IPv6
a new protocol
a new routing table

LACNIC XI, May 29, 2008, Salvador, Brazil

Iljitsch van Beijnum

Sorry.

Today, we're out of IPv4 addresses.

Legend

- ★ Not usable
- ★ Given out to end-user
- ★ "Various registries"
- ★ RIPE NCC (Europe and more)
- ★ ARIN (North America)
- ★ APNIC (Asia, Australia and Pacific)
- ★ LACNIC (Latin America + Caribbean)
- ★ AfriNIC (Africa)

182 of 221 usable /8s given out, 82.4%

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111
112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

Not really out of v4 yet

- IANA still has 654 million (39 /8s)
- AfriNIC: 9 M unused of 17 M
- APNIC: 64 M unused of 470 M
- ARIN: 71 M unused of 487 M
- LACNIC: 39 M unused of 101 M
- RIPENCC: 39 M unused of 436 M
- Legacy: 180 M unused of 1544 M
- Total: 402 M + IANA = **1056** M (28.5%)

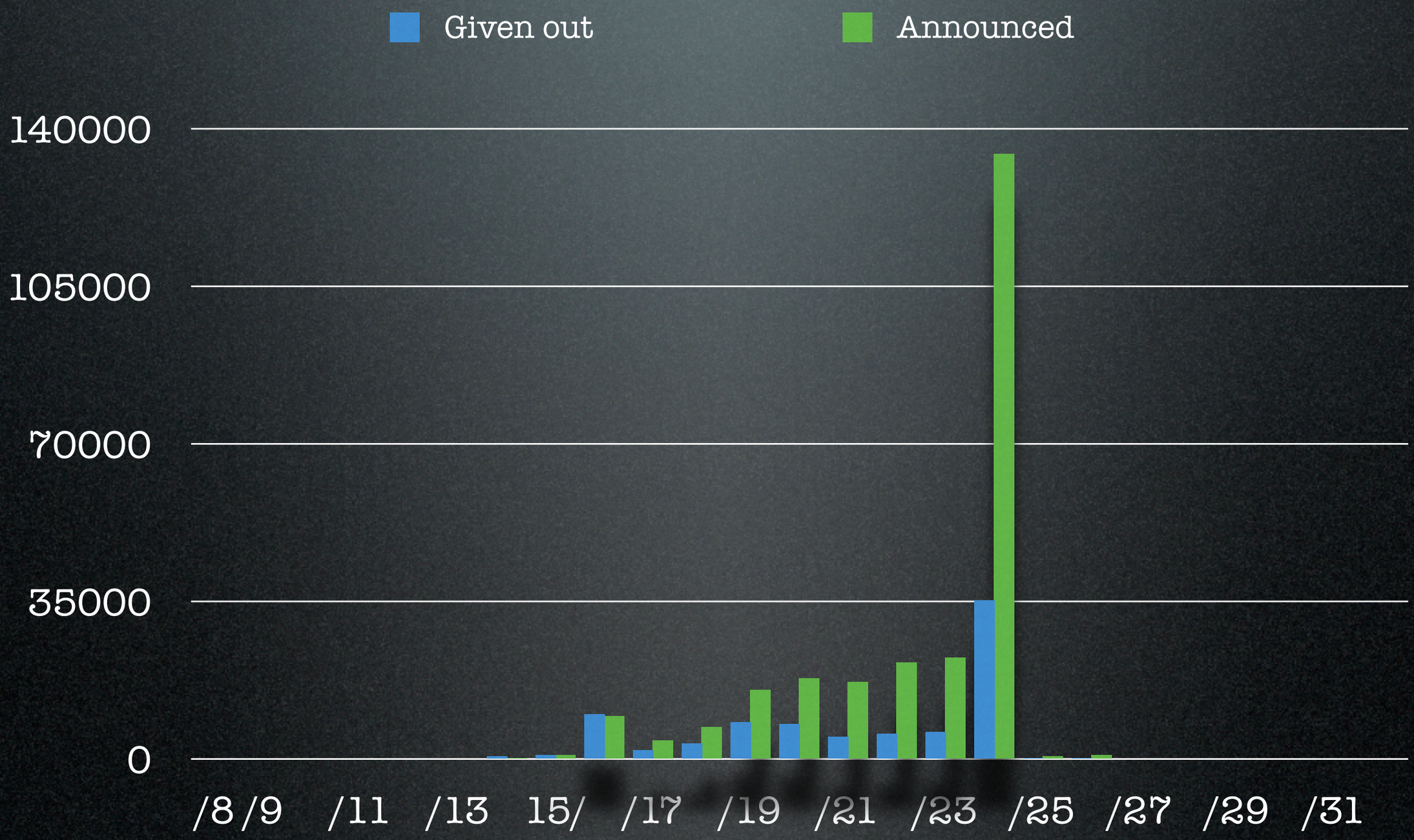
However...

- Geoff Huston says:
  - Projected RIR Unallocated Address Pool Exhaustion: 25-Nov-2011
    - That's 3.5 years from now!
- In the IETF, new work can easily take 3.5 years
- In government and even some businesses, too

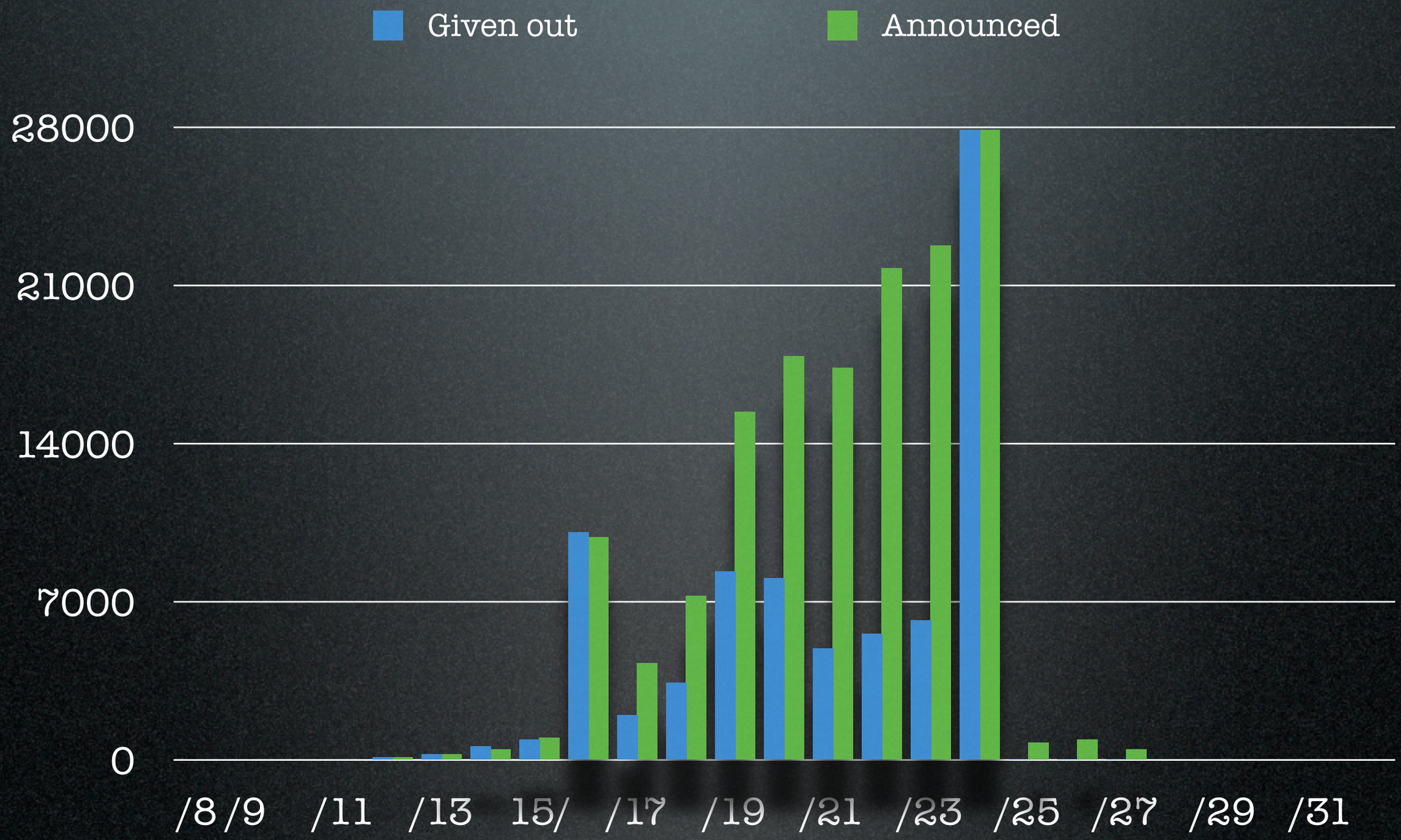
So if you want to do something new, plan on doing it without IPv4 addresses!

What does this mean for the routing tables?

IPv4 table


Given out
Announced
140000
105000
70000
35000
0
/8
/9
/11
/13
15/
/17
/19
/21
/23
/25
/27
/29
/31

IPv4 table (5 x zoom)



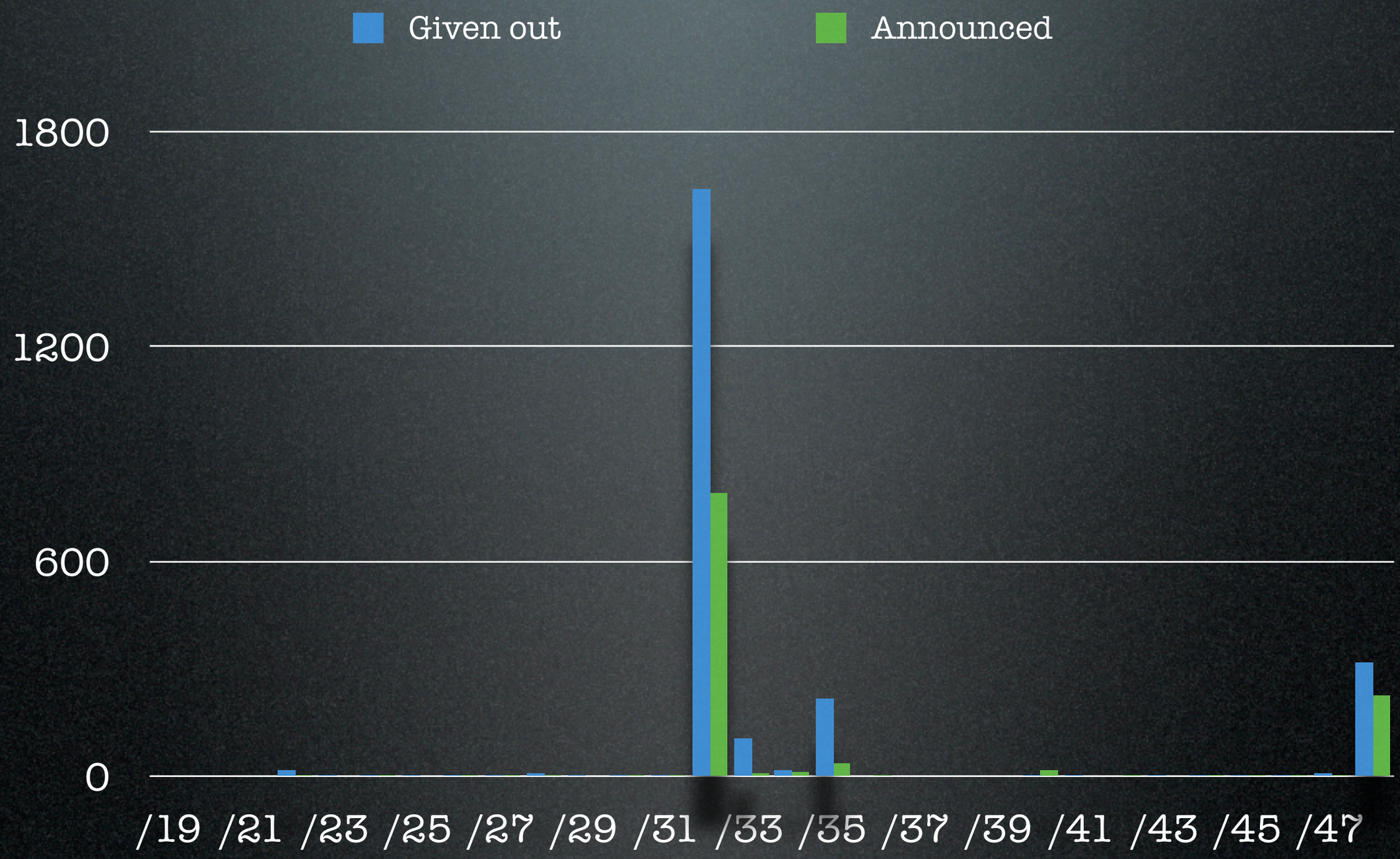
Addresses announced

- Total: 1870 million (256k pfx)
- < /8 319 million (19 pfx)
- / 9 - /15: 793 million (2003 pfx)
- = / 16: 650 million (9919 pfx)
- > /16: 108 million (244k pfx)
  - 95% of prefixes = 6% of address space

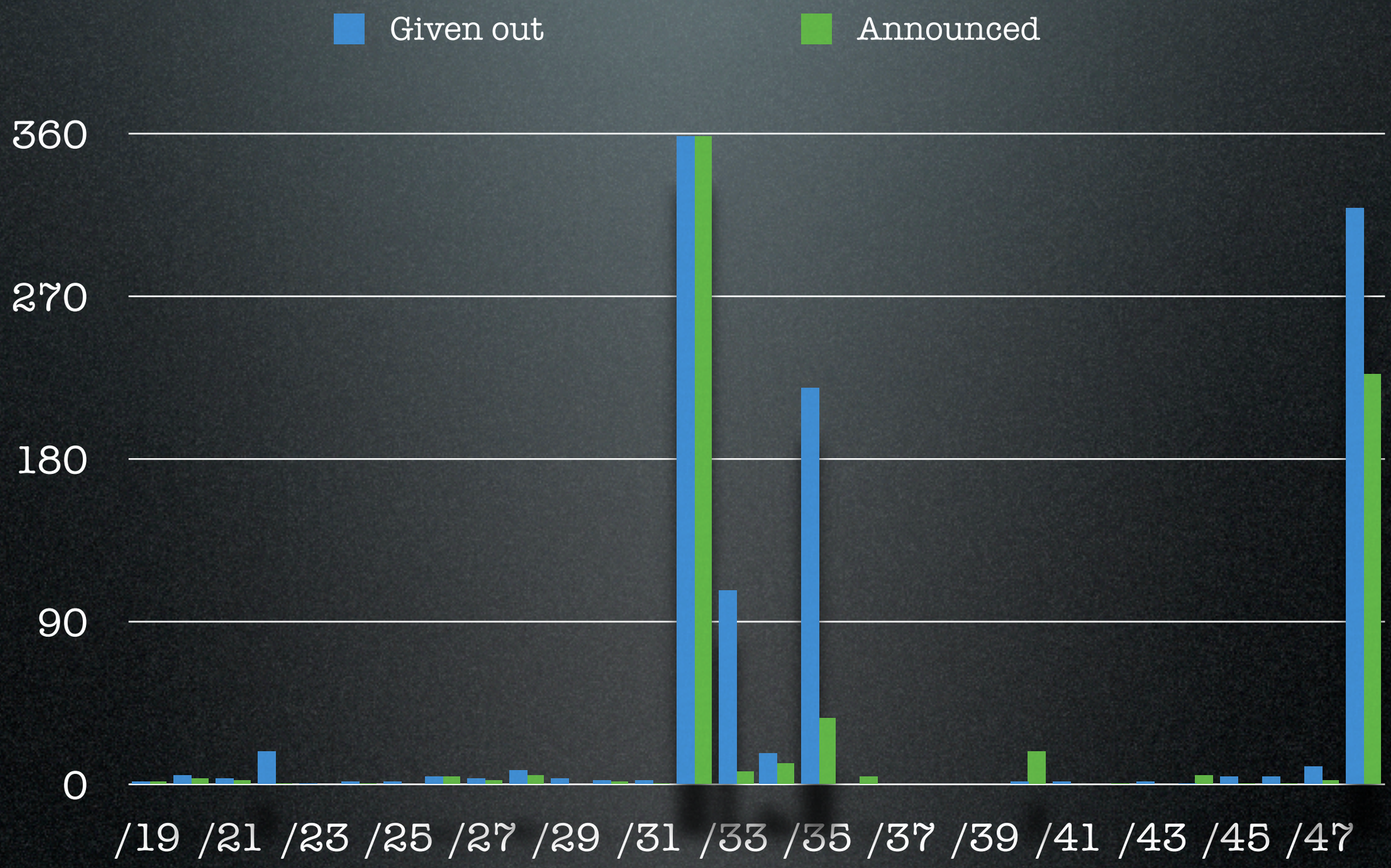
Allocation/assignment weirdness

```
mysql> select num, 32 - log(num)/log(2) as len1,
ceiling(32 - log(num+3)/log(2)) as len2, count(*) from
addrspace where type='ipv4' group by len1 order by len1
limit 12;
+----------+--------------------+------+----------+
| num      | len1               | len2 | count(*) |
+----------+--------------------+------+----------+
|      256 |                 24 |   24 |    35225 |
|      240 | 24.09310940043915 |   25 |        1 |
|      192 | 24.41503749927 88 |   25 |        1 |
|      140 |  24.870716983055 |   25 |        1 |
|      128 |                 25 |   25 |       82 |
|       96 | 25.41503749927 88 |   26 |        1 |
|       64 |                 26 |   26 |       53 |
|       48 | 26.41503749927 88 |   27 |        1 |
|       36 | 26.83007499855 77 |   27 |        1 |
|       32 |                 27 |   27 |       44 |
|       16 |                 28 |   28 |       11 |
|        8 |                 29 |   29 |        8 |
+----------+--------------------+------+----------+
```

IPv6 table


Given out
Announced
1800
1200
600
0
/19
/21
/23
/25
/27
/29
/31
/33
/35
/37
/39
/41
/43
/45
/47

IPv6 table (5 x zoom)



Reserving extra space

```
mysql> select num, country, type, descr, num from
addrspace where descr like "2620:0:%" limit 125, 10;
+------+---------+------+--------------+------+
| num  | country | type | descr        | num  |
+------+---------+------+--------------+------+
|   48 | US      | ipv6 | 2620:0:ba0:: |   48 |
|   48 | US      | ipv6 | 2620:0:bb0:: |   48 |
|   48 | US      | ipv6 | 2620:0:bc0:: |   48 |
|   48 | CA      | ipv6 | 2620:0:bd0:: |   48 |
|   48 | US      | ipv6 | 2620:0:be0:: |   48 |
|   48 | US      | ipv6 | 2620:0:bf0:: |   48 |
|   48 | US      | ipv6 | 2620:0:c00:: |   48 |
|   48 | US      | ipv6 | 2620:0:c10:: |   48 |
|   48 | US      | ipv6 | 2620:0:c20:: |   48 |
|   48 | US      | ipv6 | 2620:0:c30:: |   48 |
+------+---------+------+--------------+------+
```

- ARIN reserves a /44 per assigned /48!

Why is the IPv6 table so different?

- No classful legacy
- Global agreement on /32 and /48 minimums
- More space, so ISPs can get a very large block immidately rather than return for more regularly
  - or at least, few people have grown out of their first block yet

Filtering the IPv6 routing table

Purposes of filtering

1. Reject unallocated space
   - used by hard-to-trace spam runs
2. Protect against routing table explosion from (accidental) deaggregation
   - /16 → 256 /24s is fairly common
   - /32 → 65536 /48s would be **deadly**

Filtering strategies

1. No filtering: life is good. Until it isn't. ☺
2. Reject /64s: catches almost nothing, proects against almost nothing
3. Reject > /48s: catches very little, protects against very litle
4. Reject > /32s: would work, except many legitimate /48s

More complex filtering

5. Filter > /32 from /32 space, > / 48 from / 48 space
    - separation isn't tight enough and reservations get in the way
6. Filter based on allocations/assignments
    - still doable today, probably not forever
7. Filter based on address certificates

Get rid of reservations!

- IPv4 and /35 → /32 change shows many people use both old and new, don't grow
  - so no gain, but can't make tight filters
  - A /48 + /44 is better than growing a /48 into a reserved /44
  - Even giving everyone /44 is better
  - What we need to do:
    - /48s come from /48 block, /47s from /47 block, /46s from /46 and so on

IPv6 multihoming

Multihoming

- Connect to two or more ISPs
  - if one link or ISP fails, still connected through the other
    - advantages: better uptime, easy to switch ISPs "make before break"
  - Traditionally done by getting AS number and provider independent (PI) prefix, then run BGP

Multihoming in IPv6

- For a long time, 6bone routing guidelines wouldn't allow PI space
- What if by 2050, 10% of 10 billion people wants to multihome?
  - internet increasingly important!
  - 1000000000 prefixes in routing table
- IETF started work on "scalable multihoming in IPv6" (multi6 wg)

multi6 → shim6

- After many years, multi6 chose an approach and became shim6:
  - get regular provider aggregatable (PA) address space from **each** ISP
  - negotiate extra addresses with remote system
  - switch to other addressses when current addresses stop working

shim6 limitations

- **Both ends** must support shim6 to gain multihoming benefits
- Can only protect ongoing communication
  - new sessions must try all addresses until one works
- (Currently) no mechanisms for central traffic engineering control, hosts make their own decisions

Reactions

- Some ISPs don't like it
  - strange, they don't have to run it!
    - like centrally assigned unique site local (cULA) addresses...
- Enterprises want PI anyway
- PI now possible in most regions

Current status

- Shim6 documents close to being published as RFCs
- A few preliminary implementations
- Little excitement now that IPv6 PI is possible
- Internet Research Task Force routing research group is taking on the routing scalability issue

The routing tables as we run out of IPv4 address space

The easy stuff

- IPv6 table will grow, from 1145 prefixes and less than 1000 ASes to...?
  - with 1.2 prefixes per AS, little cause for concern in the short-to-medium term
    - some people think IPv4 more specifics will also appear in IPv6, though...
- IPv4 table growth will continue until about a year after depletion based on current factors

The hard stuff

- People will find ways to keep using IPv4 after the moment of depletion
- One scenario is that rather than get a big block from a RIR, people will scrape together a lot of small blocks
  - this would be bad for the IPv4 routing table...

Small block explosion

- I don't think this is very likely
  - see yesterday's talk by Alain Durand: this is expensive and a lot of work
    - in addition: it gets worse every year
    - using existing space more intensively (more users behind one address with NAT) or IPv6 is more attractive
- **But it could happen!**

Normal post-deplection growth

- Address space is returned when people go out of business
  - will probably be given out as several smaller blocks then
- Trading will also be smaller blocks
- So small block growth will largely continue after depletion
- And of course more deaggregation

What to do?

- Migrate to IPv6 before the IPv4 tables (may) explode...
- Think whether you really need a full IPv4 BGP feed, or if a partial/filtered feed is sufficient
  - (but then you need a default route)

How to get rid of IPv4

- How can I move to IPv6 if the rest of the world is still IPv4-only?
- NAT-PT: network address translation / protocol translation (RFC 2766)
- This allows IPv6 hosts to access IPv4 services
- However, "deprected" by the IETF last year (RFC 4966)

Reviving NAT-PT / NAT64

- Some of us trying to address issues (mostly DNS-related) with NAT-PT
- Tomorrow morning a chance to try (existing) NAT-PT out for yourself!
  - I'll explain the requirements document that we wrote and ask for your feedback
- (Note that you still need some IPv4...)

Participating in the IETF

The IETF can use your help!

- IETF, huh?
  - Internet Engineering Task Force
    - (the people who write all those RFCs)
  - IETF standards, requirements, best practices and informational documents are **very** important to the future of the internet

Participating

- "We reject: kings, presidents and voting. We believe in rough consensus and running code."
- Participation is easy:
  - just join the mailinglists for the working groups you're interested in
  - three meetings per year, but no requirement to attend

How it works

- When you jump in, you will feel lost
  - highly technical discussions already going on hard to follow for newcomers
  - Just look for new discussions and join those
  - Especially keep an eye out for requirements discussions, these are higher-level, user/operator feedback is especially appreciated here

Questions?

- I'll be here today and tomorrow
- Have a look at:
  - www.bgpexpert.com
  - http://www.bgpexpert.com/ianaglobalpool.php
  - http://www.bgpexpert.com/addressespercountry.php
  - http://www.bgpexpert.com/ipv6addressespercountry.php