BGP Traffic Engineering

Andy Davidson
2Connect UK, LONAP, IXLeeds
.
andy@2connectintl.com

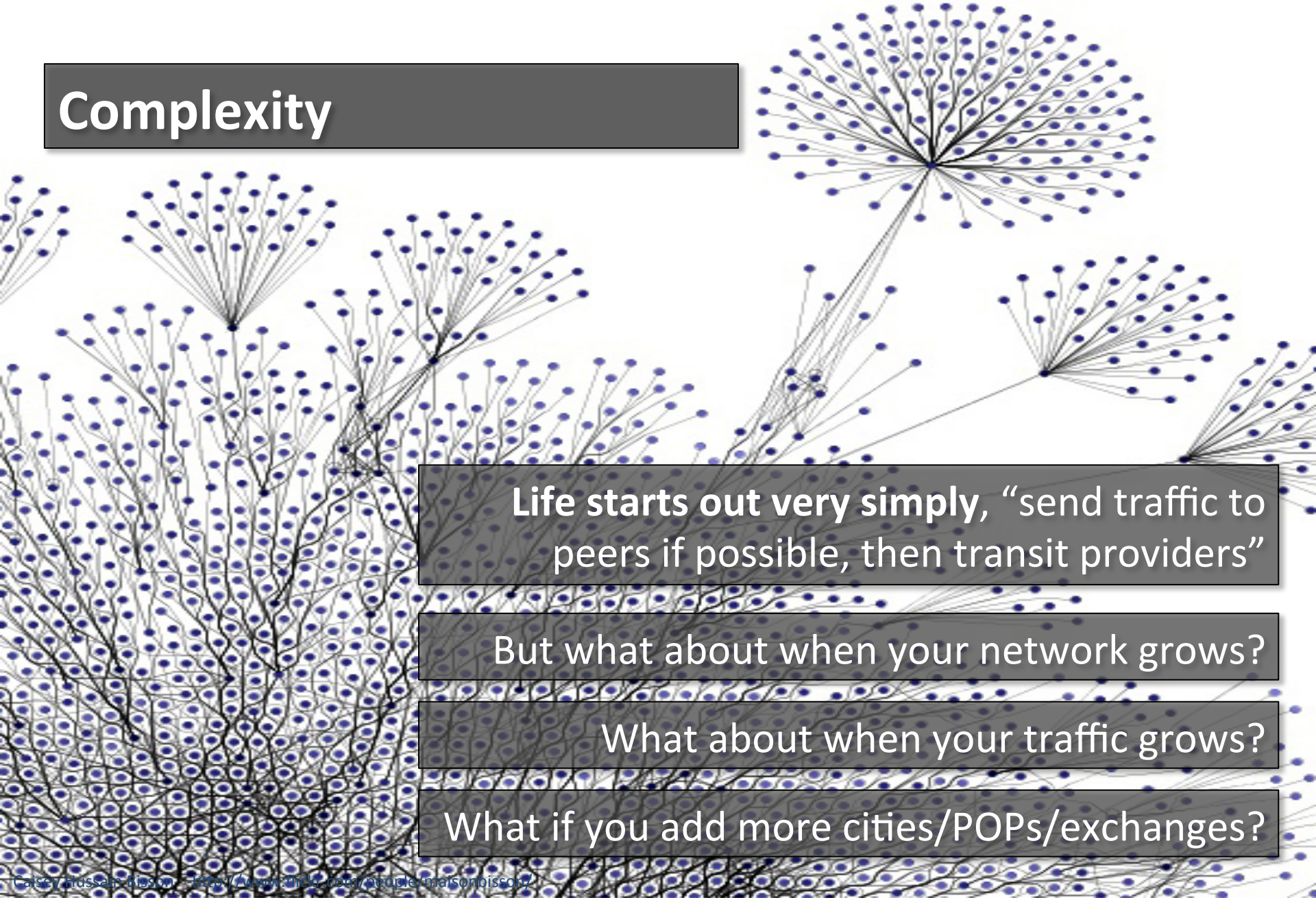
APRICOT2013, 25 Feb '13, Singapore
BCIX Tech Roundtable, 19th Nov '12, Berlin

Why do Traffic Engineering?

- Manage your capacity demands
- Ensure service quality
- Recover from Failures
- Manage service/circuit costs
- Handle traffic growth

James Cridland http://www.flickr.com/photos/jamescridland/

Complexity



Life starts out very simply, "send traffic to peers if possible, then transit providers"

But what about when your network grows?

What about when your traffic grows?

What if you add more cities/POPs/exchanges?

Caisey Hussain Bisson: http://www.flickr.com/people/maisonbisson/

Real examples

- Circuits with **cost difference** > $100/Mbit
- Regional networks - **poor local peering**
- Circuit failure causing **congestion**
- Changing **customer demand**/behaviour
  - Increased quality expectation

Internal network TE

- **Simple** compared with Interdomain TE

- You administrate both sides
  - You know the **price** of all paths
  - The IGP knows the **capacity** of all paths
  - IGP protocols let you map price, capacity to shape routing using **cost**.

Inter-domain TE

- You do NOT control both sides
  - Path vector protocols hide metric, capacity, cost
  - Simplicity of BGP protocol imposes **limitations**
  - **Volume of traffic** matters, not # of routes
- However, large volume of traffic is usually with a small number of other ASNs

BGP Best Path Selection Algorithm

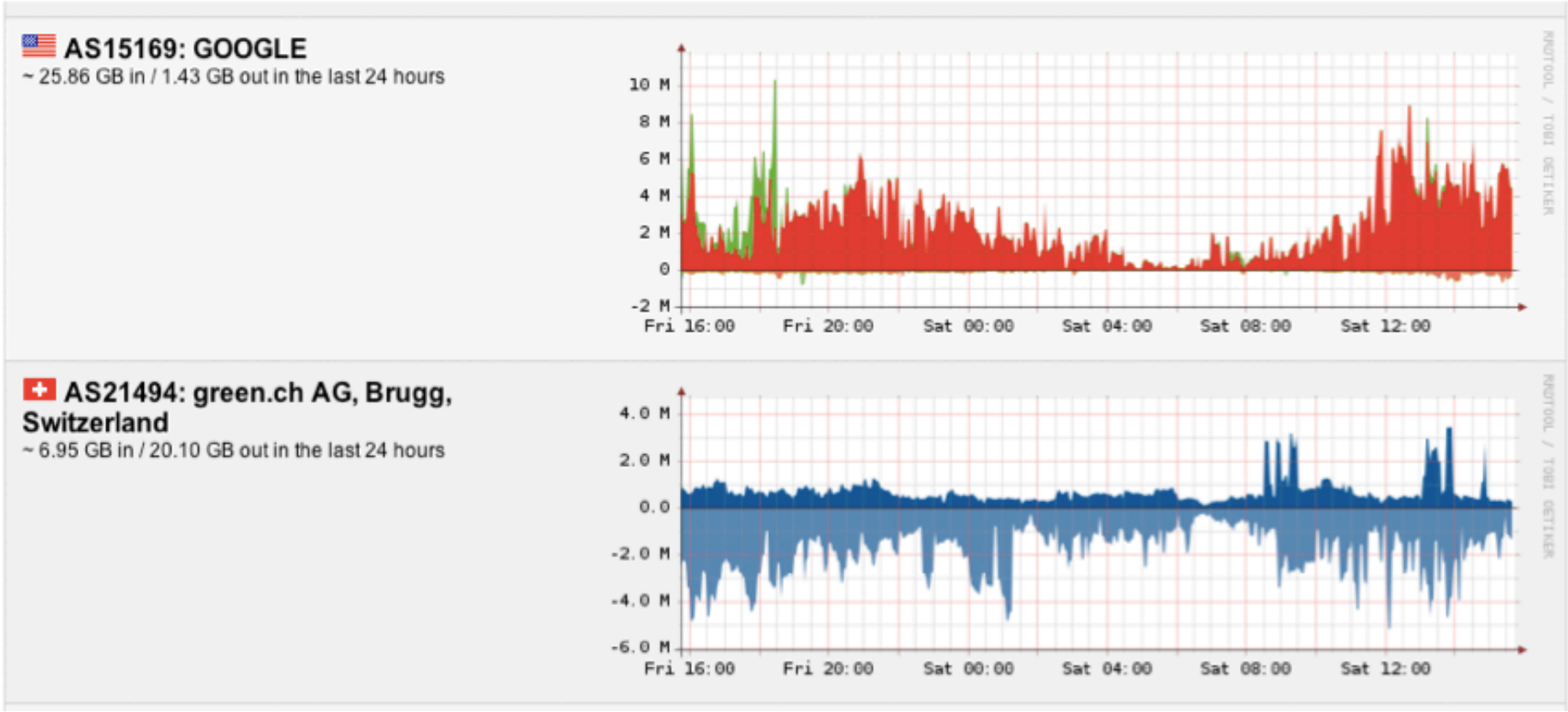
- Affects traffic in outbound direction
  - Local Preference
  - AS PATH length
  - Lowest Origin Type
  - Lowest MED
  - Prefer eBGP paths
  - Lowest IGP Metric
  - Oldest route

Outbound vs Inbound

- Outbound heavy networks
  - Somewhat **easier life**
- Inbound heavy networks
  - You must **trick** the Best Path Selection methods of networks sending you traffic.
  - *Their* config change will move ***your*** traffic.

Elliot http://www.flickr.com/photos/pointnshoot/677657225/

You need **data**



Manuel Kasper - https://neon1.net/as-stats/as-stats-presentation-swinog16.pdf

Other ways to get data

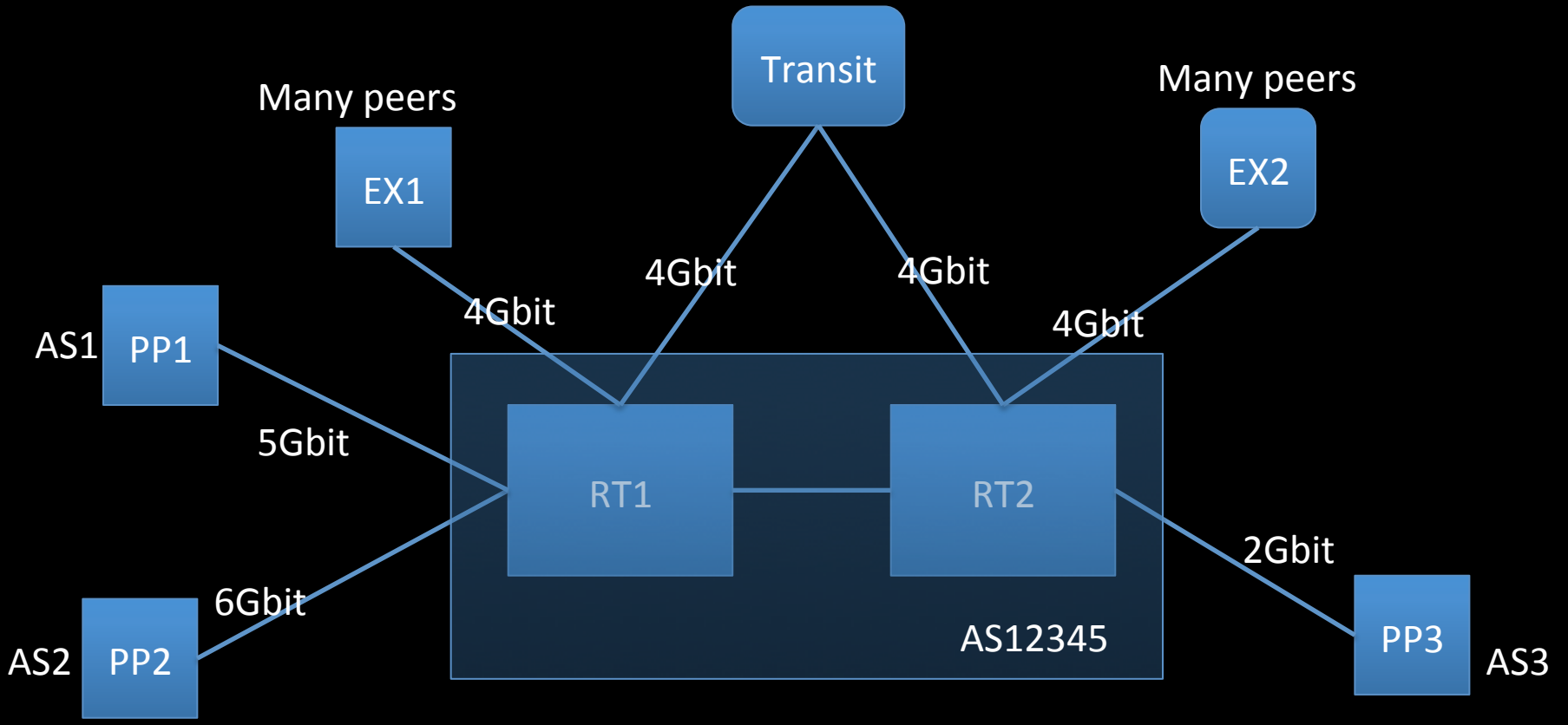
- Log file analysis
  - Useful before you have a network, for working out the benefit of building a network/peering.
  - Best for ‘single service’ networks, webshops…
- Wild Guess
  - Your instinct is better than you think, confirm with top talkers, etc.
  - But you should use Netflow. ☺

Data tells you

- Your traffic **direction**
  - Mainly inbound
  - Mainly outbound
  - Balanced
- Your **top traffic originators** or **destinations**

Mainly outbound, single POP

- Localpref
  - A **hammer** – blunt tool, inflexible.. But it is a tool.
  - "Generally" prefer to send traffic to customers, then peers, then transits.
  - Manage top 'n' networks, so that there is a **preferred path**, and a **failure path**, with capacity on both circuits.



AS2 is your largest flow - via PP2 - maybe needs a second private peer backup on RT2?
AS1 via PP1, configure a backup over EX1 or EX2 for deterministic routing?
Can you move larger peers behind EX1 and EX2 onto private peering?
If there is an exchange failure, where will the traffic go? How big a flow should you care about?
If you lose RT2, how will traffic to PP3 and traffic volume via EX2 be delivered?
If you lose RT1, how will traffic volume via PP3 and EX1 be delivered?

Localpref – blunt hammer

10.0.0.0/8 Localpref 100 via 100 123
10.0.0.0/8 Localpref 500 via 300 200 200 200 200 123

Which link will you prefer ?
AS123 here is trying to shape inbound traffic via AS100. Why ?
- Higher capacity link ?
- More reliable ?

What should you do ?

Answer: It depends on the **volume** of traffic, **cost** of capacity, **value** of traffic

Configuration Example

Larger flows are in ASNs
Listed in as-path 30 and 40

Deterministic exits configured

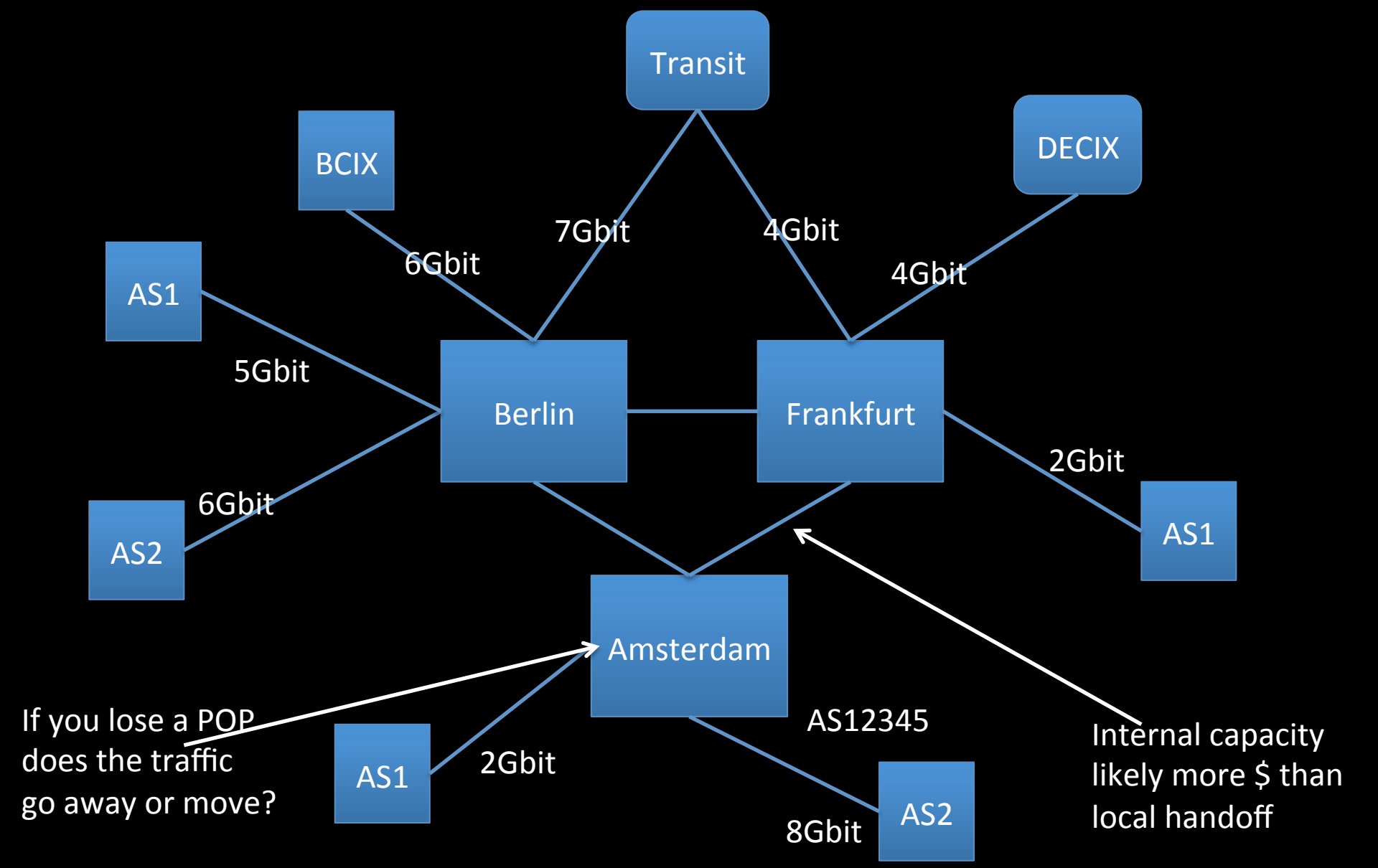
```
ip as-path access-list 30 permit _7018_
ip as-path access-list 30 permit _2828_
ip as-path access-list 30 permit _4323_
ip as-path access-list 30 permit _3561_
ip as-path access-list 30 permit _1668_
ip as-path access-list 40 permit _3330_

route-map PEER_EX1 permit 10
 match as-path 30
 set local-preference 300
route-map PEER_EX1 permit 15
 match as-path 40
 set local-preference 200
route-map PEER_EX1 permit 20
 set local-preference 150

route-map PEER_EX2 permit 10
 match as-path 40
 set local-preference 300
route-map PEER_EX2 permit 15
 match as-path 30
 set local-preference 200
route-map PEER_EX2 permit 20
 set local-preference 150
```

Mainly outbound – Many POPs

- Use hot potato routing to best effect
  - Nearest exit routing
  - Understand who your top traffic sinks are and peer at all POPs
  - Ignore MEDs from others – unless you want to carry the traffic on your backbone



If you understand your top flows, you will cope with traffic growth and failures

Deterministic routing

- Local Preference
- AS PATH length
- Lowest Origin Type
- Lowest MED
- Prefer eBGP paths
- Lowest IGP Metric
- Oldest route

Top flows should leave your network via deterministic means, and not left to BGP Best Path selection (or to chance).

If you are relying on oldest route to make the decision, you risk traffic taking unpredictable routes.

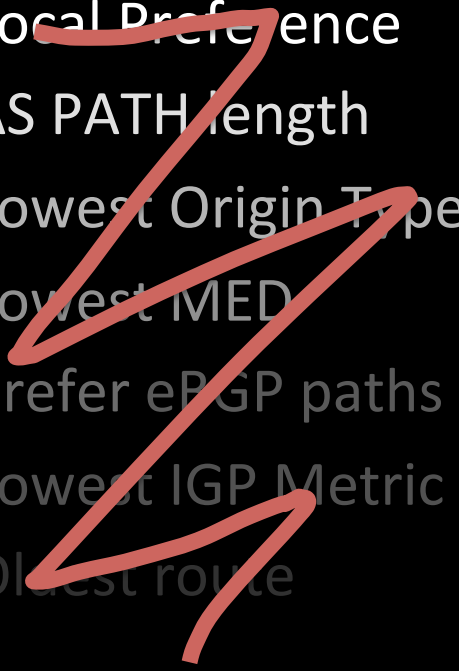
However, oldest routes do break the 'flapping sessions' problem. You need to monitor and manage your top flows constantly.

Inbound traffic engineering

- Much harder
  - Trick others' Best Path calculations
  - You do not administrate origin party router
- But remember…
  - Largest flows come from a small number of networks
  - Content networks want to deliver traffic to you as well as possible!

Selective Announcements

- Shortest prefix
- Local Preference
- AS PATH length
- Lowest Origin Type
- Lowest MED
- Prefer eBGP paths
- Lowest IGP Metric
- Oldest route



Prefix length considered before BGP.

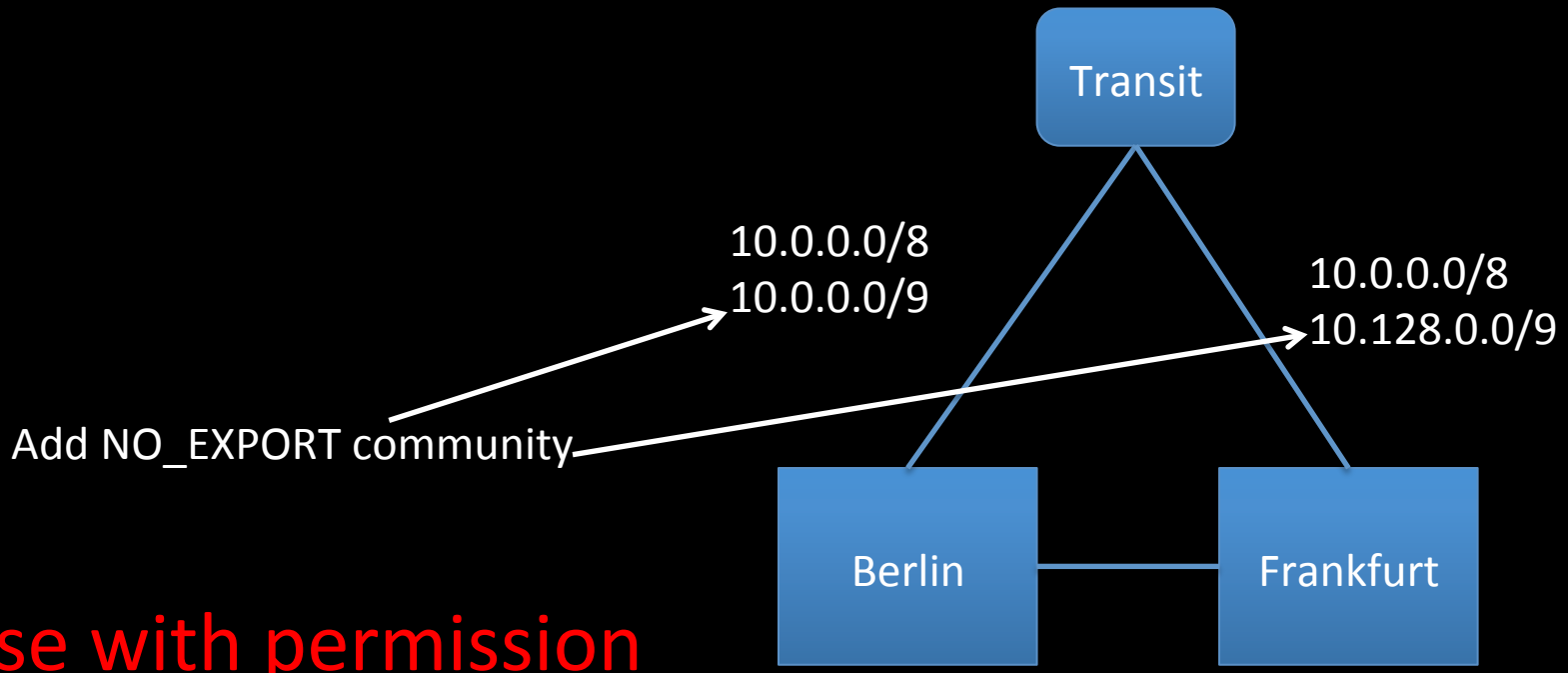
10.0.0.0/16 vs 10.0.0.0/17 & 10.129.0.0/17

Problem of Selective Announcements

- Often filtered
- Considered rude – might lead to depeering
- Never announce ‘globally’

…But can be used to great effect

- To the same peer or transit provider, announce aggregate and regional pfx



Use with permission

AS_PATH prepending

- Signal preferred path by growing AS_PATH on less preferred paths
- Marginal effect which degrades quickly
- Signal backup link to a single AS, but load-balancing capacity is much harder
- May not be heard at ‘distant’ ASNs
- Another ‘blunt’ tool, but can move some traffic.

2.5 AS Path Prepending

AS path prepending is a common way of making routes less attractive since AS path length is usually one of the BGP path selection criteria. A customer network may use these communities to selectively request AS3320 to insert additional copies of the AS number 3320 when propagating the customer routes to neighbors.

Community Value	Name	Description
65012 : X	AS Prepend 2x to AS X	Prepend 3320 two times to named peer (ASN=X)
65013 : X	AS Prepend 3x to AS X	Prepend 3320 three times to named peer (ASN=X)
6501n : 65001	AS Prepend by Class: Peer	Prepend 3320 n times to peers. n=2 or 3.
6501n : 65002	AS Prepend by Class: Upstream	Prepend 3320 n times to upstream.
6501n : 65003	AS Prepend by Class: Peer & Upstream	Prepend 3320 n times to peers and upstream.
6501n : 65004	AS Prepend by Class: Customer	Prepend 3320 n times to customers.
6501n : 65005	AS Prepend by Class: Customer & Peer	Prepend 3320 n times to customers and peers.
6501n : 65006	AS Prepend by Class: Customer & Upstream	Prepend 3320 n times to customers and upstream.
6501n : 65007	AS Prepend by Class: All	Prepend 3320 n times to all AS3320 neighbors.

Community Value	Name	Description
65001 : 100	Standard Local Preference	Set Local Preference to 100 (default).
65001 : 50	Low Priority Local Preference	Set Local Preference to 50.
65001 : 150	High Priority Local Preference	Raise Local Preference value to 150. Requires authorization from AS3320 backbone engineering.

2.4 Restrict Route Propagation

A customer network may use these communities to restrict propagation of its routes to AS3320 peers. However, the well known community NOPEER should be employed instead of these where appropriate.

Community Value	Name	Description
65010 : X	No Export to AS X	Do not advertise route(s) to named AS3320 peer (ASN=X)
65010 : 65001	No Export by Class: Peer	Do not advertise route(s) to AS3320 peers.
65010 : 65002	No Export by Class: Upstream	Do not advertise route(s) to AS3320 upstream.
65010 : 65003	No Export by Class: Peer & Upstream	Do not advertise route(s) to AS3320 peers and upstream.

MEDs

- Lowest MED wins.
  - Opposite of Nearest Exit routing, “carry traffic to me”
  - Only works to the same peer in multiple regions
  - Copy IGP metric to MED
  - Normally subject to negotiation
- Sometimes honoured, often when network traffic is latency or loss sensitive.

MEDs are often filtered

- Many networks set MED to 0 when they learn prefixes, so that hot potato routing will override MED.

```
route-map peers-in permit 10
    set local-preference 200
    set metric 0
```

Origin changing

- IGP
- EGP
- Incomplete

```
route-map PEERS permit 10
 set origin igp

route-route-map TRANSIT permit 10
 set origin incomplete
```

Often peers set to ‘igp’ or ‘egp’ statically on routers to nullify effects of Origin changing.

Inbound – what does work well?

- Overprovisioning
- Peer with top networks widely (buy options!)
- Build relationships
- **Constantly monitor and manage**
- If you care about your traffic, let it go. ☺
  - Playing games with peering hurts your customers' traffic

- Affecting distant ASNs is very hard – a region may only see a single next-hop ASN.

Constantly manage

- Peering on the Internet changes every day.
- Capacity on the Internet grows every day.
- Small networks become large.
- Large networks become larger (consolidation)
- A “bad” path might become good overnight

Questions?

Andy Davidson

andy@2connectintl.com

+44 114 3190605