

Don't Forget Today's Infrastructure!

IAB Routing & Addressing Workshop
and
Do we need a new network model?

Elwyn Davies
Folly Consulting
Terena European Future Networks Initiatives Workshop
Schiphol, Netherlands
22 February 2007

Foreword

- I am a member of the Internet Architecture Board
- I was at the IAB Routing and Addressing Workshop RAWS in October 2006
- Draft report...
<http://www.ietf.org/internet-drafts/draft-iab-raws-report-00.txt>
- ... but ...
- This is entirely a personal view!

Why am I here?

- Report what we learnt at the IAB RAWS
- Report about what the IETF/IRTF are planning partly as a result of RAWS
- Remind the workshop that we need to keep the infrastructure running while we experiment
- Present some personal thoughts on interesting areas of research
- Encourage work to help improve the

Agenda - Part 1

IAB Hat On (sort of)

- Why Hold RAWS Workshop?
- Workshop Objectives
- Info Nuggets
- Key Findings
- Workshop Recommendations
- Where next?

Why did IAB organize RAWS?

- 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

The Challenges

- Phenomenal growth in
 - Number of Hosts and Sites
 - Traffic per Host
- The need for resilience and redundancy
 - Commercial
 - Critical Infrastructure
- Dilution of operational skills vs
increased complexity

RAWS Objectives

- Gather the views of (backbone) operators
- 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 researchers

Info Nuggets

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

Sources of Routing Table Growth

- ***Currently... 1M routes in DFZ routers (200K external)***
 - Organic Growth (more hosts, sites)
 - Deaggregation due to...
 - Multihoming
 - Traffic Engineering for
 - load spreading
 - policy routing (economics & politics)
 - Incompetence
- ***Future... 1M external routes in the DFZ in 5 years?***
 - Use of IPv6
 - Organic growth
 - Parallel (dual-stack) deployment

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

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
 - Bigger tables, Fatter pipes, (More servers)

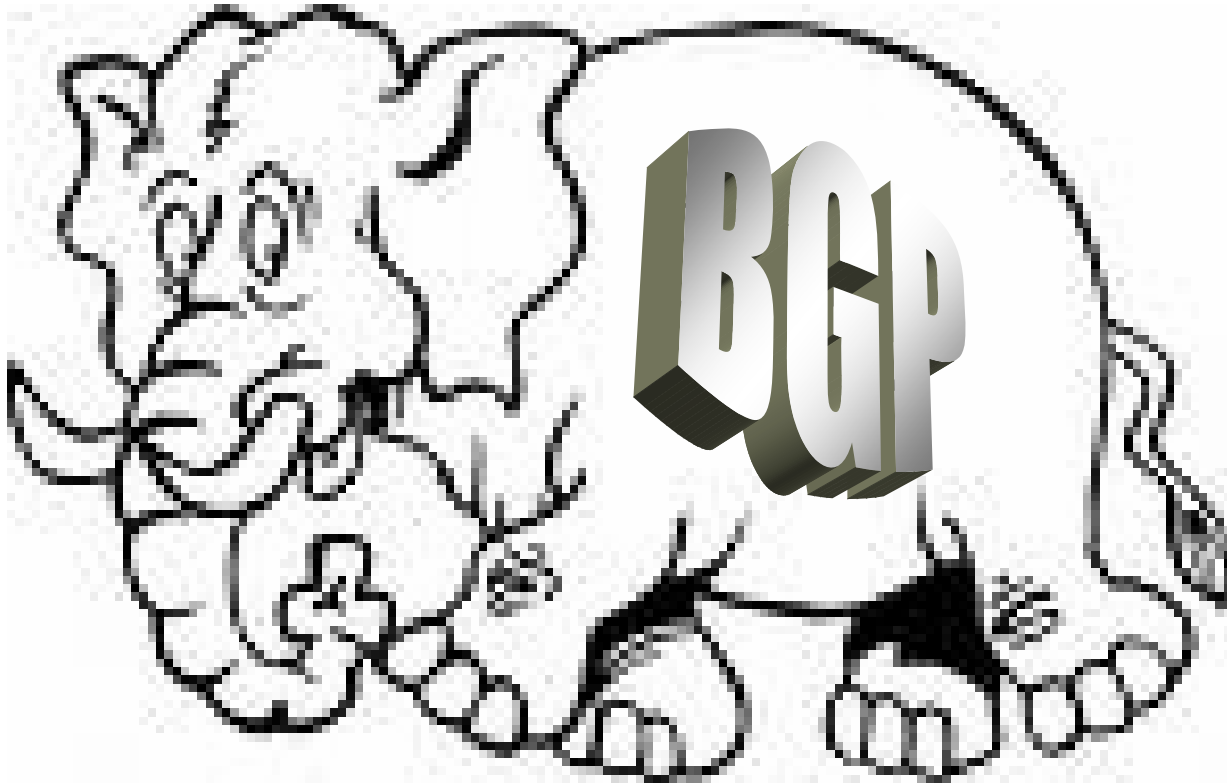
Key Workshop Findings

- **The scalability of the routing system is an urgent problem**
- **The use of IP addresses for both ID and Locator is a problem**
- **Participants felt a solution to id/loc split might help solve multihoming and mobility**
- **Costs and Benefits in current practices are not aligned (think: multihoming)**
- **Costs/benefit tradeoffs vary according to stakeholder type (think: enterprise vs content provider)**

What next?

- Degree of urgency!
 - Feeling that we are 'riding the tiger'!
- Reach out to *all* stakeholders
- Series of area meetings at IETF 68 (Prague)
 - Id/loc split, new rtg/addr arch, interim solns...?
- Coordinate multiple strands of response
- *Engage the research community*

Obstacles



SCIENCE



Agenda - Part 2

IAB Hat Off

- Architectural Thoughts
- Identifier/Locator Split proposals
- Traffic Engineering and Multihoming
- Do we need a revised network model?
- Mismatch between routing with BGP and the current real network?
- What might we do?

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 RAWS workshop agreed that there was a scaling problem with routing, the community needs 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/addressing problem

Identifier/Locator Split

Meaning of Identity

- **What if we do try to untangle identities and locators?**
- **Alleged object: reduce routing table size**
 - **Need to ensure that the solution...**
 - **solves the right problem**
 - **doesn't actually make other things worse**
- **Hazards:**
 - **Aggregates get too large for a single pipe**
 - **We just moved the problem closer to the DFZ!**
 - **Solutions often involve tunnels or encapsulation (YAN*)**
 - **Data needed for ECMP now buried deep in the packet**
 - **Makes more work for core routers**
 - **Risks of multiple solutions**

* YAN =

Yet Another NAT

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
 - Need better TE tools
- Do the assumptions of the (implicit) network model still apply?

Multihoming and TE

Two faces of the same coin?

- Multihoming not just an edge phenomenon
- Providers are multihomed in the meshy net
- TE only really interesting when there are multiple (nearly) equally useful traffic paths
- Who decides the path for traffic?
 - Tussle: hosts/end sites vs providers
- Is it really essential for all a connection's packets to follow the same path?
 - Makes the providers' job harder!

Network Modelling

- (Original) Internet Layer Assumptions
 - There is a path across the cloud
 - No statements about capacity
- Transport Layer Assumptions (driven by TCP)
 - Capacity is limited
 - Nearly in-order delivery is possible/desirable

Routing/TE Mantras

- All the packets from A to B must follow the same path
 - Conflicts with meshy network
 - Limits the choices of providers
 - Ultimately limits capacity from A to B

Consequences for Routing

- Need to know more about the distant parts of the route than is necessary
- Inability to distribute traffic from a large stream across multiple paths
 - (not just ECMP paths)
- Militates against map-style algorithms
- Drives homogeneity - difficult to connect to DTNs and mobile networks because they can't export exact routes

Changing the Model?

- The Internet is inherently multi-pathed
 - Original IP model encompasses this
- Accept that paths are bandwidth limited
- There is no longer one best path...
 - but there may be several that are much the same
 - Embrace and use all the paths
 - Delegate the path decision per packet
 - Restore the original intent of transport protocols (reduce the need for in-order delivery)

Contrasts and Analogies

- Contrast:
 - SCTP: Embraces multiple paths and tries to use them all
 - SHIM6: Disguises the multiple paths and only uses one at a time
- Analogy:
 - MIMO wireless networks - abandoned the fight against multipath reception and used it to drive more bandwidth!

IDR Unchained?

- Existing research tends to try to do the same job as BGP - but more efficiently
- Is changing the network model a way to find a better routing system?
 - Removes a constraint in the best traditions of research.
- May mesh better with similar problems in wireless networks.
- See Rexford's MIRO paper
 - <http://www.cs.princeton.edu/~jrex/papers/multipath06.pdf>

More Information and Discussion Venue

- Info from workshop (work in progress)
 - <http://www.iab.org/about/workshops/routingandaddressing/index.html>
 - <http://www.ietf.org/internet-drafts/draft-iab-raws-report-00.txt>
- Geoff Huston's "Identity" presentation at IETF 67
 - <http://www3.ietf.org/proceedings/06nov/slides/intarea-1.pdf>
- Discussion currently on routing and addressing mailing list ram@iab.org
<https://www1.ietf.org/mailman/listinfo/ram>

Questions/Comments

?

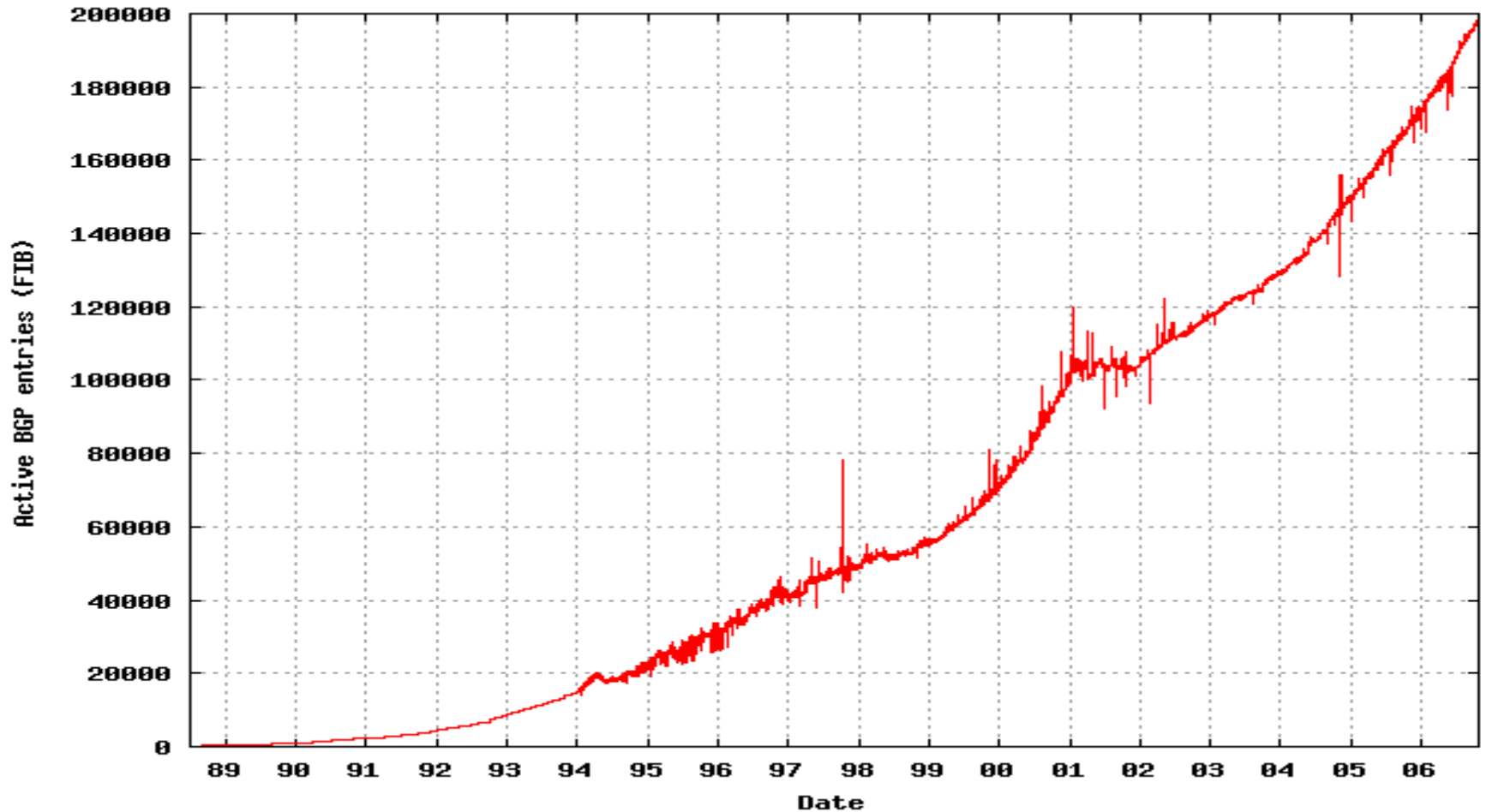
Thanks!

***Looking Forward
to your Input!***

BACKUP

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

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

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
- Hence enterprises want PI addresses
- See <http://tools.ietf.org/html/draft-chown-v6ops-renumber-thinkabout>

Support for TE

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

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?*

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