BGP Developments Update

David Freedman – LINX71

claranet

Introduction to IETF-IDR

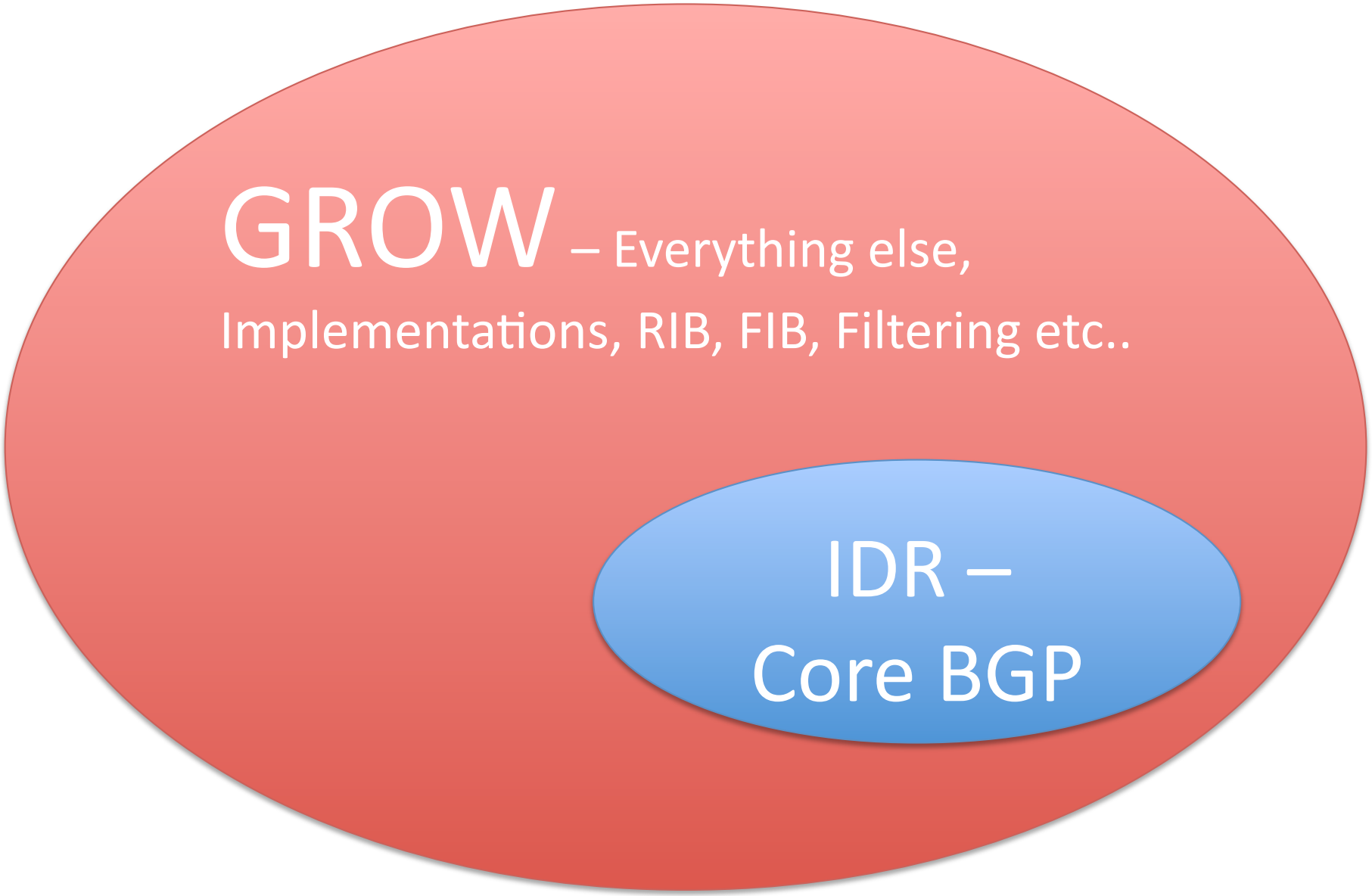
"The Inter-Domain Routing Working Group is chartered to standardize, develop, and support the Border Gateway Protocol Version 4 (BGP-4) [RFC 4271] capable of supporting policy based routing for TCP/IP internets." - http://datatracker.ietf.org/wg/idr/charter/

- Drafts go out to mailing list (idr@ietf.otg)
- Much discussion, arguments etc..
- Adoption as Working Group draft
- Last call and then discussed at IETF meeting (Last was IETF79 in Beijing)
- < insert rest of IETF process here> → IESG → RFC
- Eventually screaming customers get vendor off their backside to implement OR vendor sponsored the draft so was going to implement anyway OR it only turns up in s/w releases for time being (i.e Quagga)

Also clashes with GROW

"The purpose of the GROW is to consider the operational problems associated with the IPv4 and IPv6 global routing systems, including but not limited to routing table growth, the effects of the interactions between interior and exterior routing protocols, and the effect of address allocation policies and practices on the global routing system. Finally, where appropriate, the GROW documents the operational aspects of measurement, policy, security, and VPN infrastructures." - http://datatracker.ietf.org/wg/grow/charter/

- i.e Global Routing Operations


GROW – Everything else,
Implementations, RIB, FIB, Filtering etc..
IDR –
Core BGP

claranet

So what major changes were made to BGP since RFC4271 was out (2006)?

- Better MIBs (various)
- Extended Communities (RFC4360, RFC5668, RFC5701)
- Additional CEASE NOTIFICATION codes (RFC4486)
- BGP for 6VPE (RFC4659)
- RT-CONSTRAIN (RFC4684)
- Graceful Restart (RFC4724)
- BGP based VPLS (Autodiscovery) (RFC4761)
- ASN32 (RFC4893+Various)
- ORF (RFC5291)
- FLOWSPEC (RFC5575)

Working Group Drafts in Progress

draft-ietf-idr--best-external-02

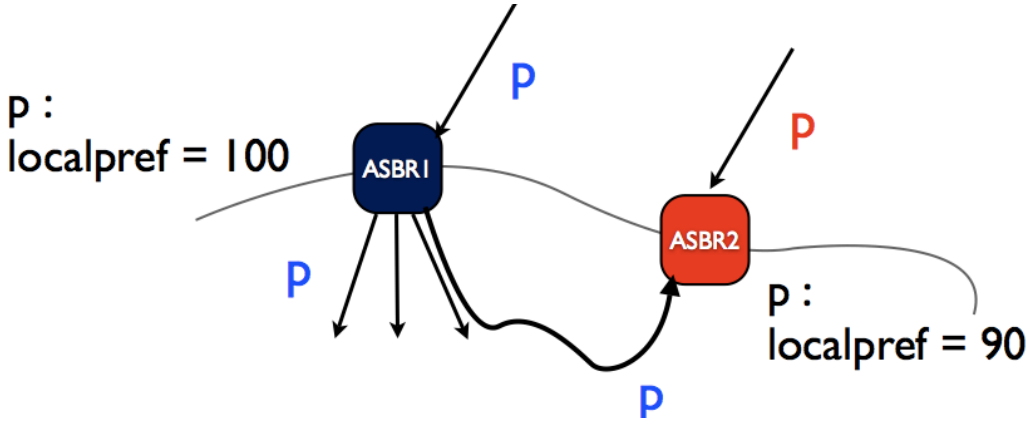


diagram: pierre.francois@uclovain.be

- For a prefix P, received by a pair of ASBRs where one ASBR depreferences by policy, the advertisement of the path P_{90} is supressed by ASBR2 and not available to the rest of the network for unequal cost load-balancing or fast convergence
- Best-external allows the **external** path to be advertised into IBGP even though the **internal** path (P_{100}) is selected for forwarding on ASBR2

claranet

Working Group Drafts in Progress

draft-ietf-grow-diverse-bgp-path-dist-02

- Deal with the internal side of best-external
- No more changes to the protocol, just implementation tweak on RR to take additional paths and distribute them on dedicated sessions (yellow)
- New RR sessions per pathset
- Primary drivers were to combat MED oscillations, improve convergence and improve load balancing capability

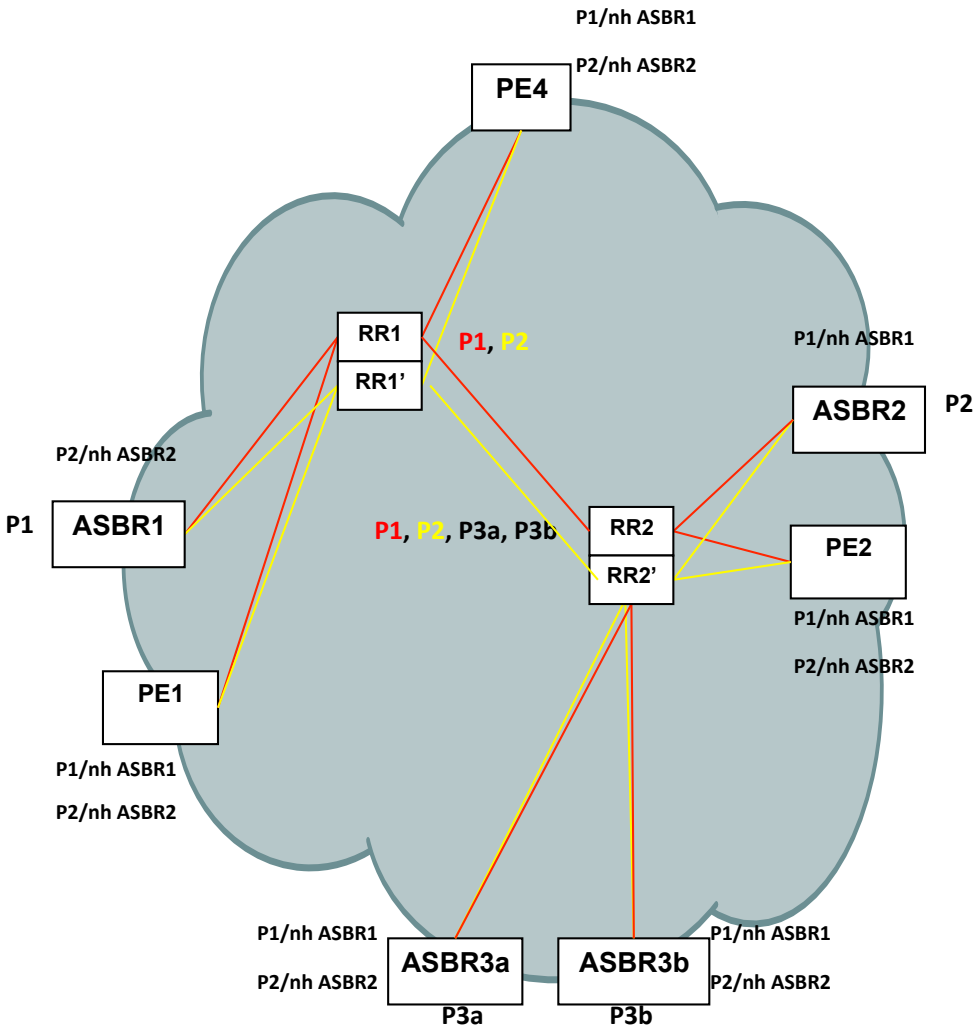


diagram: keyupate@cisco.com

claranet

Working Group Drafts in Progress

draft-ietf-idr-add-paths-04
& draft-uttaro-idr-add-paths-guidelines-03

- Unify benefits of best-external and diverse-path through core protocol modification
- Allow ASBR and RRs to pass on multiple prefixes natively ignoring the IGP tie breaker, over the same session
- Delegate authority to select bestpath to ingress PE who understands IGP topology and is given choice of both ASBRs (and hence can hot potato, ECMP or UECMP in FIB if need be)

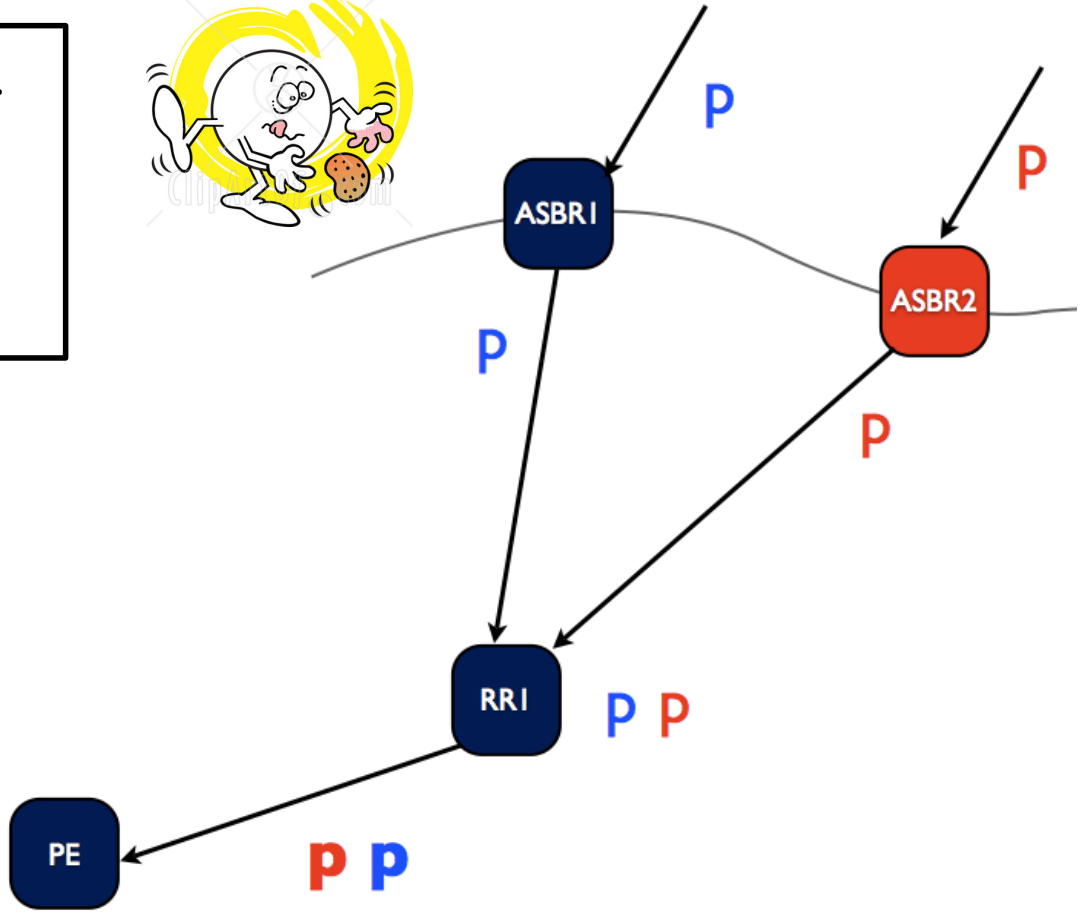
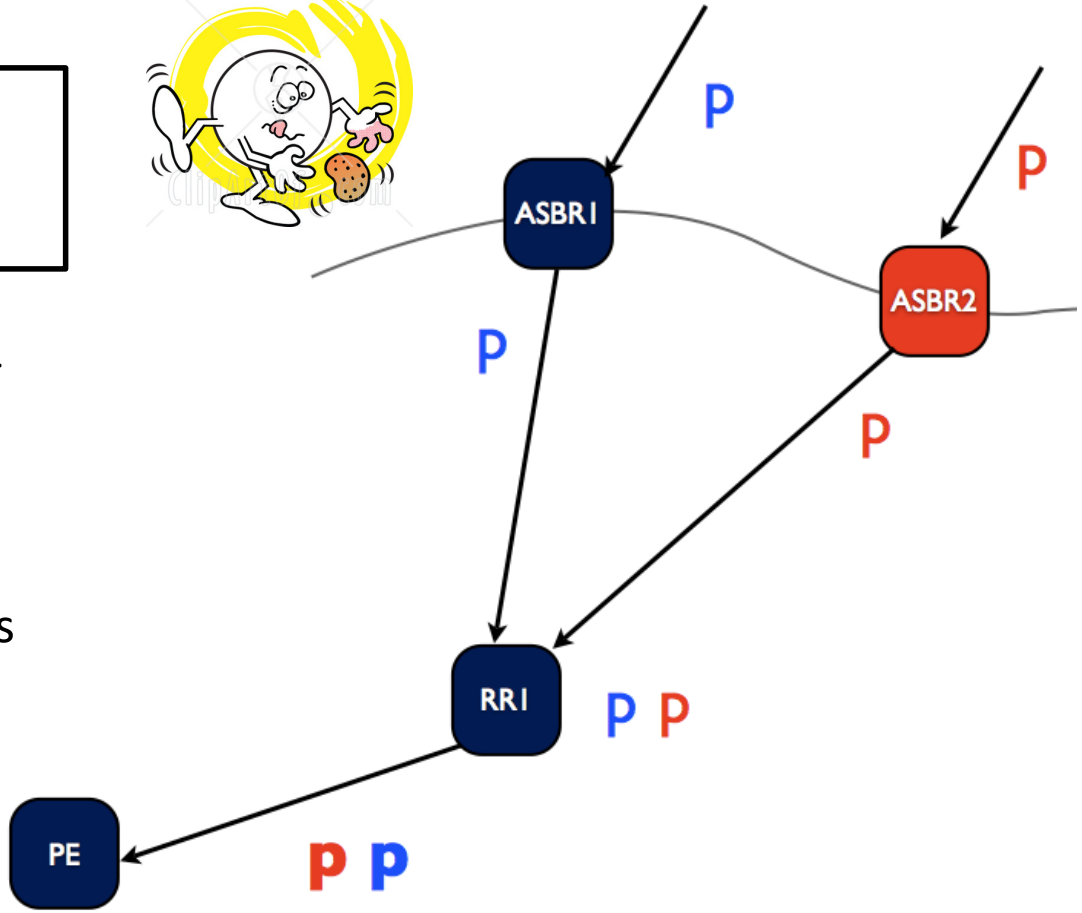


diagram: pierre.francois@uclovain.be

claranet

Working Group Drafts in Progress

draft-uttaro-idr-add-paths-guidelines-03



- Many approaches covered in add-paths-guidelines to describe how one would constrain number of advertised paths
- Two *sensible* approaches from this document, **AD-ALL** (advertise everything at expense of memory and CPU, like full iBGP mesh) and **AD-N** (advertise only N paths)
- Suggest add-paths takes precedence over best-external when configured to honour **AD-N**

diagram: pierre.francois@uclovain.be

claranet

Working Group Drafts in Progress

draft-ietf-grow-bgp-gshut-02 &
draft-decraene-idr-reserved-extended-communities-00

- gshut (Graceful Shutdown) comes from GROW
- Allows prefixes to be re-routed prior to BGP session maintenance (start reconvergence) through use of a reserved “this prefix will soon go away” non-transitive extcommunity in the AS (and between inter-AS peers)
- GROW draft defines GSHUT, IDR draft asks IANA to reserve extcommunity type registries for BGP control plane from existing IANA maintained **standards** extcommunity registries (transitive and non-transitive) – Like “Well Known” (see RFC1997)
- GSHUT then plans to reserve a non-transitive from this (suggested 0xFFFF0000)
- This kind of WG interdependency is **bad** because it can result in **deadlock**

Working Group Drafts in Progress

draft-ietf-idr-advisory-00 &
draft-raszuk-bgp-diagnostic-message-00

- Advisory presented at LINX67, “BGP SMS/Twitter”, unstructured plain text messages between peers, has become a working group item and an implementation is available for Quagga

- Diagnostic proposes structured message types, defined in the draft are messages for exchanging adj-in/out RIB information and doing prefix diagnostics to aid engineers when troubleshooting

- Advisory could possibly travel as a fixed diagnostic type, (though we’ve not been asked!)

Working Group Drafts in Progress

draft-chen-ebgp-error-handling-00

- Malformed Optional Transitives gave us headaches recently (see *"Handling BGP attribute errors"* – LINX65)
- Error handling in BGP needs overhaul, current response is to send NOTIFICATION to connected peer and drop the session
- Limit the destructive power of such an update by **treat-as-withdraw** mechanism whereby if we find an attribute we can't parse in the update we just withdraw the prefixes it mentions as opposed to tearing down the session.
- treat-as-withdraw generally safe except in some internal edge-cases

Working Group Drafts in Progress

draft-keyur-bgp-enhanced-route-refresh-01 &
draft-zeng-one-time-prefix-orf-00

- Route-Refresh (RFC2918) does not do consistency checking, if the refresh is interrupted (i.e truncated), prefixes may be missing which may lead to inconsistent forwarding states
- draft-keyur demarcs refresh with start and end markers to help us determine if we ran into problems
- Combined with draft-zeng, missing prefixes can be re-sent in new refresh with ORF, draft-zeng proposes the ability to just re-send problematic prefixes (without having to refresh the entire adj-rib-out)

Working Group Drafts in Progress

draft-raszuk-wide-bgp-communities-01 & draft-raszuk-registered-wide-bgp-community-values-00

- Bigger, richer communities, holding more attributes
- Not intended to replace standard or extended communities
- Two types, fixed length and variable length
- registered-wide registers some useful variable length types, some specific to dealing with DoS attacks (i.e **ATTACK_TARGET**), rest will be held in IANA FCFS registry.

Working Group Drafts in Progress

draft-jasinska-ix-bgp-route-server-01

- Describes technical and operational requirements for route-servers at IXPs
- Introduces description of how per client prefix-filtering may lead to prefix-hiding and suggests solutions
- Vendor authorship (Cisco), IXP operators (AMS-IX, INEX) and Service Providers (Limelight)
- Started life in GROW, debate as to whether should cross to IDR based on a change to the core protocol (conflict with RFC4271 sec 5 re: update attribute transparency)
- Current implementations (Quagga, BIRD, OpenBGPD) have been patched and should fall into line

Questions?