BGP, where are now?

John Scudder and David Ward

IETF-68, March 22, 2007

Agenda

- Trivia
- Dynamic behavior
- Convergence properties and pro
- Convergence/stability work item

Goals and Priori

- Goal: Maximize connectivity of Ir
- Convergence and stability are sul this
- Implication: Priorities
  - First: fastest service restoration

Focus

- This talk focuses on performance a
- There are other very important as BGP
  - Services
  - Operations
  - Weird behaviors (wedgies, etc)
  - S

Shalt Not's

- BGP uses ASes for loop suppress
  nothing else!
  - Speaking of "overloading things"… A
    locators. No topological significance
- Auto-aggregation appears to be a
  starter
  - Even proxy aggregation is tricky, but

MP-BGP

- BGP carries data for multiple adc families (AFs)
  - Plain old IP (v4, v6)
  - VPNv4
  - Other things
- Not all AFs need to be present o

VPNs

- Often observed that VPN tables Internet table
  - True, in aggregate
  - But, not true of any *single* VPN table
- Inherently parallelizable
  - No single PE or RR holds all VPN ta

BGP dynamic beha

- Confusion even among routing ex
- Of course, surprising emergent b
  are possible
- … but important to understand l
  conditions

BGP and TCP

- BGP runs over TCP
  - Flow control: important implications dynamics
  - Intuition about TCP is usually wrong

BGP under loa

- When uncongested, BGP will pas
  as fast as they are received
  - Modulo MRAI, dampening
- Degradation mode under (CPU)
  congestion: state compression
  - "Adaptive low-pass filter" behavior

BGP under load

- BGP adapts to speed of peer
  - Slow peer gets routes as slow as it wa
    state compression)
  - Fast peer gets routes as fast as it want
  - Implication: One slow peer does not h
    convergence
- Update packing

BGP convergen

- At least O(n) in the size of the D
  - Fundamental to how BGP transport
- But full convergences don't happ
  - At startup ("initial convergence")
  - On rare occasions otherwise
- Hard to "fix" completely — but i

BGP convergence

- Techniques to avoid full converge
  - Graceful Restart
  - Nonstop Routing
- … or to cover them up
  - Different flavors of fast reroute
- … or to pre-converge by adverti

Route Reflectic

- RRs hide backup paths
  - Reduce RIB sizes (but less than you
  - Bad for convergence
- Convergence:
  - State reduction/data hiding
  - Faster convergence

Known Algorithm Deficiencies

- Path hunting
- Nonconverging policies
- At least O(n) in DFZ size

Path Hunting

- Well-known amplification effect
- Approaches to reduce
  - Root cause notification
  - Propagation of backup paths

Propagation of Ba
Paths

- Transit ASes seldom fully partitic each other
- However, when a single AS-AS li down, border router temporaril routes
  - Due to aggressive data hiding by less

Propagation of Ba Paths [2]

- Speculation: many "path disturba events caused by this effect
- Intra-domain backup propagation today
- Cost: some additional RIB state v
- Benefit: faster internal convergen

Some Possible T

- As-pathlimit
- Aggregate withdraw
- Best-external
- Better instrumentation reusing WRD infra
- BGP free core (pick your encap)

Moving Forwar

- Narrow down (or expand!) “possib
  list
- Align costs and benefits
  - Those who pay, must benefit, or solutic
    be deployed
  - Many examples of existing technically-e
    “solutions” to current problems… but
    exist. Example: BCP-38

Dampening

- Misused in past (we were wrong default parameters)
- Heavy contribution of few sites t suggests very generous paramete only penalize egregious flappers
  - Study needed to validate what const "egregious"

Punch Line

- BGP not in danger of falling over
  - Lots of runway
- IDR
  - Near-term improvements
- RRG
  - Fundamental changes, e.g. new routi