Building an IXP

Andy Davidson
andy@lonap.net

2Connect / LONAP / IXLeeds
APRICOT2013, Singapore, February 2013

Steal this presentation! ☺

Downloadable notes

Before you buy a **switch**

Planning

Photo courtesy
Orange County Archives

Community

Nomadic Lass - http://www.flickr.com/photos/nomadic_lass/5416543588/

Legal Structure

Legal Entity

Commercial

Association

Define your offering

Peering LAN

Closed User Groups?

Transit OK?

Number Resources

Europe

- End User Contract
- /24 PI only for IXPs
- Special IX v6 rules

Asia/Pac

- Become member
- /22 for all members

Jumbo Frames

> 1500/1522

Peering LAN / CUG

ISLs

Spanning Tree

MRP

Layer 2 Resilience Protocols

VPLS

EAPS

Trill
one day!

Flex Links

Port Security

‘n’ MACs per port

Storm Prevention

Leaky ISP Layer 2

Monitor

Rate Limit

Unknown Unicast

Other Switch Features

Colocation

Independence

Support

Partnership

Ian Harvey, AQL - Photograph of Salem DC2, Leeds - colocation of IXLeeds

Collector
(BGP Router)

Debug

Monitor

Support

After installation

Monitoring

- Availability
- Performance
- Ethernet OAM?

[Availability report completed in 0 min 0 sec]

Hostgroup 'all' Host State Breakdowns:

Host	% Time Up	% Time Down	% Time Unreachable	% Time Undetermined
gw-airportbp-1	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
gw0	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
localhost	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
sw1	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
sw2	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
Average	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%

Hostgroup 'http-servers' Host State Breakdowns:

Host	% Time Up	% Time Down	% Time Unreachable	% Time Undetermined
localhost	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
Average	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%

Hostgroup 'i_cisco_infra' Host State Breakdowns:

Host	% Time Up	% Time Down	% Time Unreachable	% Time Undetermined
gw-airportbp-1	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
gw0	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
sw1	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
sw2	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
Average	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%

Hostgroup 'i_unix_srv' Host State Breakdowns:

Host	% Time Up	% Time Down	% Time Unreachable	% Time Undetermined
localhost	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%
Average	100.000% (100.000%)	0.000% (0.000%)	0.000% (0.000%)	0.000%

Plan for a **dense** cabling environment

SFlow

Customers

Top Peers

Exchange

Marketing
Resilience
Missed Ops

Service Levels

Cost to Exchange funds

Commercial v Association

Communications

Tech Support

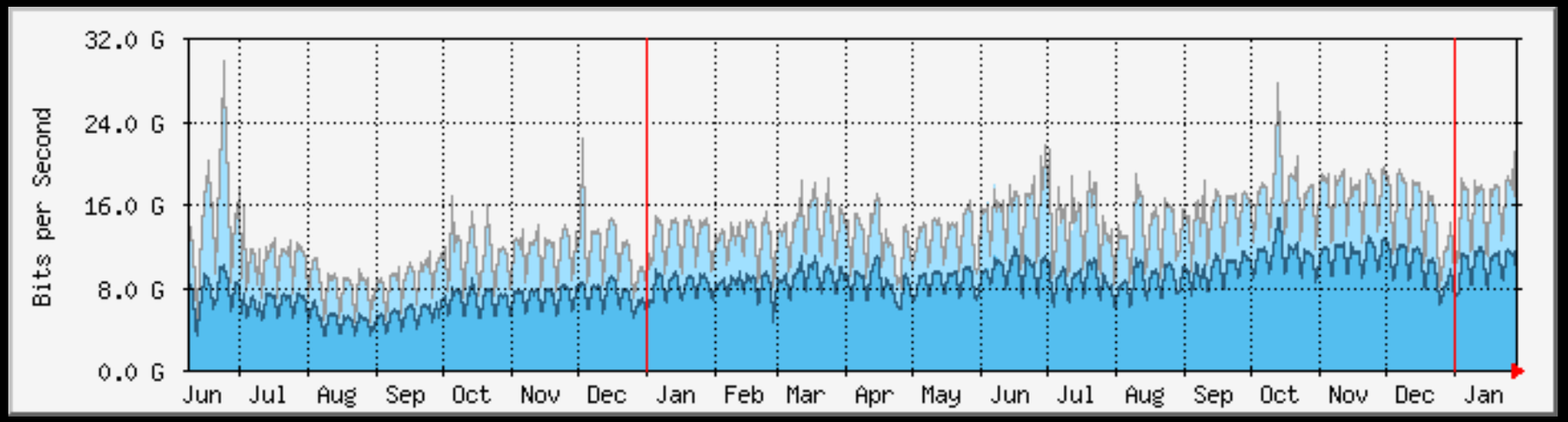
Organisational

New Member

Operations

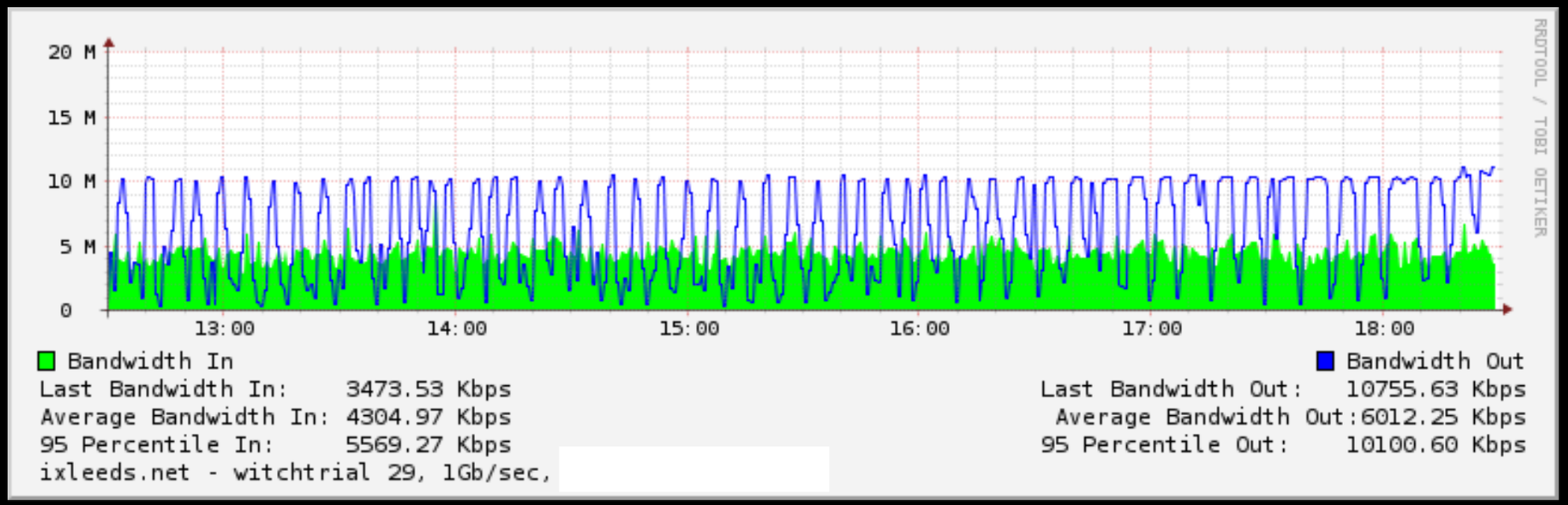
Aggregate Stats

Public
5 Min Interval



Port Stats

Private
1 Min Interval



Growing your Membership

Carrier

Enterprise

Content

Access

Turn up customers!

Announce

Monitor

Connection

Quarantine

Cabling

New customers!

Quarantine Process

New customers

'Naughty' customers

Prohibited Traffic

BGP Clue

Magnus http://www.flickr.com/photos/magh/2159613408/

Offer IPv6

From day one

Same service level

Congestion

Member Ports

ISLs

Interswitch Links

ARP Sponge

When larger

Limits spurious broadcasts

Route

Servers

Motivation

Instant Traffic on Connection

Encourage Peering

Sell Ports, Grow Traffic

sharyn morrow - http://www.flickr.com/photos/sharynmorrow/1923985105/

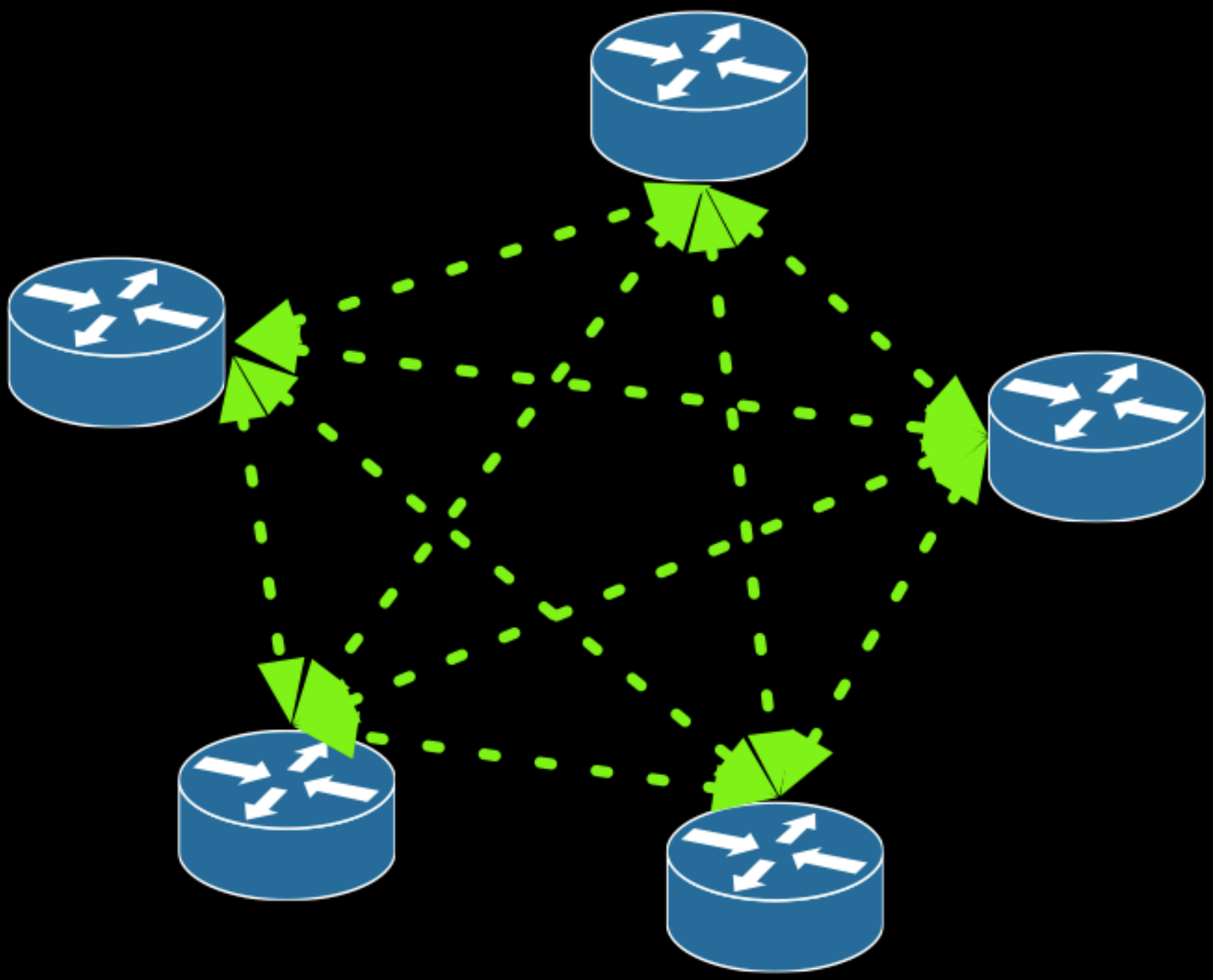
Route Servers / MLP

1 Session, Many Peers

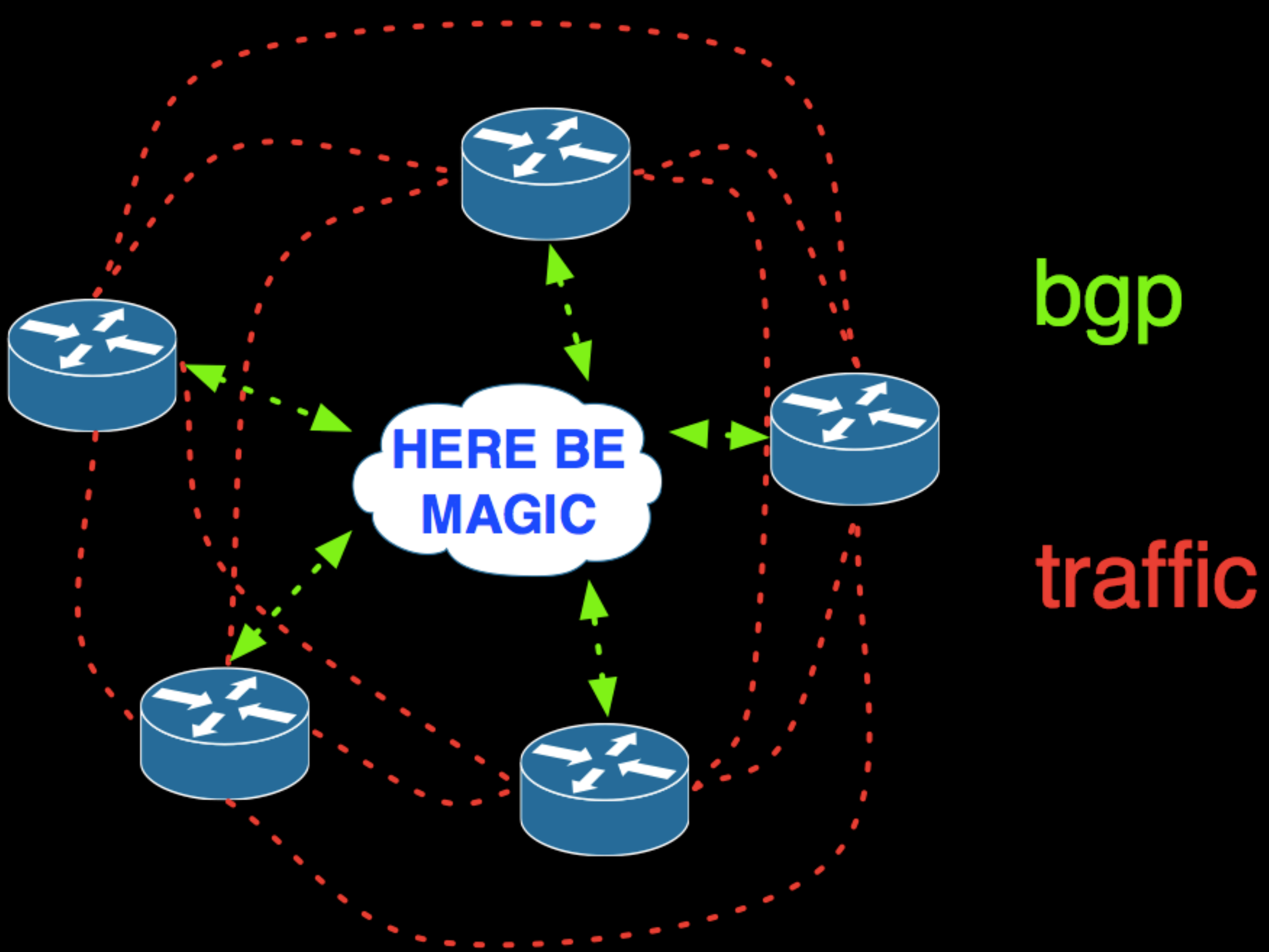
Make it reliable

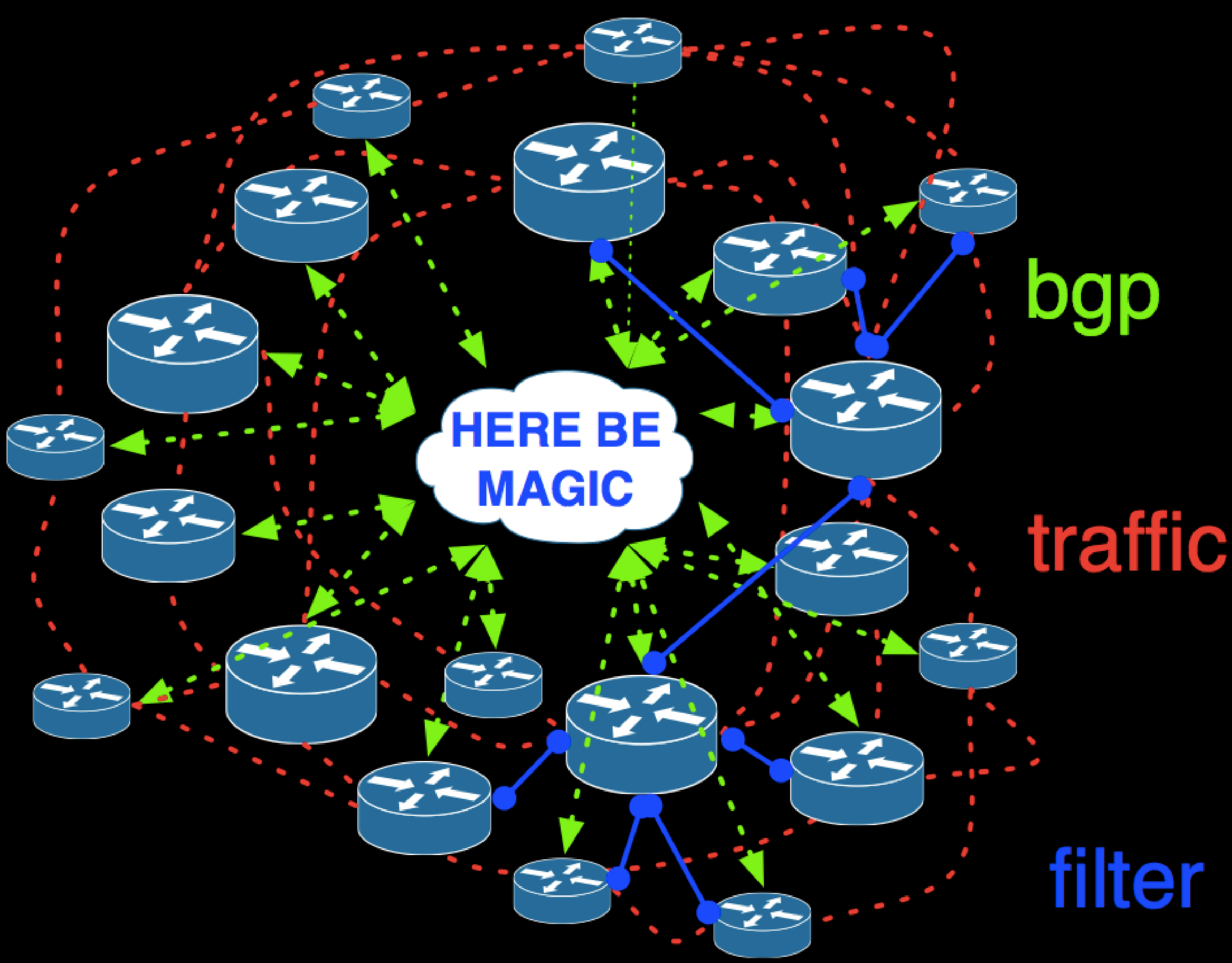
Make it **optional**

Election night crowd, Wellington, 1931. Alexander Turnbull Library

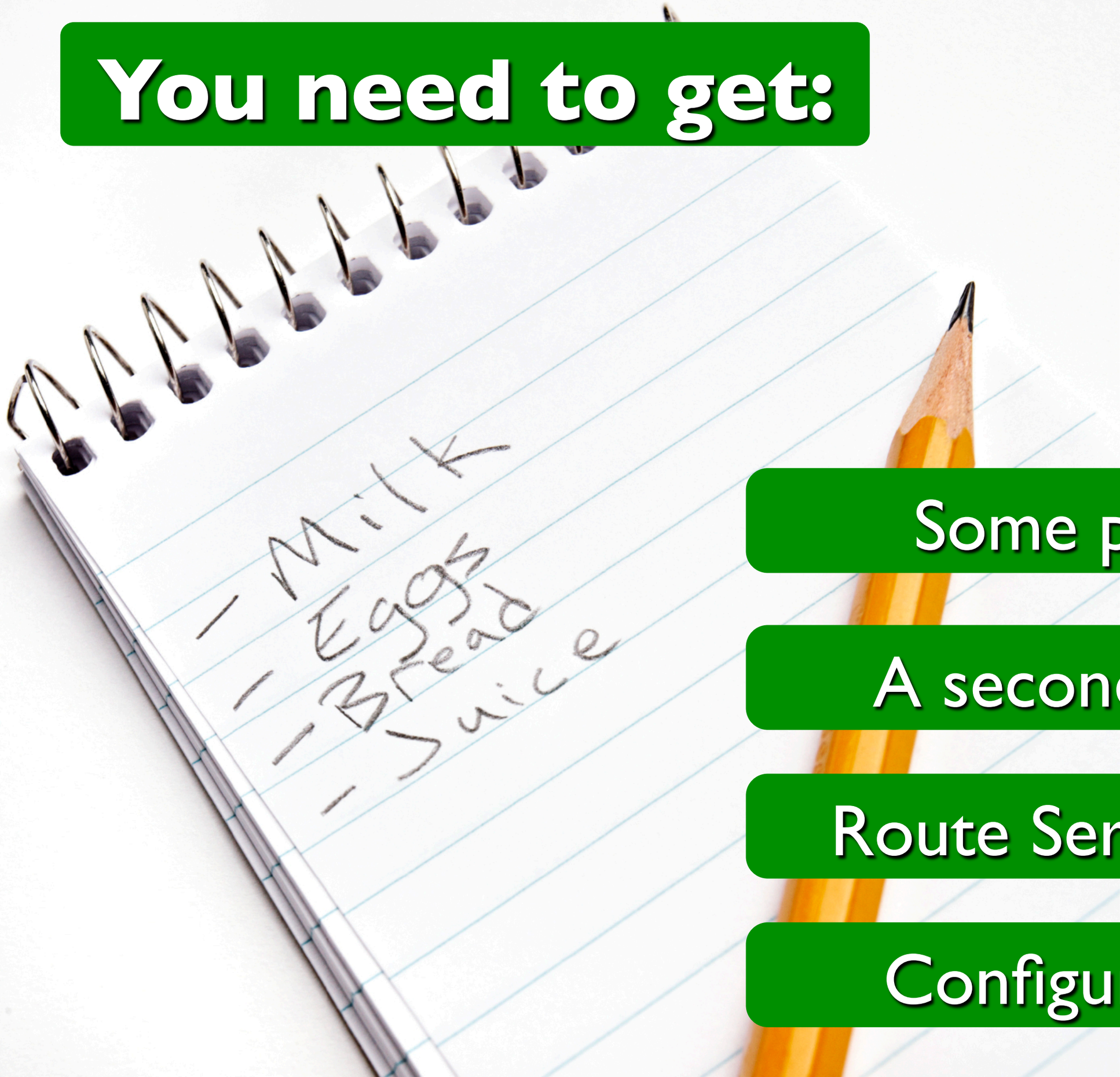


bgp


HERE BE
MAGIC
bgp
traffic


HERE BE
MAGIC
bgp
traffic
filter

You need to get:



- Some peers
- A second ASN
- Route Server s/w
- Configuration

From: noc@clevernet.net.uk
To: support@lonap.net
Subject: Clevernet Route Server

Dear Colleagues,

Please can I ask that AS65534 turn up sessions facing the IXP route server. My AS-SET is AS-CLEVERNET for v4 and v6 routes.

My router is at 193.203.5.x, 2001:7f8:17::FFFE:1.

Yours, CleverNoc

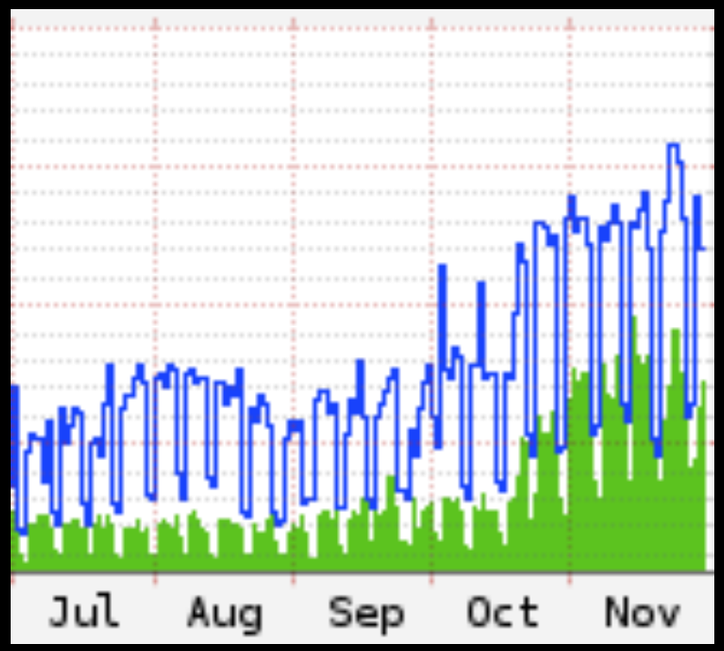
```
router bgp 123
     no bgp enforce-first-as      (- very important for route servers)
     neighbor lonaprs peer-group    (- own group recommended)
     neighbor lonaprs remote-as 8550
     neighbor lonaprs description LONAP MLP
     neighbor lonaprs route-map lonap-rs-out out
     neighbor lonaprs route-map lonap-rs-in in
     neighbor lonaprs maximum-prefix 20000
     neighbor 193.203.5.1 peer-group lonaprs
     neighbor 193.203.5.2 peer-group lonaprs

route-map lonap-rs-out
     match as-path 10          (- or however you prefix filter)
     set community xxx

route-map lonap-rs-in
     set local-preference 1000     (- or whatever you use for peers)
```

```
protocols {
    bgp {
        group lonap-rs {
            peer-as 8550;
            description “LONAP Route Servers”;
            family inet {
                unicast {
                    prefix-limit {
                        maximum 20000;
                        teardown 99;
                    }
                }
            }
            import [ lonap-in rejectpolicy ];
            export [ as65534 bgp_customers aggregate rejectpolicy ];
            neighbor 193.203.5.1;
            neighbor 193.203.5.2;
        }
    }
}
```

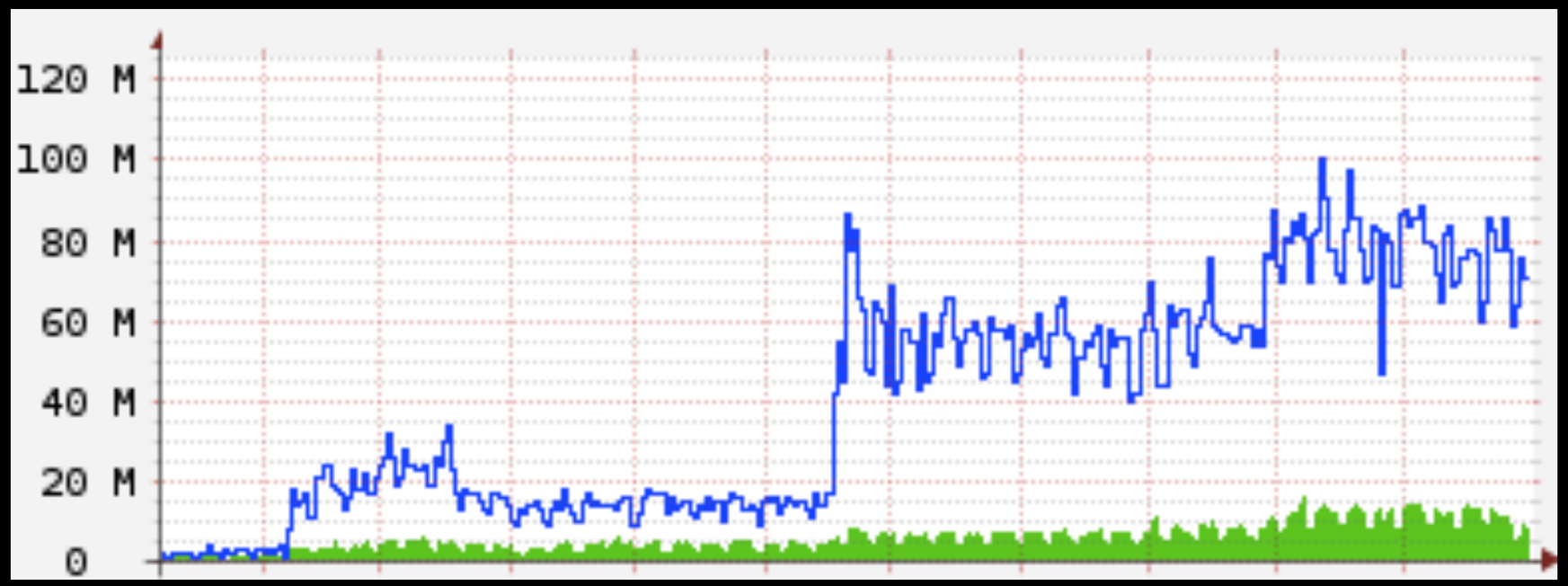
```
operator@thdo-bdr-8> show bgp summary | match 8550 193.203.5.1           8550      42732      40417
0       1     1w5d13h 243/3185/3185/0      0/0/0/0193.203.5.2           8550      48666      40795       0
1     1w4d14h 1567/3064/3064/0     0/0/0/0
```

```
peering-3#sh ip bgp sum | i 8550193.203.5.1     4  8550   39530   39417
141140751    0    0 1w5d          3120193.203.5.2     4  8550   22866   18437
141140751    0    0 1w4d          3048
```

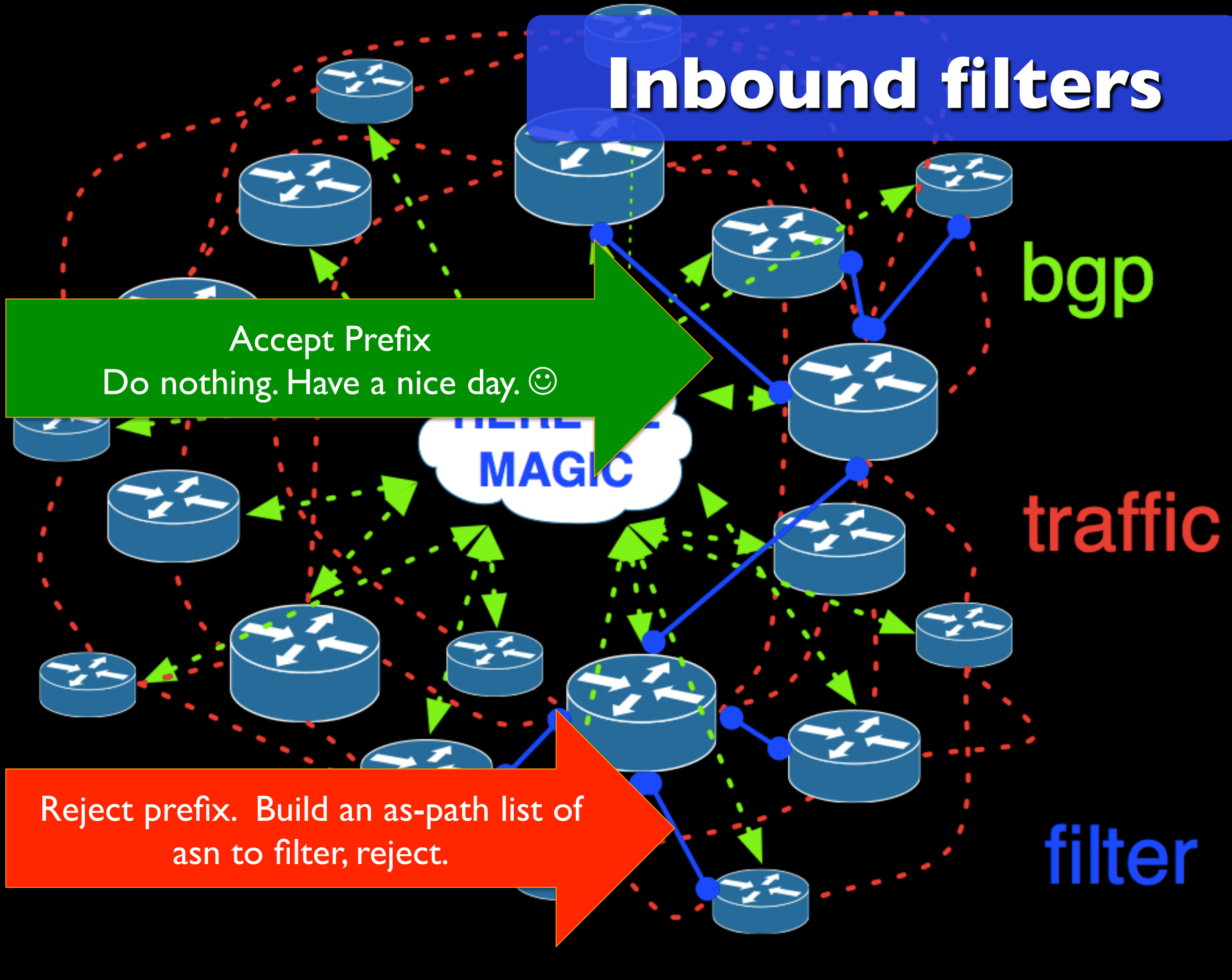
Effect on existing members:



Effect on new members:



Inbound filters



Outbound filters

bgp

traffic

filter

Send my prefix (e.g. deaggregate, or selective policy) **only** to certain peers
65534:peer-as

Hide my prefix from transit customer 0:peer-as

32 bit ASN filtering

Hide pfx from specific peers.
Community 0:peer-as

Send specific pfx to peers.
Community 8550:peer-as

32bit?
RFC5668 – still too small

Rich Renomeron - https://www.flickr.com/photos/rrenomeron/2834296612/

Publish in RIR DB

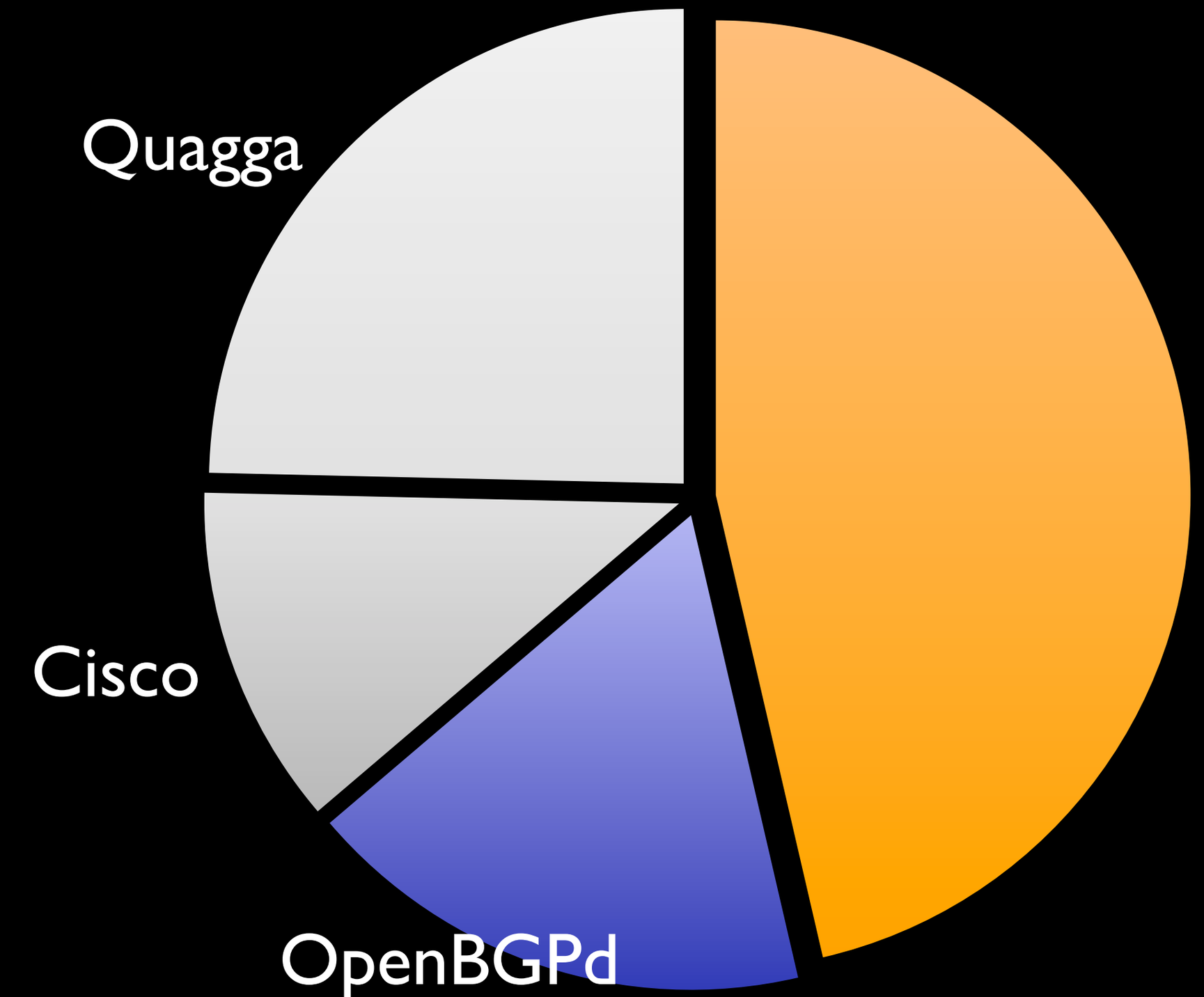
```
% Information related to 'AS8550'

aut-num:        AS8550
as-name:        LONAP-MLP
descr:          LONAP Ltd.
org:            ORG-LLUL1-RIPE
admin-c:        AS8330-RIPE
tech-c:         AS8330-RIPE
mnt-by:         AS8330-MNT
mnt-by:         RIPE-NCC-END-MNT
source:         RIPE # Filtered
remarks:        For connected networks see AS-LONAP-MLP or AS-LONAP-MLP-v6
remarks:        ------------------------------------------------
remarks:        Communities to control route re-distribution:
remarks:        8550:8550 - propagate to all destination networks
remarks:        8550:n    - propagate to AS n
remarks:        0:8550    - deny to all destination networks
remarks:        0:n       - deny to AS n
remarks:        Default policy is DENY ALL, so for an open peering
remarks:        policy set 8550: 8550 on all routes.
remarks:
remarks:        ------------------------------------------------
```

Hygienic Peering

- Enforced IRR database filtering on MLP (you do not need to filter the MLP)
- Required to peer with several MLP servers
- Config change via automation system to reduce human error impact

Software use in Europe

PS, Use Bird!



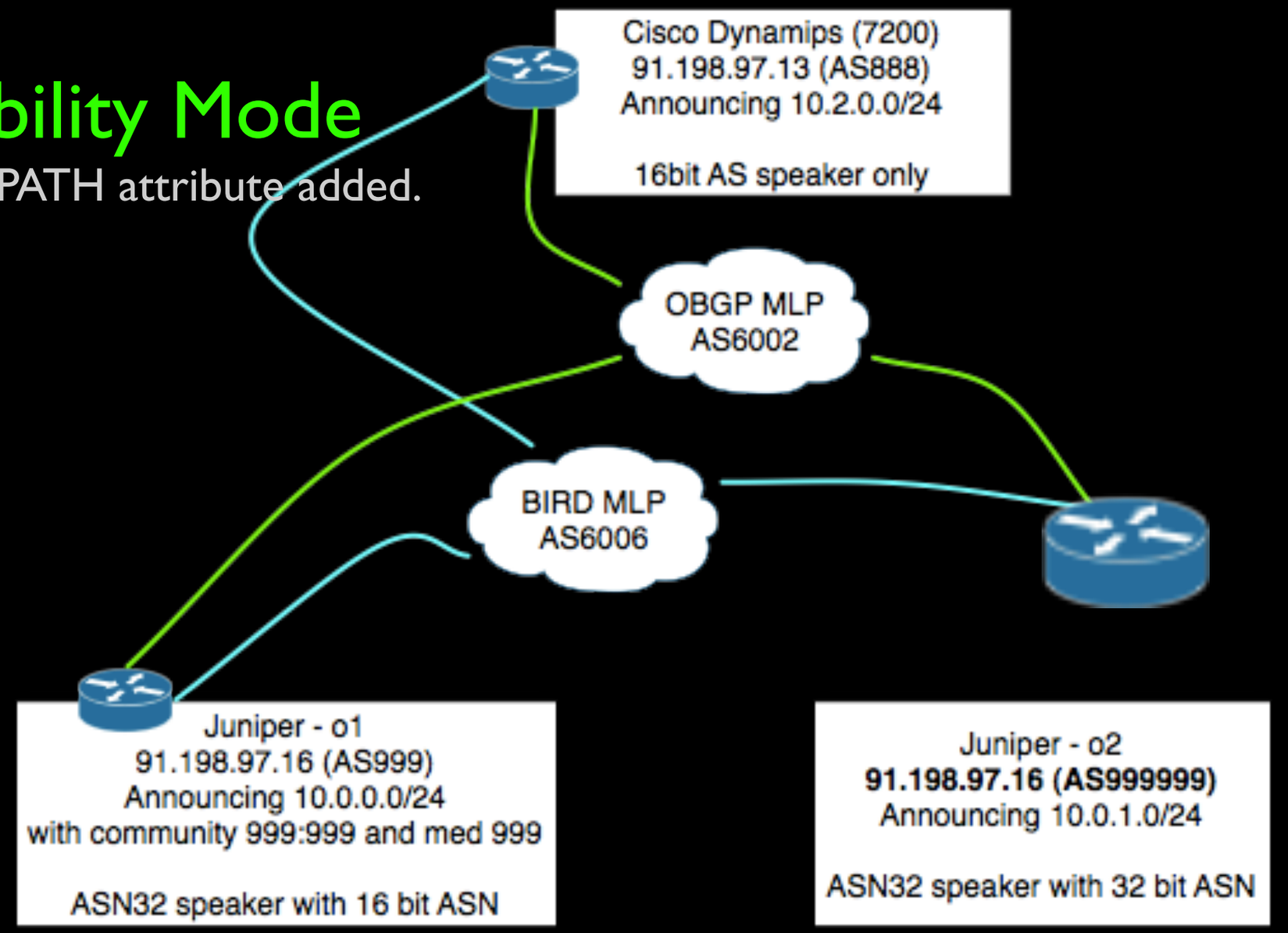
Source: Euro-IX

Correct 25rd Feb 2013

ASN32 - Just works

Correct Compatibility Mode

AS23456 in AS_PATH. NEW_AS_PATH attribute added.



Correct 32 bit ASN properties

(AS4_PATH, adjacent to AS999999)

http://tools.ietf.org/html/
draft-jasinska-ix-bgp-route-server-03

How we do it at LONAP……!

How we do it...

ssh

```
andy@chilli > pwd
/home/andy/lonap-svn/rs-config
andy@chilli > cat cfgSamplePeers.cfg
193.203.5.4,65530,AS-SUPERNET,Supernet
193.203.5.5,65540,AS-MAGICNET,Magicnet
193.203.5.6,65530,AS-GLOBALNET,GlobalNet
193.203.5.7,65520,AS-CONNECTNET,Connectnet
2001:7F8:17::10:1,65530,AS-SUPERNET,Supernet
2001:7F8:17::20:2,65540,AS-MAGICNET,Magicnet
andy@chilli > cat cfgTplConfigs.cfg
tplBIRD.tpl,bird.conf
tplBIRD6.tpl,bird6.conf
tplOpenBGP.tpl,bgpd.conf
andy@chilli >
```

Configure AS/AS-Set

```
andy@chilli > cat tplOpenBGP.tpl
# Config version: $Id: tplOpenBGP.tpl 1246 2011-11-17 14:14:23Z andy $

AS 8550
router-id 193.203.5.1
transparent-as yes
fib-update no

[% FOREACH i IN uniqueasn %]rde rib peer[% i %]
[% END %]

group "RS" {
  announce all
  set nexthop no-modify

  [% FOREACH neigh IN neighbors %]
  neighbor [% neigh.neighbor %] {
    descr "[% neigh.desc %]"
    remote-as [% neigh.asn %]
    passive
    max-prefix [% neigh.maxpfx %] restart 15
    enforce neighbor-as yes
    rib peer[% neigh.asn %]
    announce IPv4 unicast
  }
  [% END %]

  [% FOREACH neigh IN neighbors6 %]
  neighbor [% neigh.neighbor %] {
    descr "[% neigh.desc %]-6"
    remote-as [% neigh.asn %]
    max-prefix [% neigh.maxpfx %]
    enforce neighbor-as yes
    rib peer[% neigh.asn %]
    announce IPv6 unicast
    announce IPv4 none
  }
  [% END %]

}
```

Iterate over configured peers

```
andy@chilli > perl RSBuild.pl -c cfgSamplePeers.cfg -b cfgTplConfigs.cfg

        group "RS" {
          announce all
          set nexthop no-modify
          neighbor 193.203.5.5 {
            descr "Magicnet"
            remote-as 65540
            passive
            max-prefix 100 restart 15
            enforce neighbor-as yes
            rib peer65540
            announce IPv4 unicast
          }

          neighbor 193.203.5.6 {
            descr "GlobalNet"
            remote-as 65530
            passive
            max-prefix 100 restart 15
            enforce neighbor-as yes
            rib peer65530
            announce IPv4 unicast
          }

        allow quick from 193.203.5.5 prefix 10.0.0.0/23 prefixlen <=24
        allow quick from 193.203.5.5 prefix 10.1.1.0/24 prefixlen <=24
        allow quick from 193.203.5.6 prefix 172.16.2.0/24 prefixlen <=24
```

Generates neighbors per peer

Use the FULL CONFIG in the notes

And filters

Shortcut for peer filtering

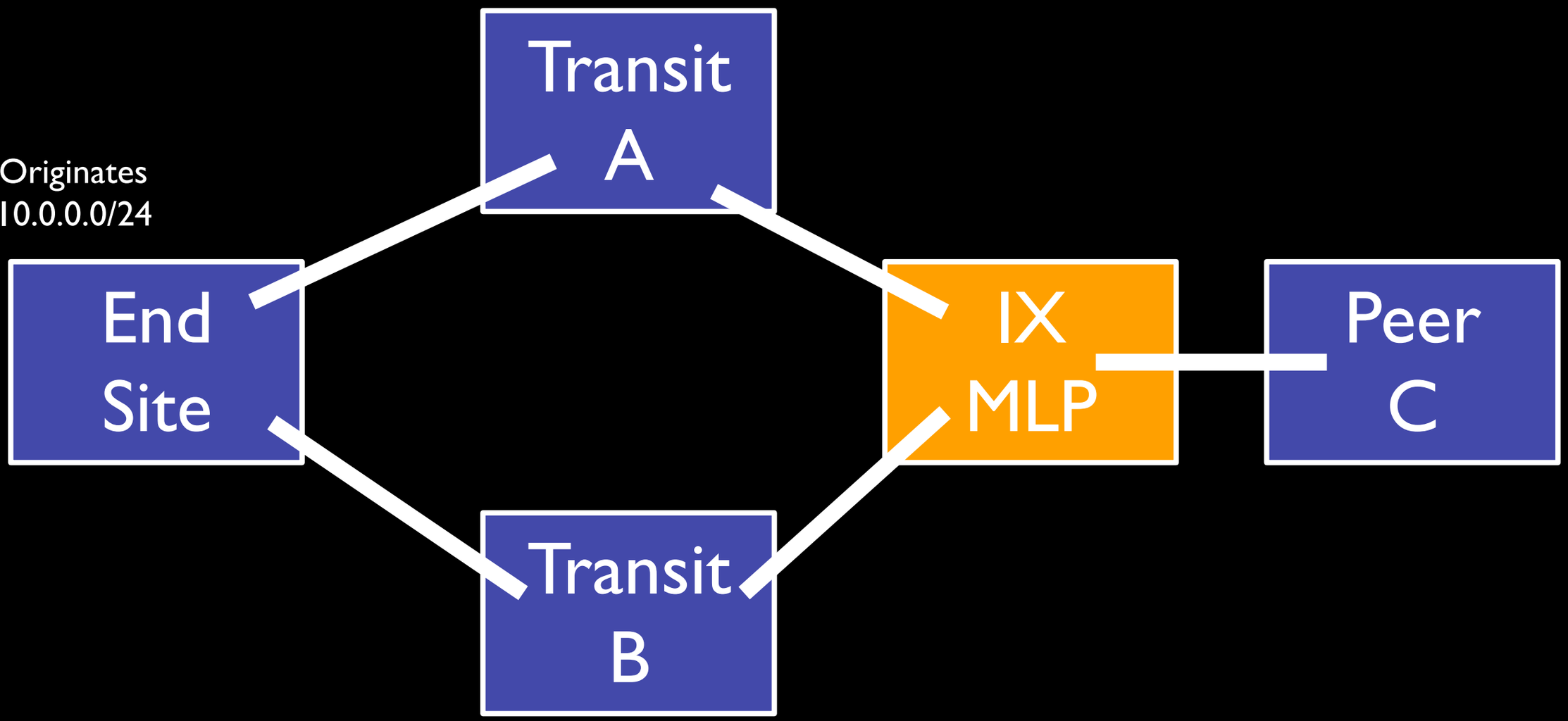
OpenBGP

```
deny quick to group RS community 0:neighbor-as
deny quick to group RS community 0:8550
allow to group RS community 8550:neighbor-as
allow to group RS community 8550:8550
```

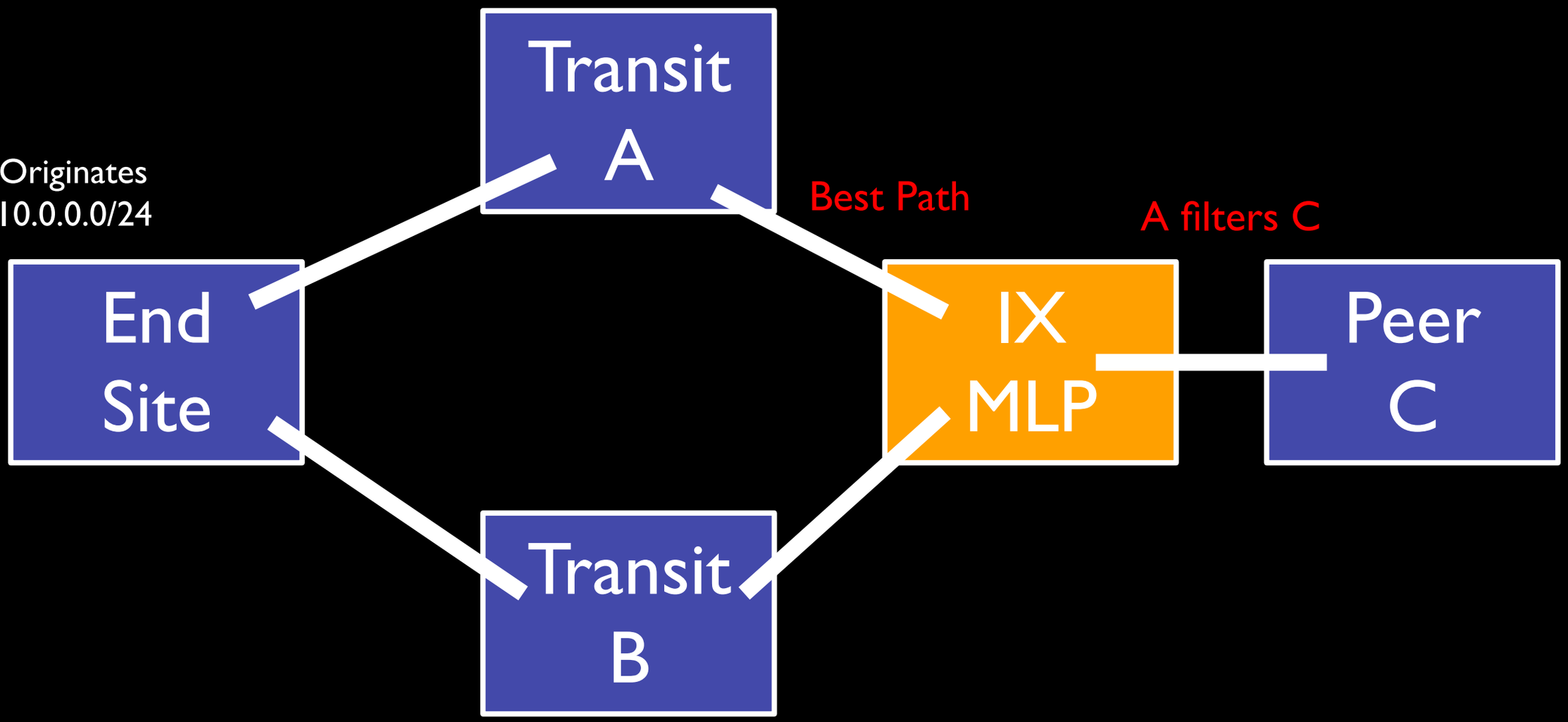
Bird

```
function bgp_out(int peeras)
{
        if ! (source = RTS_BGP ) then return false;
        if (0,peeras) ~ bgp_community then return false;
        if (myas,peeras) ~ bgp_community then return true;
        if (0, myas) ~ bgp_community then return false;
        return true;
}
```

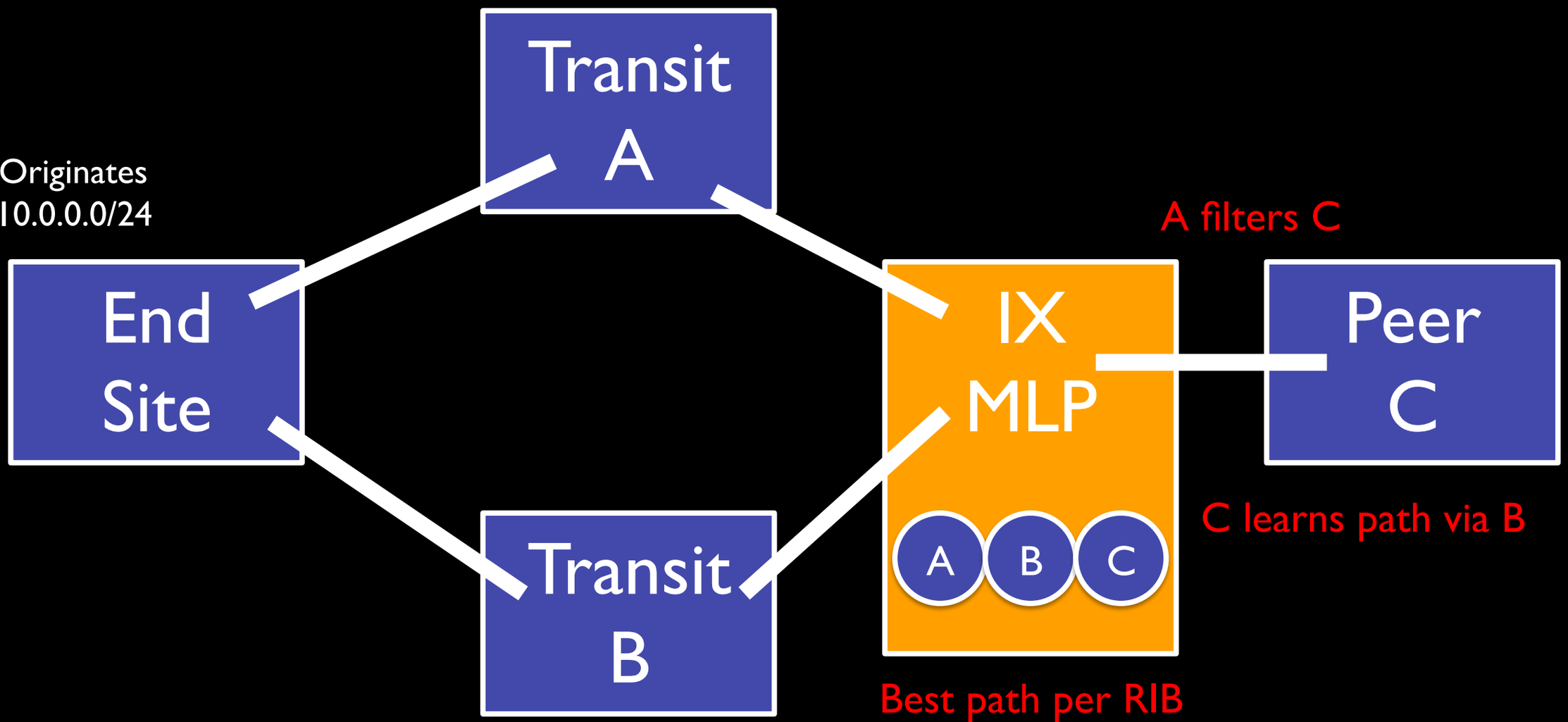
The Tale of the Single RIB



What will happen here?



What will happen now?



```
/* build a list of bgp tables */
table T65520;
table T65530;
table T65540;


/* build a list of bgp pipes */
protocol pipe P65520 {
        table master;
        mode transparent;
        peer table T65520;
        import where bgp_in(65520);
        export where bgp_out(65520);
}
protocol pipe P65530 {
        table master;
        mode transparent;
        peer table T65530;
        import where bgp_in(65530);
        export where bgp_out(65530);
}
protocol pipe P65540 {
        table master;
        mode transparent;
        peer table T65540;
        import where bgp_in(65540);
        export where bgp_out(65540);
}
```

```
protocol bgp R65530x4 {
        local as 8550;
        neighbor 193.203.5.4 as 65530;
        passive on;
        import all;
        export all;
        route limit 100;
        table T65530;
        connect retry time 6000;
        rs client;
}
```

Bird

```
rde rib peer65520
rde rib peer65530
rde rib peer65540
```

```
neighbor 193.203.5.4 {
  descr "Supernet"
  remote-as 65530
  passive
  max-prefix 100 restart 15
  enforce neighbor-as yes
  rib peer65530
  announce IPv4 unicast
}
```

OpenBGP

Any Questions?

andy@lonap.net
+44 114 3190605
twitter: @andyd