IETF/RRG Internet redesign discussion & BGP stability proposals JANOG 20

Robert Raszuk
raszuk@juniper.net

Agenda

- What problems are we solving here ?
- Proposals for long term solutions
- Proposals for short term solutions
- Summary & conclusions
- Acknowledgements
- References

What problems are we discussing ?

draft-irtf-rrg-design-goals-01

- **Is routing table size an issue ? →** *No - in a sence that it will not fit – it will, could - in a sence of the consequences.*

- **Is fib size an issue ? →** *Pehaps not with aggregation/compression ... Prognosis for 10M flat FIB entries are real.*

- **Is growing churn/instability an issue ? →** *There is scope for improvement !"*

- **End user portability between providers without site renumbering**

- **Multihoming without leaking more specifics to DFZ**

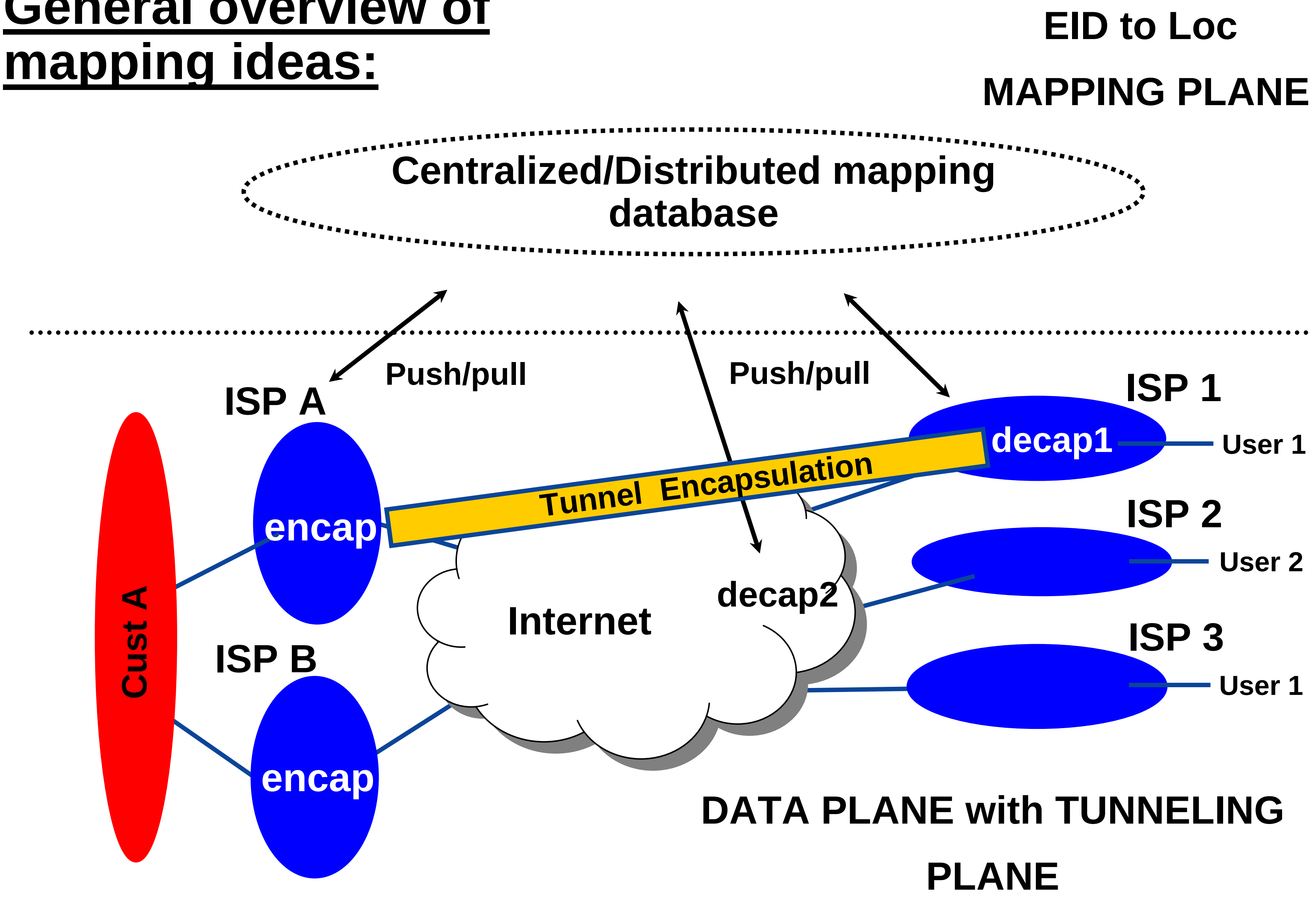
- **Internet TE without leaking more specifics to DFZ**

- **Internet wide convergence proportional to number of routes.**

Proposals for long term solutions

- **LISP**
- **LISP-CONS**
- **I-VIP**
- **NERD**
- [AIRA/RIRA]
- **CRIO**
- **eFIT**
- **HLP**
- **IPvLX**
- **TAMARA**
- [RiNG]
- **NIRA**
- **Teredo**
- **GSE/RFC1955**
- **Shim6**
- **HIP**
- **MULTI6**
- **and many more …**

General overview of mapping ideas:



Some terminology

- Identifier: identifies a host. Says nothing about how to reach the host.
- Locator: Tells how to reach a host or group of hosts.
- Current IPv4 and IPv6 address are both identifiers and locators ssimultaneously. .
- ETR – router which encapsulates packets for tunneling
- ITR – router which performs decapsulation

Major points of solutions

- Network based mapping – host transparency
- Host based mapping – network transparency
- Hybrid architecture

Network-based mapping:

- Transparent to host (smooth migration)
- Host network, transport, application layers unaffected
- Can't help with IPv4-IPv6 transition
- Can't help with IPv4 address space depletion

Major points of solutions

Host-based mapping:

- Transparent to network (tunnel starts and ends on host)
- One or more host layers affected
- 100's of millions of hosts == hard to deploy!
- Could help with transition, depletion
- No need to push/query mapping plane to routers

Hybrid solution:

- Encap, decap on either host or network
- No requirement to upgrade hosts on both sides
- Can coexist together
- Has all benefits of network and hosts based mapping

Major points of solutions

- All solutions are TUNNEL based
- TUNNEL not as point to point pipe
- TUNNEL as just automatic packet encapsulation

WHY ?

- Reduction of forwarding state in transit routers
- Selecting the best single interdomain encapsulation across OS stacks and router vendors
- Line rate encapsulation & decapsulation

Major points of solutions

- **New EID to RLOCs mapping plane needs to be designed, implemented and deployed**
- **Models:**
  - Simple database
  - DNS extensions
  - DHT based algorithm
- **Static vs dynamic nature of information changes**
- **Methods of information query:**
  - Push
  - Pull
  - Incremental updates

Major points of solutions

- BGP Traffic Engineering is not well defined in current proposals
- How to reach ETRs still depends on classic BGP routing
- When ETRs are using anycast addresses (multiple of them have the same ETR address) Traffic Eng to them becomes very difficult.
- ETR address can be injected into upstream ISP just as today more specifics. But this is only neigbor AS – one hop BGP Traffic Engineering – performed by ETR owner not always equal to EID owner !

Major points of solutions

- User mobility has already been solved by Mobile IP WG
- Current proposals reg mobility reuse many concepts from mobile IP WG
- Most require middle helper nodes to serve as transit tunnel routers

- Intra-domain BGP free core is already possible for long time
- Even so not many operators have decided to remove full BGP from their P routers (BGP/RIB/FIB)

Major points of solutions

- “Any problem in computer science can be solved with another layer of indirection.”

  —David Wheeler

- “But that usually will create another problem.”

  —rest of the quote

Proposals for short term solutions

- Draft-li-bgp-stability-01.txt
- Aggregate withdraw
- Attaching „location" to updates and cause/failure location to withdraws
- Fully utilize concept of BGP recursion (double recursion) by introduction of hierarchy with existing protocol and deployed code

draft-li-bgp-stability-01.txt

- Path hunting → Normal BGP (path-vector protocol) behaviour to select the best possible path to the destination during update or withdraw events.
- Results in Internet wide instabilities, excessive CPU processing for routers as well as increase in BGP update traffic.
- Issues are aggravated by the random propagation and very different processing delays within different routers of the Internet. (refractions)
- Can be very accurately compared to a wave model.
- Draft is in early research/brainstorming stage
- Not all ideas in it will work out
- Authors have asked for collaboration and discussion

draft-li-bgp-stability-01.txt

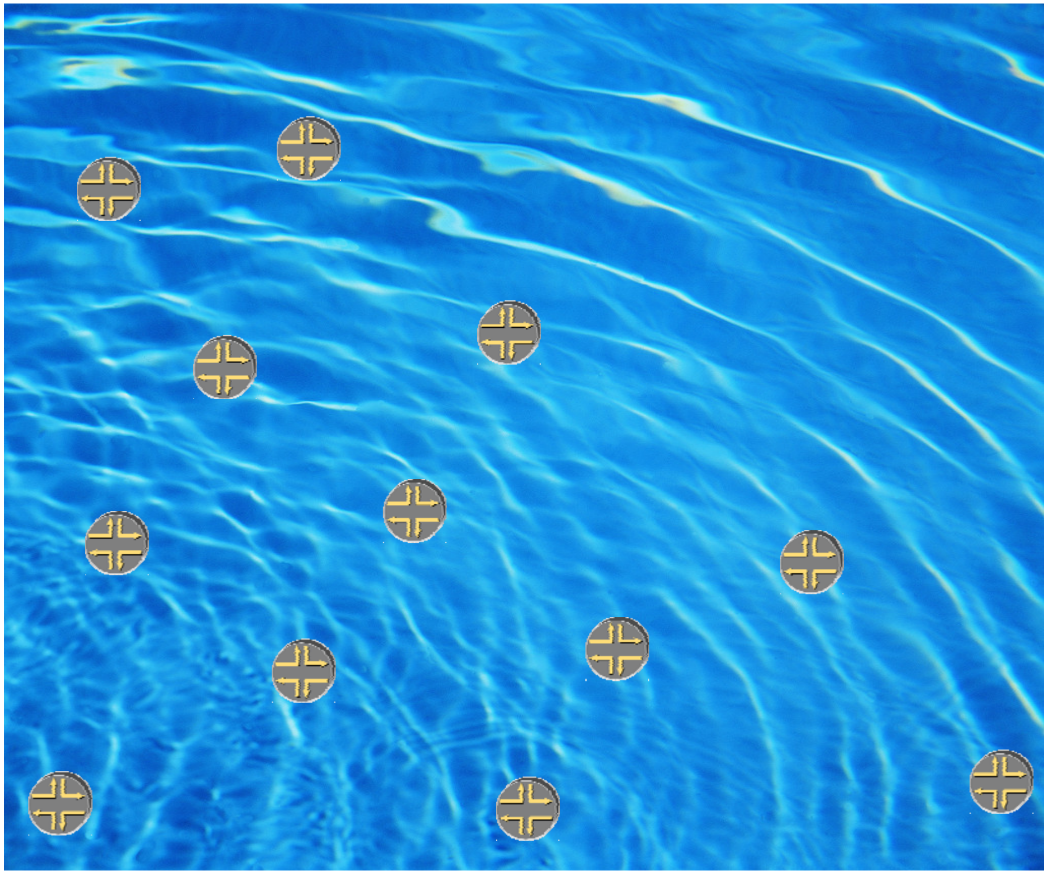
- **The wave model – Ideal situation – no refractions and diffractions:**

draft-li-bgp-stability-01.txt

- The wave model – Reality:

- Different implementations,
- Different CPUs
- Different I/O processing,
- Different MRAI timers,
- Different policies,
- Different processing seq.,

Each router on the way creates its own wavefront.



draft-li-bgp-stability-01.txt

- ***Band-stop filtering:*** To allow a timer based window in which we do not accumulate penalty due to given instability of a prefix. After that time subsequent events related to such prefix are considered as bad.

- ***Path Length Damping:*** The idea to suppress advertisement of new best path for some timer if the resulting new best path has AS_PATH longer then the previous one. Could suppress 20% of BGP churn especially for tail circuits failures of stub sites. Drawback is that for dual homed sites the other valid & permanent path may suffer from long convergence.

draft-li-bgp-stability-01.txt

- ***Optimal path hysteresis:*** A proposal to cache permanently optimal path (long stable path to a prefix) or perhaps just its length. If after withdraw it again reappears for given prefix it is installed and advertised immediately. If some other inferior path is selected (hopefully temporarily) it is not advertised or perhaps not installed for some timer value.

- ***Delayed best path selection:*** The idea to delay best path calculation for any path other then the first one for a given prefix. The delay should be set to allow subsequent calculations when the paths have stabilized.

Summary & conclusions

- Growing number of proposals for redesign the Internet overall architecture
- Some of redesign proposals requires a new mapping plane between EIDs & Locators
- Current Internet as well as BGP protocol is far from the cliff yet there is room for increasing its stability.
- BGP is flexible, supports recursion since day one and does evolve with time.
- Room for contributions both from customers and academia on improving BGP protocol and it's network wide behaviour.

Acknowlegments:

Many thx for review, opinions & contributions to those slides to: Miya Kohno, Yakov Rekhter, Tony Li, Dino Farinacci, Robin Whittle & John Scudder.

References:

- RRG Web page http://www3.tools.ietf.org/group/irtf/trac/wiki/RoutingResearchGroup
- IViP Home page http://www.firstpr.com.au/ip/ivip/
- IAB workshop report: draft-iab-raws-report-00.txt + http://www.iab.org/about/workshops/routingandaddressing/
- LISP: draft-farinacci-lisp-00.txt + www.dinof.net/~dino/ietf/lisp1.ppt & lisp2.ppt
- NERD: draft-lear-lisp-nerd-01.txt
- ROAP: http://www.isoc.org/tools/blogs/ietfjournal/?p=133
- RING: http://www.ist-ring.org/

Thank you !

<u>Questions:</u>
raszuk@juniper.net