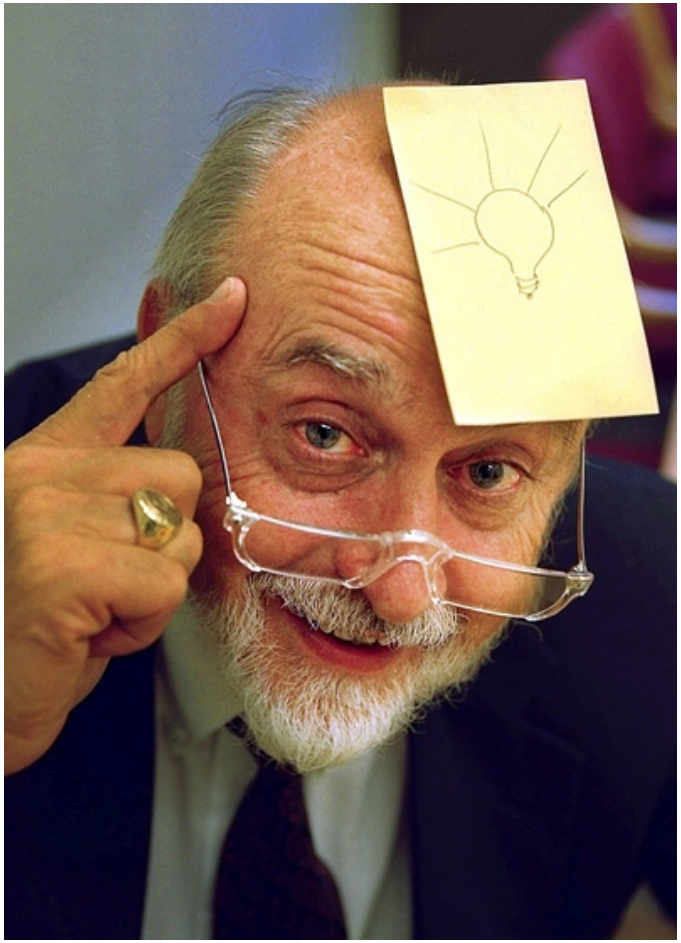
BGP Large Communities – Simple!

Martijn Schmidt
martijnschmidt@i3d.net
i3D.net – Performance Hosting

Source: https://xkcd.com/547/

i3D.NET PERFORMANCE HOSTING

BGP Communities – Sticky Notes!

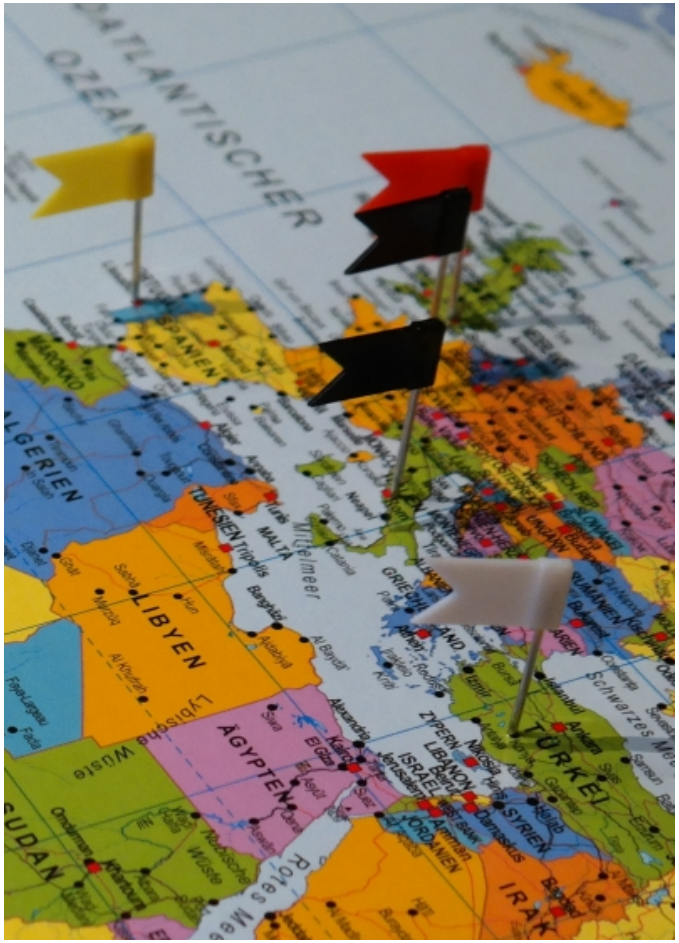


- “Sticky notes” for BGP routes
- Have been around for a long time!
  - First standardized as RFC1997 back in August 1996
  - Supported by nearly every BGP implementation in the wild
- Usually displayed as number:number
- Blank slate, user defines meaning
- May be forwarded to your friendly neighborhood network operator(s)
- Contains information or instructions
- Widely used in today’s routing policies

Source: https://commons.wikimedia.org/wiki/File:Fry-lightbulb-on-forehead1.jpg

i3D.NET PERFORMANCE HOSTING

BGP Communities: Information

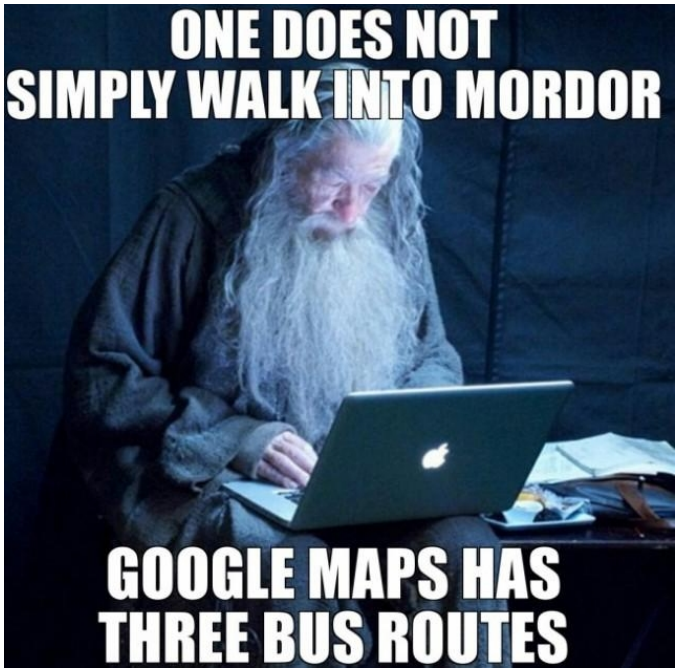


- Could be useful for anyone, if publicly documented in an easy-to-find place
- Location: where did the network learn this route?
- Relation: did I receive this route from my upstream, my peer, or my client?
- Security: complements (or replaces) a prefix-list in your route export policies: no assumptions about origin!

Source: https://www.pexels.com/photo/atlas-continent-country-creativity-269850/

i3D.NET PERFORMANCE HOSTING

BGP Communities: Instructions

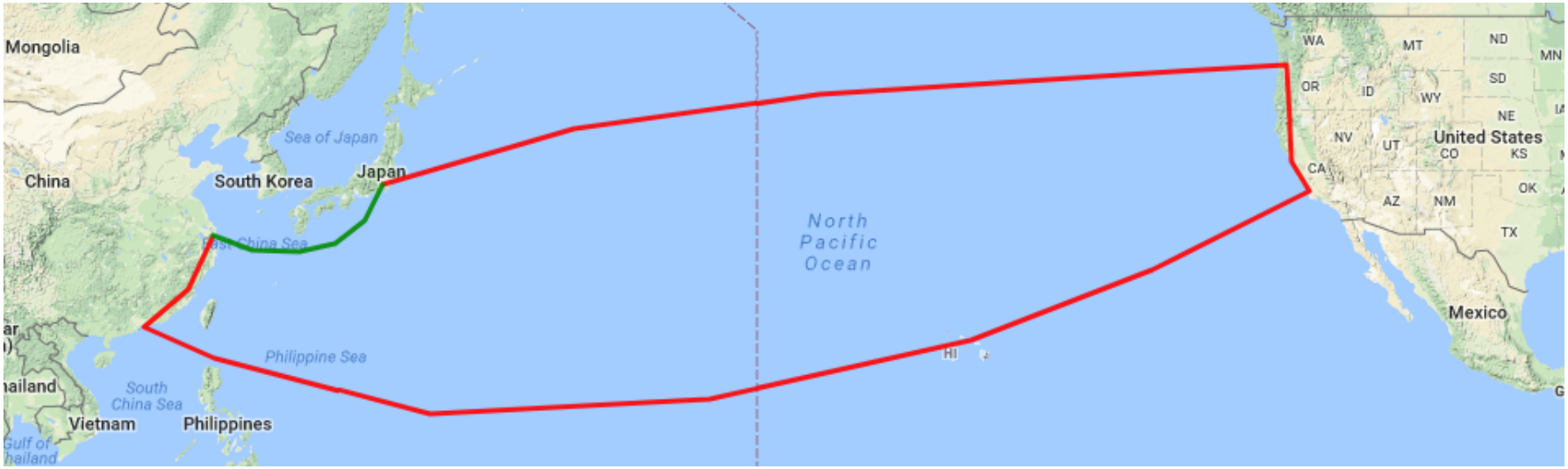


- Provides the ability to influence some BGP behaviour selectively, such as:
  - Request an artificial increase of your path length to “nudge” inbound traffic over another link
  - Request a router to stop sending your route to an adjacent network
  - Request a non-standard route preference in a remote network
- Trigger a network-wide blackhole for a target route: locally, or even upstream
- Networks are free to choose which policy features they (don’t) implement

Source: http://knowyourmeme.com/photos/720446-tech-support-gandalf

Path length: why “nudge” your traffic

- One of your upstream providers doesn’t interconnect with an important local network within your region, only in the EU or US
- BGP doesn’t know about latency, beautiful “scenic routing” ensues
- Artificially lengthen that path to avoid it if an alternative is available
- Selective – zero impact for upstream’s other interconnections



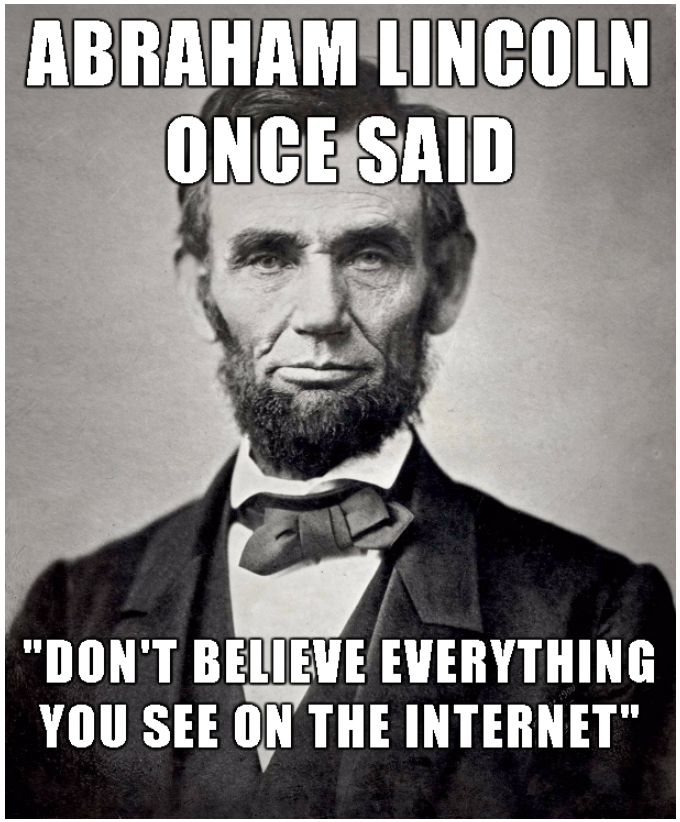
Source: http://www.gpsvisualizer.com/draw/

Path length: also useful in Europe

- Working around sub-optimal upstream backbone design – e.g. availability of a peering in Paris, but absence of direct backhaul
- ~9ms vs ~23ms will be noticed by latency-sensitive customers
- Previously equal path in Paris now looks longer due to “nudge”
- 3rd party then selects the shorter path through your other upstream

Source: http://www.gpsvisualizer.com/draw/

A word of caution: clean all the things!



- Never blindly accept BGP communities from external networks
  - Don’t risk providing “free transit” for non-customer networks that tag *their* routes with *your* label
  - Don’t let other networks blackhole (parts of) routes they didn’t register
- Extra consideration: after performing an action, strip the BGP community?
  - Pro: avoid risking that your action is repeated by an external policy
  - Con: useful when debugging why a route looks different than the rest

Source: http://imgur.com/gallery/DYyg3z2

BGP Communities: a few problems

- Notation of number:number for instructions means sacrificing either readability of target or unique scope
  - 3549:8160, don't send to 3356?
  - 65500:3356 seems clear enough
  - Should 2914 or 49544 do this?
  - Zero paths from 3356 via 49544?!
- BGP Wedgie! See also RFC4264
- Insufficient size for new ASNs because the maximum number is 65535
- Mapping larger ASNs to 64512-65534
  - Risk of collision, lacks consistency
  - Confuses humans, rather hack-like

Source: https://www.facebook.com/BuzzFeedVideo/videos/1631492713658271/

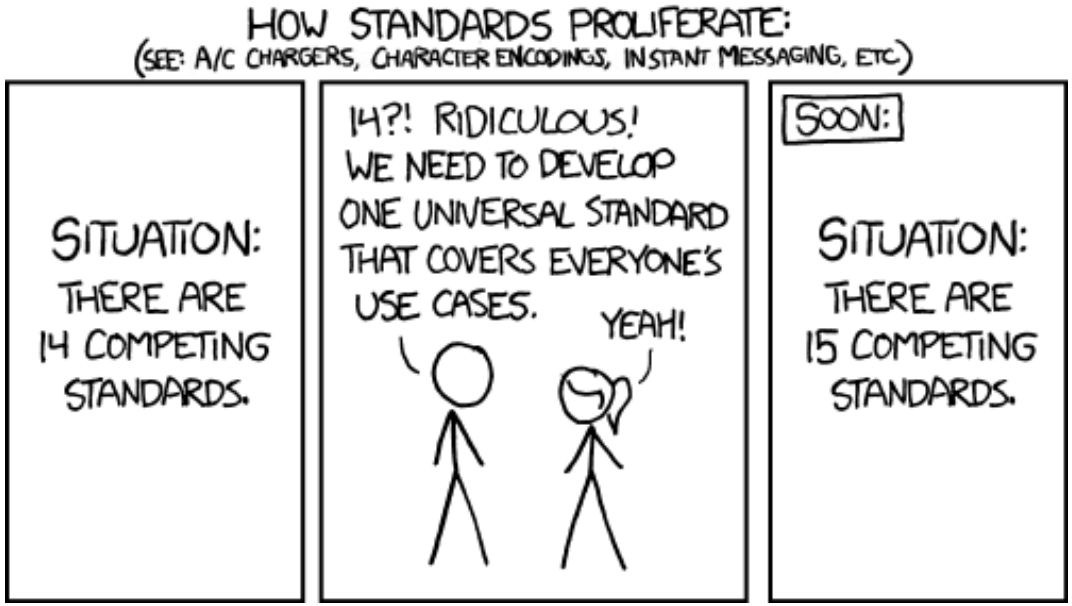
Communities: Attack of the Clones

- "Incomplete" solution defined in RFC4360 BGP Extended Communities
  - Still uses number:number notation
  - 1st number now sufficiently large
  - 2nd number still maximum 65535
- New solution was drafted by IETF IDR workgroup: BGP Wide Communities
  - First draft in 2010, still no RFC
  - No longer a "blank slate" design
  - Complicated, requires retraining for support teams and the community
  - Not implemented by major vendors

Source: http://www.imdb.com/title/tt0086190/

Large Communities: A New Hope

- "More of the same" solution in RFC8092 BGP Large Communities
  - Uses a simple "blank slate" number:number:number notation
  - Each number field goes up to 4294967295 – large enough
  - Place your ASN in the first number: no more collisions
  - Co-exists with legacy RFC1997 in your policies and routes



Source: https://xkcd.com/927/

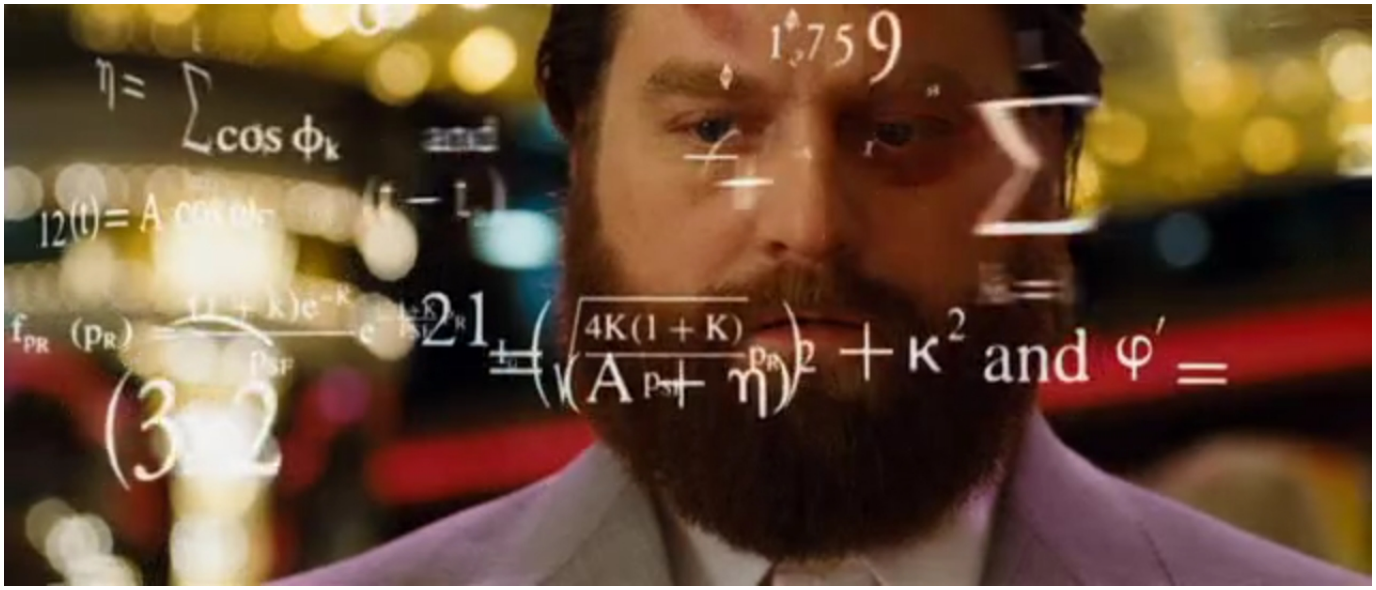
Large Communities: Companion RFC

- BGP Large Communities are still a blank slate, user defines meaning
- RFC8195 contains examples for users
  - “Use of BGP Large Communities”
  - Suggests ASN:function:parameter mapping for the three fields
  - Uses incremental numbering of the function field, though please define your own function numbers instead!
  - Networks are free to choose which features they (don’t) implement
- Try to keep things simple, use as many BGP Large Communities as you want

Source: https://static.giantbomb.com/uploads/original/0/7025/350438-companion_cube_720_credit.png

Keep it simple, stupid: Informational

- Old “regex style” communities like 49544:CTLLL can be confusing
  - Would mean: C: continent, T: type/relation, LLL: location code
- If you define a function identifier in your BGP Large Community, multiple informational codes won’t have overlapping values
- Don’t cram all your info into one parameter, just use multiple tags



Source: http://www.imdb.com/title/tt1119646/

i3D.NET PERFORMANCE HOSTING

Info: Whatever floats your boat!

- Networks can encode very detailed information in the larger fields
- Combine several location codes with a relation code, perhaps add internal classifications like a public “account number” or “circuit ID”
- For example, AS64497 might label routes learned from a customer in the Netherlands with 64497:1:528, 64497:2:150, and 64497:3:2
- Need more precision? USA has ANSI FIPS, Canada has SGC, etc

ISO 3166-1 Country ID		+	UN M.49 Region		+	Relation	
Large Community	Description		Large Community	Description		Large Community	Description
64497:1:528	Netherlands		64497:2:2	Africa		64497:3:1	Internal
64497:1:392	Japan		64497:2:9	Oceania		64497:3:2	Customer
64497:1:840	USA		64497:2:145	Western Asia		64497:3:3	Peering
			64497:2:150	Europe		64497:3:4	Transit

Source: Job Snijders’ & Greg Hankins’ slides at http://largebgpcommunities.net/talks/

Instructions: Some new possibilities

- Feel free to do things beyond what has been described in RFC8195
- Now also room to combine location-based and ASN-based actions
  - UN M.49 always contains 3 numbers, for example Europe 150
  - 2914:1150:6453, where 1150 contains action 1 & location 150
  - Requests from 2914, action prepend 1, location Europe, to 6453

Source: http://www.gpsvisualizer.com/draw/

i3D.NET
PERFORMANCE HOSTING

Blackholing: selectively, please!

- Dutch corporation may see 99% of its legitimate traffic originate in Europe
- Combine a blackhole function with a UN M.49 code to target region/country
  - For example, 49544:666:840 could ask 49544 to blackhole packets that were received in the USA
  - Must have backbone, consistent route announcements, remote ISP with hot potato routing policies
  - Doesn’t look at source IP, can’t spoof your way around this one!
- Deployed by NTT, KPN, but RFC1997

Source: https://www.nasa.gov/topics/universe/features/radio-particle-jets.html

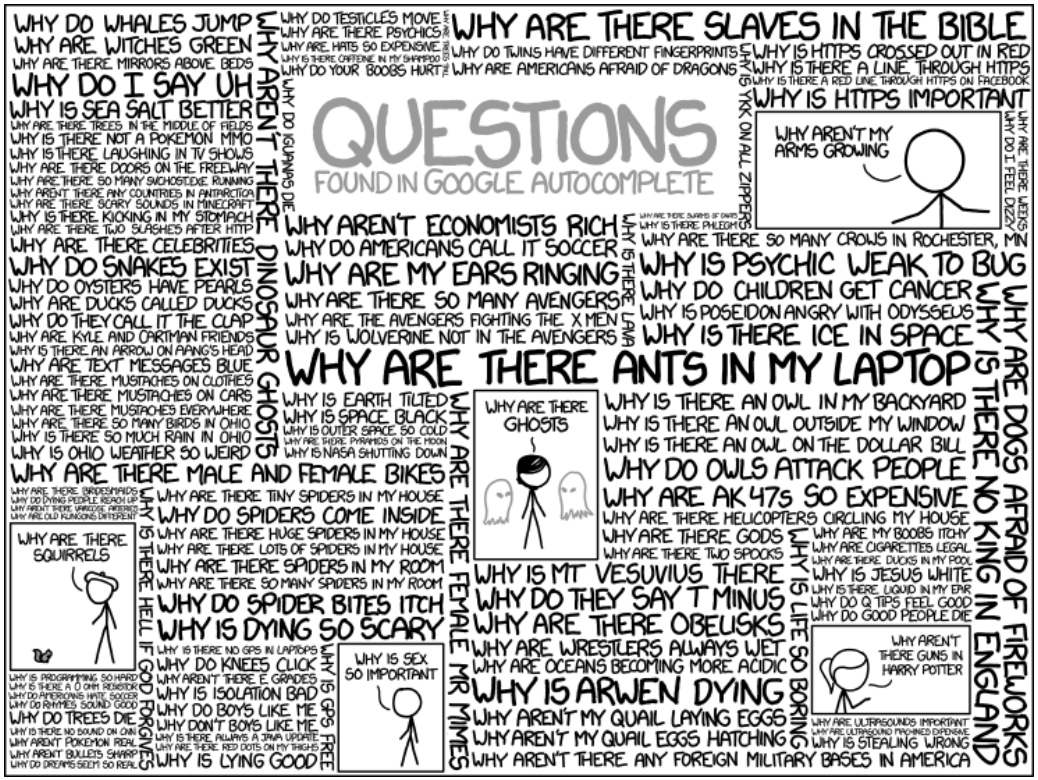
i3D.NET
PERFORMANCE HOSTING

Dear vendors: resistance is futile

- Implementations with running code
  - Cisco IOS XR beta, release 6.3.2?
  - Juniper JunOS, release 17.3R1
  - Nokia SR OS beta, release 16.0.R1
- Implementation committed / scheduled
  - Arista EOS, see BUG169446
  - Brocade IronWare, in 1H2018
  - Brocade SLX-OS, in 1H2018
  - Cisco IOS XE, release 16.9.1
- Open-source releases available for BIRD, ExaBGP, FRRouting, freeRouter, OpenBGPD, GoBGP, Quagga, etc
- More info at: http://goo.gl/TzzQUM

Source: http://www.imdb.com/title/tt0708785/

BGP Large Communities – Questions?



Martijn Schmidt
martijnschmidt@i3d.net
i3D.net – Performance Hosting

Source: https://xkcd.com/1256/