# The Tier1, LHCONE and LHCOPN

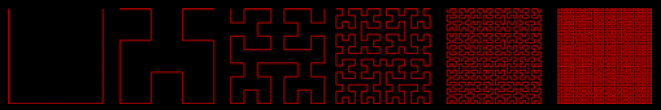
2019-04-24

James Adams
STFC, UKRI

# Background

- The RAL Tier 1 currently connects to:
  - The world
    - via Janet[1] via the RAL Campus Network
  - Tier 0s and other Tier 1s
    - via the LHCOPN[2] via a private router (OPNR).
  - Tier 2s
    - via Janet[1] via a private router (OPNR)
- Real-Soon-Now™ we want to be connecting to:
  - Tier 2s
    - via LHCONE[3] (via …some stuff…)

NB. Throughout this talk I will be displaying IPv4 space on a 12th order Hilbert Curve (ala https://xkcd.com/195/), these maps have been rendered with a modified version of ipv4-heatmap by Duane Wessels of The Measurement Factory (https://github.com/jrha/ipv4-heatmap ).



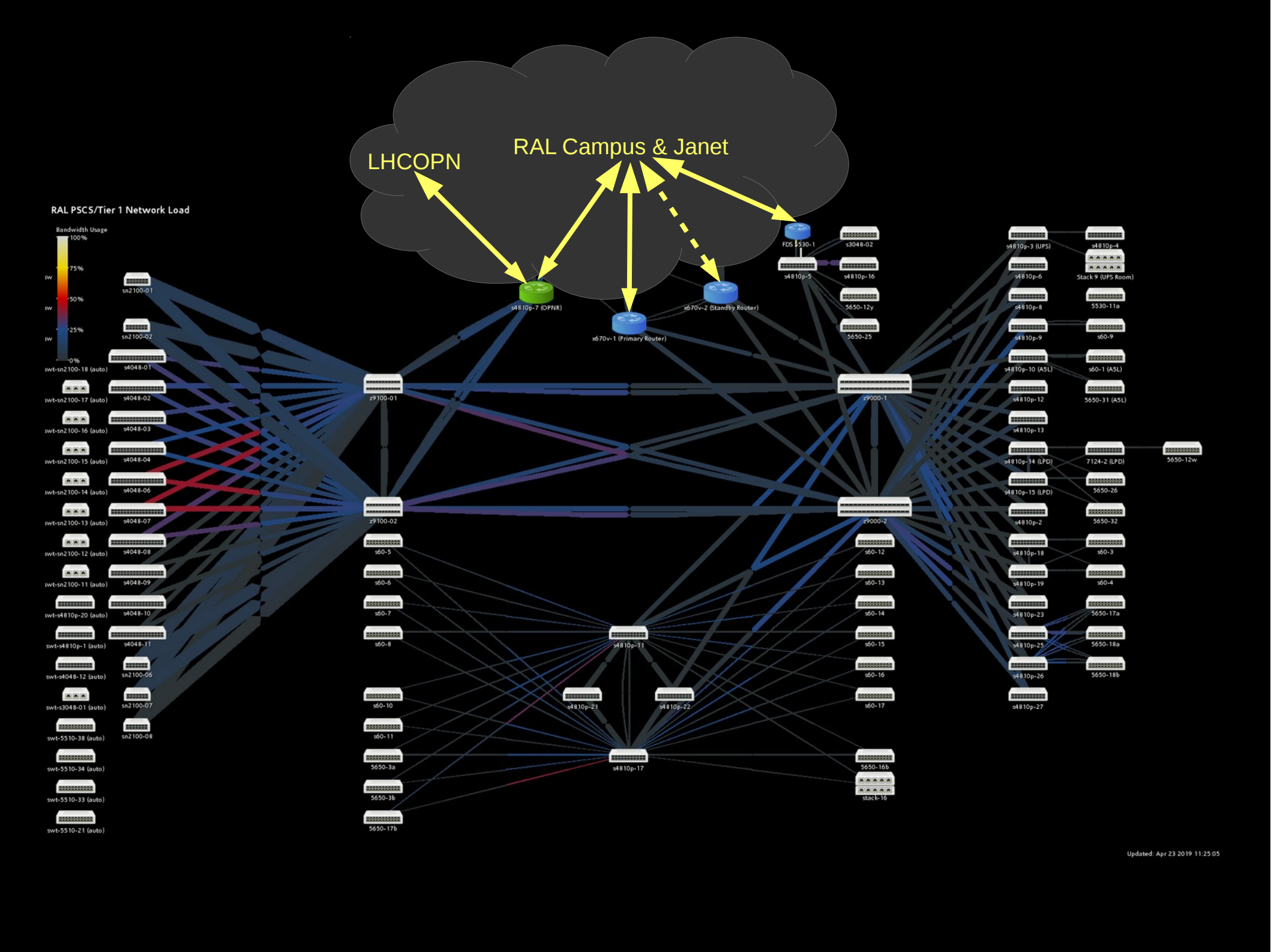
[1] https://www.jisc.ac.uk/janet

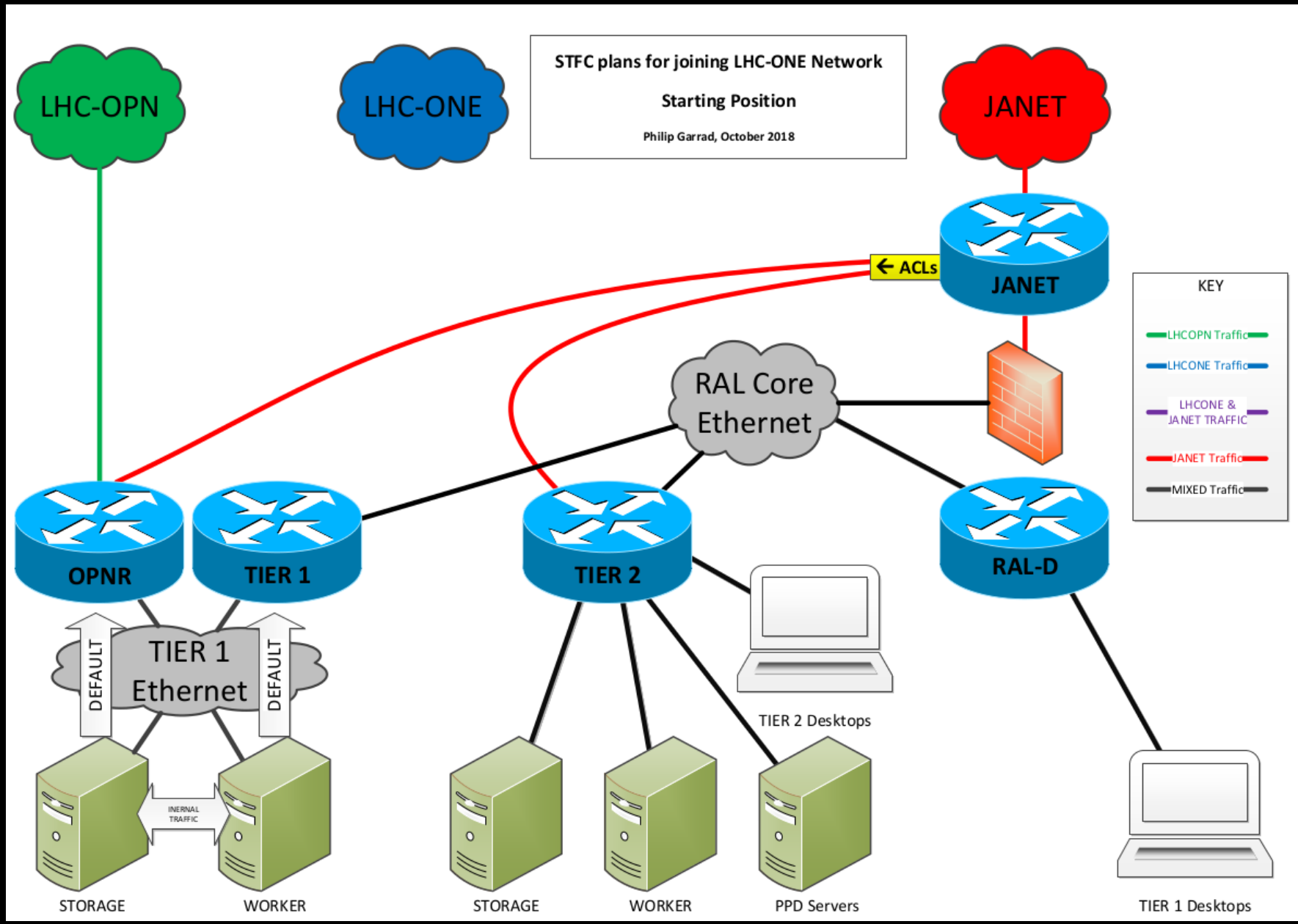
[2] Large Hadron Collider Optical Private Network http://lhcopn.web.cern.ch/lhcopn/

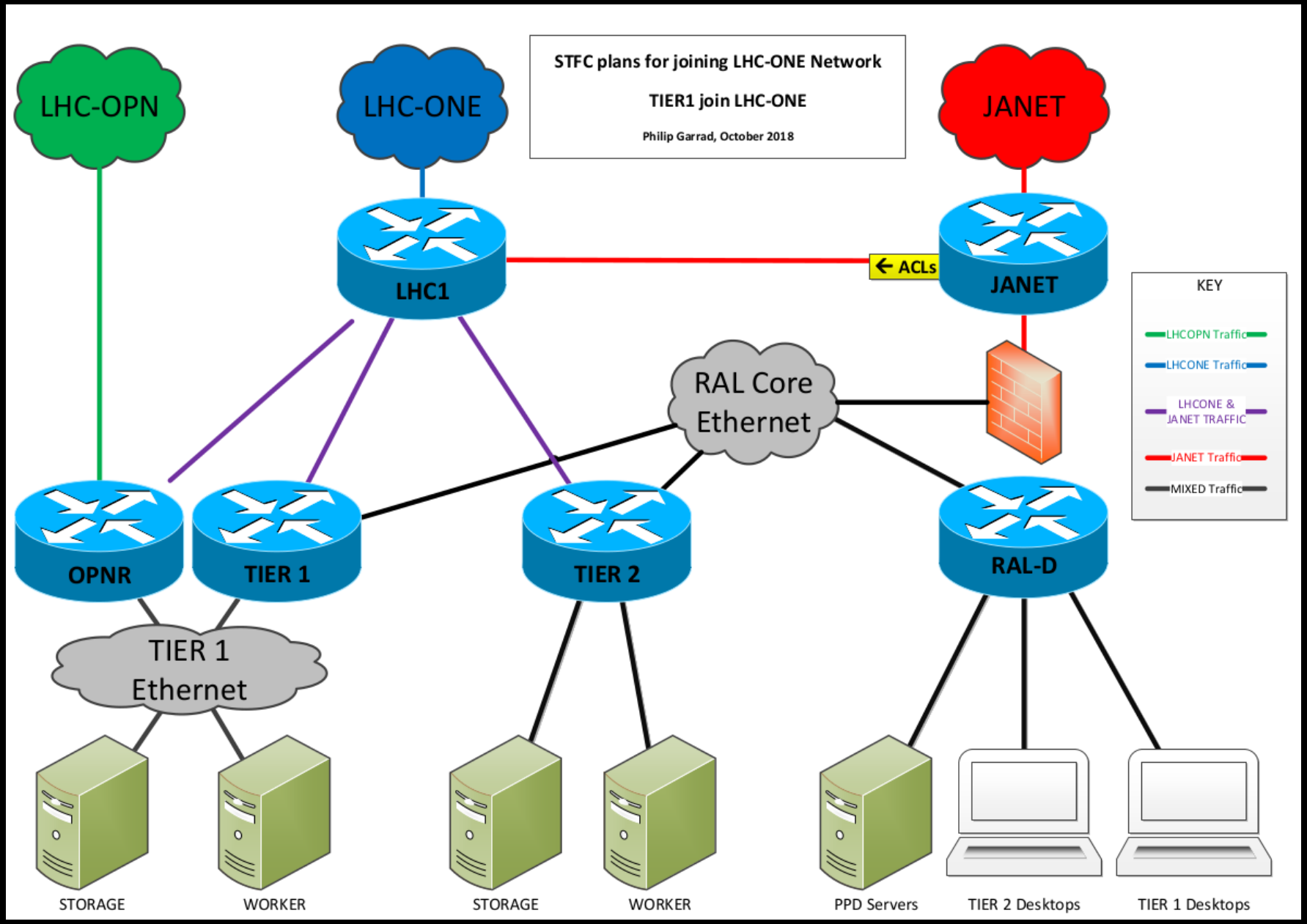
[3] Large Hadron Collider Open Network Environment http://lhcone.web.cern.ch/

# More Background

- Currently the Tier 1 provides compute, storage and services over both IPv4 and IPv6 on a single L2 segment.
  - 3× IPv4 subnets (one for LHCOPN)
  - 2× IPv6 subnets (one for LHCOPN)
- Routing to deal with this is a little arcane…
  - 3+1 physical routers
    - ~7 virtual routers
  - Nodes have ~16 IPv4 and ~8 IPv6 routing table entries
    - More default route (gateway) options than subnets


RAL PSCS/Tier 1 Network Load
Bandwidth Usage
100%
75%
50%
25%
0%
LHCOPN
RAL Campus & Janet
s4810p-7 (OPNR)
x670v-1 (Primary Router)
x670v-2 (Standby Router)
FDS 5530-1
s4810p-5
s3048-02
s4810p-16
5650-12y
5650-25
sn2100-01
sn2100-02
s4048-01
s4048-02
s4048-03
s4048-04
s4048-06
s4048-07
s4048-08
s4048-09
s4048-10
s4048-11
sn2100-06
sn2100-07
sn2100-08
swt-sn2100-18 (auto)
swt-sn2100-17 (auto)
swt-sn2100-16 (auto)
swt-sn2100-15 (auto)
swt-sn2100-14 (auto)
swt-sn2100-13 (auto)
swt-sn2100-12 (auto)
swt-sn2100-11 (auto)
swt-s4810p-20 (auto)
swt-s4810p-1 (auto)
swt-s4048-12 (auto)
swt-s3048-01 (auto)
swt-5510-38 (auto)
swt-5510-34 (auto)
swt-5510-33 (auto)
swt-5510-21 (auto)
z9100-01
z9100-02
z9000-1
z9000-2
s60-5
s60-6
s60-7
s60-8
s60-10
s60-11
5650-3a
5650-3b
5650-17b
s4810p-11
s4810p-21
s4810p-22
s4810p-17
s60-12
s60-13
s60-14
s60-15
s60-16
s60-17
5650-16b
stack-16
s4810p-3 (UPS)
s4810p-6
s4810p-8
s4810p-9
s4810p-10 (A5L)
s4810p-12
s4810p-13
s4810p-14 (LPD)
s4810p-15 (LPD)
s4810p-2
s4810p-18
s4810p-19
s4810p-23
s4810p-25
s4810p-26
s4810p-27
s4810p-4
Stack 9 (UPS Room)
5530-11a
s60-9
s60-1 (A5L)
5650-31 (A5L)
7124-2 (LPD)
5650-26
5650-32
s60-3
s60-4
5650-17a
5650-18a
5650-18b
5650-12w
Updated: Apr 23 2019 11:25:05

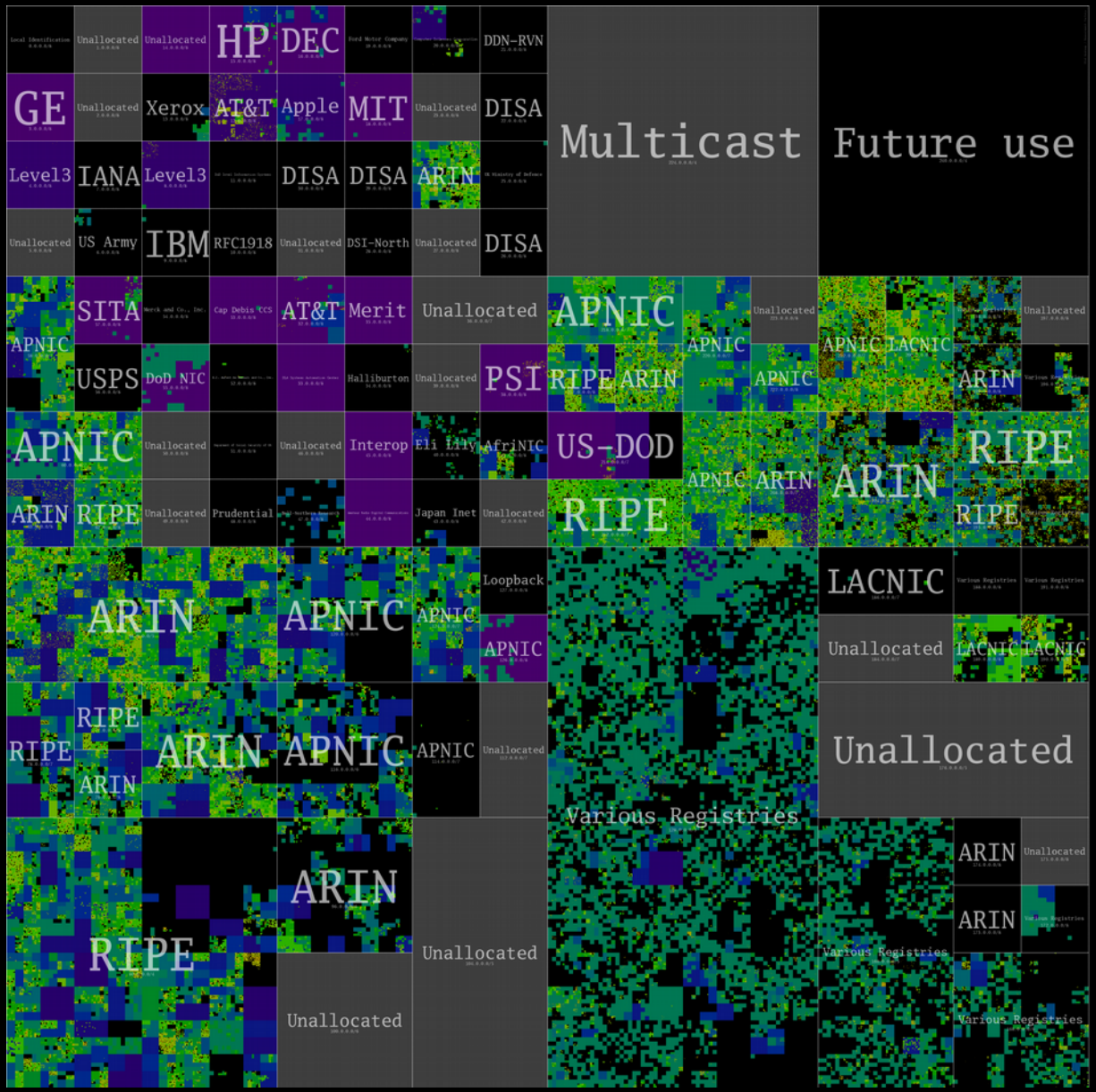

LHC-OPN
LHC-ONE
STFC plans for joining LHC-ONE Network
Starting Position
Philip Garrad, October 2018
JANET
ACLs
JANET
KEY
LHCOPN Traffic
LHCONE Traffic
LHCONE & JANET TRAFFIC
JANET Traffic
MIXED Traffic
RAL Core Ethernet
OPNR
TIER 1
TIER 2
RAL-D
DEFAULT
TIER 1 Ethernet
DEFAULT
INERNAL TRAFFIC
TIER 2 Desktops
STORAGE
WORKER
STORAGE
WORKER
PPD Servers
TIER 1 Desktops


STFC plans for joining LHC-ONE Network
TIER1 join LHC-ONE
Philip Garrad, October 2018
LHC-OPN
LHC-ONE
JANET
LHC1
ACLs
JANET
KEY
LHCOPN Traffic
LHCONE Traffic
LHCONE & JANET TRAFFIC
JANET Traffic
MIXED Traffic
RAL Core Ethernet
OPNR
TIER 1
TIER 2
RAL-D
TIER 1 Ethernet
STORAGE
WORKER
STORAGE
WORKER
PPD Servers
TIER 2 Desktops
TIER 1 Desktops

# It’s complicated…

- Clearly this could benefit from some simplification.
  - But adding more routers and routes to join LHCONE will make things even more complicated.
- GOAL: Tier 1 services should be consistent.
  - i.e. everything should have the same access to:
    - LHCOPN
    - LHCONE
    - Janet
- Best possible option would be to declare:
  - A single IPv4 subnet to the outside world.
    - Non trivial… two however…?
  - A single IPv6 subnet to the outside world.
    - Can do this now!
  - Lets just look at IPv4 for now…

# The IPv4 Internet



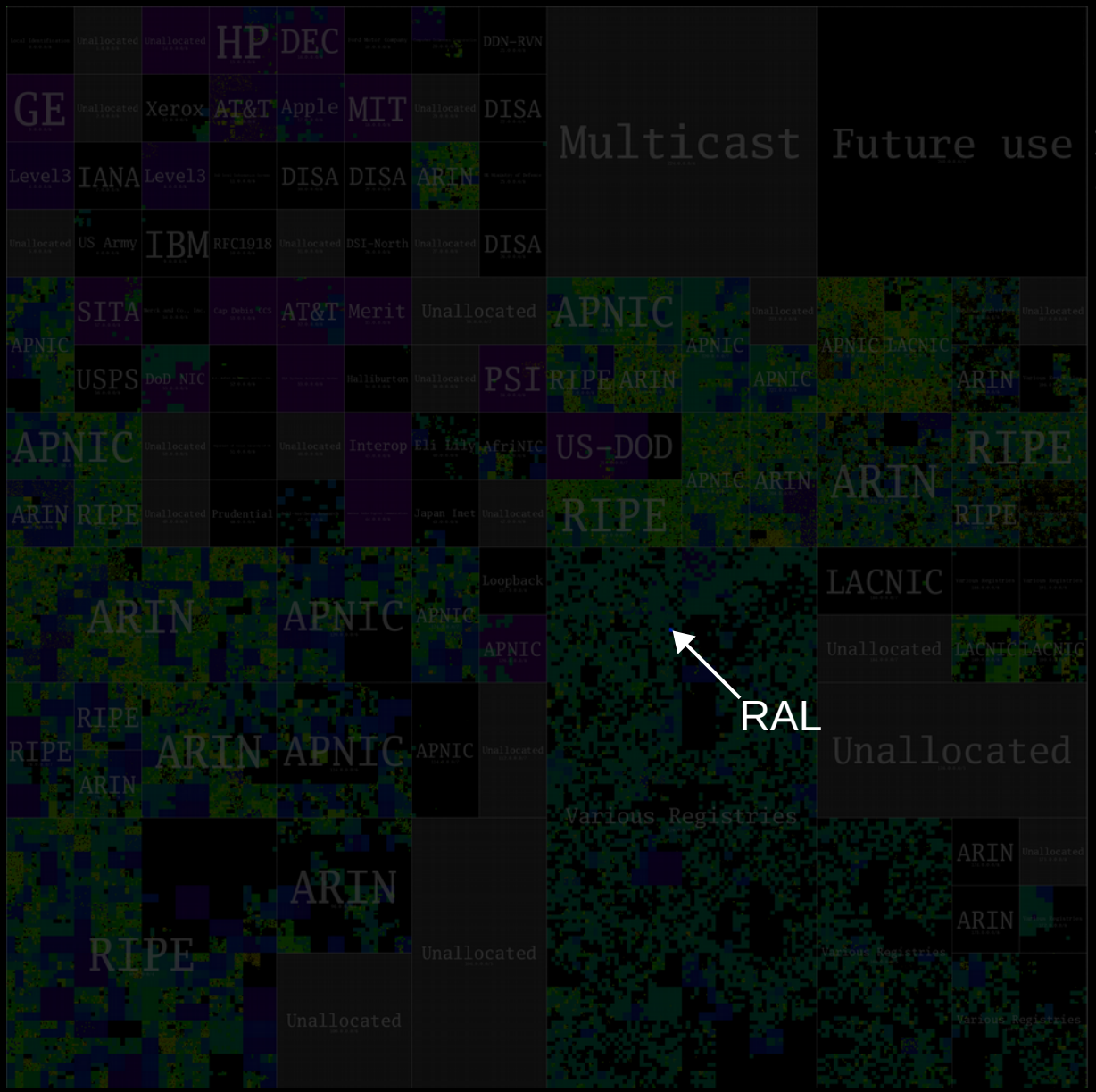

IPv4 BGP Routing Table 2008-03-01

Announcement Size

/8
/12
/16
/20
/24
/28
/32

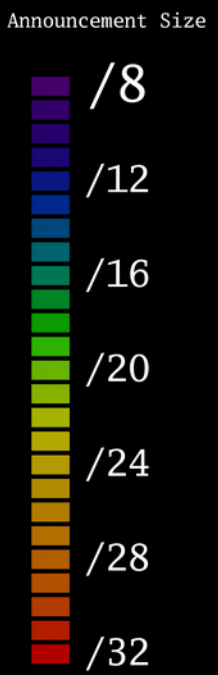
http://maps.measurement-factory.com/

# The IPv4 Internet



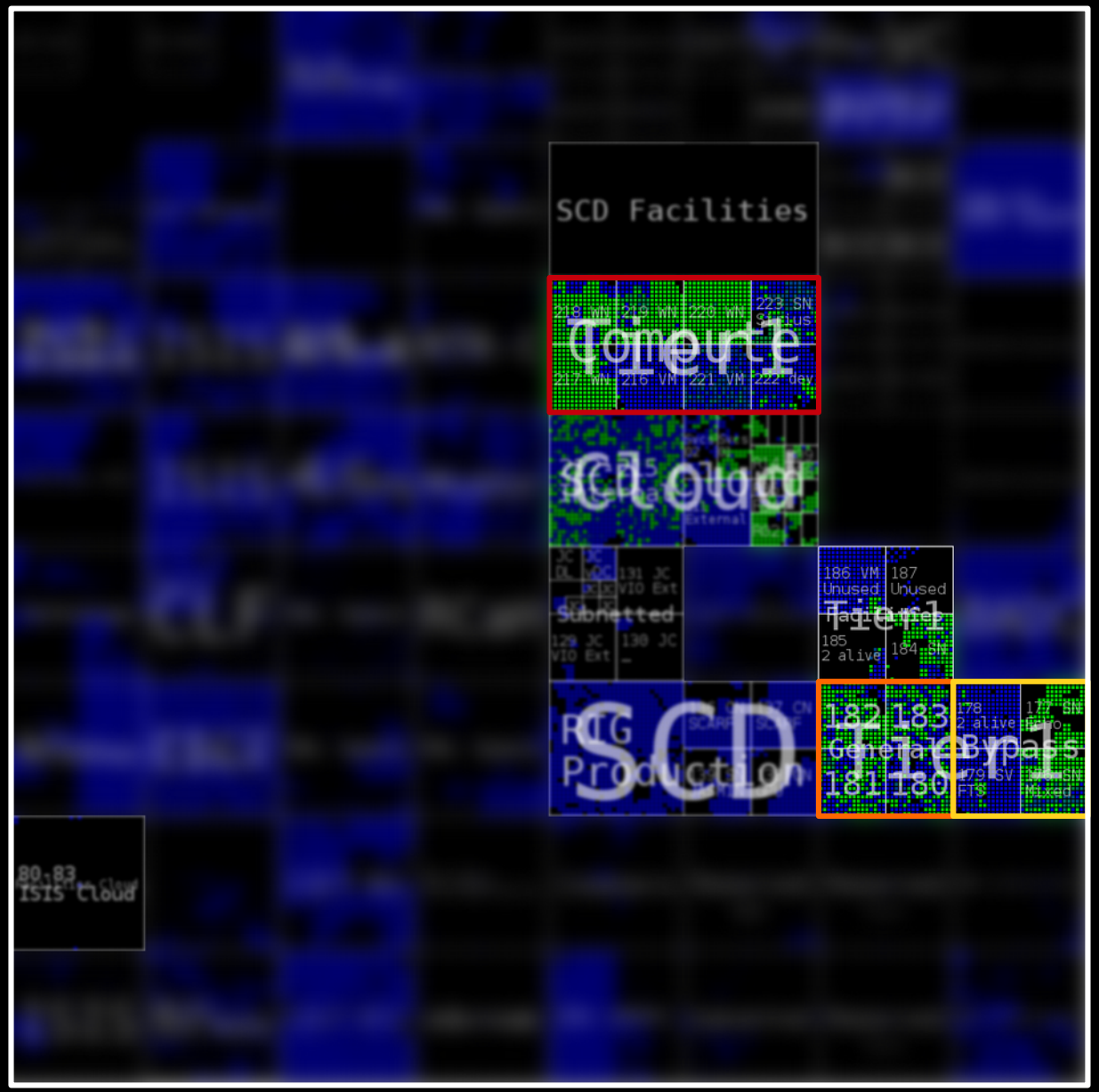

IPv4 BGP
Routing Table
2008-03-01

Zoom and enhance…



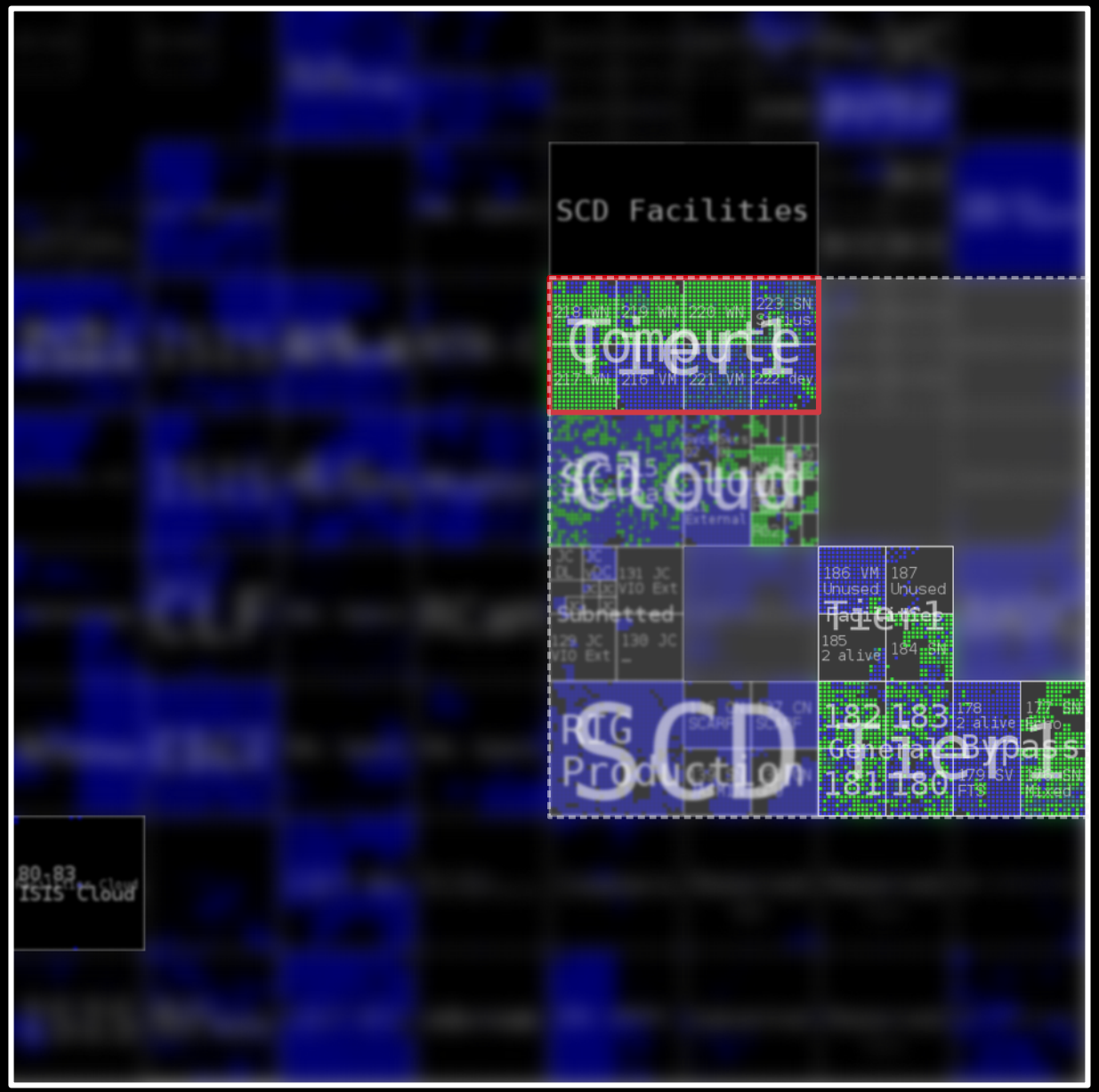

http://maps.measurement-factory.com/

# RAL, aka 130.246.0.0/16



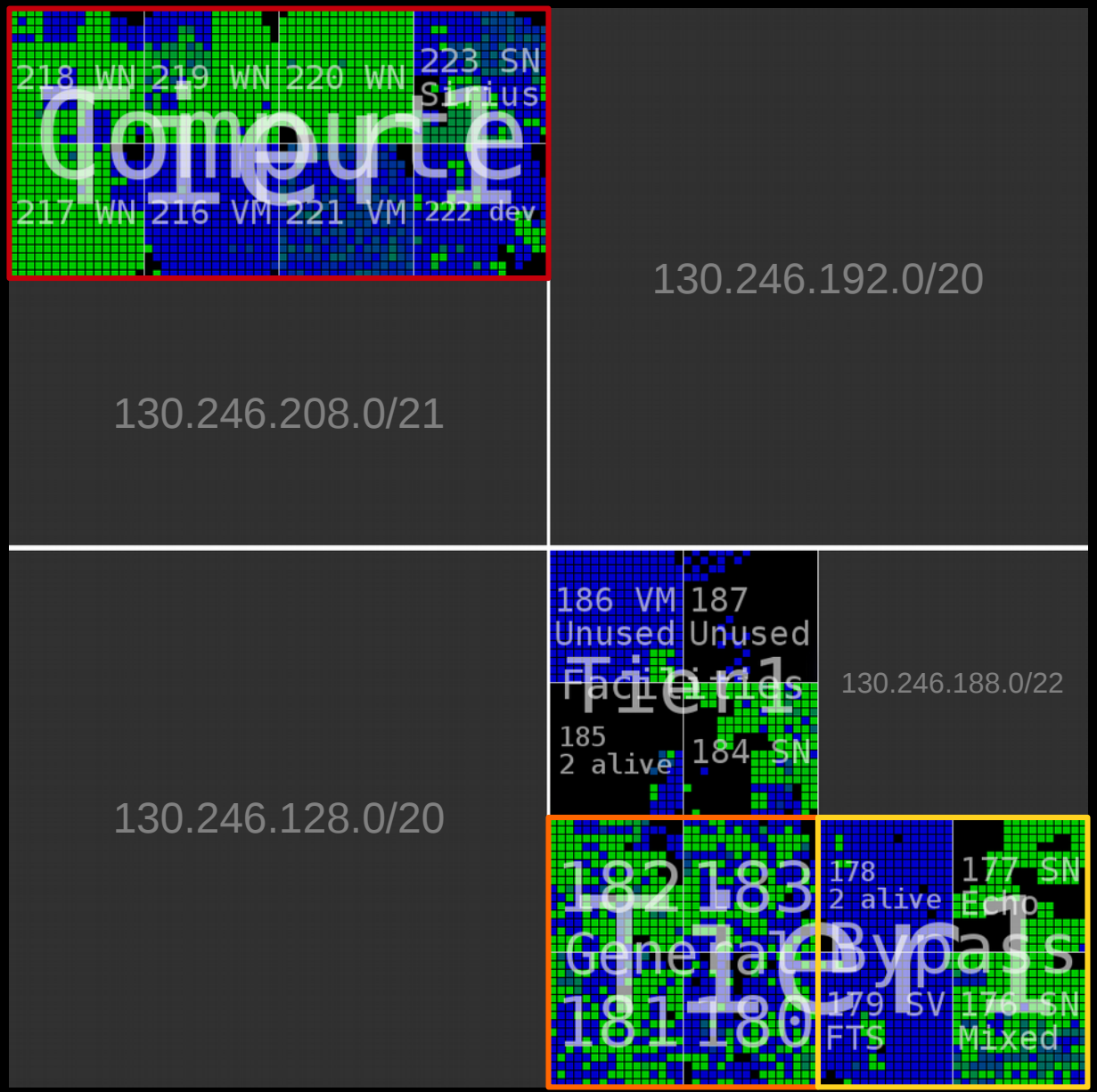

- SCD manages ~20% of RAL IPv4 space
- 3 Tier 1 subnets:
  - OPN: 130.246.176.0/22
  - Services: 130.246.180.0/22
  - Compute: 130.246.216.0/21
- Also manage
  - Facilities Services 130.246.184.0/22

# RAL, aka 130.246.0.0/16



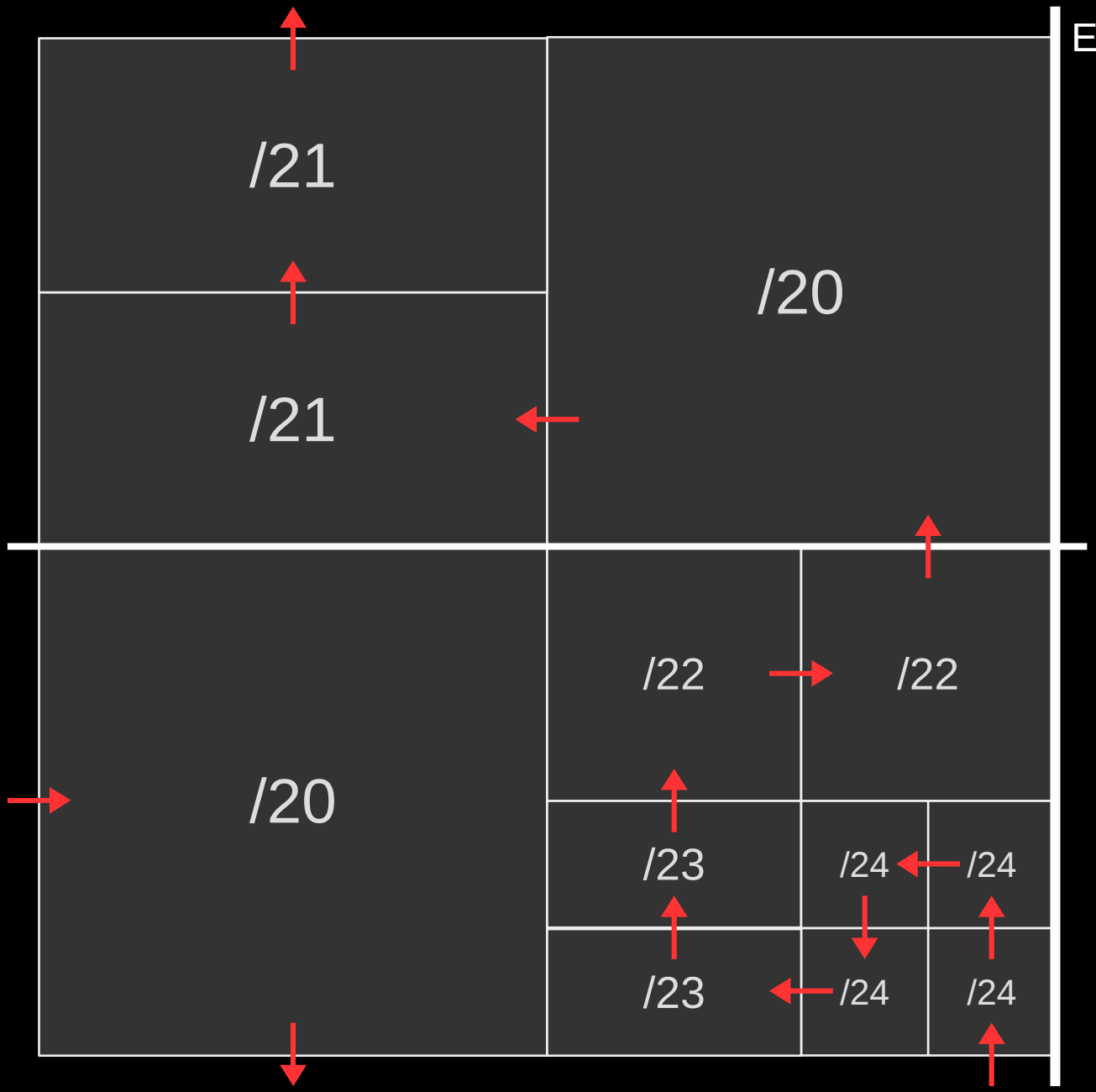

Zoom and enhance…

# Tier 1 Up Close




- Define a supernet of all subnets?
  - 130.246.0.0/16
- Surrounded by other projects.
  - Could start horse trading, but very slow.
- Will have to move hosts.
  - Cannot move Facilities Services.
  - Hard to move storage nodes.

# Reading Guide



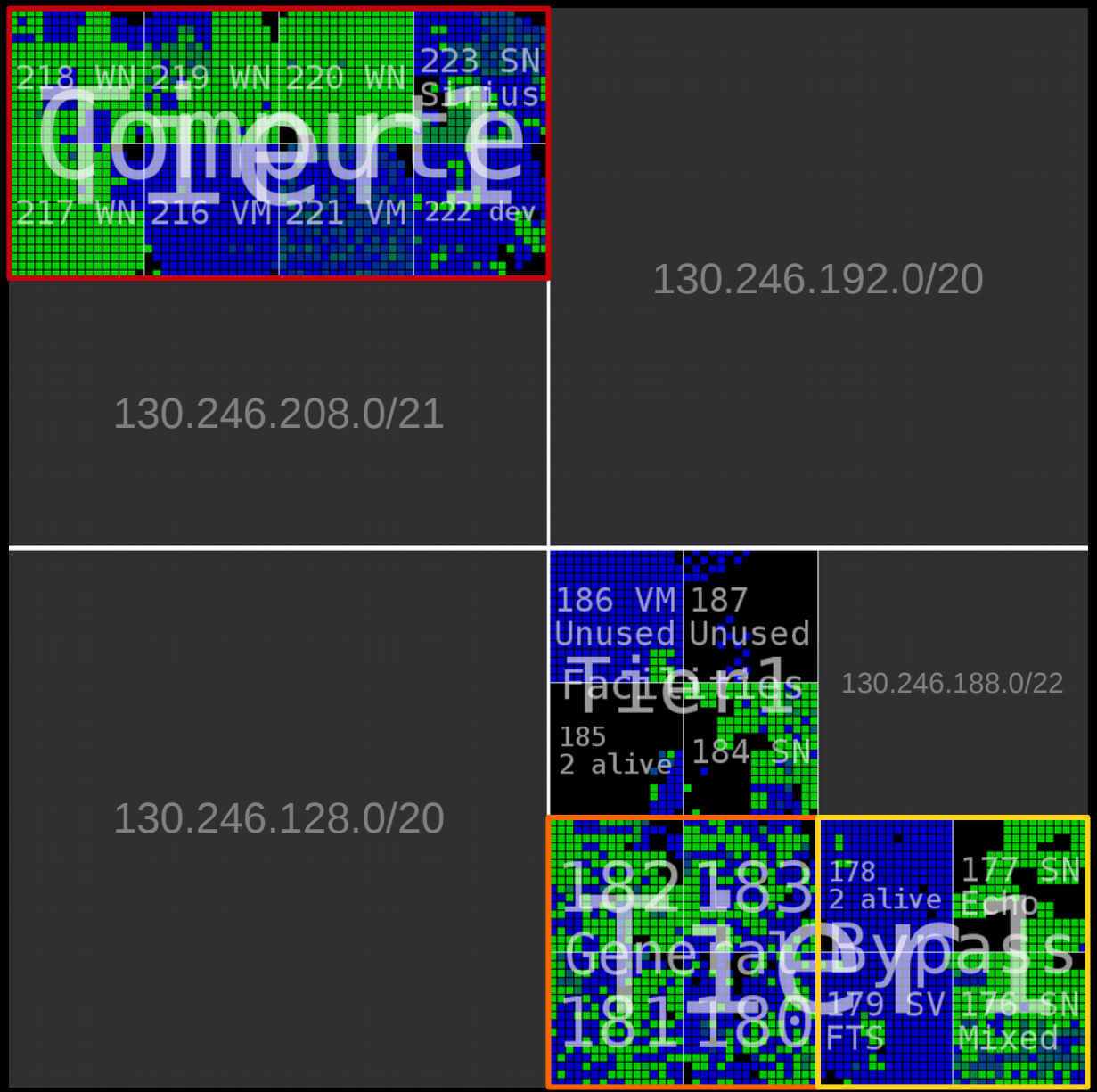

Edge of /16

Even numbered prefixes are squares and can be trivially split and fairly easily combined.

Border between adjacent /18s

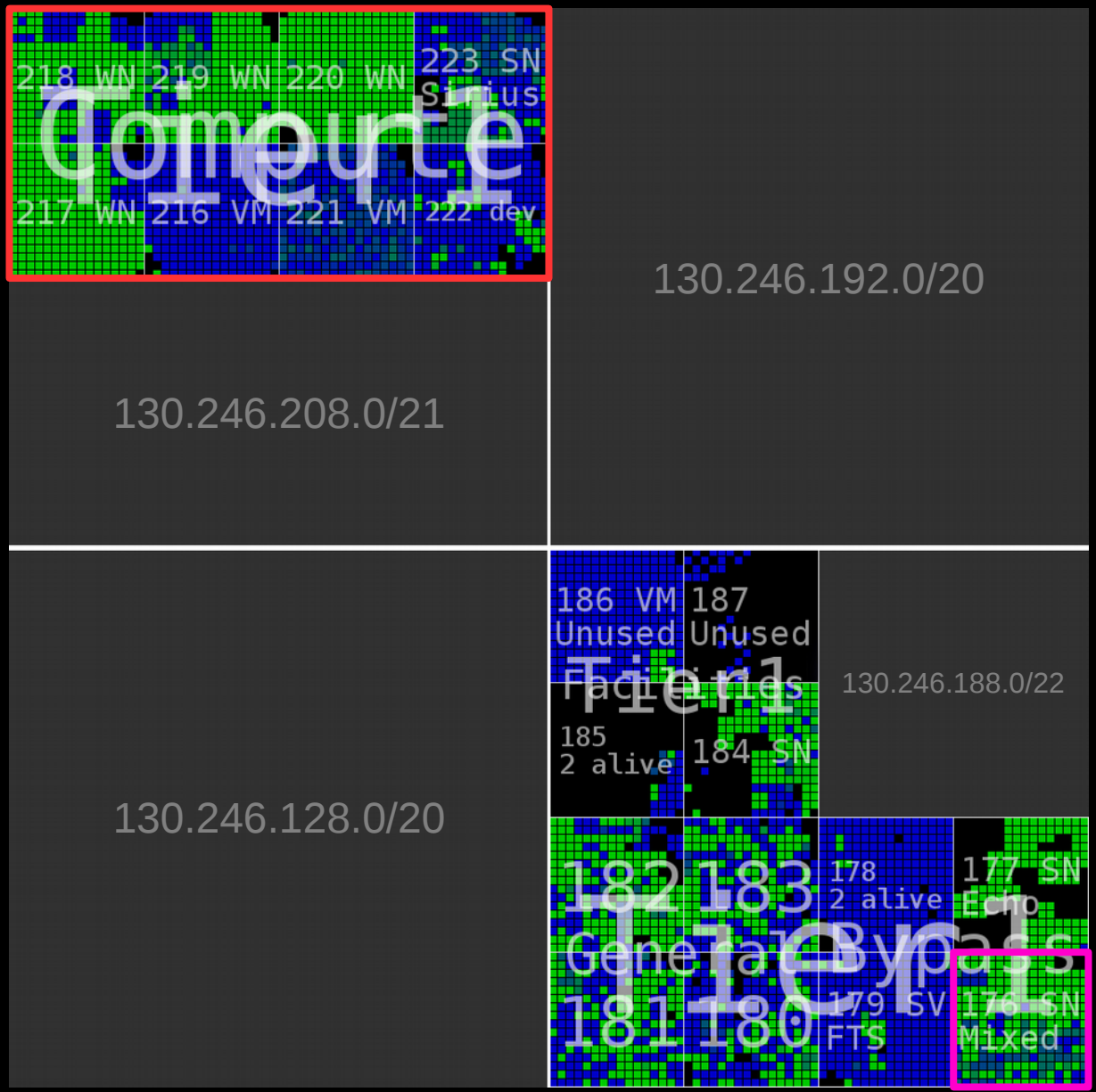
Odd numbered prefixes are rectangular and follow the orientation of the underlying curve, here they just happen to be mostly horizontal.

# Tier 1 Up Close




- A dwindling number of existing hosts in 130.246.0.176/22 need LHCOPN/ONE.
  - Echo gateways, all CASTOR storage nodes.
  - 99% are within 130.246.0.176/24
- Worker nodes ***do*** need LHCOPN/ONE access.
- Many nodes in 130.246.0.176/20 ***should not*** be on LHCOPN/ONE.

# Moving forward...




- Consolidate on: 130.246.216.0/21, 130.246.176.0/24.
  - Small number of hosts to move.
  - Move front-end load-balancers.
  - Do not move internal service components.
- Eventually drop 130.246.176.0/24.

# Next steps

NB. We currently announce 130.246.176.0/22 and 2001:630:58:1820/64 to LHCOPN.

1) Give the handful of hosts with IPv6 addresses in 2001:630:58:1800/64 addresses in 2001:630:58:1820/64.
2) Clear out top end of 130.246.216.0/21.
3) Announce 130.246.216.0/21 to LHCOPN.
4) Move front-end load-balancers.
5) Move echo gateways.
6) Announce 130.246.216.0/21 and 130.246.176.0/24 to LHCOPN.
7) Announce 130.246.216.0/21 and 130.246.176.0/24 to LHCONE.

# Conclusions

- With a series of fairly simply steps we can be ready to connect to LHCONE.

- The Tier 1 can finally appear to be a single entity.
  - All Tier 1 services get a uniform routing topology.
  - We can start moving towards a simpler configuration now.