# Inter-provider Cooperation

NANOG / SF

Monday February 10, 1997

http://psg.com/~randy/970210.nanog/

# Communities

- Can be used to give clues to peers and beyond
- Sometimes the clue transits intermediate ASs
- How to let some through and block others?
- cisco experimental code is being played with
- Clean myAS: from in and out, let others through
- Filter aggressively at the customer edge
- Bay does not yet implement communities, but seems to survive, gated resets session
- Could be making the world more complex for a limited win

# Limited use of MEDs at IXs

- When I have multiple routers on the mesh, it would allow me to influence your next hop if you listened but did not propagate
- E.g. at Pensauken, I have one router toward Europe and one to DC, I want you to use the right one
- So accept MEDs from peers at IX, but don't transit them within your AS
- BGP multi-path is an alternative, but is still gets 50% correct
- Some folk are playing with this

# Peering policy enforcement

- We need tools to detect someone making anti-policy announcements, e.g. inconsistency, ...
- We need tools to stop/detect someone pointing default
- One can use inconsistent route announcements to ‘encourage’ correct behavior
- Folk have some detection tools they might share
- Maybe router vendors would give us some tools
- Later, automatic policy change might be nice

# Port/MAC filtering at IXs

- Some provide, some are experimenting, some refuse
- Should always be bi-lateral, and never half-duplex
- When installed without warning, one gets surprises
- When debugging, one needs to know reachability
- IXs should have exchange off by default, require agreement of both parties to enable, and either may disable. This is the reverse of current policy
- If state changes, both parties should be notified by IX operator
- Are shared media interconnects a mistake?

# Local exchanges

- There will likely be more, though most small
- The goal is to make local-local traffic more efficient
- Intra-US data show that if 15% of traffic is local then that is quite high
- Japan finds 25-33% of traffic is local
- Sweden 10% local, 15% European, 65% outside
- Deutschland, external traffic much larger than local (E1 intra-Europe 2xDS3 to US)

# Local exchanges (contd)

- Local-only announcement goes against current homogenous announcement convention
- Local-only announcement is anti-aggregation
- Configuration is difficult for an NSP, and does not scale well with current tools
- So for the NSPs, the 10% benefit is not worth the peering risk/scaling problem
- Possibly not as true for local providers or some countries, or if technology changes

# Announcing the IX mesh prefix

- As a courtesy by you so I can get to my router when my external connection is down
- Don't, if the IX op has something interesting, then they should put it behind a router
- So **do not** announce exchange points to peers
- Do not accept announcements of IX prefixes from peers
- We could develop a well-known IX filter list

IX operators should set the
in-addr as their customers request

Please remember who is the
**Customer**

# Peering policies

- Few NSPs' policies are published these days
- One can publish conventions for peering without publishing criteria for deciding whether to peer
- Some folk require NDAs while discussing, especially for associated information
- There are providers selling transit across exchange points. Unless it is switched, it is not cool.
- Can provider P0 prevent a customer C from multi-homing and announcing space gotten from P0 to a new provider P1? Not unless it was in a contract.

# Route filtering policies

- Bozo filters (default, Net-10), are well known but could be better published
- There seems to be consensus to filter at /24
- More folk are filtering </24 than you might think, and the trend is more filtering
- We should publish our filters, so techniques are spread and debugging is aided

# Using NAT for multi-homing

- Allows multi-homing using space from each provider
- Some problem with stateful sessions when one line goes
- But they can be restarted immediately
- Problems with protocols which embed IP addresses, e.g. CUSeeMe

# Alternatives to using BGP with multi-homed customers

- AS space: do we really need more than 64k objects at the top of the internet?
- Is there something between BGP and an IGP to soften the edges, i.e.
- Related routing domains which do not share an IGP but appear as one AS
- Is it bad to have an inconsistent origin AS for a route? Yes. Tools will break, ownership issues, ...
- We need protocol requirements and then design proposals

# IDRP vs. BGP4++

- BGP4++ has complexity at v4/v6 borders
- Neither seems to address border softening
- The timeframe to implement and deploy could be too looooong

# Inter-provider Multicast

- Social problem because the bandwidth impact is not at the source of the flow, but at the forks and ends
- To beat streaming unicast, which is worse, we need inter-provider gateways
- Need a reliable border protocol where the multicast support is as good as the unicast
- Not enough thought has gone into it, and it is a tough problem, and routers need more horsepower
- We have problems enough keeping unicast working
- OK within one's border, inter-provider is serious pain

# Inter-provider Multicast (contd)

- Try just a few multicast exchanges (e.g. Sprint NAP) and see if it is a good thing
- MIX must be a dedicated mesh because switches are multicast ignorant
- We will need special contacts at the providers for multicast? Suggested: multicast@pro.vi.der
- If an NSP has a customer wanting to multicast, how do they get an address? IANA? SDR? ARIN<joke>?
- Try beta inter-provider multicast backbone and use one or two beta content providers to test this facility

# NOC to NOC communication: Problems and Tactics

- How to contact your routing folk should there be a serious inter-provider problem?
- When you call a peer NOC, make clear that you are a routing peer
- Peers might have unique identifiers in your customer database?
- Router geeks should exchange pager numbers
- How to get consensus on peering/routability issues?
- Give your NOC the list of your friends' names

# How to gather statistics?

- Data collection is needed for our own short and long term benefits
- How do we support /relate to kc, Vern, …?
- Need probes at NAPs to see inter-provider routing
- Need probes within network too
- Give kc/Vern machines at the IXs
- Give kc/Vern a machine within your network
- Serious concerns about data privacy and integrity
- Flap chasers may need your help to chase it through your territory, be kind

Please put ***ip classless*** in your CPE routers

# NetFlow Statistics

- One NSP says 70% of flows and 80% of traffic is web, and it shows negligible locality
- For operational hints on the how and what of gathering NetFlow stats, CICnet and others are publishing the tricks of the trade
- Have a lot of hard disk

# Interesting Parameters - Dampening

- BGP damping thresholds, suggested by European, Sprint/ICM, and others
  - /24 45 125 2000 255
  - /19-23 30 750 2000 45
  - /rest 15 750 2000 30 (cisco default?)
- Vendors might make these the default
- Will it need a flag day?

# Interesting Parameters - SPD thresholds

- Being used because the tradeoff is increased latency for **greatly** reduced packet loss
- Most ISPs seem to be using default
- Tune max to match largest hold queue length, and set min a bit below that
- Still a lot of trial and error

# Interesting Parameters - MBONE bandwidth

- Nobody seems willing to say they do any backbone bandwidth limitation
- Very little MBONE is being done, period

# Cooperation to block net abuse

- Please source filter when you configure CPE
- Not enough CPU to do it on aggregation or backbone routers
- Filter abusers at the victim's CPE
- Except for child porn and copyright, you are liable if you shut customers off sans process
- It may be time to use MD5 auth for BGP peers

# Tracing SYN attacks through multiple backbones

- These problems seem to need engineering, as they are above NOC capabilities
- Hard to trace across shared-media in presence of source address shifting
- Does your NOC know who to invoke when they get an inter-provider SYN attack call?
- cisco has TCP intercept code in 11.2 which does early drop