CABLE & WIRELESS

IPv6 - Evolutional Issues & challenges

Presentation by
Udo Steinegger

Public IP Engineering, Munich

May, 2005

Agenda

Prelude

How C&W have deployed IPv6 in AS1273

How vendors successfully prevent a global roll out of IPv6

Break on through to the other side (…of the AS border)

Conclusion

Section one

Prelude

Messages Of The Day

- there is a global movement to IPv6
- IPv6 is inevitable for continuing market growth with new applications (mobile markets, etc)
- Our customers are demanding IPv6 services today.
- Require vendors to provide full support of IPv6 in their HW / SW today (yes, we are in 2005 and not in 1999) .
- No excuses anymore for intercontinental IPv6 tunnels, as there are already enough ISPs that offer native IPv6 connectivity between the continents.
- Swaps of full IPv6 routing tables are **not** needed any longer.
- Strong focus on technical sanity of upstream and peering relationships necessary.

Section two

IPv6 deployment in C&W's AS1273

Addressing

- C&W uses 2001:5000::/21 in the core network and for static customers, mobile (phone) services, etc
- Internal addressing plan for IP access uses a hierarchy of /32 per country and /40 per site (allows for IBGP aggregation and confederation designs if required).
- IP access customers do get /48 by default per site unless very good reason to do otherwise.
- Hosting customers do get /48 assigned, but only /64 configured on their (VLAN-) interface. In case of Firewalls or other routing devices used in the customer‘s setup, the full /48 gets routed towards the customer.
- All links, regardless of interface type, do get /64 assigned to simplify operations, IP admin, DNS and management.
- No stateless autoconfig anywhere.

IS-IS

- **Multitopology IS-IS** is being used to seperate IPv4 and IPv6 topologies:
  - to be able to „route around“ devices that are not IPv6 aware (important for the roll-out phase, to not blackhole IPv6 traffic).
  - to have the opportunity for **different link costs** for IPv4 and IPv6 (differing topologies).
- IS-IS carries only loopbacks and backbone links, while the rest is in IBGP.

BGP

- Same local preference structure as in IPv4 is used for IPv6 (customers > peers > transit)
- Support for BGP communities to enable customers and peers to steer their traffic (blackholing currently unsupported but available soon™)
- aggregates, static routes and customer links are carried in IBGP.
- Seperate IBGP mesh for IPv6 AFI, no 6PE

6PE or not?

- C&W have decided not to use 6PE, as 6PE is considered as a ugly hack to work around certain vendor gear lacking stable/complete IPv6 implementation.

- 6PE might become a necessity at a later point in time, because unfortunately economics may require C&W to deploy core hardware which isn‘t (yet) IPv6 capable (It is not justifyable to shell out 20-50 times the money for 10G equipment only for the added benefit of IPv6 support).

- Downside: no MPLS-TE/VPN for IPv6 as there is no LDP and RSVP for IPv6 available from Juniper and Cisco (and nobody else).

Hardware

- IPv6 has been rolled out onto:
  - All core devices, as the core is up to now Juniper gear only (M-Series).
  - Juniper aggregation devices (M-Series).
- IPv6 has not (yet) been rolled onto:
  - Juniper E-Series aggregation routers.
  - Cisco aggregation routers.

Section three

How vendors successfully prevent a working global IPv6 roll-out

Juniper E-Series (ERX)

- No IS-IS support for IPv6 at all (not even speaking of multitopology IS-IS).
- Juniper charges a premium amount per IPv6 license. (IPv6 is a **basic requirement** nowadays and not just some sort of optional exotic feature)

☛ **successfully prevents deployment of IPv6 in large scale residential DSL market**

Cisco

There is no IOS code, that we have tested, for the platforms 7500 and 12000, without severe issues.

- 12.0S has IPv6 support only for GSR
  12.2S only for 7500
  ☛ no common code base possible for 7500 and GSR
- 12.2S with all required features (12.2(25)S*) has new CEF code wich shows serious problems („`show cef table consistency-check`“ is your friend)
- Bad to no performance on the deployed GSR E0/E1 linecards.
- Infamous BGP „ghost route bug“ (IOS forgetting to send BGP withdrawl; still unfixed for years)

Others (random examples)

- Alcatel
  - Does not have support for IPv6 in the current product portfolio.
  - IPv6 support planned for Q4/2006.
  - No support for multitopology IS-IS yet.
- Tellabs
  - Does not have support for IPv6 in their current product portfolio
  - IPv6 support planned for Q1/2006.
  - No support for multitopology IS-IS yet and not planned up to now.

It‘s 2005 now, and we can not deploy planned features

Section four

Break on through to the other side

…of the AS Border

US Internet Exchange Points vs. The World

- While in Europe and Asia the exchange points already have IPv6 on the same LAN/VLAN structure as IPv4, some of the US IXPs still have IPv6 physically seperated from IPv4 and require an extra physical interface from the ISPs that want to peer IPv6.
  - ☛ Unnecessary costs for hardware
  - ☛ No budget for extra hardware only for IPv6
    (We are speaking 5-figure plus maintenance)

US Internet Exchange Points vs. The World

- Some of the US IXPs still use 6BONE (3ffe::/16) address space and refuse to convert to available production address space that is reserved (ARIN: 2001:504::/30) for IXPs.
  - ☛ 6BONE address space will become invalid on 2006-06-06 and IXPs *have* to renumber by then.

- Altough some IXPs on the west coast do operate a shared layer 2 infrastructure and merged their IPv4 peering LANs, there seems to be political resistance to do so for the IPv6 peering LANs.
  Unfortunately this seriously hinders native IPv6 connectivity between the ISP and the research community.

Observed Problems: Insane Tunnels

Tunnels that don‘t align with your physical infrastructure.

- Don‘t throw transcontinental and trans-country tunnels to external parties.
  - ☛ if you don‘t have a network with physically large coverage, then don‘t pretend you do! Use transit/peerings who have the necessary IPv6 footprint!

- Your IPv4 transits might provide IPv6 transit only via tunnels from some central hubs.
  - ☛Those tunnels are ok, as they encourage your IPv4 transit to further invest into native IPv6 deployment.
    - ☛ demonstrated demand.

Observed Problems: Misconfigured Tunnel MTU settings

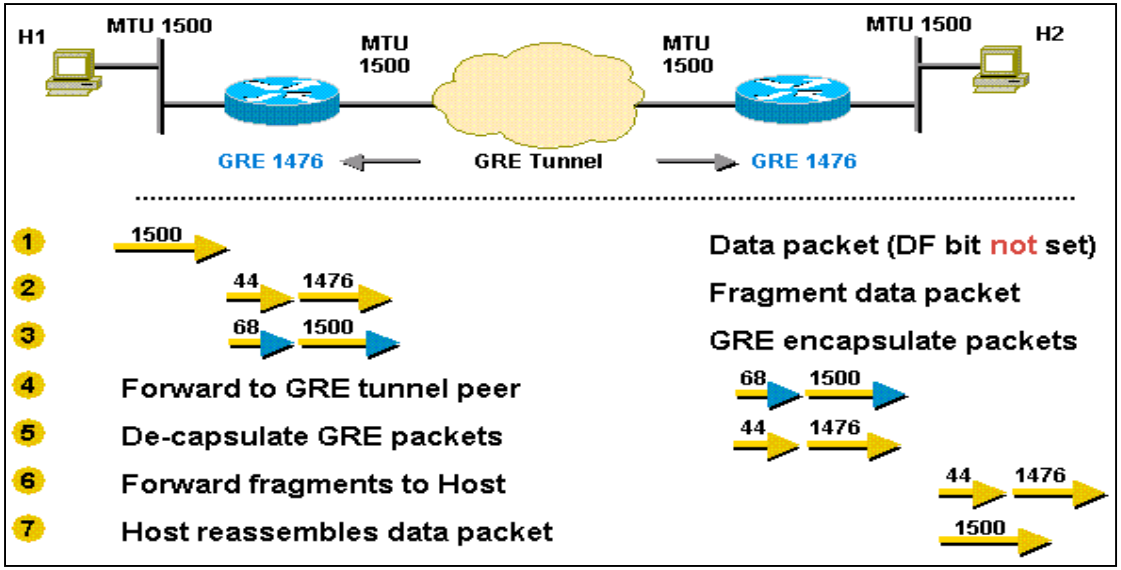
- Both IOS and JunOS derive the tunnel interface payload MTU from the egress physical interface payload MTU.
- Problem occurs if the egress interface MTU is larger than the end-to-end path MTU between the tunnel end points.
- Extremely hard to debug from a remote point of view.
- Troubleshooting requires cooperation of all operators on the (possibly asymmetric) paths between the applications.

Observed Problems: Misconfigured Tunnel MTU settings

Example:

Egress interface POS (MTU 4470)
☛ calculated GRE payload tunnel MTU 4470-24=4446

end-to-end path MTU actually 1500 bytes only, allowing for only tunnel payload MTU of 1500-24=1476!
☛ blackholing of IPv6 packets larger than 1476 bytes.



http://www.cisco.com/warp/public/105/pmtud_ipfrag.html#t

Observed Problems: Misconfigured Tunnel MTU settings

Always configure explicit tunnel payload MTU.

IOS: `ipv6 mtu ...`,

Junos: `family inet6 mtu ...`)

Unless path MTU between tunnel end-points is larger than 1500 for sure, use a maximum of:

GRE (w/o seqnr or keying, etc options): **1476**

IPv6inIPv4 (proto 41): **1480**

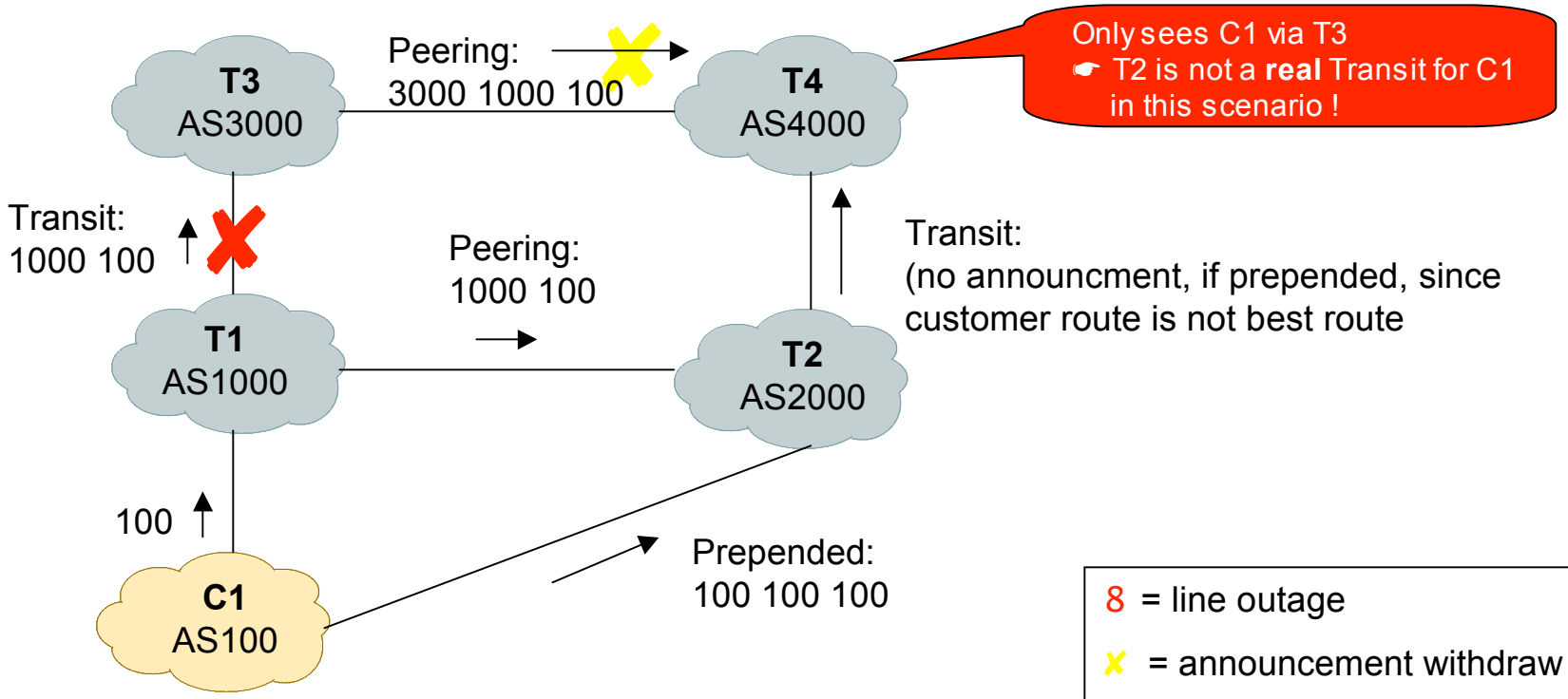
Always verify path MTU between tunnel end-points at least when setting up the tunnel. (ping with DF bit set)

Observed Problems: Careless BGP

- People not filtering their customer announcements
  - ☛ IBGP and foreign transit leaking into the DFZ
- No internal tagging of routes as customer / peer / transit routes
  - ☛ no consistent advertisement to peers / transits.
- Accidential (or „don‘t care!“) route leaking

Observed Problems: Careless BGP

- No localpref structure (customers same as peers/transit)
  - -> no guaranteed upstream provided to customers!



Observed Problems: Misguided „anti-bogon“ filtering

- It took the best of about 3 months to have people fix their anti-bogon filtering in order to allow the full propagation of `2001:5000::/21`
  - Still finding places where it is being filtered.

- People still filtering valid /48 „PI“ prefixes

  ARIN: `2001:500::/29`
  RIPE: `2001:7f8::/32`
  APNIC: `2001:7fa::/32`
  LACNIC: `2001:12f8::/32`

Observed Problems: Misguided „anti-bogon“ filtering

Less-specifics are not harmful.
-> DON‘T filter them!

Until the „IPv6 multihoming issue“ is settled, be liberal in accepting /48 (being permissive rather than strict).

Discard anything longer than /48 on EBGP
(people happily leak /64 peering LANs).

Problems Observed: Free Transits

There are several reasons why people provide free Transit.

- to help others bootstrapping if the IPv4 uplinks don't support IPv6
  - That‘s by all means encouraged
- to provide redundancy uplink if ISP in question has only one IPv6-supporting uplink
  - As long as it is ensured to be **backup-only** (localpref, prepends, etc), it is fine as well.

Problems Observed: Free Transits

- to artificially enlarge the „weight“ of the own ASN in terms of number of routes being advertised to peers.
  - There are folks that do so to gain a better standing in peering negotiations. Obviously this is not OK.
  - We have seen folks, that announce a large number of routes and in the worst case less than 25% of their routes are from actual paying IPv4 customers that also do IPv6.
    - This actually prevents people from using their existing upstream providers for IPv6 (if existing). In turn to that, there will be no budgets for IPv6 on the upstream providers side, as the customer traffic travels differently.
- To remain „Tier 1“ by doing routeswaps (often uncontrolled).

Problems Observed: Misc

- DNS
  - ip6.int ➡ ip6.arpa transition.
- Allocations / BGP announcements:
  - /35 ➡ /32 transition.
    - There are still ISPs out there, that still announce their allocation as /35 instead of /32.

Section five
Conclusion

Conclusion
General

- The vendors must better support IPv6 in their gear NOW.
- Some of the IXP must be more flexible and adjust their setup to the needs of today.
- The now active IPv6 players should clean up their setups in order to move forward with proper routing pathes.

Conclusion
Hints on how to get quality peers

- Select your IPv6 peers carefully based an a few criteria
  - The peer should have a **BGP local-pref structure** in order to be able to distinguish between their customers and peers/Transits.
  - The peer should **not** be a „**full route swapper**“, but if they are, then:
  - The peer should support **BGP communities**, that enable you to control, where they propagate your peering routes to.
  - Last but not least, the peer should not give transit to non-adjacent ASNs.
  - Do **not** set up **inter-continental tunnels**, especially not to non-adjacent networks for peerings.
  - Use the routes of your Transit provider instead.

Conclusion: Pointers

- Mailing-list for operational discussions on IPv6:
  http://lists.cluenet.de/mailman/listinfo/ipv6-ops/

- Getting started with IPv6 in the US:
  http://www.occaid.org/

- Getting started with IPv6 in Europe:
  http://www.sixxs.net/

Thank you

CABLE & WIRELESS