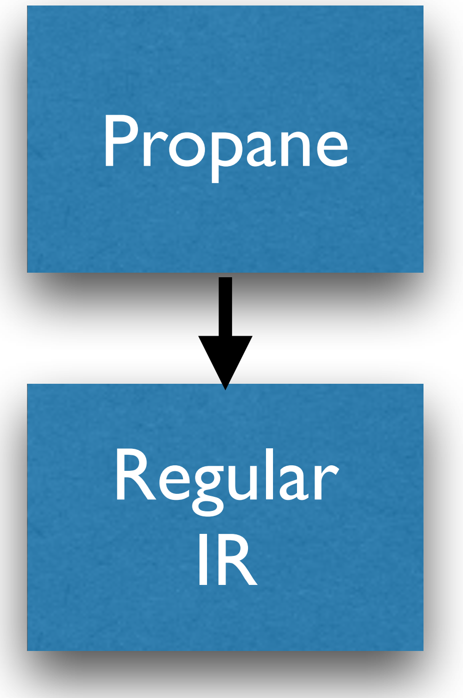
Vendor-independent BGP

CISCO

Juniper

Vendor-specific configurations

Propane Regular IR



Expand constraints in to regular expressions. EG:

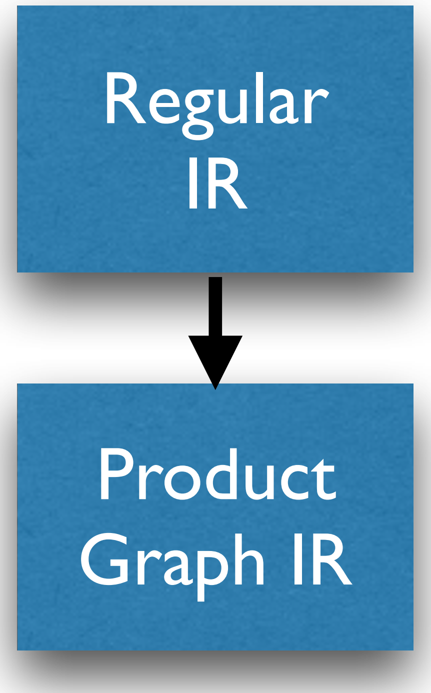
$$\texttt{end}(X) = (\Sigma^{*}.X)$$

$$\begin{aligned}\texttt{exit}(X) &= (\texttt{out}^{*}.\texttt{in}^{*}.(X \cap \texttt{in}).\texttt{out}^{+}) \cup \\ &\quad (\texttt{out}^{*}.\texttt{in}^{+}.(X \cap \texttt{out}).\texttt{out}^{*})\end{aligned}$$

A few other simple transformations:

- conjunction of constraints ==> intersection of regular expressions
- conjunction of policies ==> prefix-by-prefix intersection
- nested preferences lifted: (x >> y) . z ==> (x.z) >> (y.z)

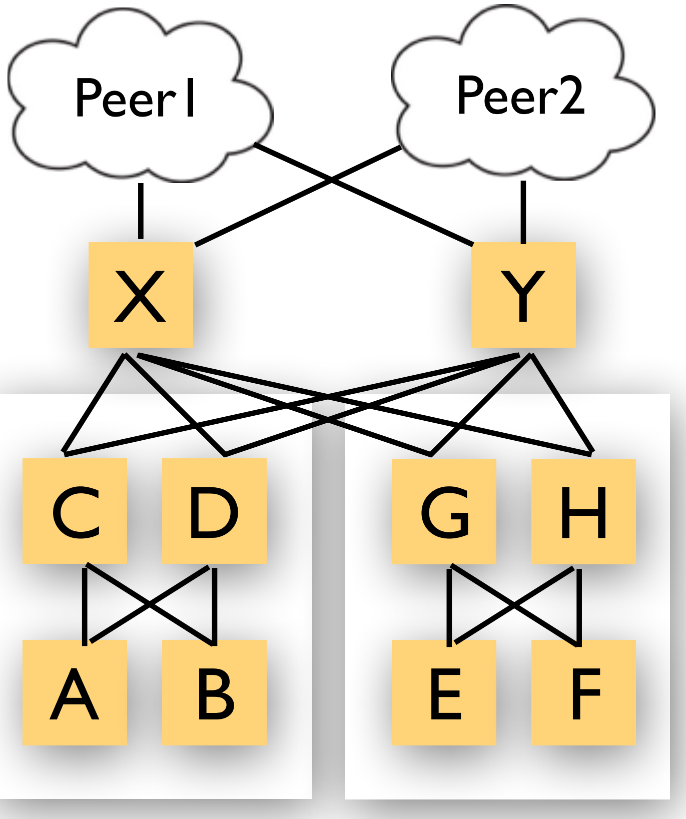
Constructing the Product Graph (PG)



Regular IR

Product Graph IR

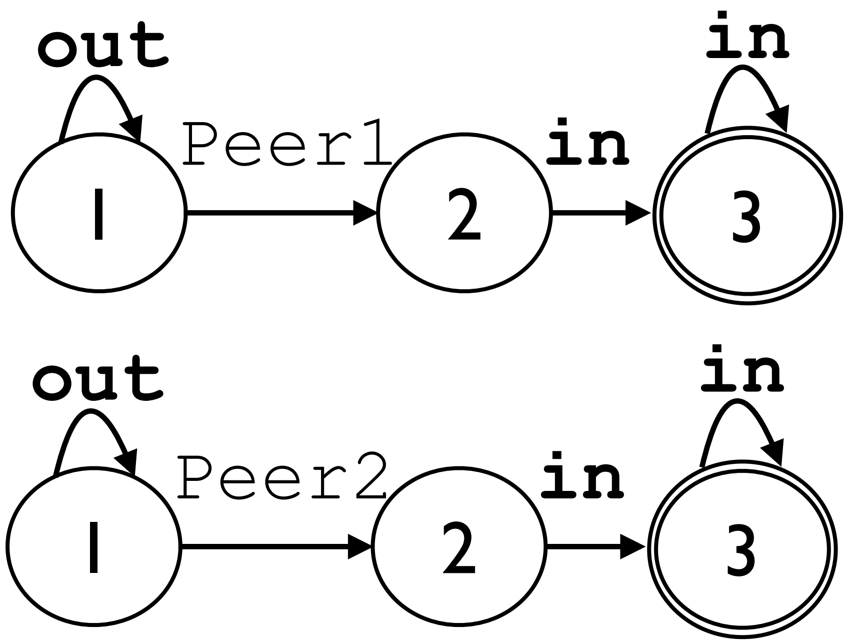
(a) Topology



(b) Policy

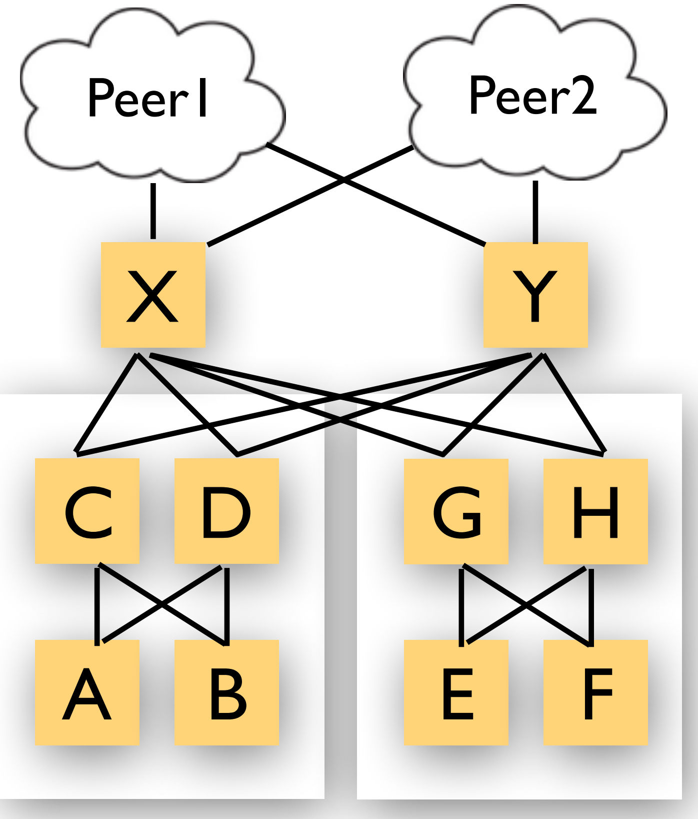
1)

2)

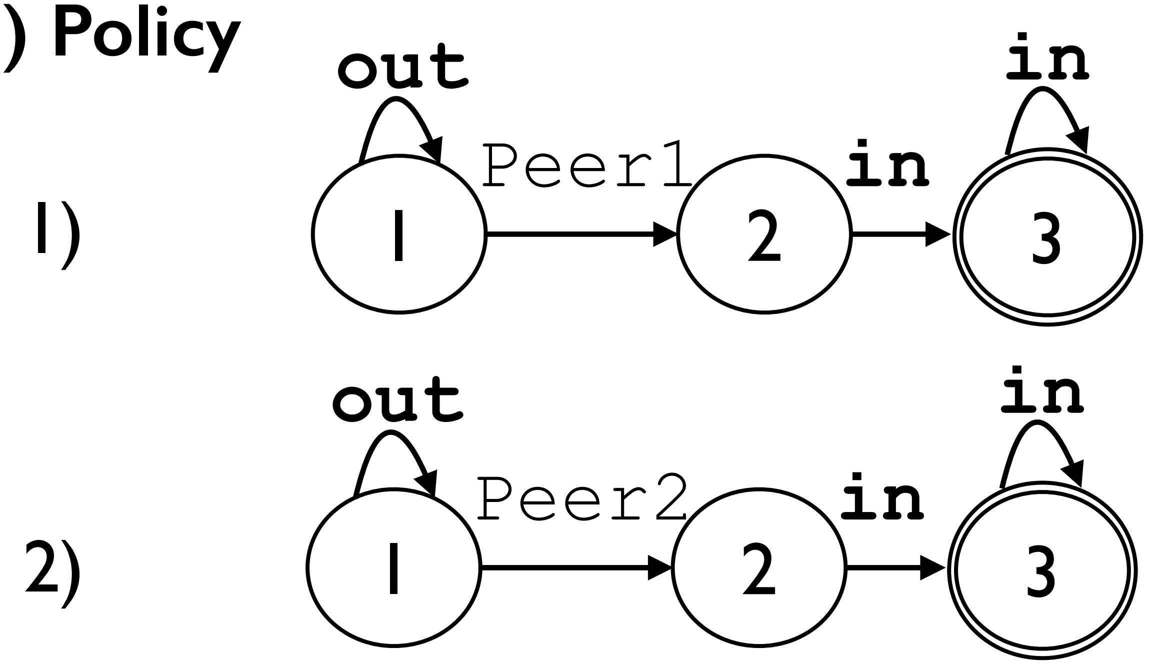


Constructing the Product Graph (PG)

(a) Topology



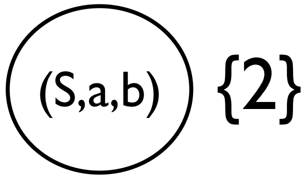
(b) Policy



General Idea:
PG represents locations reachable in the topology while following the policy

Each PG node contains:

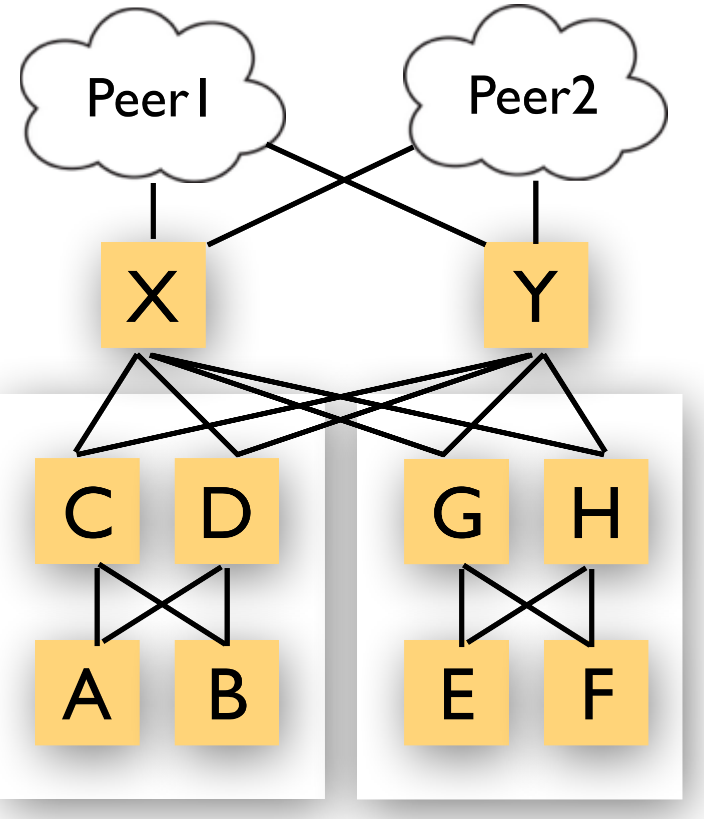
- topology node (S)
- state of automaton 1 (a)
- state of automaton 2 (b)
- set of preferences achieved



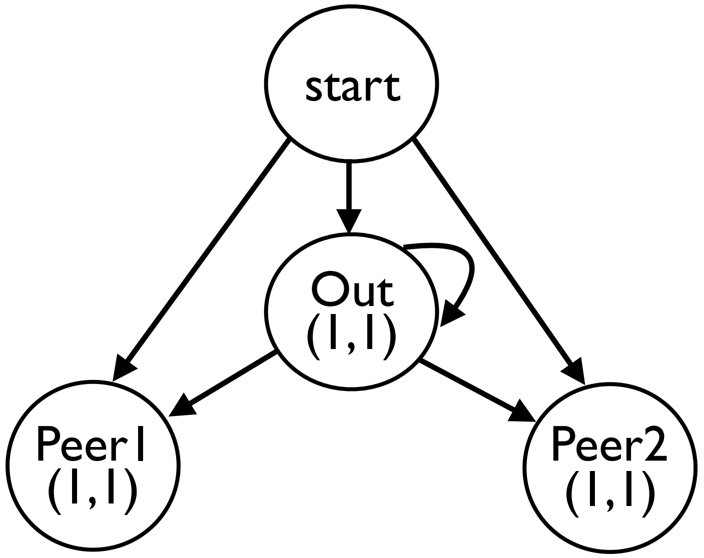
Two PG nodes are connected if topology nodes are connected and the automata make the specified transition

Constructing the Product Graph (PG)

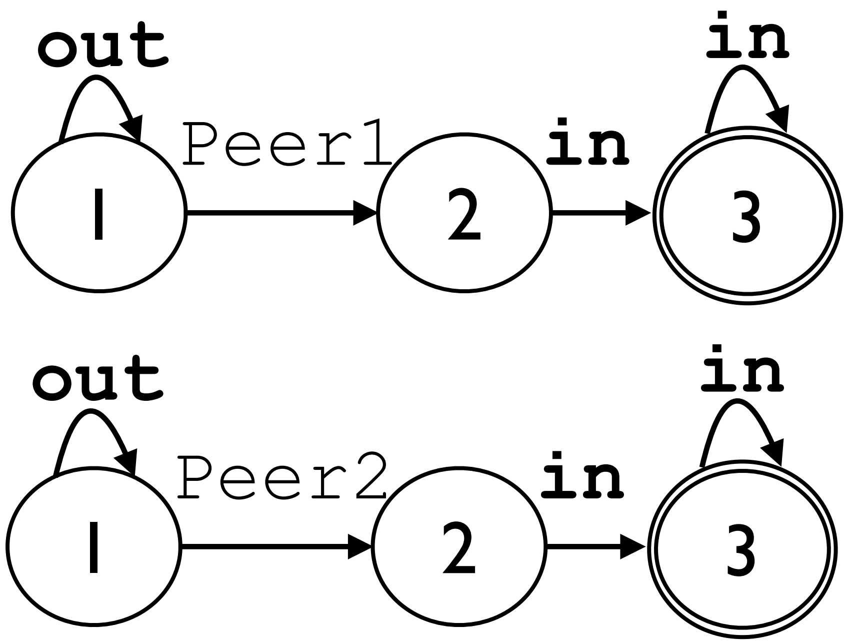
(a) Topology



(c) Product Graph

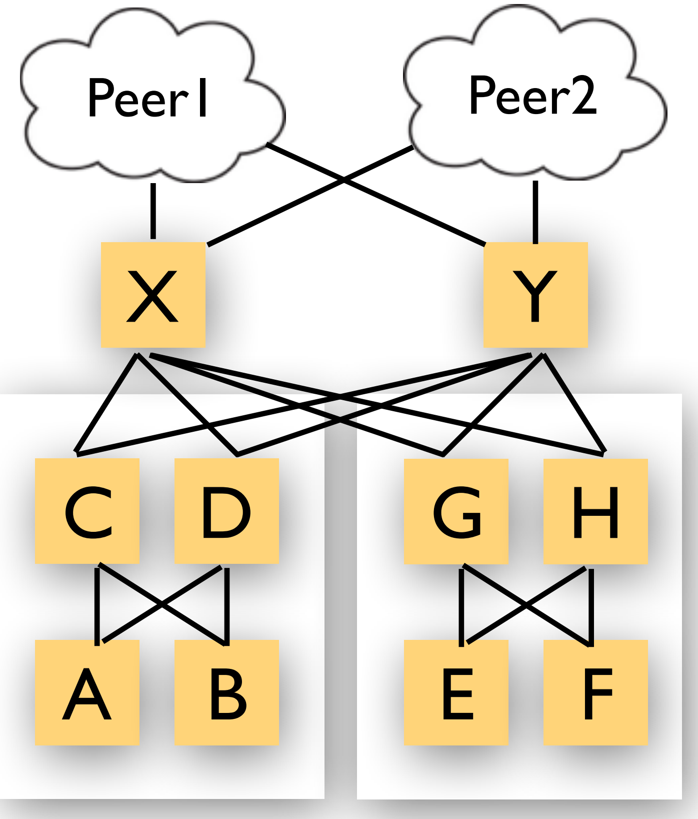


(b) Policy

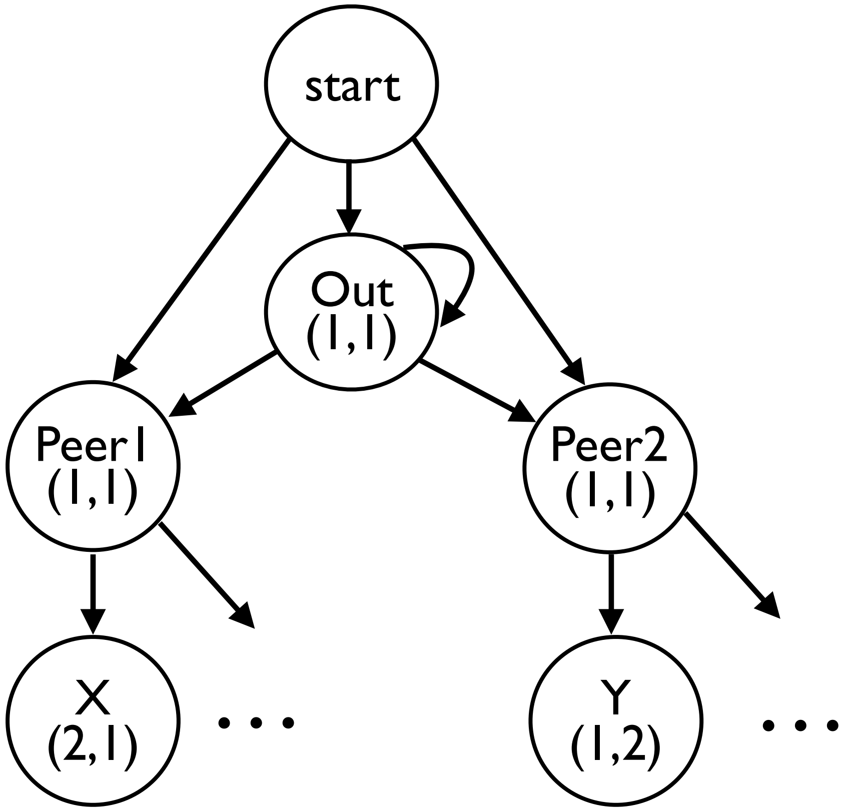


Constructing the Product Graph (PG)

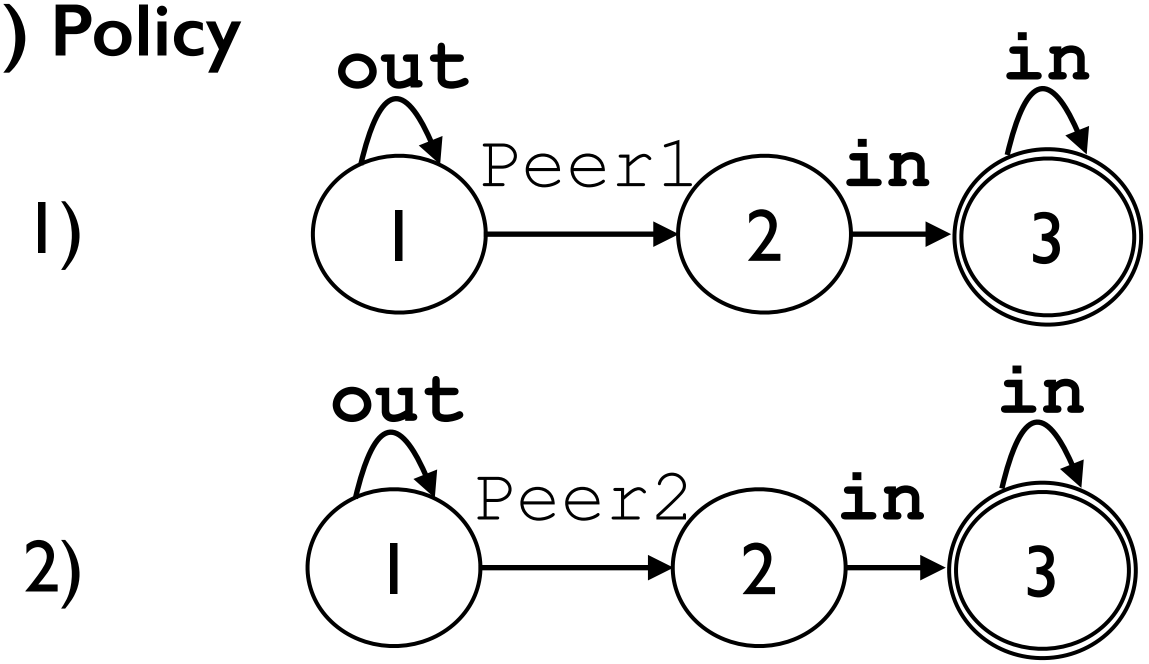
(a) Topology



(c) Product Graph

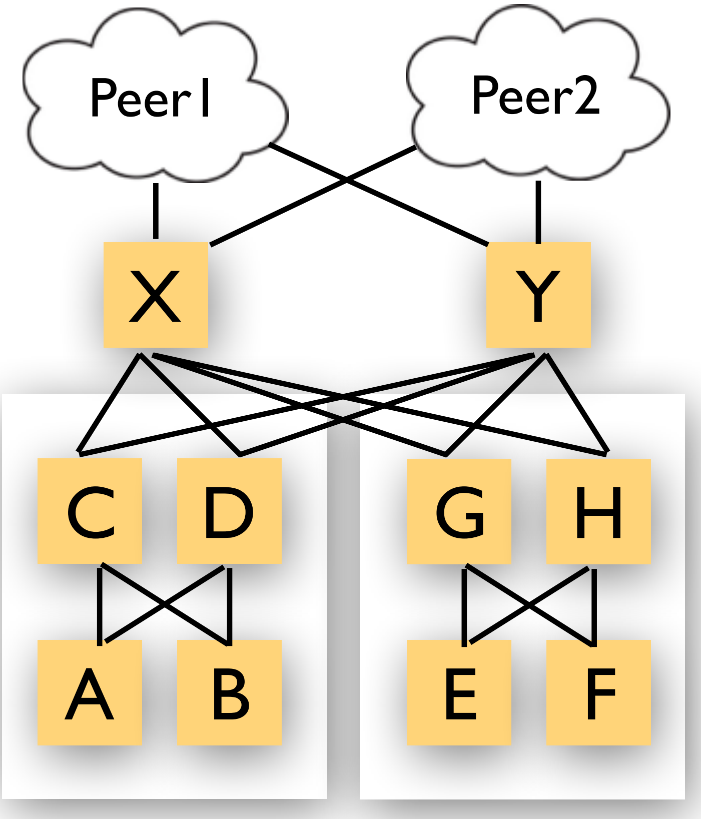


(b) Policy

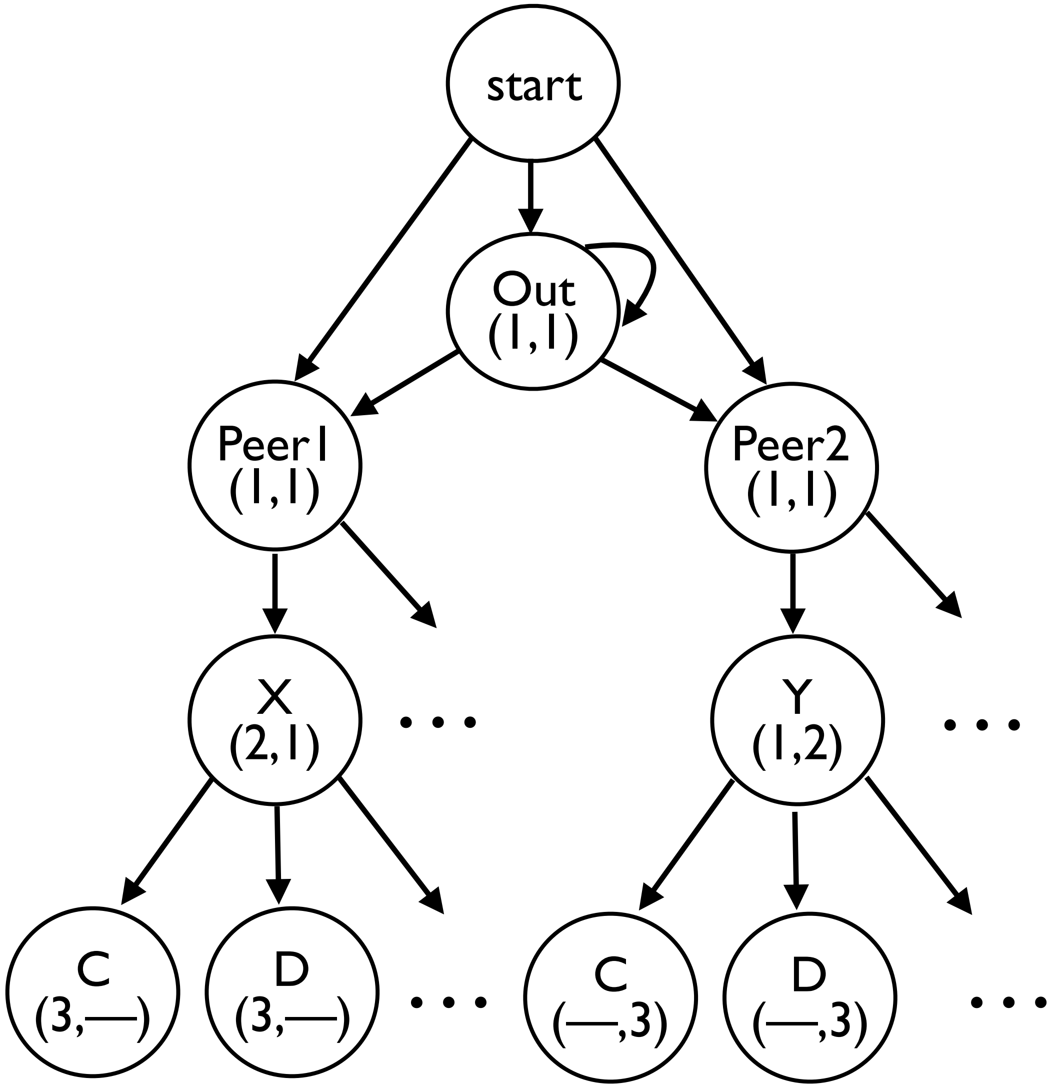


Constructing the Product Graph (PG)

(a) Topology

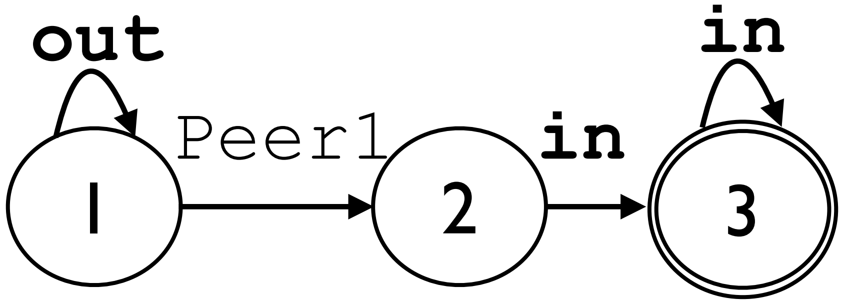


(c) Product Graph

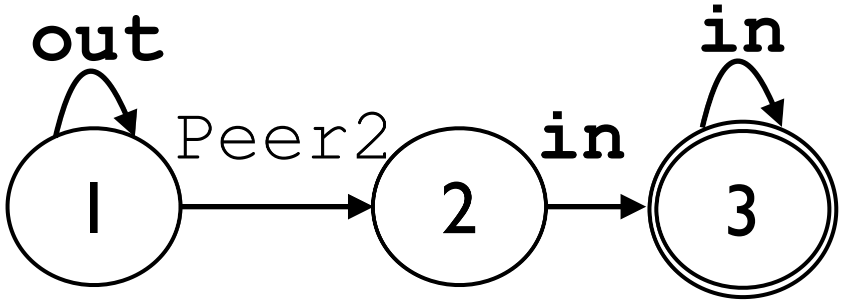


(b) Policy

1)

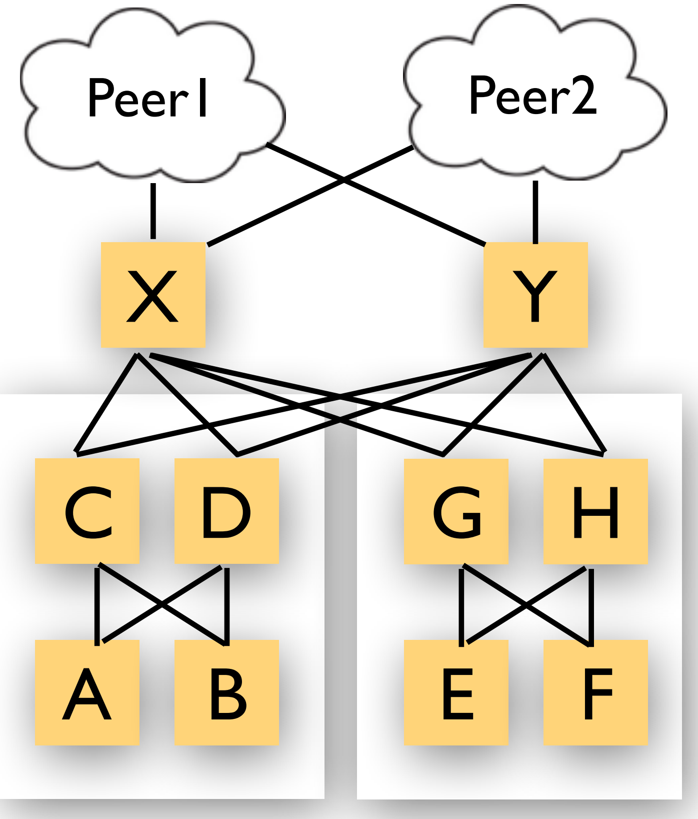


2)



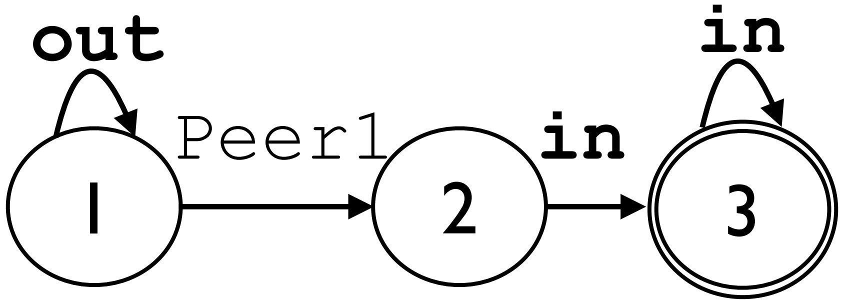
Constructing the Product Graph (PG)

(a) Topology

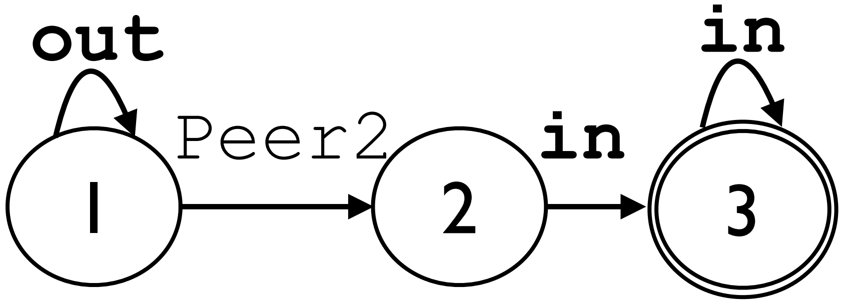


(b) Policy

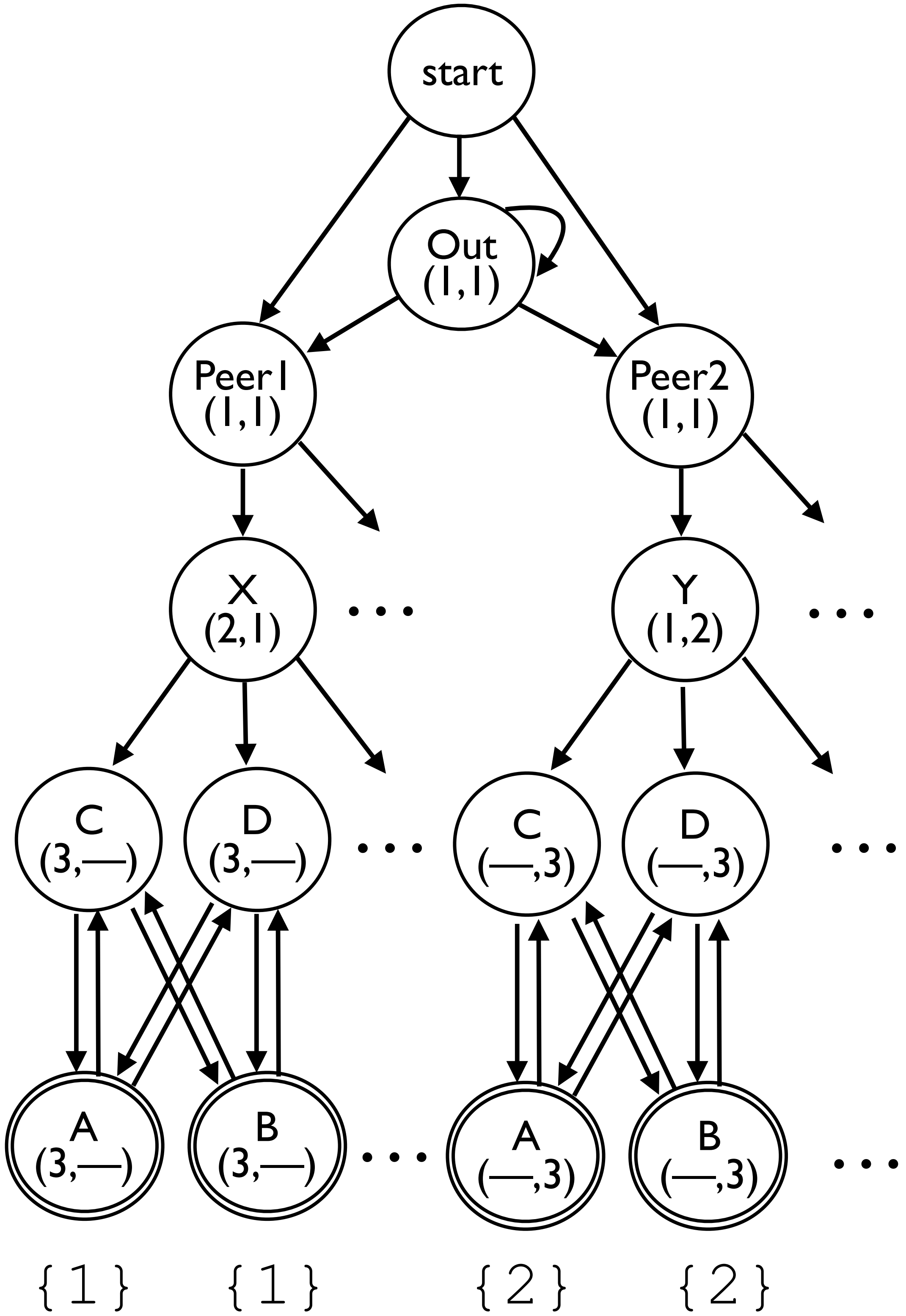
1)



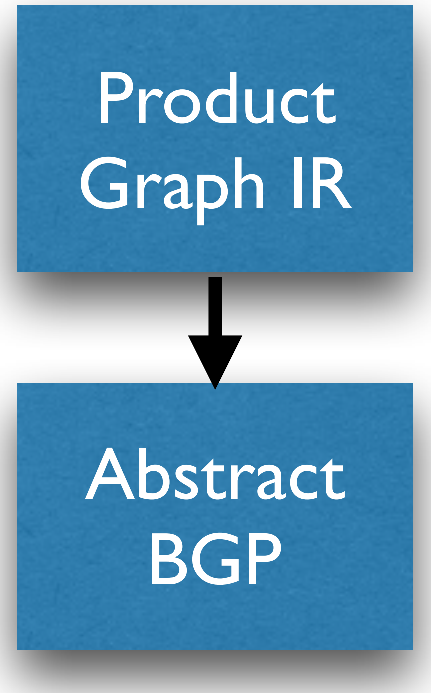
2)

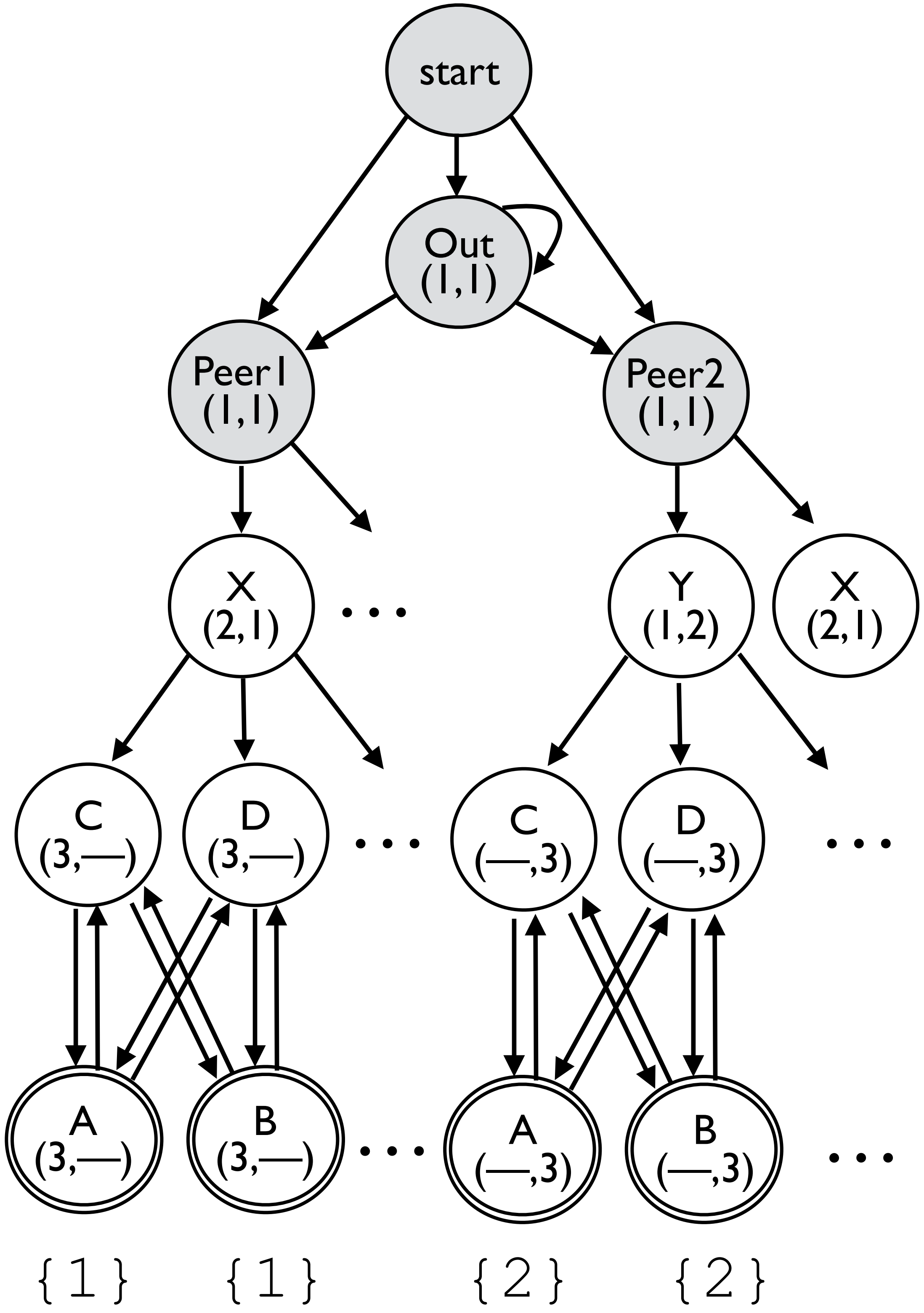


(c) Product Graph



Compilation to BGP:

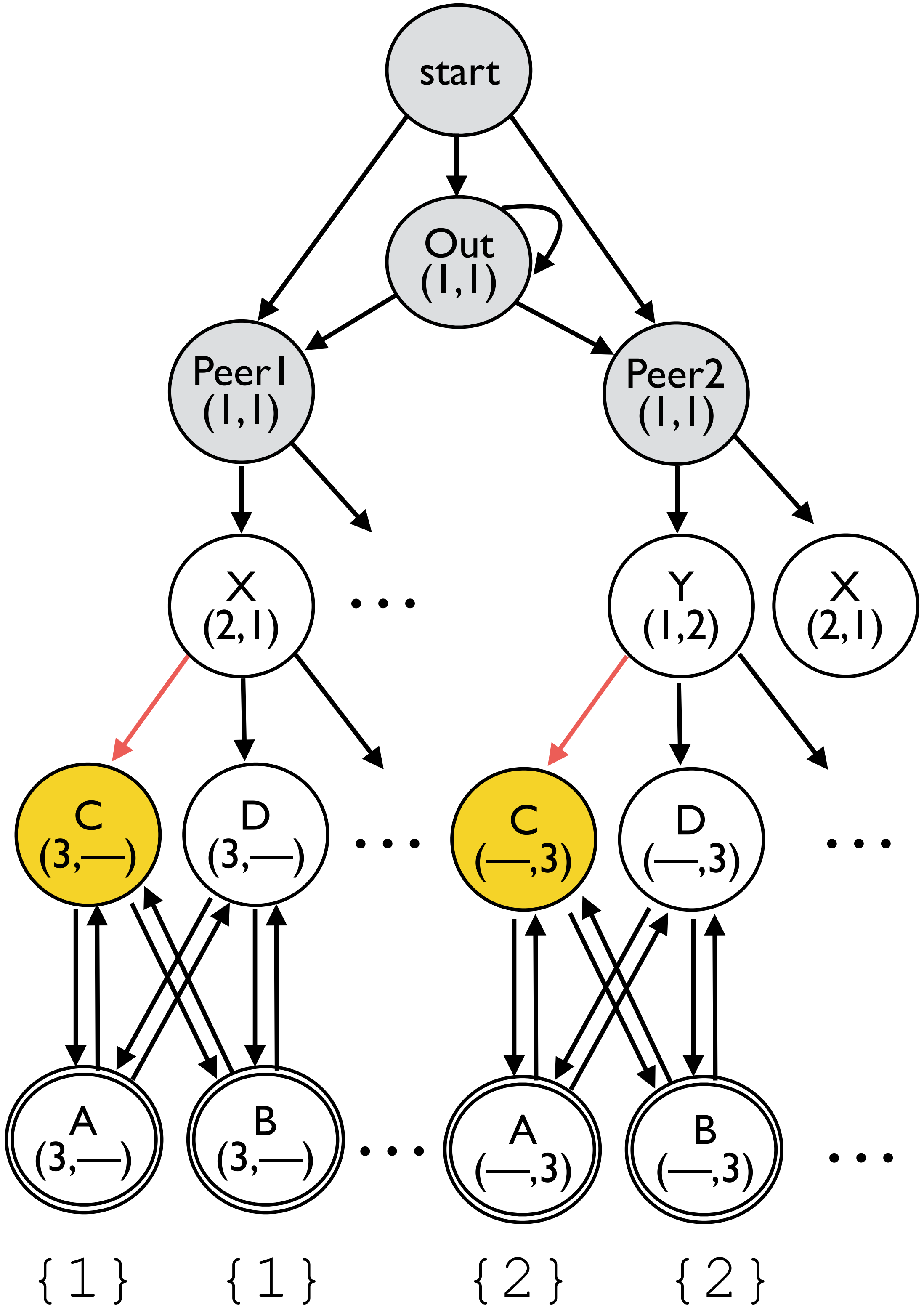




Idea:

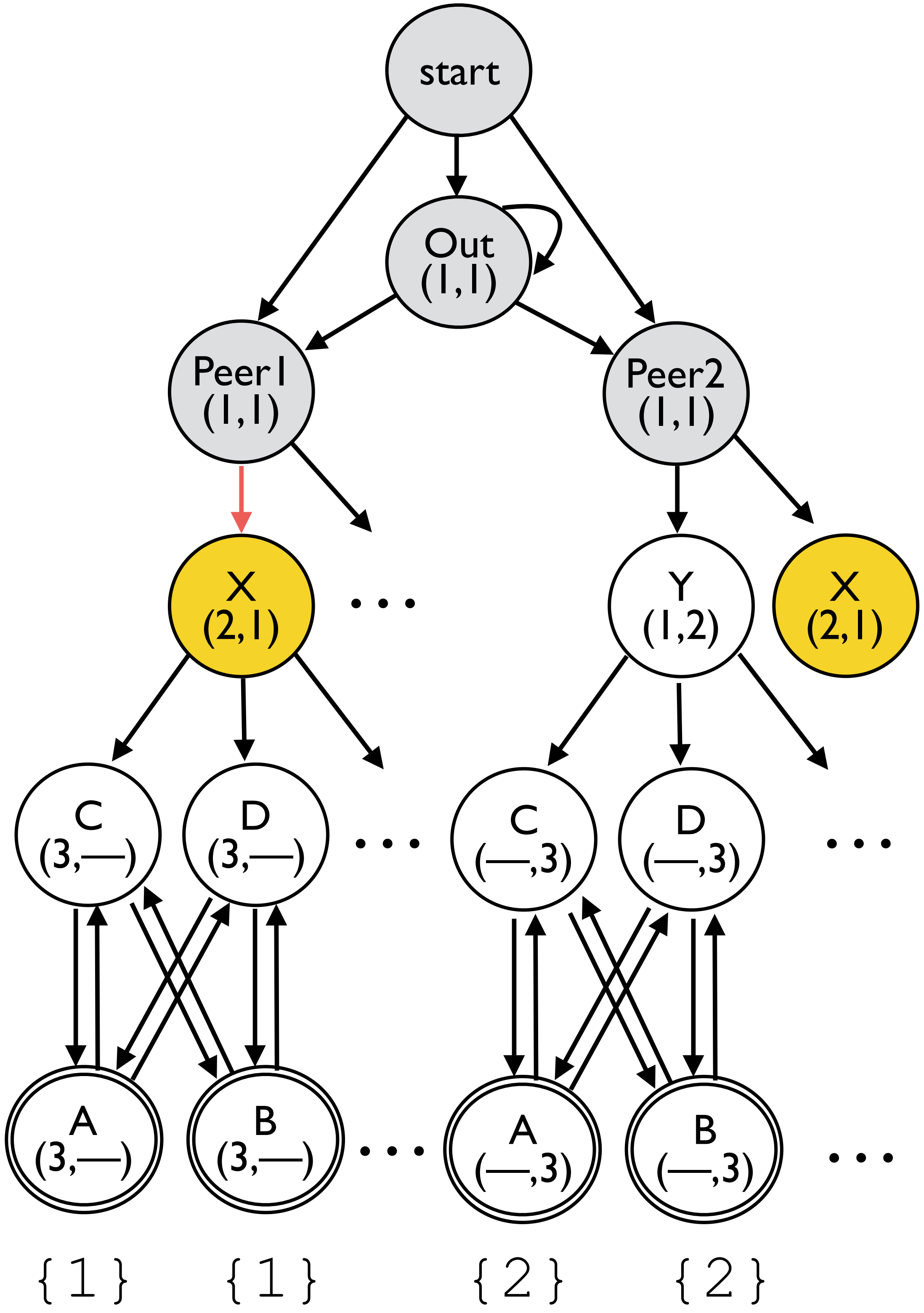
- Filter import messages according to incoming PG edges.
- For each internal location, decide which announcements to prefer, forward messages along PG edges
- Use a community value to tag the state of the automata
- For each external location, do nothing

Compilation to BGP:



```
Router C
    allow peer=X comm=(2,1)
        peer ←{A,B} comm ← (3, —)
    allow peer=Y comm=(1,2)
        peer ←{A,B} comm ← (—, 3)
```

Compilation to BGP:



Router C

allow peer=X comm=(2,1)

peer ←{A,B} comm ← (3, —)

allow peer=Y comm=(1,2)

peer ←{A,B} comm ← (—, 3)

Router X

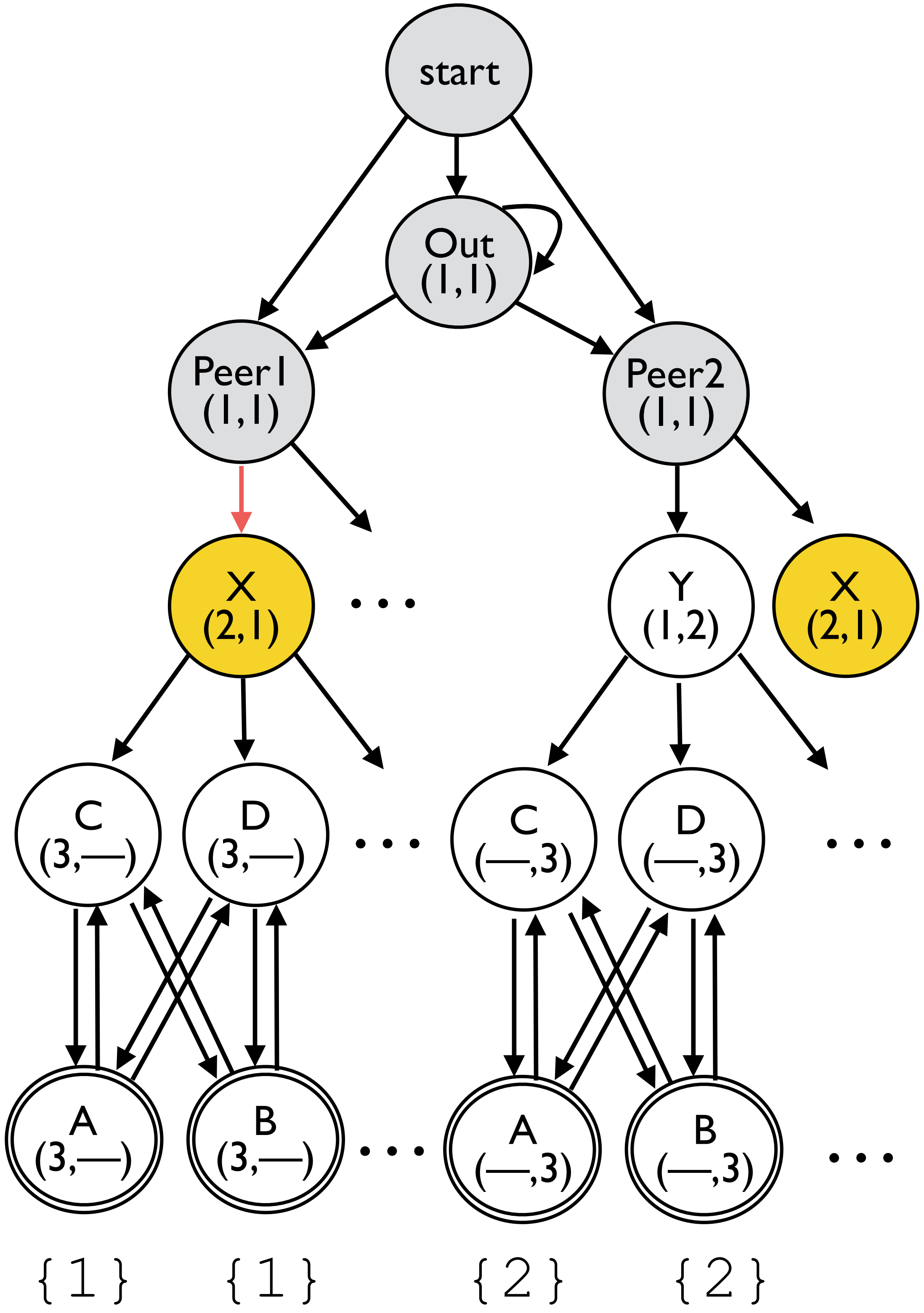
allow regex(Peer1 . out*)

peer ←{C,D,G,H} comm ← (2, 1)

allow regex(Peer2 . out*)

peer ←{C,D,G,H} comm ← (1, 2)

Compilation to BGP:



Router C
```
allow peer=X comm=(2,1)
    peer ←{A,B} comm ← (3, —)
allow peer=Y comm=(1,2)
    peer ←{A,B} comm ← (—, 3)
```

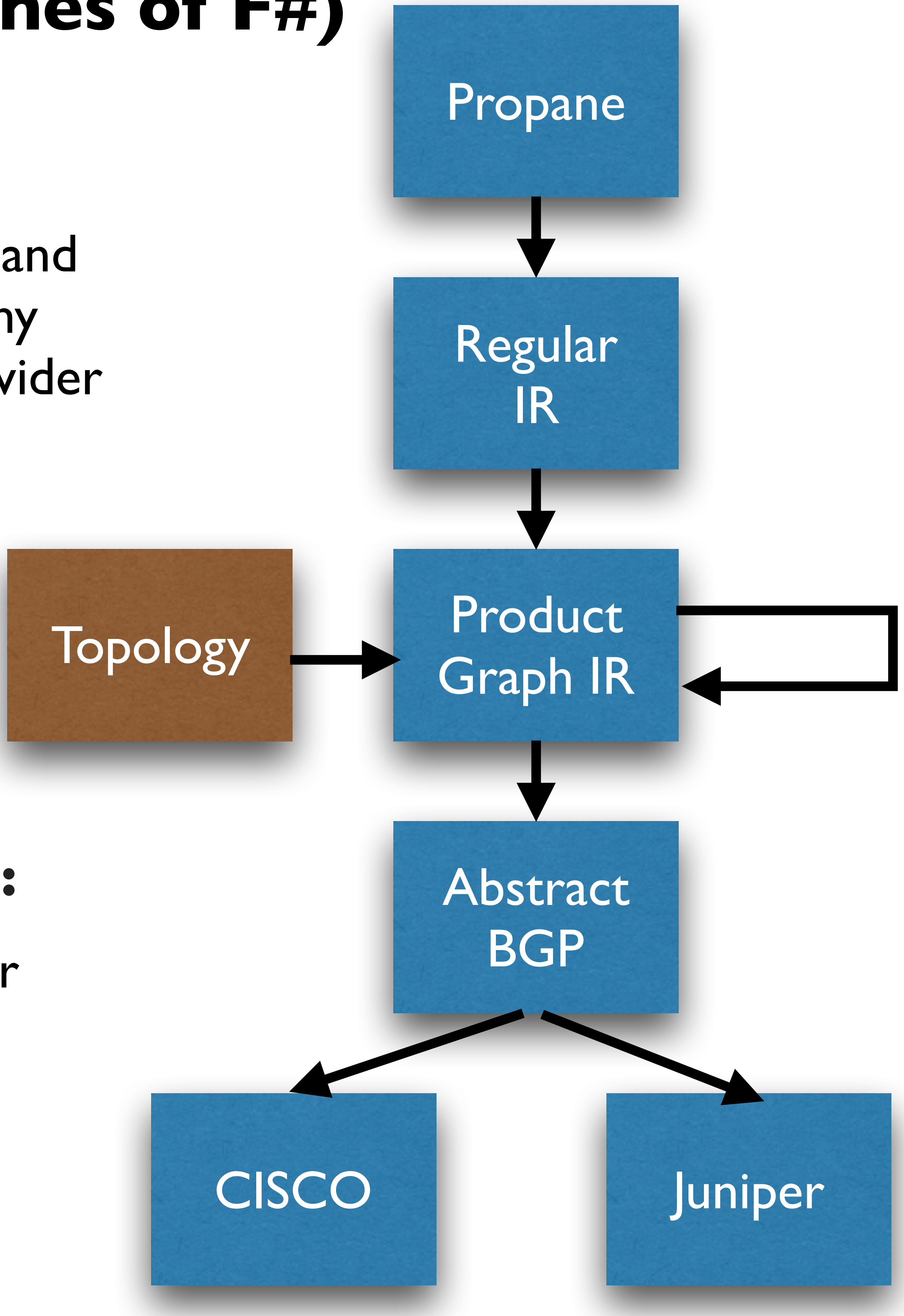
Graph Analysis

Router X
```
allow regex(Peer1 . out*) with lp = 100
    peer ←{C,D,G,H} comm ← (2, 1)
allow regex(Peer2 . out*) with lp = 99
    peer ←{C,D,G,H} comm ← (1, 2)
```

Implementation (5,500 lines of F#)

Benchmarks:

- data center policies (~1600 routers) and backbone policies (~200 routers, many peers/router) from a large cloud provider
- policy from English docs
- Ignoring prefix, customer group and ownership definitions:
  - 31 lines for data center
  - 43 lines for backbone



Scaling (8 core Windows machine):

- 10s/pfx (mean) for largest data center
- 45s/pfx (mean) for largest backbone
- 3 minutes total for the backbone
- 9 minutes total for the data center

Thanks