Project GOLF – 3 years of EVPN in the Datacentre

David Freedman / David Overton / Sandy Breeze – Claranet Group Engineering - Networks

claranet | helping our customers do amazing things

Introduction

- **We only have 30 minutes, so….**
- **This talk is about:**
  - Long term EVPN/VXLAN in real-life datacentres.
  - What our technology solution looks like.
  - Testing and automation to support this.
- **This talk is <u>not</u> about:**
  - Capacity planning.
  - Change (and customer!) management.
  - Monitoring.
- **Introduction to the team.**

claranet | helping our customers do amazing things

Introduction

- **The story starts in 2014**
  - **Like many other good stories, this one starts with ageing equipment.**
    - Specifically Cat6500 SUP720-3BXL
    - The proud workhorse of the datacentre, it did everything, from L2 switching to MPLS PE.
    - But 10+ year old technology at this point.
  - **Constant operational headaches**
    - FIB , QoS and VLAN scale
    - Fragile management plane
    - Sub-par IPv6 feature-set
    - Extremely challenging L2VPN feature-set
    - STP – Need I say more?

Introduction

- **Our fiscal year runs July to June**
  - **The July 2014 budget (for fiscal year 2015) assumed linear growth.**
  - **The normal trend was go bigger**
    - Base assumption was made to keep Cat6k, and move to SUP-2T/DFC4
  - **But bigger != better**
    - We needed stability AND features.
    - No assurances of both
  - **We knew this meant a technology change**
    - But what?
    - And how major?
  - **We had to pay attention to the driving forces in the DC .**
    - Namely virtualisation and cloud
    - The same people demanding innovation in the network

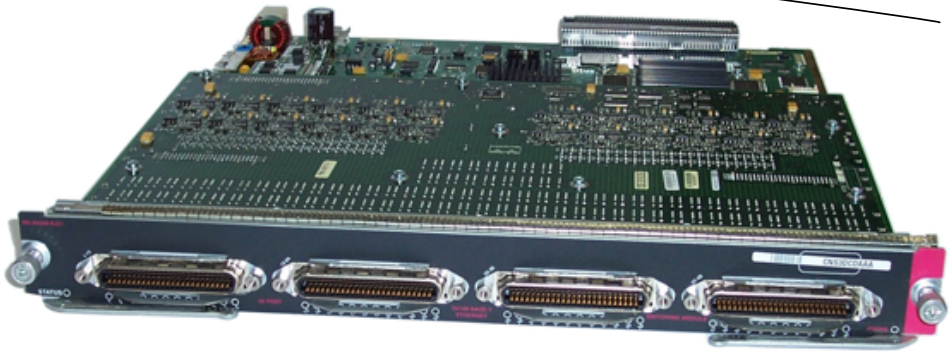
- **We had to do our research:**
  - What features do our customers need?
  - What would it take to do something new?
  - What opportunities would this give us?
  - What principles would we build it under?
  - How could we implement it? and by when?
  - What about monitoring and automation?
  - What about tooling and documentation?

Kit replacement 2015

- Mix of catalyst 6509 and 6513 chassis
- Lots of catalyst 2900, 3650, 3750 etc..
- Vast estate of 10/100 (Base-T) ports.

14.4”

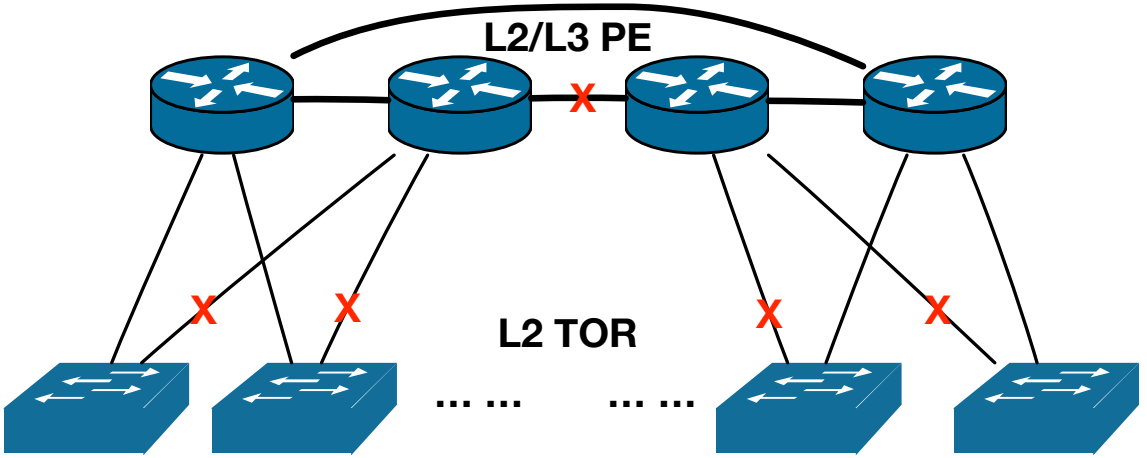
6”



WS-X6548-RJ21 (banana for scale)

claranet | helping our customers do amazing things

Legacy L2

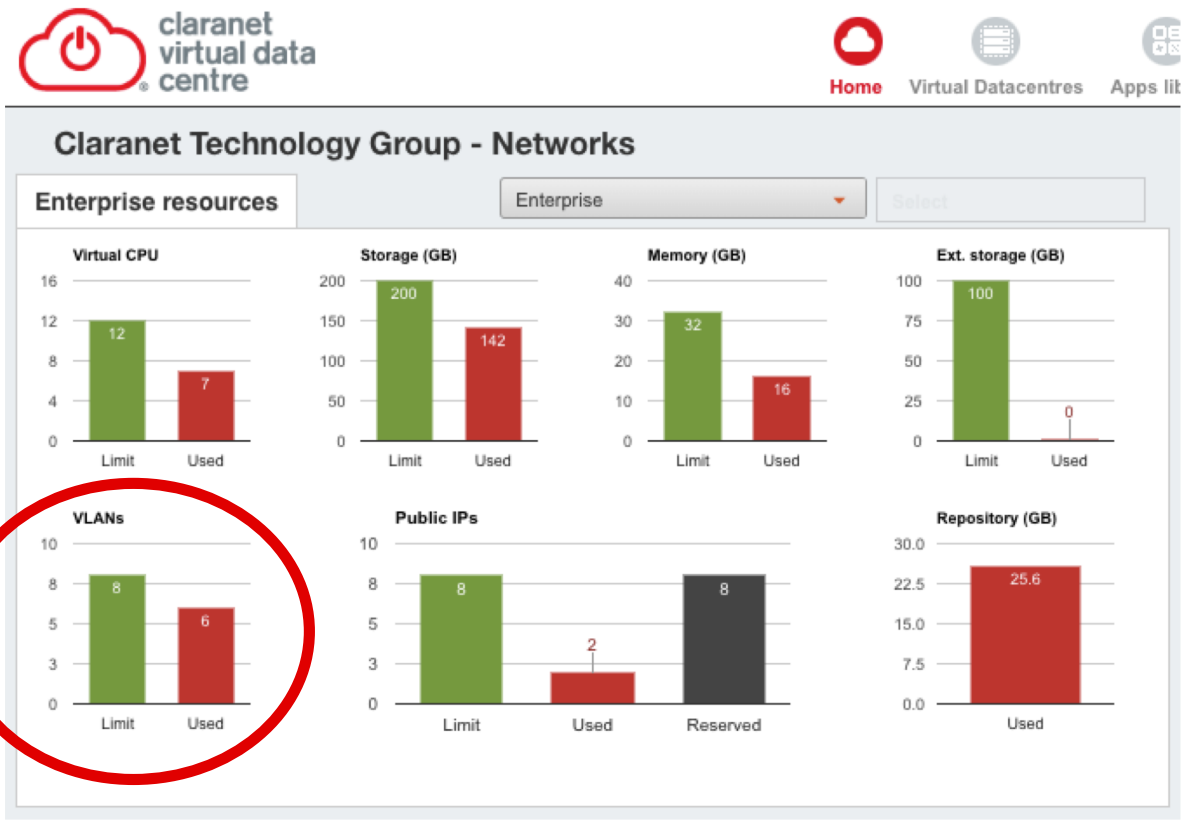


STP

MST0 - root PEx
MST1 - root PEy
MST2 - root PEz
X - BLK

Capacity

- #VLANs
- STP
- Interfaces
- Uplinks

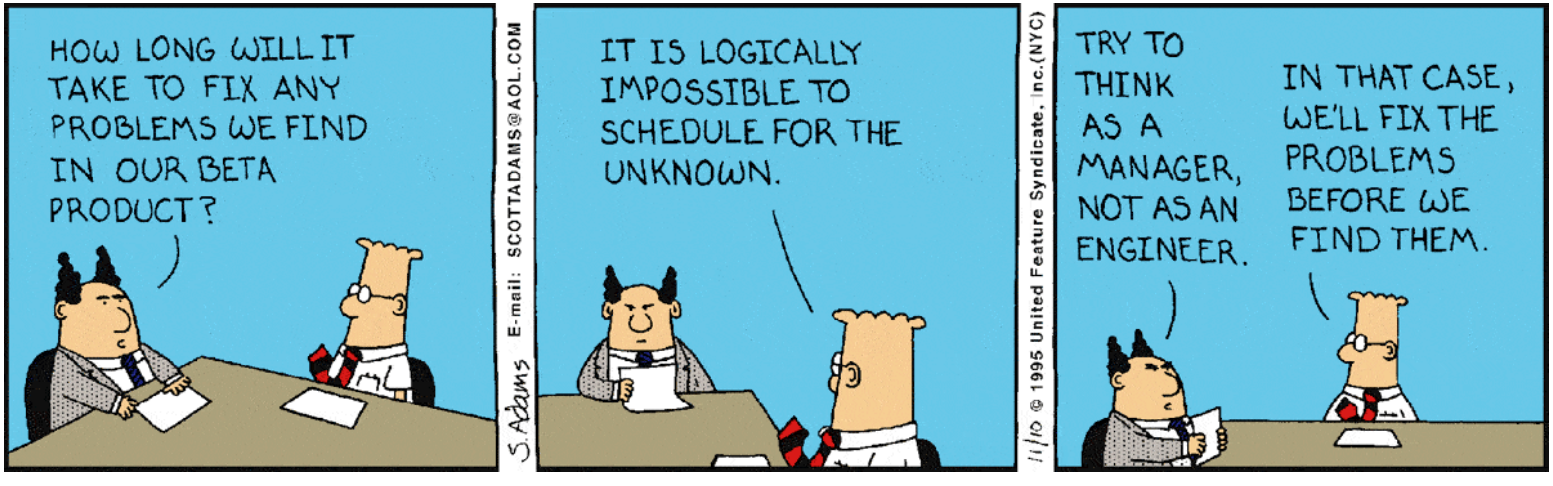


claranet | helping our customers do amazing things

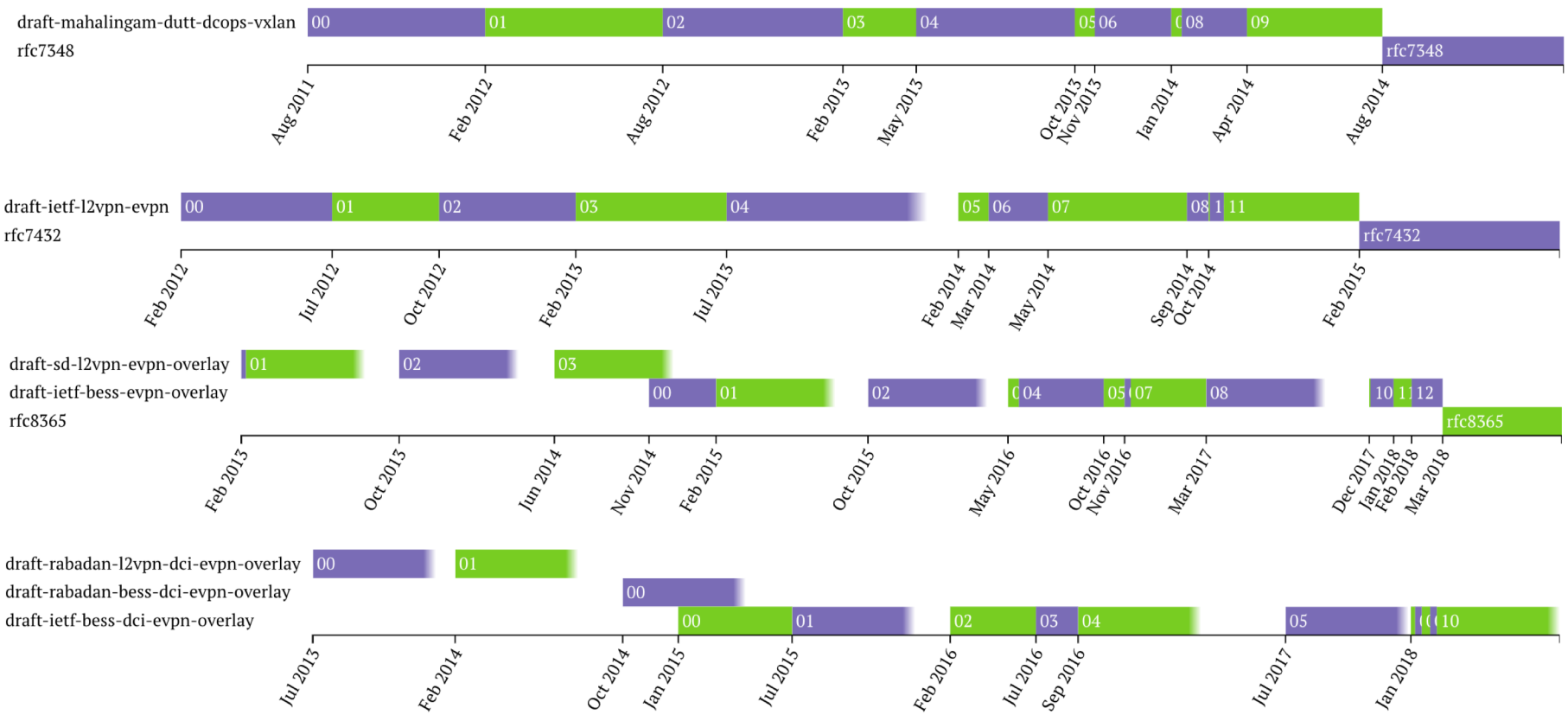
Modern L2

- **MUST:**
  - Be standards based
  - Scale (to our expectation)
  - Interop with existing MPLS L2/L3VPN
- **MUST NOT:**
  - Rely on a controller
  - Be based on proprietary features
  - Rely on any data-plane flood & learn

Time for VXLAN:

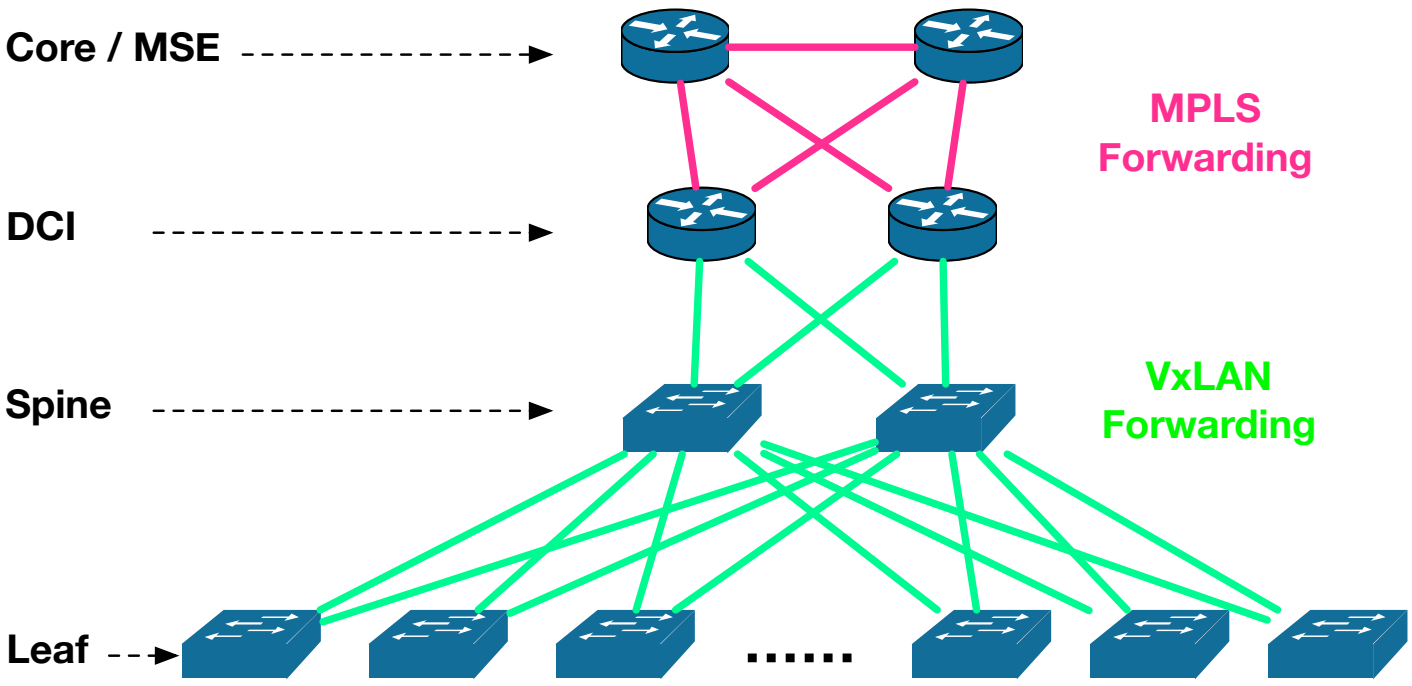


Standards timelines



claranet | helping our customers do amazing things

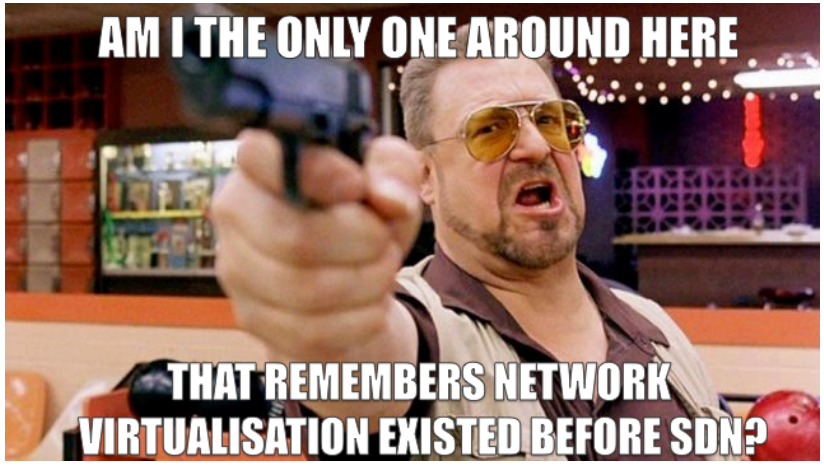
Today



ACI vs Standalone NX-OS

ACI

- Turnkey DC network and hypervisor automation
- Proprietary control plane…

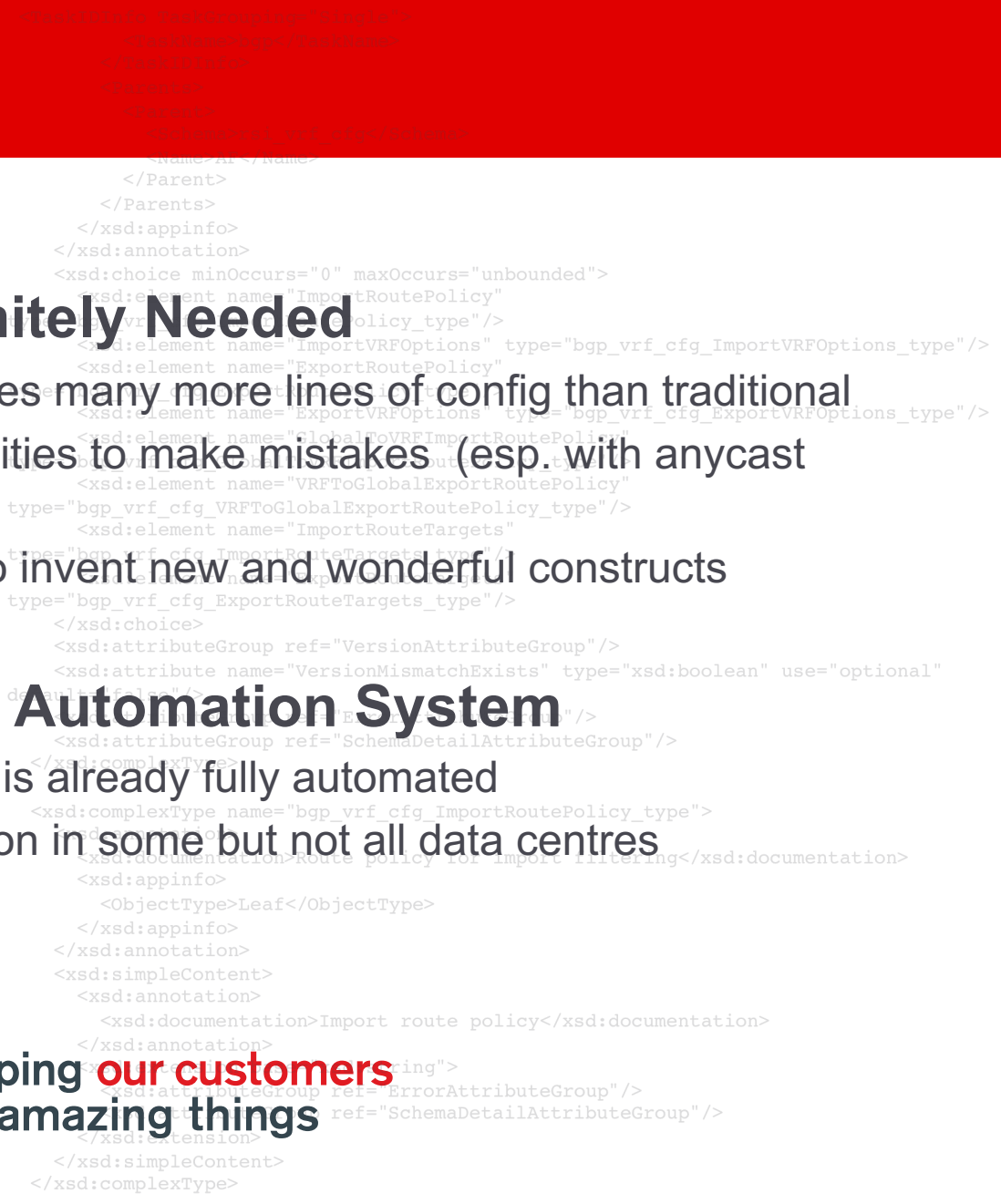


Standalone

- More traditional ‘IOS-like’ OS
- Our feature requests made it here
- We had ~12 years of network automation behind us
- Forced to follow open standards (good)
- Possibility to inter-op.

Automation

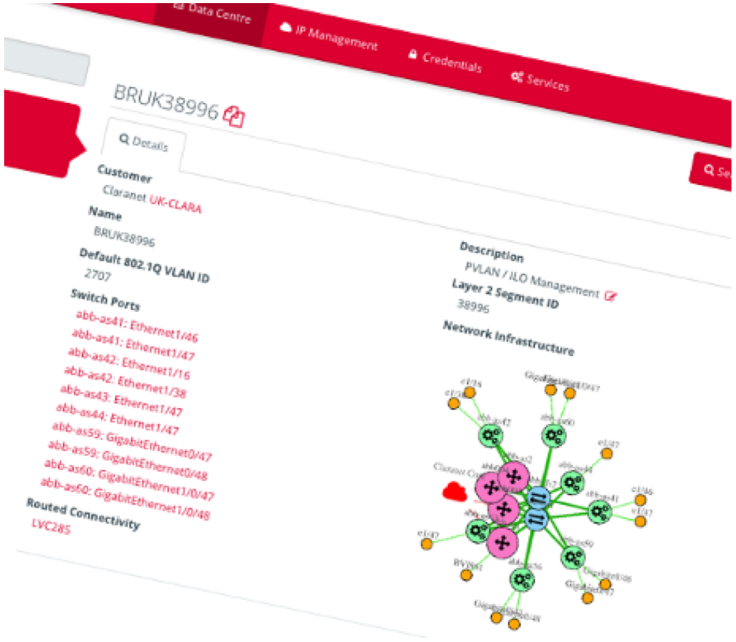
- **Automation Definitely Needed**
  - VXLAN/EVPN requires many more lines of config than traditional
  - Many more opportunities to make mistakes (esp. with anycast gateway)
  - More opportunities to invent new and wonderful constructs
- **Existing Network Automation System**
  - Much of the network is already fully automated
  - Existing L2 automation in some but not all data centres



claranet | helping our customers do amazing things

Our automation stack

- Entirely in-house developed
- System encompasses IPAM, VLANs, Network Topology, etc
- Modular components called “policers” which build and sync configuration to network devices
- Web based user interface for end-users with REST API available for scripting



claranet | helping our customers do amazing things

Auto Provisioning

- **POAP (Nexus' ZTP)**
  - Switch is added into the system using a single command line tool (*golf-config*)
  - Switch boots, DHCPs, upgrades, reboots, downloads base configuration
  - Templating engine is used to serve up a base configuration (IS-IS, uplinks, etc)
  - Policer connects to switch and completes configuration (BGP, etc)

- **Result!**
  - Zero involvement required for adding new switches
  - This lets field engineering teams get on with it

Talking to the Nexus 9000

- **REST API Obviously!**
- **Or not…**
  - Produces XML output which is awkward to work with
  - Not possible to secure API
    - but can be dropped into a VRF
  - Requires the use of CLI commands to make configuration changes
- **Let's look elsewhere..**
  - It's not nice to have to read configuration in one format but write it in another
  - Offers few advantages over working directly with the CLI configuration

Talking to the Nexus 9000

- **Model based API (NX-API REST MDP)**
  - Evolution of the internal ACI API.
    - Which didn't model VXLAN.
  - NX-API REST has a VXLAN model.
    - But, not YANG, proprietary.
  - None of this existed at the time!

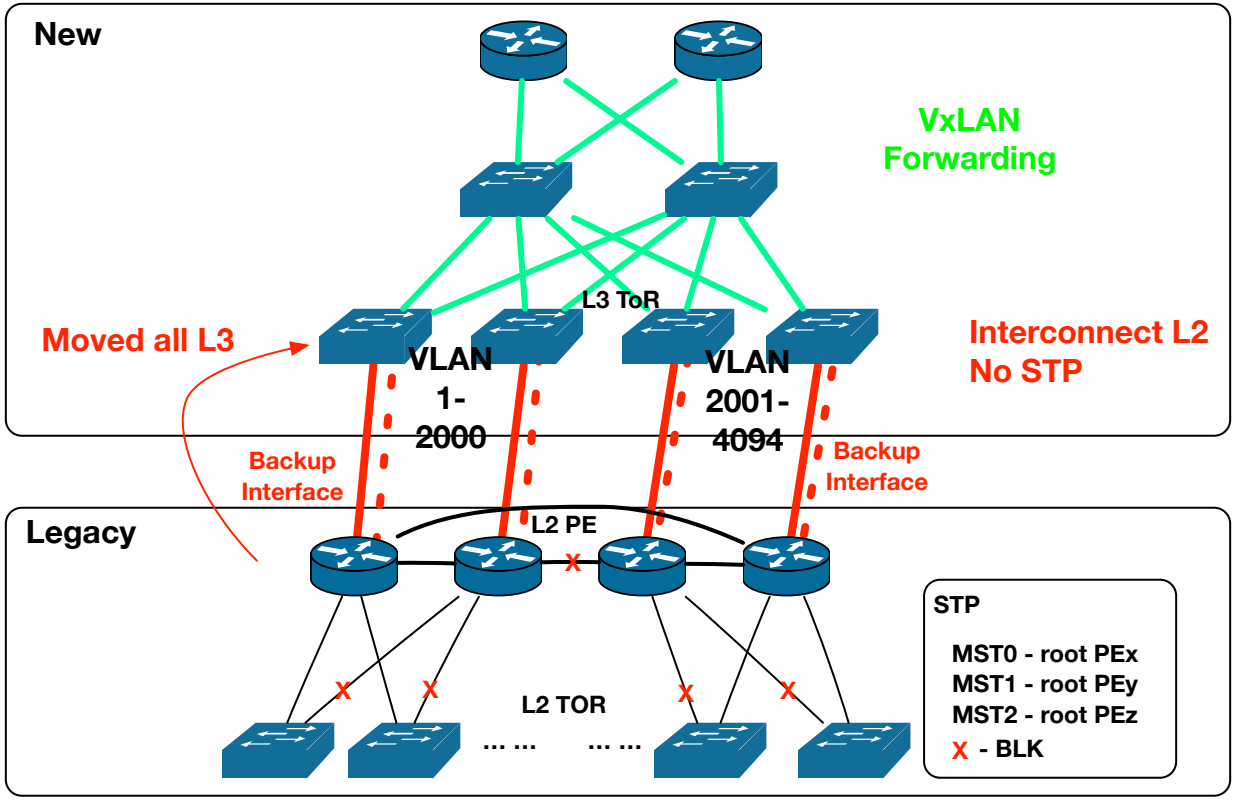
- **Solution: Enhance existing code**
  - Already supports some NX-OS (Nexus 5K already in production)
  - Already has concept of Layer 3, just not in the Data Centre

Talking to the Nexus 9000

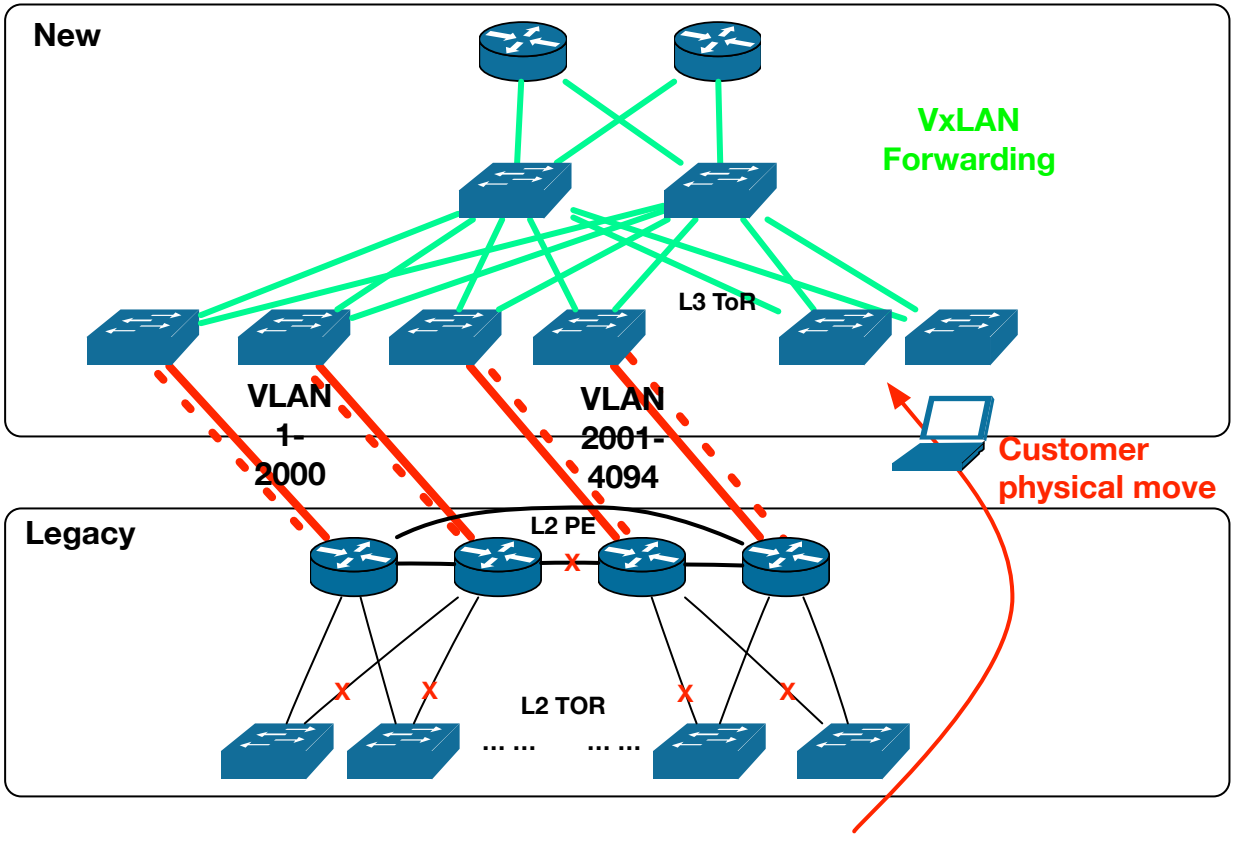
- **Enhancements**
  - Rewrote NX-OS parser to make it more robust
  - Tweaked existing L2 support and added VXLAN/L2VNI support
  - Modify existing L3 code to support L3VNI+AC-GW

- **Unfortunately, we still scrape these devices :(**
  - But it works!
  - And they are very tolerant toward it

Migration example – Phase 1



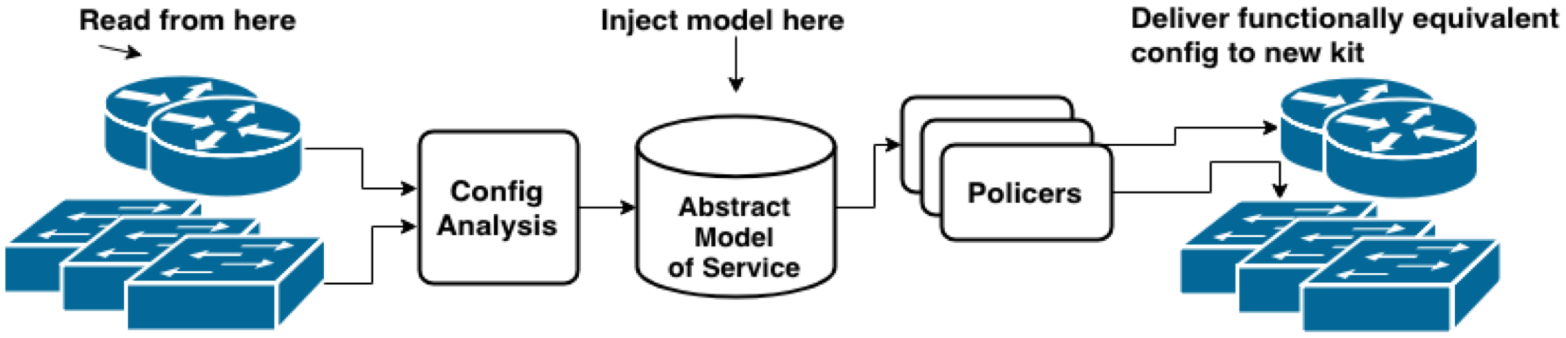
Migration example – Ph2



claranet | helping our customers do amazing things

Migration – Software

- **Enhanced a tool we wrote previously (*ios-suck*)**
  - Originally written for migration of ethernet aggregation platform from Cisco 7600 to ASR9K. This meant it already understood VLANs, SVIs, BGP, QoS, etc
  - Works through static analysis of configuration files
  - Highlights inconsistencies in manually written configuration



Pain Points

- **Reserved VLANs, L3VNI VLANs**
  - These seem to differ on every new device. Fortunately, most can be moved
  - L3VNI eats a VLAN for each VRF, causing allocation pressure
- **CLI Bugs (predictably)**
  - Stuck configuration (e.g. VLAN maps)
  - Phantom configuration (e.g. EVPN)
  - Inconsistent visibility of state (e.g. shutdown/no shutdown)
- **Features not working properly**
  - Requires version-specific workarounds in automation
- **TAC**
  - It's all new to them too!

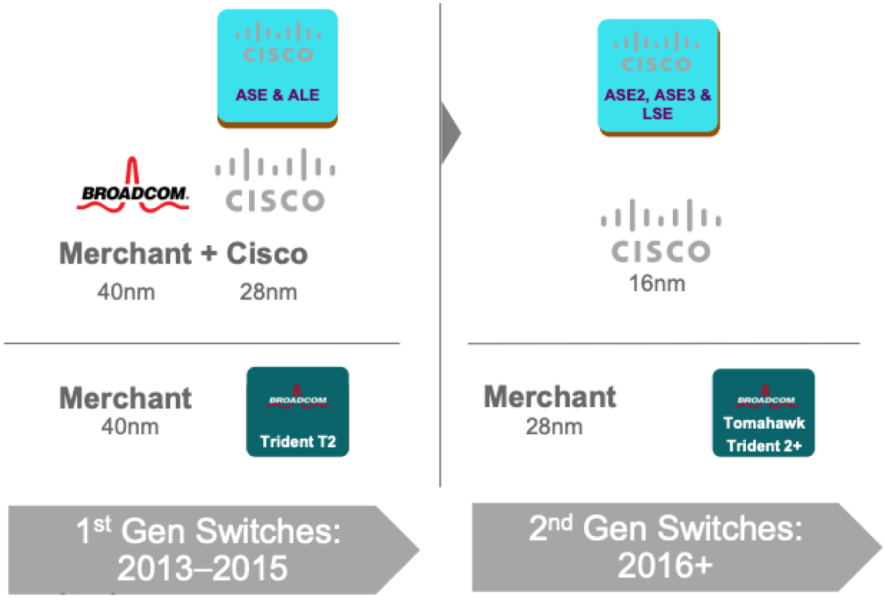
Testing reveals

- How FIB can get inconsistent with RIB
- Scale docs are sometimes wrong when they say x ACE's and y QoS policies (and other things in TCAM)
- Really useful features like VLAN translation or ESI-MH do not work with all these other mandatory boring features

etc…

Merchant vs custom silicon

- Launched on merchant (T2) silicon to get the features
- Vendor silicon (LSE) soon developed to drive new features and scale
- Vendor silicon missed some features in merchant – feature parity issues (e.g. standards based ESI-MH)
- Saw less bugs with vendor silicon (maybe because code matured too)



Summary

- **Know what you're getting yourself in for**
  - Watch presentations like this
  - Talk to others
- **How can you automate?**
  - Even doing ZTP, this will drastically change your experience
- **Understand your customer requirements**
  - Model some of their L2 topologies in your lab
  - Speak to them, have understanding of what they do
- **Get to grips with new behaviour**
  - Make every team troubleshoot issues
- **Stick to your guns and drive the vendor**
  - Don't take no for an answer
  - Be wary of the restrictions of merchant silicon

Conclusion

- **EVPN is a thing, it's not even a new thing**
  - Almost all vendors in the DC space now shipping EVPN implementations
- **We've been doing it for years**
  - Title of the presentation says it's three, but thinking about it, we've been working with it for much longer
- **It's stable**
  - We've been offering production services with SLA on it
- **It's best served with automation**
  - Allows you to reap of benefits of scale

EVPN

claranet | helping our customers do amazing things

Questions now – if we have time

claranet | helping our customers do amazing things

We're hiring!

https://www.claranet.co.uk/about-us/careers/jobs/all

claranet | helping our customers do amazing things

claranet

helping our customers
do amazing things