Routing with Open Source Technology

Andy Davidson
NetSumo Ltd / LONAP Ltd
WYLUG - 6th April 2009

Agenda / Aims

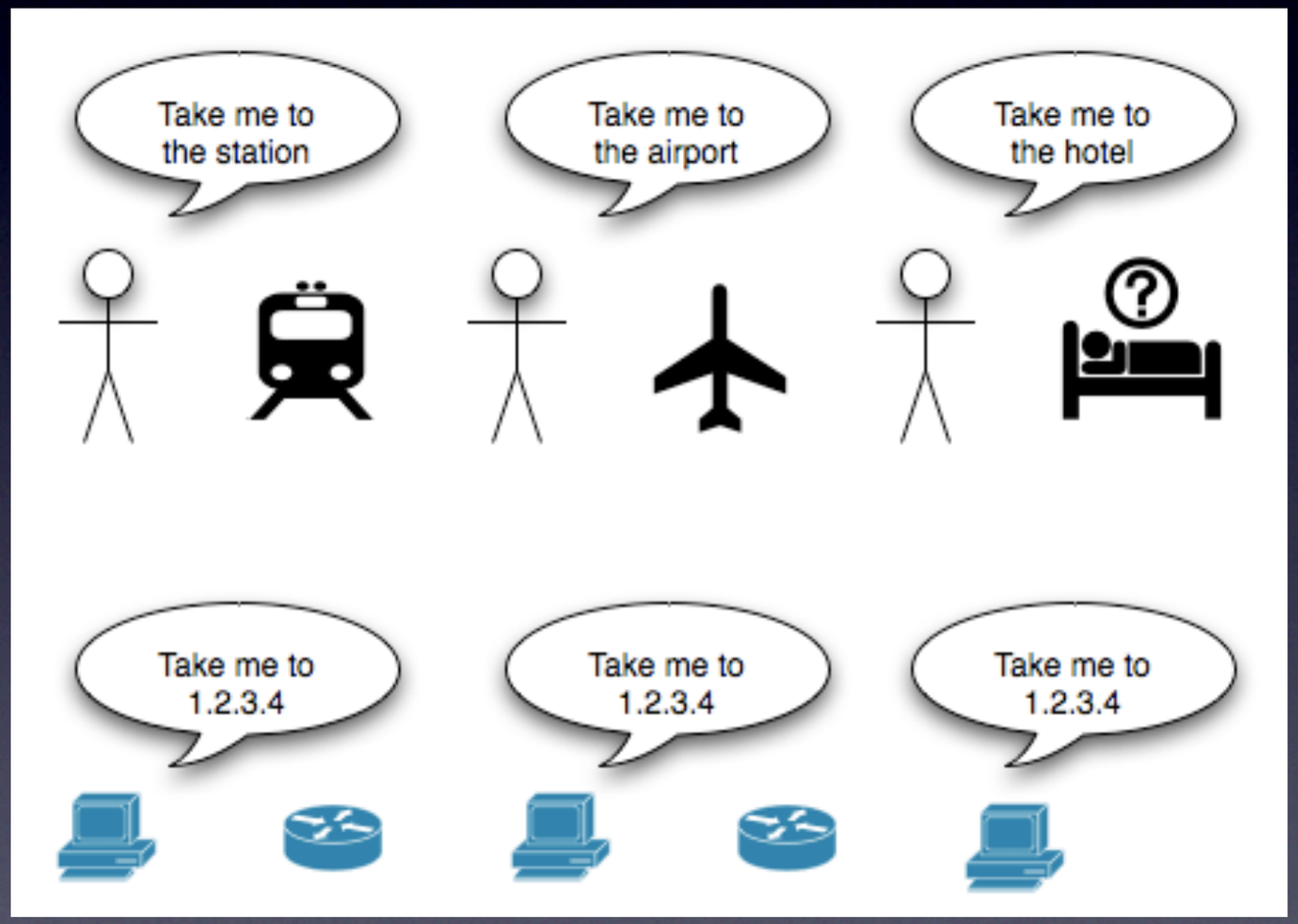
- To explain the concept and uses of routing
- To introduce the open source routing tools and show how they work
- To demonstrate how they are used in industry

An example of routing

- Walk to bus stop
- Bus to train station
- Train to Leeds
- Taxi to Old Broadcasting House

- Needs knowledge of next ‘hop’ and ultimate destination
- Can’t route around problems in advance
- No knowledge of ‘better’ routes

IP Routing is different



- Knows only destination
- By design:
- Path can route around problems
- Path takes advantage of new improvements

Routing Table

```
andy@kula:~$
andy@kula:~$ /sbin/route
Kernel IP routing table
Destination     Gateway         Genmask         Flags Metric Ref    Use Iface
193.239.32.0    *               255.255.255.0   U     0      0        0 eth0
default         ge0-1-0-rtr0.so 0.0.0.0         UG    0      0        0 eth0
andy@kula:~$ ip route
193.239.32.0/24 dev eth0  proto kernel  scope link  src 193.239.32.181
default via 193.239.32.1 dev eth0
andy@kula:~$ ip -f inet6 route
2a02:c30:1:1::/64 dev eth0  metric 256  expires 2592116sec mtu 1500 advmss 1440 hoplimit 4294967295
fe80::/64 dev eth0  metric 256  mtu 1500 advmss 1440 hoplimit 4294967295
default via 2a02:c30:1:1::1 dev eth0  metric 1  mtu 1500 advmss 1440 hoplimit 4294967295
```

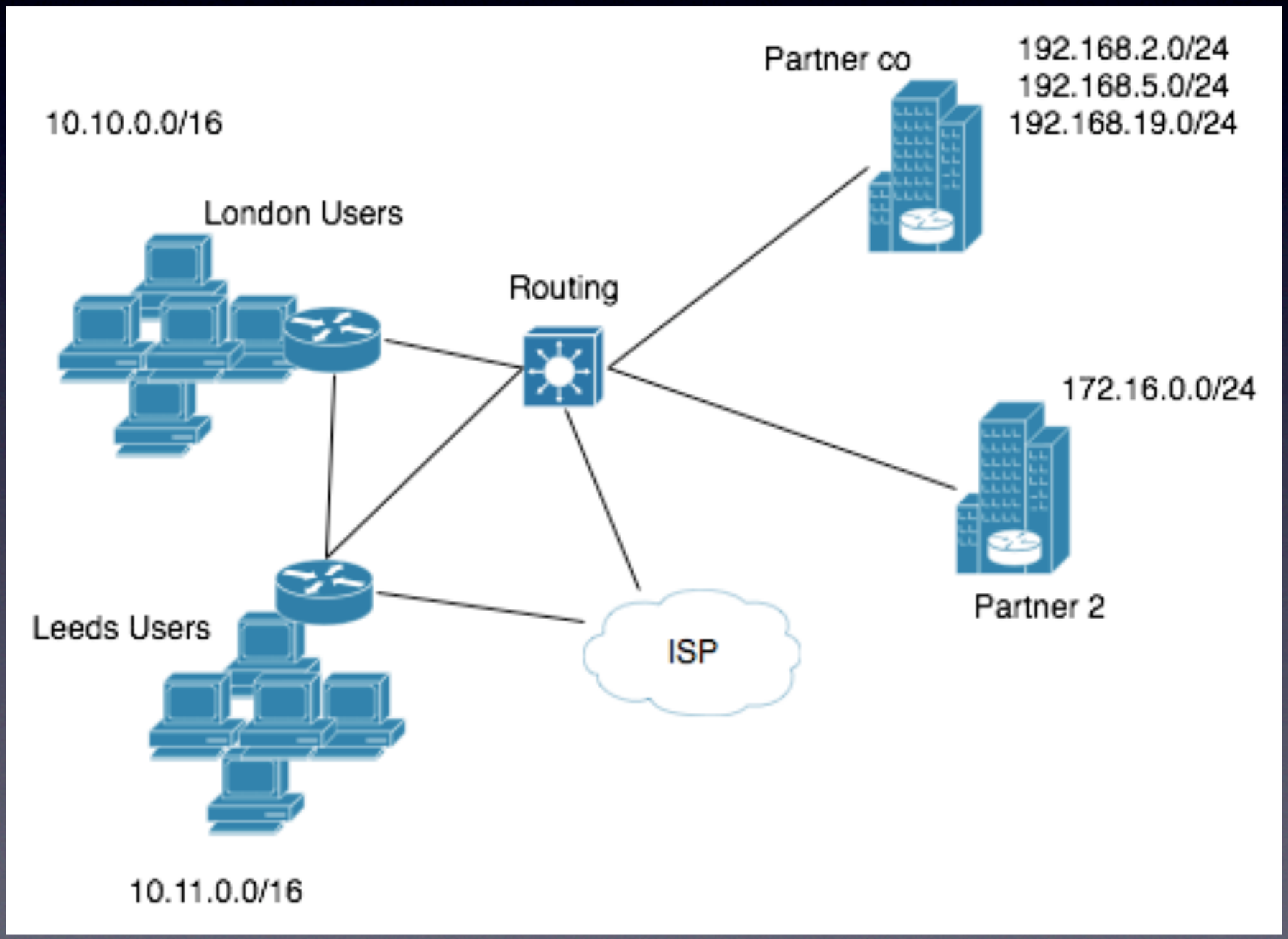
Typical single homed host running Linux

Routing Table 2

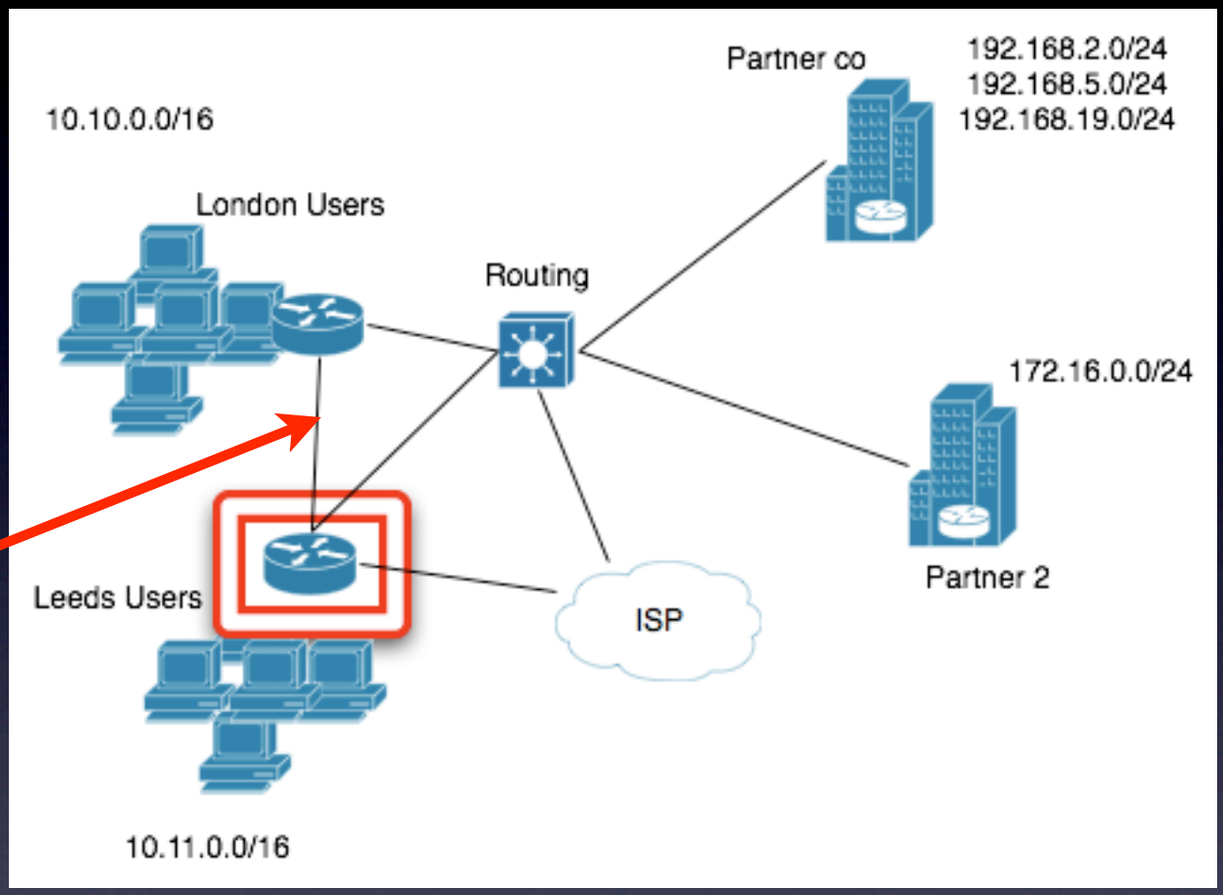
```
Kernel IP routing table
Destination     Gateway         Genmask         Flags Metric Ref    Use Iface
77.75.109.200   0.0.0.0         255.255.255.252 U     0      0        0 eth0.314
62.197.50.204   0.0.0.0         255.255.255.252 U     0      0        0 eth1
62.197.50.160   0.0.0.0         255.255.255.248 U     0      0        0 eth2
80.82.143.240   62.197.50.162   255.255.255.248 UG    0      0        0 eth2
195.177.252.160 0.0.0.0         255.255.255.240 U     0      0        0 eth0
195.177.253.176 0.0.0.0         255.255.255.240 U     0      0        0 eth0.22
195.177.252.128 0.0.0.0         255.255.255.240 U     0      0        0 eth0.6
80.82.143.224   62.197.50.162   255.255.255.240 UG    0      0        0 eth2
194.165.60.64   0.0.0.0         255.255.255.240 U     0      0        0 eth3
195.35.104.0    62.197.50.162   255.255.255.224 UG    0      0        0 eth2
208.111.149.0   62.197.50.162   255.255.255.224 UG    0      0        0 eth2
91.194.68.0     0.0.0.0         255.255.255.192 U     0      0        0 eth0.11
212.15.70.0     62.197.50.162   255.255.255.192 UG    0      0        0 eth2
194.153.156.192 62.197.50.162   255.255.255.192 UG    0      0        0 eth2
125.245.172.0   77.75.109.201   255.255.255.128 UG    0      0        0 eth0.314
206.165.148.0   77.75.109.201   255.255.255.128 UG    0      0        0 eth0.314
218.38.0.128    77.75.109.201   255.255.255.128 UG    0      0        0 eth0.314
210.99.173.0    77.75.109.201   255.255.255.128 UG    0      0        0 eth0.314
210.99.173.128  77.75.109.201   255.255.255.128 UG    0      0        0 eth0.314
```

Typical multi-homed host running Linux

Simple enterprise routing requirement



Example routing tables



Link networks / "Point to Point" links

Default route

Destination	Gateway	Genmask	Iface
10.11.0.0	0.0.0.0	255.255.0.0	eth0
10.10.0.0	1.2.3.1	255.255.0.0	eth1
192.168.2.0	1.2.3.5	255.255.255.0	eth2
192.168.5.0	1.2.3.5	255.255.255.0	eth2
192.156.19.0	1.2.3.5	255.255.255.0	eth2
172.16.0.0	1.2.3.5	255.255.255.0	eth2
0.0.0.0	91.1.2.3	0.0.0.0	eth3
1.2.3.0	0.0.0.0	255.255.255.252	eth1
1.2.3.4	0.0.0.0	255.255.255.252	eth2

Could fail over default route to eth2 if Leeds isp outage….

Static Routing

- Allows you to build a "map" of an IP network.
- Slow to respond to failure or improvement of IP networks.
- Hard to scale across your own network - complexity increases exponentially as you add routers.

```
root@gw:~ # route add -net 192.168.0.0 netmask 255.255.255.0 gw 10.1.1.1
root@gw:~ # route add -net 172.16.0.0  netmask 255.255.255.0 gw 1.2.3.1
.
.
```

Dynamic Routing

- Better
- One router announces network prefix information to another using software known as “routing protocols”
- Auto-learns when there are routine failures or improvements on your network
- Scales well when properly implemented!

Categories of RP

- Link State - ‘floods’ a network with messages to construct a ‘map’ of a network relative to the router.
- Distance Vector - Receives information about the routing tables of neighbouring nodes to construct “my” routing table.

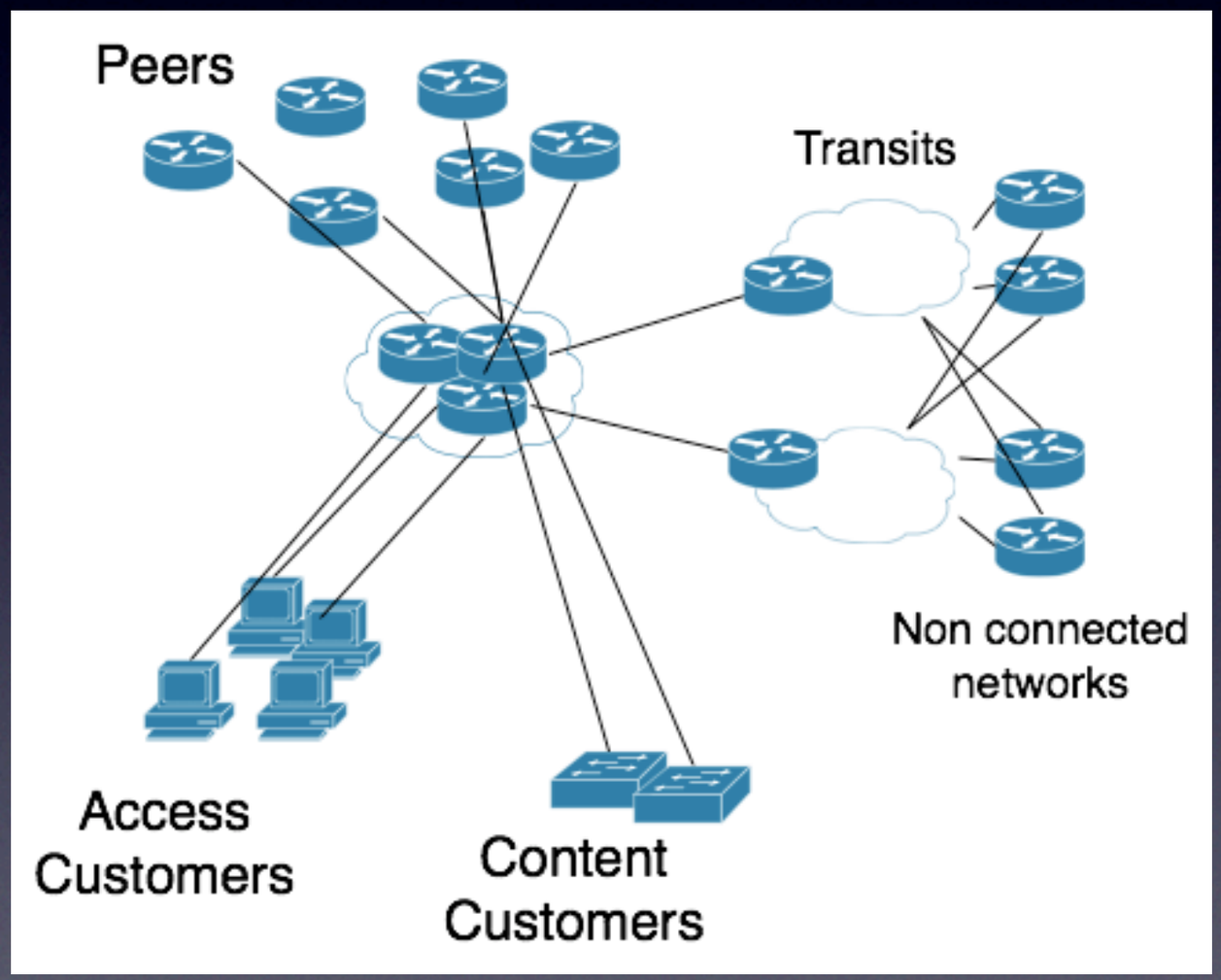
Categories of RP 2

- Internal - uses information about networks WITHIN a single network/administrative domain
- External - uses information about networks outside of your administrative domain
- Typically a network will use both.

So ....

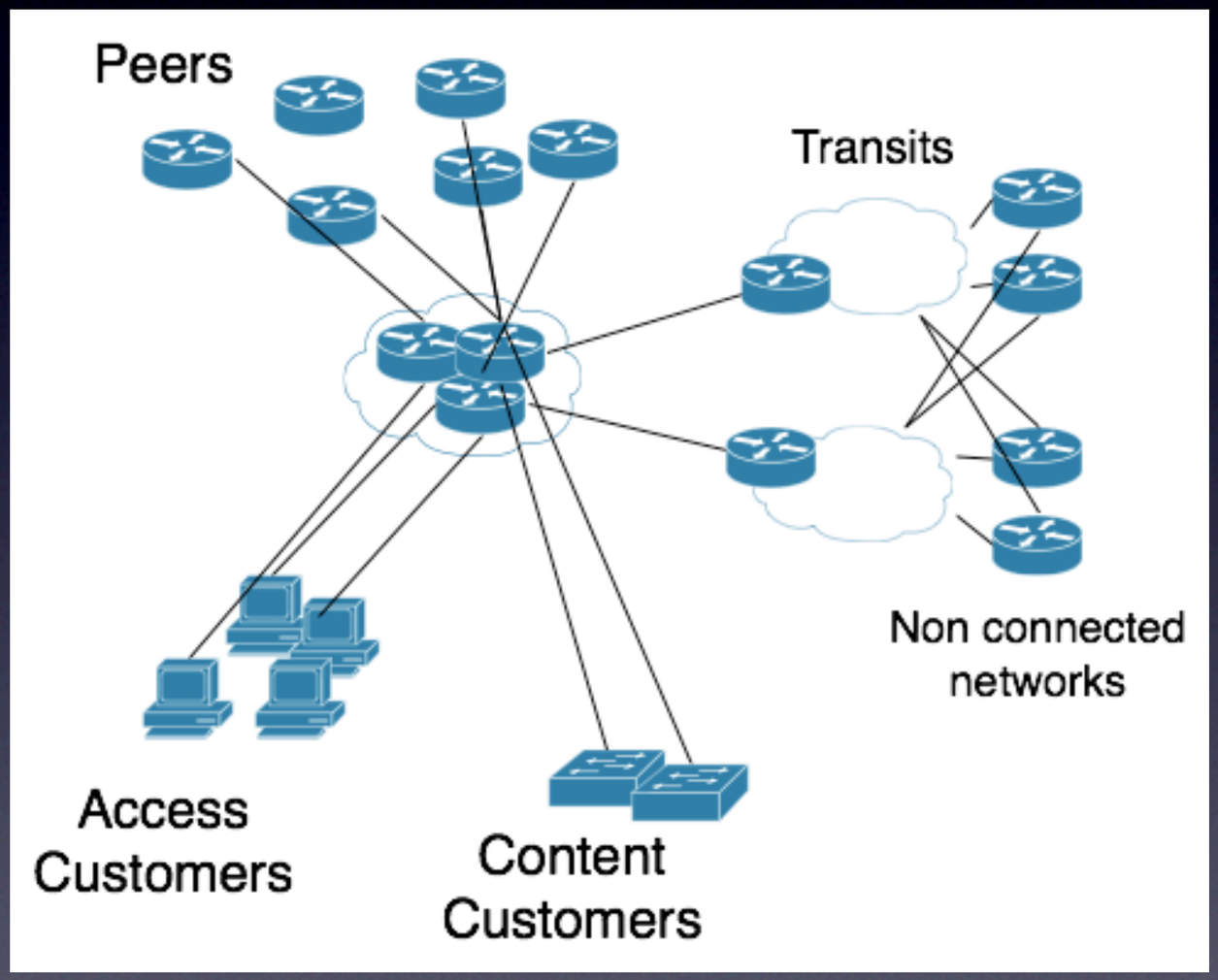
- Traffic routing decisions are made using a routing table
- Routing protocols are an efficient way to add/modify prefixes in the routing table

A more complex example



- Example ISP shown
- Many links - upstream/ customer/peer
- Requirement to establish “best” path
- Internal/External routing
- Huge number of exits/ interfaces

Routing in this scenario



- Internal
  - Customer routes
  - Link-nets
- External
  - Peers and upstreams
- “Best” path calculation different for both types

A general introduction OSPF

- used as an IGP
- carries any "**directly connected**" interfaces, so customers and linkets to peers/transits
- FAST convergence, makes for great fault tolerance on resilient network

A general introduction BGP

- Carries routing table data between different networks
- Peer-to-peer, policy driven
- Used as golden standard for route exchange between service provider networks
- ~280,000 prefixes in default free zone, April 09.

“Best” path ?

- Normally shortest path, but ….
- Lowest latency
- Cost
- Politics

Open Source Routing

- Quagga
- OpenBGPd
- BIRD
- Xorp, Vyatta

Quagga

- Fork of GNU/Zebra routing daemon
- Commonly deployed
- Cisco like interface - good for training
- Implements OSPF and BGP (and other protocols)

Demo

Link nets: Connect this Linux router to other routers

Live demo: Demonstration of BGP sessions in action

```
eth2      Link encap:Ethernet  HWaddr 08:00:20:D1:48:06
          inet addr:62.197.50.163  Bcast:62.197.50.167  Mask:255.255.255.248
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:133956977 errors:0 dropped:0 overruns:0 frame:650
          TX packets:981361819 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:2662481867 (2.4 GiB)  TX bytes:2935960398 (2.7 GiB)
          Interrupt:10 Base address:0xbc00

eth3      Link encap:Ethernet  HWaddr 08:00:20:62:6E:0B
          inet addr:194.165.60.75  Bcast:194.165.60.79  Mask:255.255.255.240
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:0 errors:0 dropped:0 overruns:0 frame:0
          TX packets:0 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:0 (0.0 b)  TX bytes:0 (0.0 b)
          Interrupt:11 Base address:0xb400

eth0.6    Link encap:Ethernet  HWaddr 00:14:22:72:48:94
          inet addr:195.177.252.142  Bcast:195.177.252.143  Mask:255.255.255.240
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:0 errors:0 dropped:0 overruns:0 frame:0
          TX packets:44742 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:0 (0.0 b)  TX bytes:1879164 (1.7 MiB)

eth0.7    Link encap:Ethernet  HWaddr 00:14:22:72:48:94
          inet addr:195.177.252.126  Bcast:195.177.252.127  Mask:255.255.255.128
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:98257077 errors:0 dropped:0 overruns:0 frame:0
          TX packets:88981110 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:2437798330 (2.2 GiB)  TX bytes:738898220 (704.6 MiB)
```

Uses in industry

- Open Source routing is not widely used within ISP community outside :
  - Training / Simulation
  - Modeling
  - Route-servers
  - Anycast deployments

Limitations

- ‘PC’ Hardware is not a capable routing platform
  - Absence of carrier grade interfaces/backplanes
    - (Leads to reduced speed of access/high latency)
    - Poor port density
    - Poor availability of hardware switching
    - Access to support

That said …..

Connected to LINX

Tiny power footprint

Any Questions ?

Any Answers ?

andy.davidson@netsumo.com