Brief Update on The IAB Routing and Addressing Workshop

How can research help?

Elwyn Davies (IAB)
Routing Research Group
IETF67
San Diego, CA

Acknowledgements

- David Meyer for workshop slides
  - See IETF 67 Technical Plenary Slides
- Geoff Huston for thoughts on Identifiers and BGP statistics
  - http://www.potaroo.net/
  - IETF 67 Internet Area Proceedings
- Pekka Nikander for architectural thoughts
  - Arcitecture discussion mailing list

Agenda - Part 1

IAB Hat On

- Why Hold This Workshop?
- Logistics
- Workshop Objectives
- Info Nuggets
- Key Findings
- Workshop Recommendations

Why Hold This Workshop?

- The Internet's routing system is facing a set of serious scaling problems, and...
  - We *are* the IAB, after all, and...
    - *"A is for Architecture"* -- Leslie Daigle
- And importantly...
  - There is a shared opinion among *many backbone operators* that none of the existing IETF efforts provides a complete set of solutions

Logistics

- The workshop was held in Amsterdam, Netherlands on October 18-19, 2006
- 38 attendees
  - Focused on backbone operators
    - Also a few h/w designers, enterprise types
  - 18 (of the 38) were IESG, IAB, or IRTF
  - One scribe
- Many thanks to ISOC/RIPE NCC/ NLnet Labs/Cisco
- And everyone who made the trip to help us think about these issues

Workshop Objectives

- To develop a shared understanding of the problems that operators are facing with today's routing and addressing system, and
- To use that information to inform the IETF process
- ***and stimulate the IRTF***

Info Nuggets

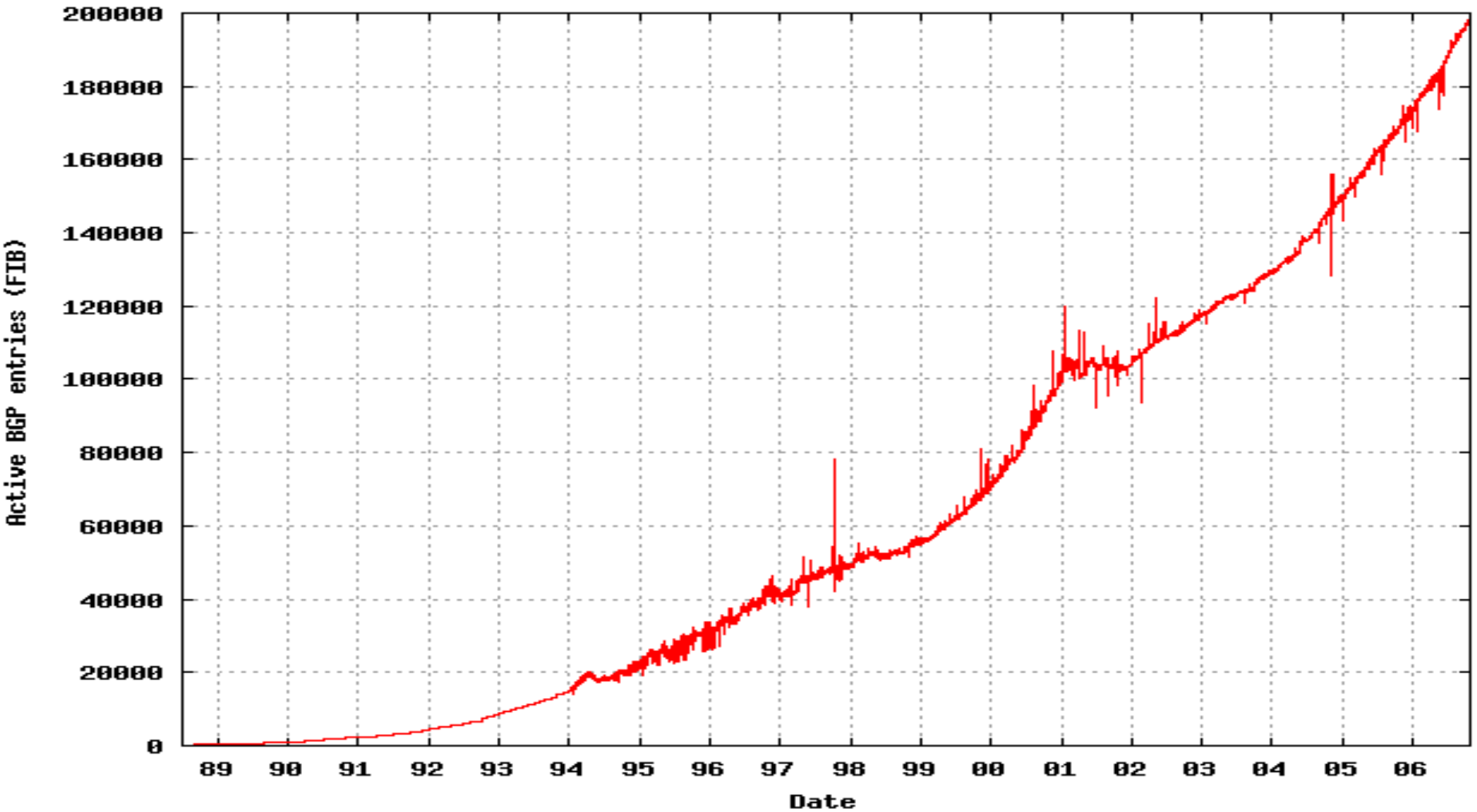
- Sources of Routing Table Growth
- Current and Future Routing Table Sizes
- The Impact of Traffic Engineering
- Instability and Convergence Time
- Renumbering is a Capital Offence
- Economics of Routers
- Power Hunger and Heat Death

Sources of Routing Table Growth

- ***Currently...***
  - Organic Growth (more hosts, sites)
  - Deaggregation due to...
    - Multihoming
    - Traffic Engineering for
      - load spreading
      - policy routing (economics & politics)
    - Incompetence
- ***Future...***
  - Use of IPv6
    - Organic growth
    - Parallel (dual-stack) deployment

Routing Table Growth

BGP DFZ Route Count



Routing Table Size *Now and Tomorrow*

- Some Tier 1 providers already have routing table sizes from 0.5M to 1M routes
  - Made up from
    - 200K External routes
    - 50K-150K Internal deaggregates
    - Remainder customer VPN routes
- Estimates by Jason Schiller indicate that adding IPv6 (worst case) will grow the routing table to 1M routes without customer VPN routes in 5 years

Impact of Traffic Engineering

- Need to steer traffic to meet business aims
  - Satisfy external policy (political constraints)
  - Meet premium customer expectations
  - Keep pipes full without traffic loss
- Partly driven by use of multiple parallel paths
  - Traffic volume growing faster than pipe sizes
  - 'Sweet spot' for price-performance is lower than maximum size
- Basic BGP routes are often inappropriate
  - Need TE deaggregation to provide precision needed
  - Needs to be fairly 'fine grained' to achieve 1% resolution

Instability and Convergence Time

- There is a good deal of instability and churn in the BGP tables
  - Lots of updates - repeated cycles
- A small number of AS's generate a large proportion of the instability
- Combined with slower convergence as tables grow and traffic engineering changes, this keeps core routers continually busy processing updates to RIBs and loading new FIB tables

Renumbering is a Capital Offence

- Currently asking an enterprise to renumber (all) their nodes is likely to result in blood on the floor
- IP addresses are embedded in far too many things
- See http://tools.ietf.org/html/draft-chown-v6ops-r

Router Economics

- Mainly about "big iron" in the core
  - especially at the edge of the core
    - edge routers don't just route - lots of other resource hungry functions - ACLs, classification, schedulers
    - and these functions scale with table size also
- Too much "bleeding edge", low volume silicon
  - Heavy duty ASICs
    - commodity processors are not fit for purpose
  - Memory - SRAM, TCAM, multiple DRAMs
- Improvements in performance overall in these categories may not be enough to offset routing table growth
- Routing table growth combined with bleeding edge technology usage may make routers more expensive per prefix/route over time

Power Hunger and Heat Death

- "Big Iron" Routers are very power hungry
- Mainly due to the heavy duty ASICs in the forwarding engines
- Already at or beyond limits of power supply in typical co-lo facilities (-48v DC)
- Getting rid of waste heat is a major problem
  - Exceeding heat removal capabilities
  - Needing extreme heat sinks locally and for whole facilities (build by bodies of water!)
- Problem is getting worse

Key Workshop Findings

- **The scalability of the routing system is an urgent problem**
- Super-linear RIB growth is a great concern
  - Increased BGP convergence times and associated costs
  - RIB (UPDATE) dynamics also an issue (cf deaggregation)
- Questions about the applicability of Moore's Law to high-end routers (in particular, FIB memories)
- And of course, along with all of the various constraints e.g., no provider lock (PA/CIDR), TE, multihoming, ...
- Shared problem between IPv{4,6}
  - Larger IPv6 address space exacerbates these problems

Key Workshop Findings

- **The use of IP addresses for both ID and Locator is a problem**
  - Workshop participants felt that a solution to this overloading may solve the mobility and multihoming problems
  - Examined the tradeoffs inherent in SHIM6 and GSE
  - Long term solutions need to consider the anticipated "orders of magnitude" growth in new mobile end devices

Key Workshop Findings

- **Costs and Benefits in current practices are not aligned**
  - Canonical example: multihoming
- **Cost/Benefit curves vary by stakeholder**
  - An enterprise may have a very different view of the cost/benefit tradeoffs of a given solution set than say, content provider might

Workshop Recommendations

- **These problems are urgent**
  - Need to start working on solutions now
- **Need to reach out to all stakeholders**
  - In addition to backbone providers, we need to reach out to the content providers, enterprises, applications folks, vendors,...

Workshop Recommendations

- **We must develop solutions in an open & transparent way, engaging the broad community**
  - ***Engaging research community as well***
- **Look into whether interim solutions are necessary to buy us a little time**

Workshop Recommendations

- **Need to develop a clear and coordinated approach to solutions development**
  - Roadmap
  - Near, intermediate, and long term steps from current state to solutions

Invitation...

- The IETF needs the help of the research community to understand the problem and produce good, deployable solutions
- ***Please ask us (I*, any other participants) about the problems, think about these problems and tell us your thoughts (and solutions)!***

Agenda - Part 2

IAB Hat Off

- What can the research community do to help?
- Thoughts on topics
  - NOT a complete set!

Architectural Thought

- Too much symptomatic fixing
- Need to find ways to a sustainable future rather than point fixes
- This is VERY Difficult
  - Ossification has set in
  - Least Common Denominator thinking
  - see DARPA NewArch report

Just what is the Scaling Problem?

- Whilst the workshop agreed that there was a scaling problem with routing, we need to be absolutely sure what we mean by this problem!
- The power issue is serious
- Interaction with packet classification and ACLs is important - not just a pure routing problem

Routing for a Meshy Net

- The network is no longer the same shape
- BGP tools are not as effective as they were
  - Path stuffing etc no longer works
- Dima Krioukov's presentation later

Support for TE

- Traffic Engineering is currently horribly ad hoc
  - Tweaking of BGP
  - Deaggregation of routes
  - Inspecting packets to spread loads
- Controllable and Manageble mechanisms needed
- Integrated in the routing system

Meaning of Identity

- **What if we do try to untangle identities and locators?**
  - **Need to ensure that the solution...**
    - **solves the right problem**
    - **doesn't actually make other things worse**
- See Geoff Huston's presentation to Internet Area session
  - http://www3.ietf.org/proceedings/06nov/slides/intarea-1.pdf
- Pekka Nikander says...
  - There are at least three alternative ways to answer to a problem requiring naming changes:
    - Overload the current name spaces with new semantics
    - Change (completely) the semantics of an existing name space
    - Add a new name space
  - Mostly we do #1 at present because others cost money

Aside: Lookup mechanisms

- If you separate id and locator...
  - You need an extra lookup/map
- How do you do this?
  - Overloading DNS ***again?***
- Is there a good way to do the lookup in a cheap, fast, *non-hierarchical*, scaleable, distributed way?
  - Gospel is that we have exhausted the possibilities... ***Is this a failure of imagination?***
  - A hard research problem still!

Short Term Fixes

- Improving iBGP
  - It has major problems!
  - e.g., Balakrishnan's paper
    http://nms.lcs.mit.edu/papers/index.php?detail=141
- Tools to help an AS apply policy from a central point
  - Maybe will reduce the instability
  - Reduce shortage of skilled BGP hackers

More Information and Discussion Venue

- Output from workshop (work in progress)
  - http://www.iab.org/about/workshops/routingandaddre
- Discussion currently on arch-discuss mailing list

  architecture-discuss@ietf.org

  https://www1.ietf.org/mailman/listinfo/architecture-discuss

Questions/Comments?

Thanks!

Looking Forward to your Input!