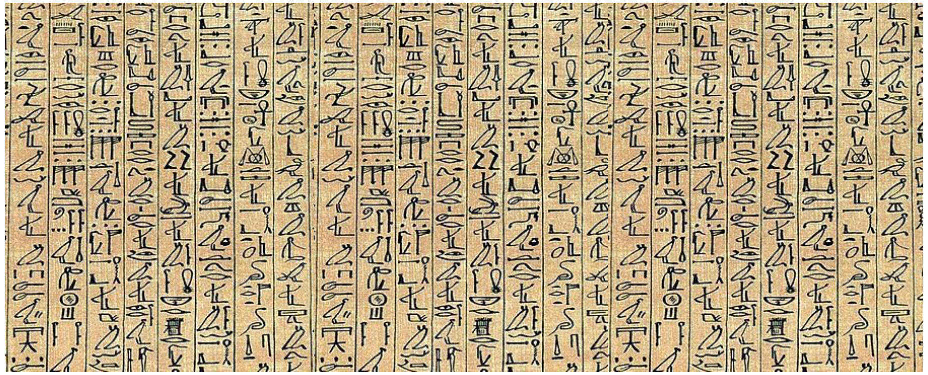
30 Years of BGP

Design Expectations vs. Deployment Reality in Protocol Development

Geoff Huston
APNIC
June 2019

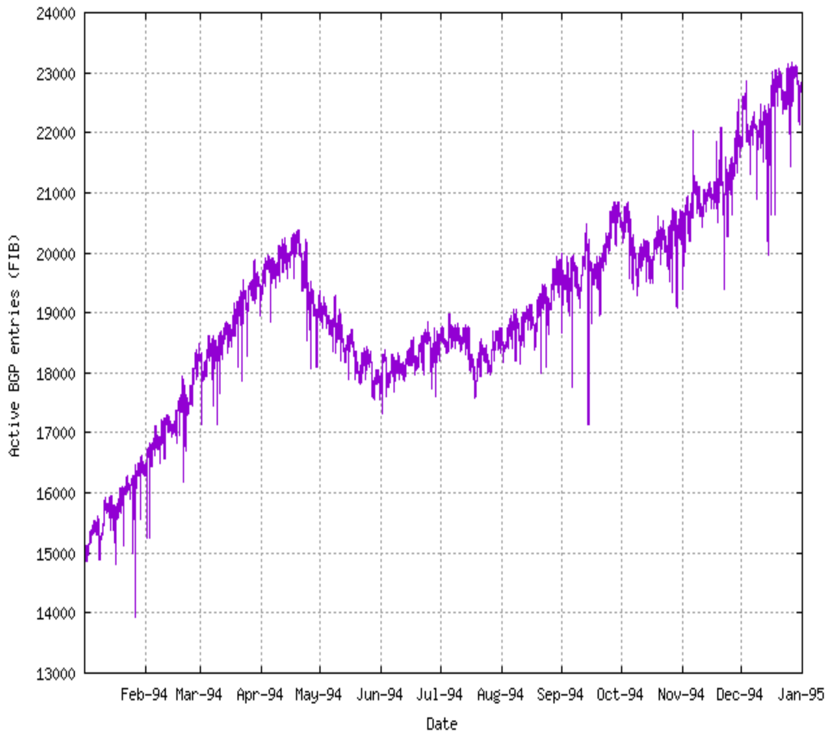
In the Beginning…



- BGP was an evolution of the earlier EGP protocol (developed in 1982 by Eric Rosen and Dave Mills)
- **BGP-1** – RFC 1105, June 1989, Kirk Lougheed, Yakov Rekhter
  - TCP-based message exchange protocol, based on distance vector routing algorithm with explicit path attributes
- **BGP-3** – RFC1267, October 1991, Kirk Lougheed, Yakov Rekhter
  - Essentially a clarification and minor tweaks to the basic concepts used in BGP
- **BGP-4** – RFC 1654, July 1994, Yakov Rekhter, Tony Li
  - Added CIDR (explicit prefix lengths) and proxy aggregation

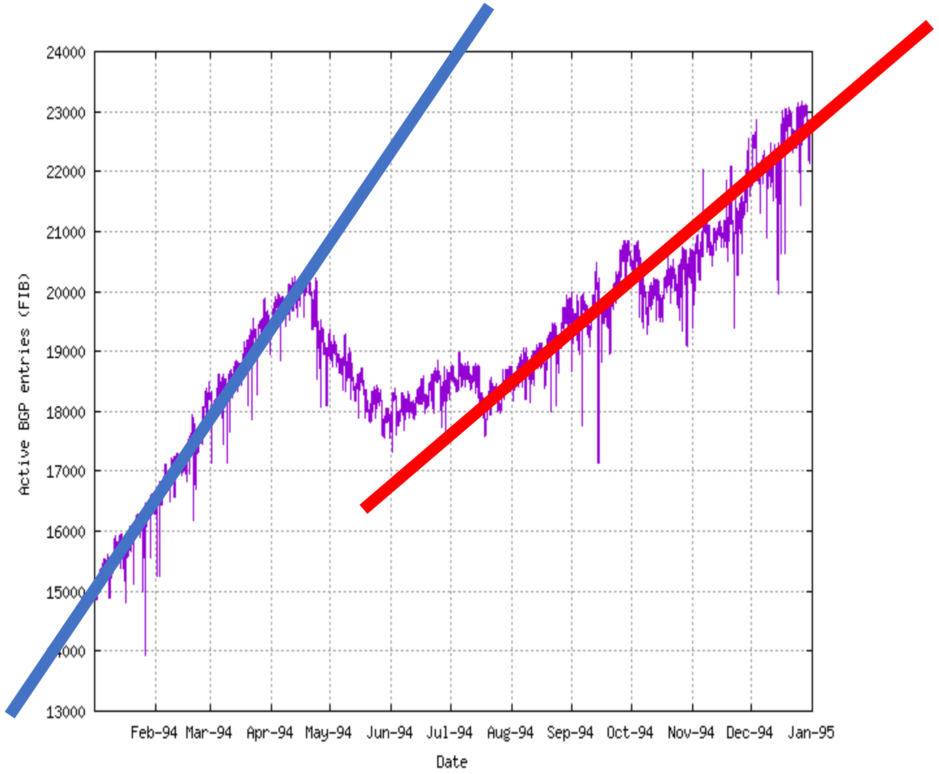
Containing the Routing "Explosion"

- IETF ROAD Efforts in 1992 (RFC1380)
  - Predicted exhaustion of IPv4 addresses and scaling explosion of inter-domain routing
- The chosen “solution” was to drop the concept of address classes from BGP
- It (sort of) worked



Containing the Routing "Explosion"

- IETF ROAD Efforts in 1992 (RFC1380)
  - Predicted exhaustion of IPv4 addresses and scaling explosion of inter-domain routing
- The chosen "solution" was to drop the concept of address classes from BGP
- It (sort of) worked



IPv6 and BGP

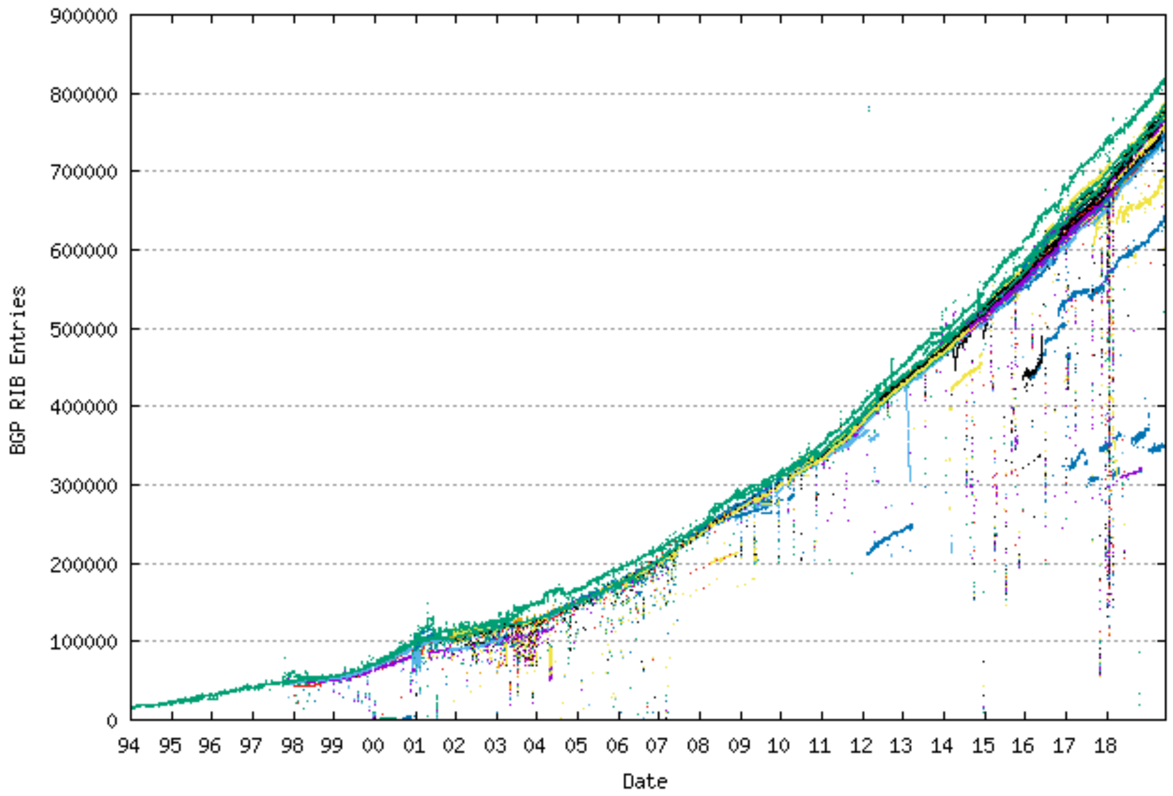
- While the IETF adopted the IPv6 address architecture for the address exhaustion issue, it was unable to find an IPv6 routing architecture that had similar scaling properties
  - IETF efforts to impose a routing hierarchy (TLAs and sub-TLAs – RFC 2928) got nowhere!
- So we just used BGP for IPv6 in the same way as we used BGP for IPv4
  - Address allocation policies that allocated ‘independent’ address blocks of /35 or larger
  - ISP traffic engineering and hijack “defence” by advertising /48s

BGP isn't perfect

- Session insecurity
- Payload insecurity
- Protocol instability
- Sparseness of signalling
- No ability to distinguish between topology maintenance and policy negotiation

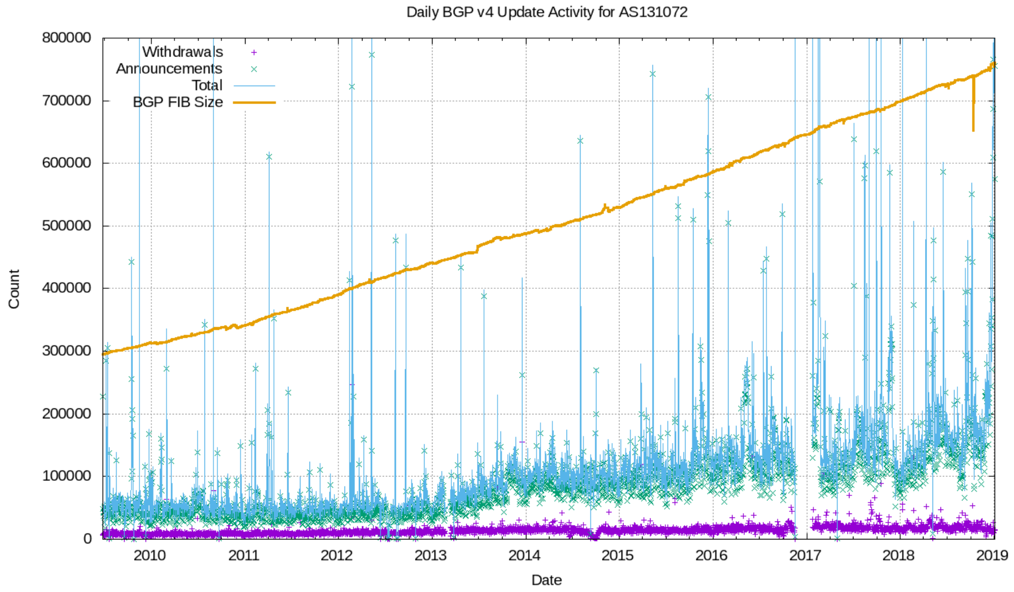
Scale generates inertia

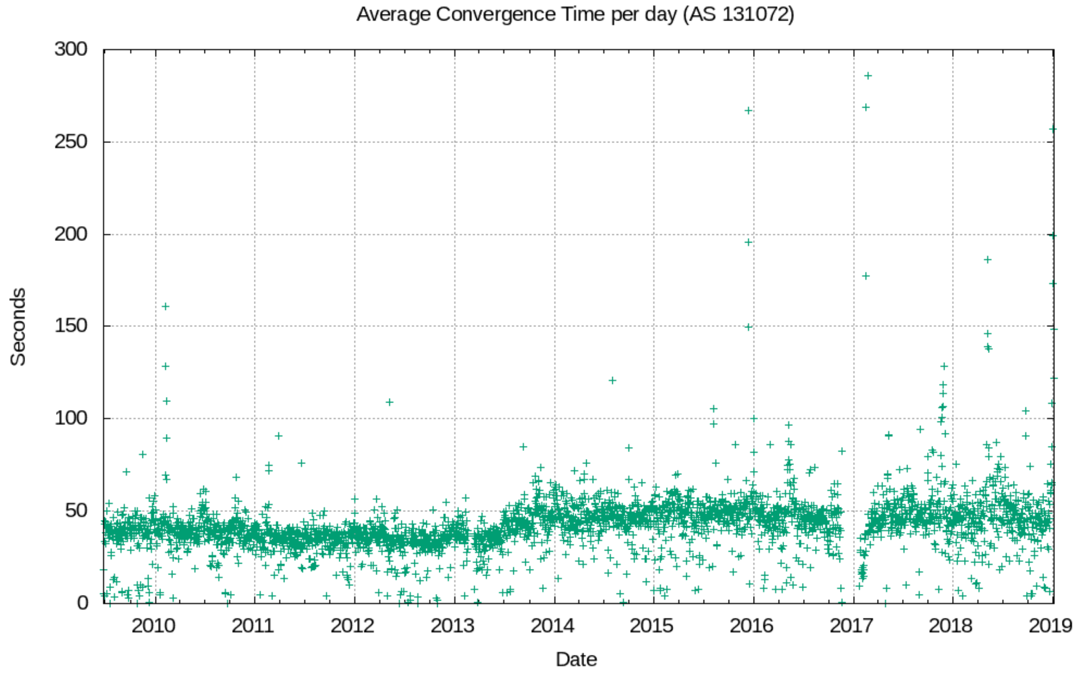
- BGP-4 was introduces when the routing table contains ~ 20K entries – it is now ~800K entries
- The network carries some 75K ASNs
- Changing the internet to use a new common IDR protocol would be incredibly challenging – something would need to break to force change



BGP Scaling

- BGP has scaled because the protocol only passes changes
- As long as the change rate is low the BGP load is low
- And the inter-AS topology of the Internet works in BGP's favor
- And this has allowed BGP to grow well beyond the original design expectations





Expectations vs Deployment

- Session lifetime
  - Expectations of short session lifetimes – experience of longevity
- Session Security
  - Expectation of routing being a public function - experience of session attack
- Payload Integrity
  - Expectations of mutual trust – experience of malicious and negligent attack
- Protocol Performance
  - Expectations of slow performance – experience of more demanding environments
- Error Handling
  - Expectations of "clear session" as the universal solution – experience required better recovery without session teardown
- Use
  - Expectations of topology maintenance – experience of traffic engineering

Why does BGP still work?

- It’s a Hop-by-Hop protocol
  - This allows new behaviours to be deployed on an incremental basis, as long as there is a “tunnelling” capability to pass through legacy speakers
    - A classic example is the 2-byte to 4-byte AS number transition in BGP
- It’s layered above a generic reliable stream transport
  - Arbitrary message sizes are supportable
  - No need to refresh sent information