COMP312-09A Communications and Systems Software

Routing 4 - BGP in Operation

Richard Nelson

richardn@cs.waikato.ac.nz

THE UNIVERSITY OF WAIKATO

BGP in Operation

- BGP Reminder
- BGP Route Damping
- Route Redistribution
- Unicast RPF
- Multihoming
- Policies for a Transit AS
- Peering, Internet Exchanges, Route Servers

THE UNIVERSITY OF WAIKATO

BGP Protocol Summary

- BGP is the only EGP in use in the Internet.
- Static configuration of peer relationships.
- Information carried over TCP sessions.
- Routing governed by policy to reflect business relationships.
- Multiple metrics available in choosing BGP routes.
- “Path vector” routing based in large part on AS paths.
- Over 300 000 routes in a full routing table now, so scaling matters.
- Can carry routing information for multiple address types.

THE UNIVERSITY OF WAIKATO

BGP in Operation

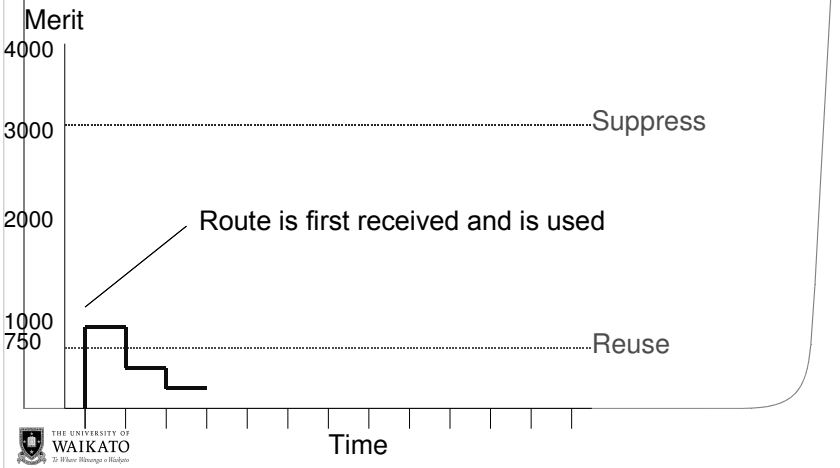
- BGP Reminder
- BGP Route Damping
- Route Redistribution
- Unicast RPF
- Multihoming
- Policies for a Transit AS
- Peering, Internet Exchanges, Route Servers

THE UNIVERSITY OF WAIKATO

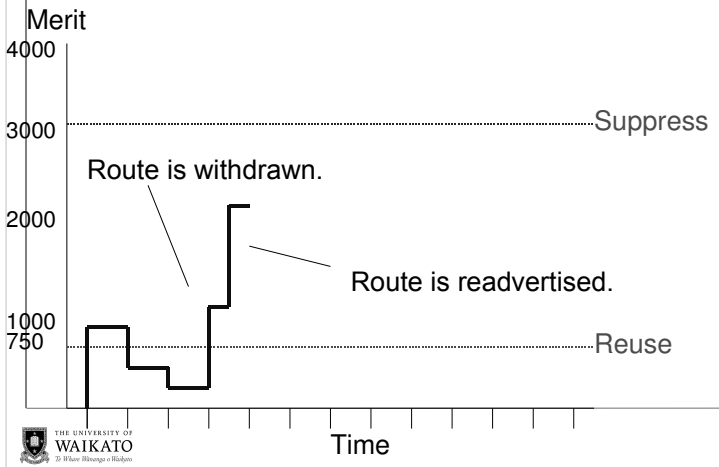
Route Damping

- Pick two numbers, say 3000 and 750.
- Every time a route is newly advertised to you or is withdrawn, add a fixed amount to a number associated with that route (“figure of merit”), say 1000. If that takes the route's number to over 3000, suppress the route.
- Every so often, say fifteen minutes, halve the number associated with that route. If that takes the number below your second limit, go back to considering the route as normal.
- Effect: Unstable routes may not be used (and hence may not be readvertised).

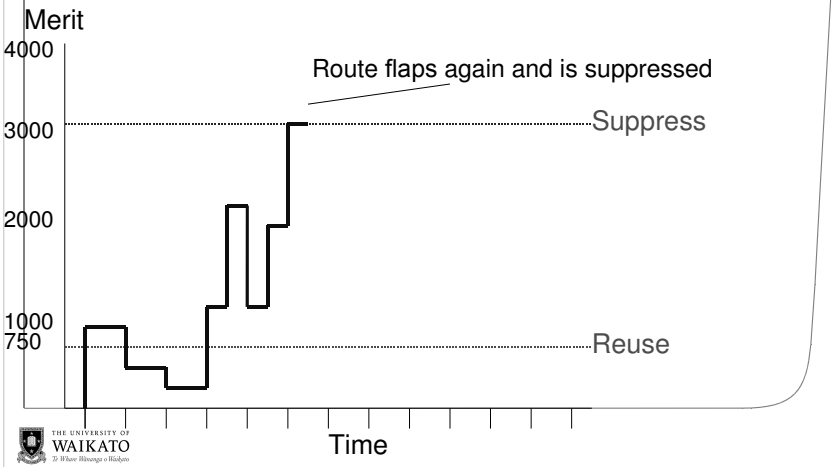
Route Damping



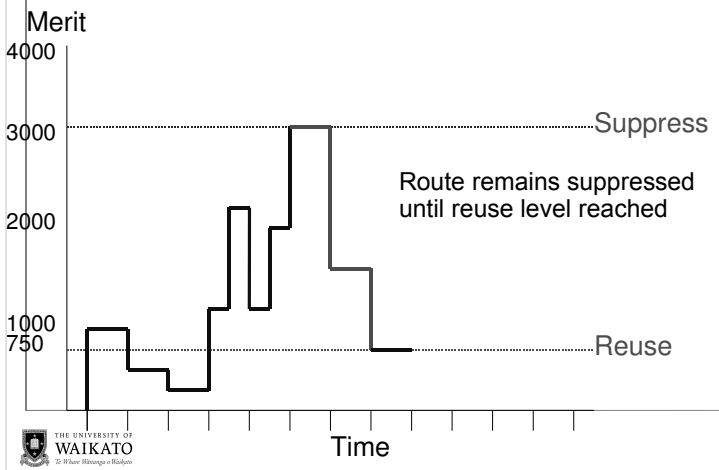
Route Damping



Route Damping



Route Damping



Route Damping

- In practice all parameters used can be configured.
- Can set specific sets of parameters for particular neighbours.
- Not applied to routes learned by iBGP.
- Also applied when route attributes change.
- So do not try to reproduce your IGP's metrics in BGP. (Remember: more stability than an IGP, less information.)

Problems

- The Internet is complex with many paths
- Different BGP implementations hold announcements for different lengths of time before passing them on
- Result is that a single “flap” can appear as multiple flaps elsewhere in the internet.
- The vendor defaults are quite severe (3 or 4 flaps leads to route suppression
- So rebooting a router or restarting a BGP session can result in routes being suppressed.
- Current recommendation is not to use route flap damping, at least not with vendor defaults.

BGP in Operation

- BGP Reminder
- BGP Route Damping
- Route Redistribution
- Unicast RPF
- Multihoming
- Policies for a Transit AS
- Peering, Internet Exchanges, Route Servers

Route Redistribution

- Can redistribute routes from BGP into an IGP. Often done, especially if only a default route is being received by BGP. (You can write a policy to ignore any other route you receive.)
- Can redistribute routes from an IGP into BGP.
  - Done by telco's providing VPN service, but in that case not widely advertised.
  - In general, redistributing IGP routes into BGP is a bad idea.

Route Redistribution

- Can redistribute routes between IGP's, e.g. From RIP into OSPF.
  - In OSPF, router doing this is ASBR.
  - Redistributed routes are seen by OSPF (or IS-IS) as External.
- Why would you want to do this?

BGP in Operation

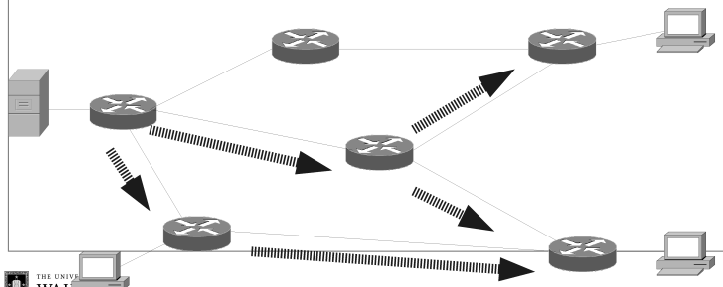
- BGP Reminder
- BGP Route Damping
- Route Redistribution
- Unicast RPF
- Multihoming
- Policies for a Transit AS
- Peering, Internet Exchanges, Route Servers

Unicast RPF

- Nothing directly to do with BGP.
- But worth knowing about before we start talking about inter-domain routing in practice.
- Security feature – protects against IP address spoofing.

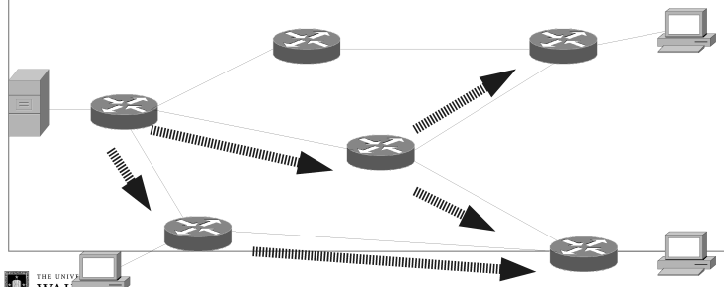
Reverse Path Forwarding

A stream of multicast packets are forwarded through an IP network from a source to the members of a multicast group. A router may receive such packets through more than one interface at a time.



Reverse Path Forwarding

Multicast router checks through which interface it would send packets to the source of the stream, and won't forward the stream through that interface. This prevents multicast loops.



Unicast RPF

Unicast Reverse Path Forwarding (uRPF) applies the same logic to unicast IP packets. The router looks up its routing table for the source address of the incoming packet, and will not forward the packet if it came in through another interface.

- Good idea for routers at the edge of a company's network. Don't allow packets appearing to come from internal addresses (e.g. from private address space) to come from the outside world.
- Works just fine with tunneled remote access.

Unicast RPF

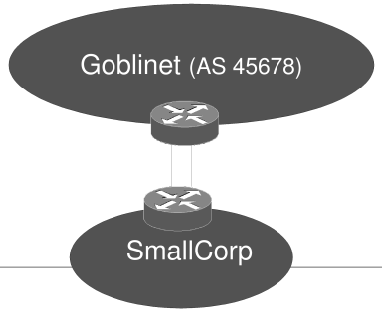
- Care required with more complex inter-AS routing. Not everyone's routing table may match yours just at this moment.

BGP in Operation

- BGP Reminder
- BGP Route Damping
- Route Redistribution
- Unicast RPF
- Multihoming
- Policies for a Transit AS
- Peering, Internet Exchanges, Route Servers

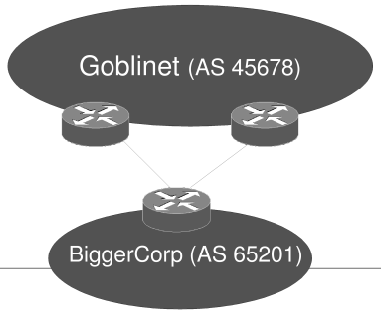
Connecting to Your ISP

Simple case: One customer router connected to one ISP router. IGP or static routing. Customer uses PA address space.



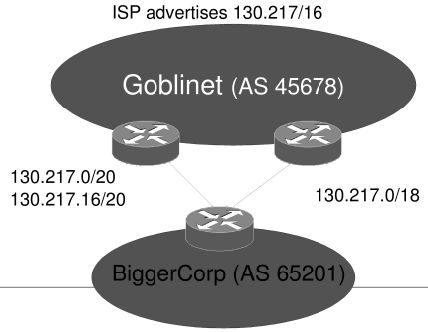
Connecting to Your ISP

Case two: One customer router connected to two ISP sites. BGP. Customer uses PA address space, private AS number.



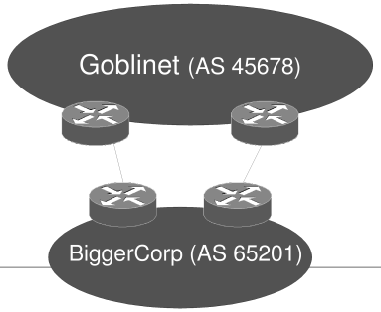
Connecting to Your ISP

Customer can receive default routes from the ISP, but still use BGP to balance load, e.g.



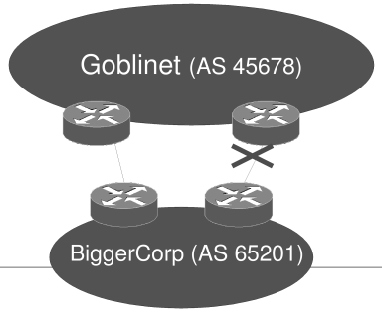
Connecting to Your ISP

Case three: As for case two, but remove customer router as single point of failure. Customer receives default routes from ISP and redistribute those into the IGP. (Why?)



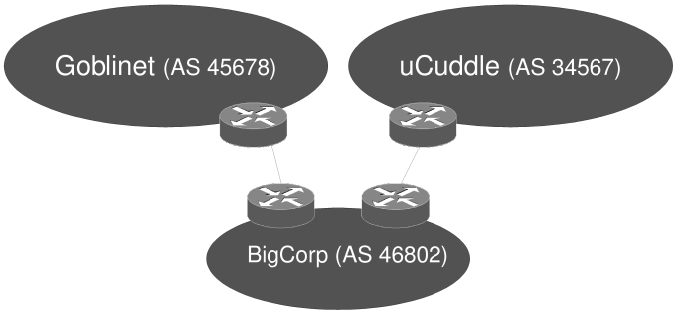
Connecting to Your ISP

- Customer routers have no use for a large route table.
- IGP will reflect any loss of an upstream link.



Connecting to Your ISP

Case four: Multihoming



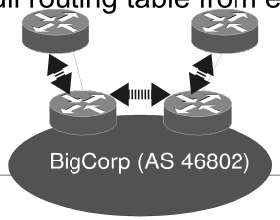
Multihoming

- Customer needs public AS number
- Customer needs provider independent IP address space
- Outside world sees customer's IP address space originating from customer's AS number.
- Key reason why CIDR aggregation doesn't reduce the default-free zone'd routing table as much as hoped.
- Simple routing configuration: Advertise all of customer's address range to both ISP's, default IGP routes point to both ISP links. BUT could then send traffic for one ISP to the other.

Multihoming

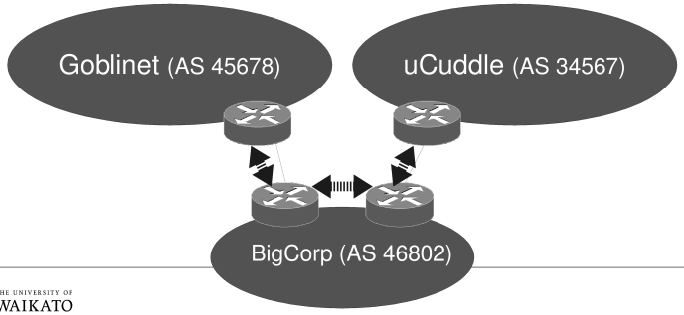
Routing options include:

- Advertise all of customer's address range to both ISP's, default IGP routes point to both ISP links.
- Receive from each ISP routes originating inside its AS (or those of others of its own customers).
- Receive a full routing table from either or both ISP's.



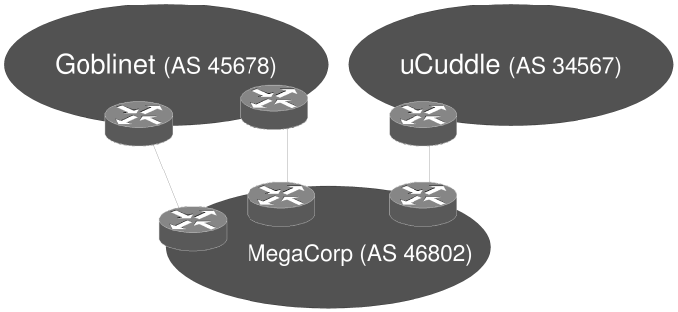
Multihoming

Customer should apply policy (filters) to its BGP sessions so that it advertises only its own address range(s) to each ISP. (What would happen if it advertised other routes?)



Connecting to Your ISP

Case five: Multihoming



Multihoming Summary

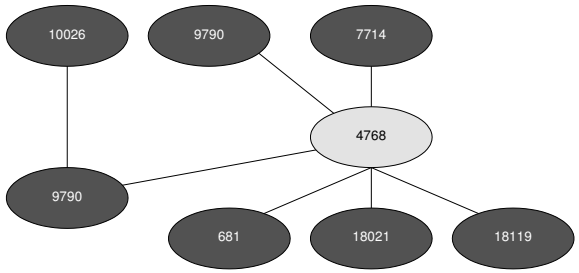
- Customer needs public AS number and provider independent address space. That, plus having to deal with more than one ISP, not worthwhile for many.
- Outside world sees customer's IP address space originating from customer's AS number.
- Key reason why CIDR aggregation doesn't reduce the default-free zone'd routing table as much as hoped.
- Allows choices about which routes the customer will see, but those need ISP cooperation.

BGP in Operation

- BGP Reminder
- BGP Route Damping
- Route Redistribution
- Unicast RPF
- Multihoming
- Policies for a Transit AS
- Peering, Internet Exchanges, Route Servers

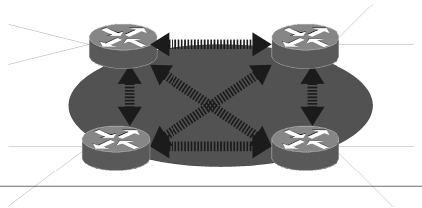
A Transit AS

Advertises routes to one AS learned from another. e.g. An ISP with customers using BGP.



Transit AS's

- IGP carries only internal information
- Advertising a route is an invitation to send traffic.
- BGP can't detect loops inside an AS, so needs a full mesh of BGP routers (or other measures not covered in this course).



Filtering Incoming Routes

- Reject Bogon Prefixes – should never be advertised. Including
  - IP addresses with special uses, e.g. 127/8
  - RFC1918 addresses, e.g. 192.168/16
  - Addresses not yet allocated by a Regional Internet Registry (See http://www.team-cymru.com)
  - Multicast addresses
  - Clog up your routing table, used by Bad People
- Filter by prefix length, e.g. Accept nothing smaller than a /24.

Neighbour-Specific Filtering

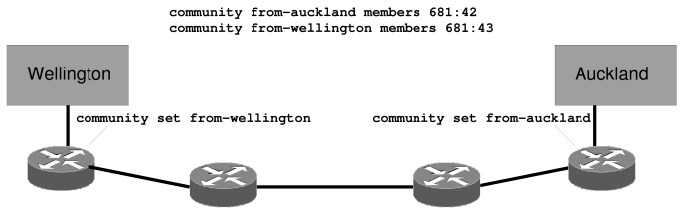
An AS may filter incoming routes in ways based on knowledge of what routes should be being received, e.g.

- An ISP accepting from a customer only advertisement of that customer's address range(s).
- Filtering out default routes from any AS except those from which a default route is expected.

More Complex Policy

One example:

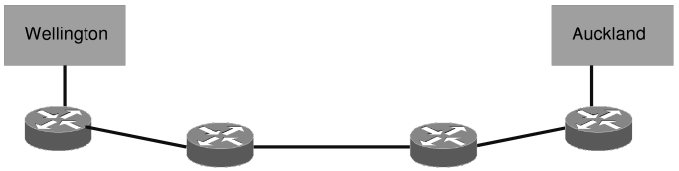
- A carrier with sites in multiple cities applies tags – communities - to each incoming route according to where it was learned..



More Complex Policy

One example:

- A carrier with sites in multiple cities applies tags – communities - to each incoming route according to where it was learned.
- Local Preference is applied to routes by routers in different cities according to whether they were learned in this city or elsewhere.




```
from {
  community from-wellington
}
then {
  local-preference 120
}
```


More Complex Policy

Why?

- "Hot Potato Routing" – Minimise cost by sending packets out of this AS as soon as possible. e.g. Routers in Auckland will prefer routes learned in Auckland, because traffic sent to Wellington that way uses someone else's inter-city links.
- "Cold Potato Routing" – Maximise control by sending packets between cities within this AS. e.g. Routers in Auckland will prefer routes learned in Wellington over those learned in Auckland.

Influencing Route Choice

Routes advertised by an AS may be made more or less preferable by

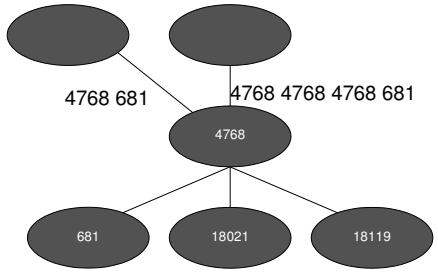
- Setting MED values for use by a neigbouring AS
- AS path pre-pending – adding your own AS number more than once to the AS Path of the route. e.g. A route with AS Path

4768 4768 4768 681

will be less preferred than one with AS Path

4768 681

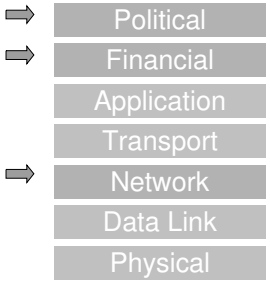
Influencing Route Choice



Transit AS Summary

- IGP carries only internal information
- Advertising a route is an invitation to send traffic.
- Routes are normally filtered on receipt.
- Routes can be altered when learned from another router inside the AS.
- Routes are often modified by policy on being advertised out of the AS.
- Policies give effect to commercial decisions.

The Other Reference Model

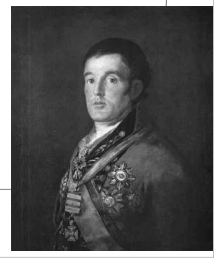


BGP in Operation

- BGP Reminder
- BGP Route Damping
- Route Redistribution
- Unicast RPF
- Multihoming
- Policies for a Transit AS
- Peering, Internet Exchanges, Route Servers

Peering and Transit

- Transit – traffic is carried for a fee ($$$)
- Peering – a reciprocal arrangement in which two carriers exchange traffic, each getting access to the other's customers. (No $$$)
  - Not to be confused with a BGP peer
  - Or any other kind

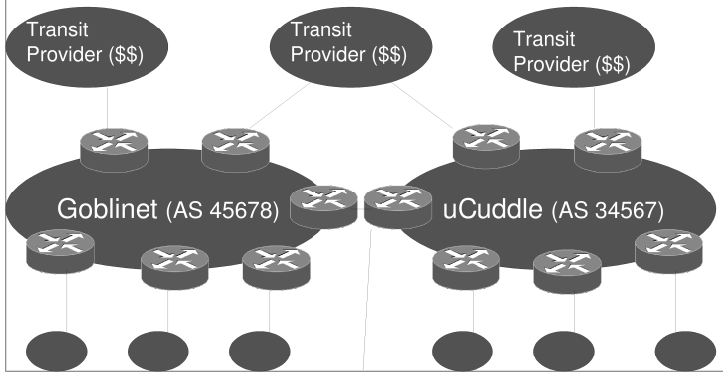


Peering Example

Two ISP's, Goblinet and uCuddle, agree to a peering arrangement in which each will carry traffic between their customers and the other's customers without charge.

Requires a connection between their networks. Could use leased telco fibre.

Peering Example



What routes are advertised here?

Peering

- These are transit AS's, which will have routing policies.
- Will favour sending traffic to peers (free) rather than to transit providers ($$$).

Peering

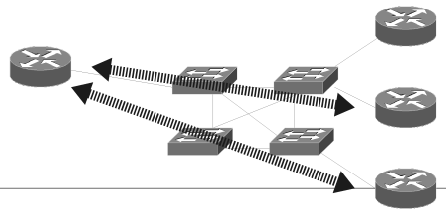
- Cuts transit costs for ISP's.
- Also done by content providers (e.g. TradeMe)

BUT

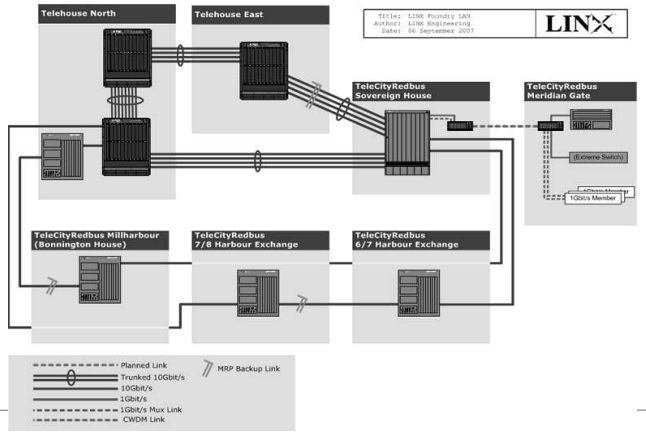
- Every peering arrangement costs something, especially if new links need to be built
- An ISP won't peer if it thinks the alternative is for it to be paid transit fees.
- Telecom and TelstraClear do not peer inside New Zealand.

A Peering Exchange

To make peering with multiple organisations easier and cheaper, all can connect to an exchange point. Could be a switching structure inside an exchange building. Participants install their own routers inside the building and make their own peering arrangements as they wish. (e.g. Auckland Peering Exchange)



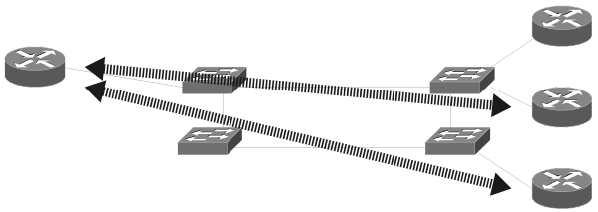
A Bigger Exchange



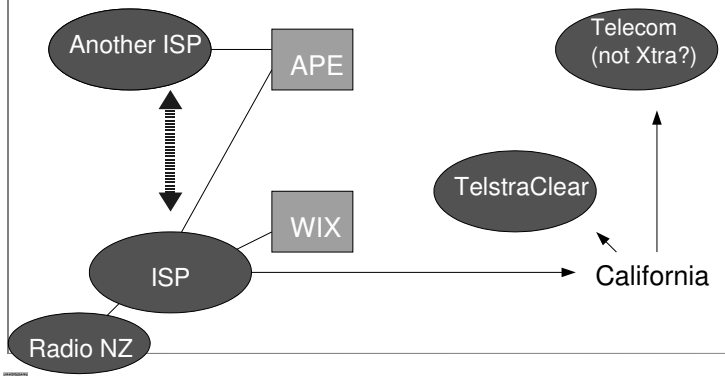
Graphic from http://www.linx.net/pubtools/topology.html

A Peering Exchange

Or the same switching can be distributed over a metropolitan area network (e.g. Wellington Internet Exchange - WIX)



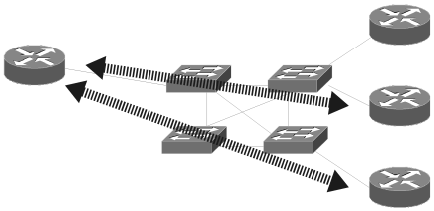
Radio New Zealand



http://computerworld.co.nz/news.nsf/NL/54FCC3B9D0B8BB96CC25702F00740446

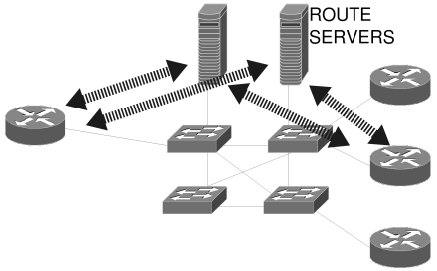
A Peering Exchange

Internet Exchanges make the cost of adding a further peering arrangement much lower, but you still need to negotiate this one a one-to-one basis. Unless ...



Route Registry

A route registry may operate at an exchange one or more route servers. Participating routers form BGP peer relationships with the route servers instead of each other.



Route Servers

- Simplify peering with multiple other organisations.
- Exchange routes with multiple organisations' routers.
- Do not carry routed traffic.
- Implement filters as specified by each organisation.
- Organisations specify their routing policy in files supplied to the registry using Routing Policy Specification Language (RPSL – RFC2622)
- Code exists to produce router configuration statements from the RPSL.

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

RPSL

```
route-set:  AS9560:RS-ROUTES:AS23754
descr:      advertised to AS9560 by CityLink - AS23754
members:    174.128.16.0/20^20-29,
            198.48.0.0/22^22-29,
            202.8.44.0/22^22-29,
            202.7.6.0/23^23-29,
            198.32.71.0/24^24-29,
            202.7.5.0/24^24-29
admin-c:    RPA1-NZRR
tech-c:     RPA1-NZRR
notify:     rpsl-admin@nzix.net
notify:     nznog@list.waikato.ac.nz
notify:     asjl@citylink.co.nz
mnt-by:     MAINT-NZRR-NZ
changed:    rpsl-admin@nzix.net 20080316
source:     NZRR
```

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Route Servers

- Simplify peering with multiple other organisations.
- Exchange routes with multiple organisations' routers.
- Do not carry routed traffic.
- Implement filters as specified by each organisation.
- Organisations specify their routing policy in files supplied to the registry using Routing Policy Specification Language (RPSL – RFC2622)
- A route registry may supply a web interface - “Looking Glass” - allowing the public to view its route servers' routing tables.

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Reading

New Zealand Internet Exchanges http://www.nzix.net/

APE Looking Glass e.g.

http://nzix.net/cgi-bin/lg.cgi?router=rs1.ape.net.nz

Further Reading

Abley, *A Short History of Peering in New Zealand*
http://www.r2.co.nz/20050203/joea-peeringnz.asx (Before 13:40 is on the DNS root servers.) Slides at
http://2005.nznog.org/slides/jabley.pdf

Norton, *Internet Service Providers and Peering*
http://www.equinix.com/pdf/whitepapers/PeeringWP.2.pdf

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato