Building a nationwide carrier for <$1M

war stories & lessons learned

Jonny Martin & Tim Hoffman

NANOG 70, Bellevue, WA

Background

- Founded by Roger De Salis
  - Tired of the over-engineered and over-priced options in the market holding NZ back
  - Attempting to change price and performance by at least an order of magnitude
- Funded by a guy who found civil works more interesting than the network (and funded as such)
- 2005: just two companies in NZ with inter-city fibre. Would not lease to others
- DWDM/L2 capacity was a duopoly - $$$$$$$$
- Over 5 years, FX laid thousands of KM's of fibre around country
- Built a nationwide network with 56 POPs
- Often started 1,000km fibre builds with only 25% funding committed
- ***Built an IP/MPLS network that ran a substantial portion of NZ's internet traffic for less than a million dollars***

Uh… what?

We’re going to share some war stories from pulling this off...

Disclaimer: FX got acquired in 2014 and the network refreshed. The current network is nothing like what is presented here.

We didn’t have

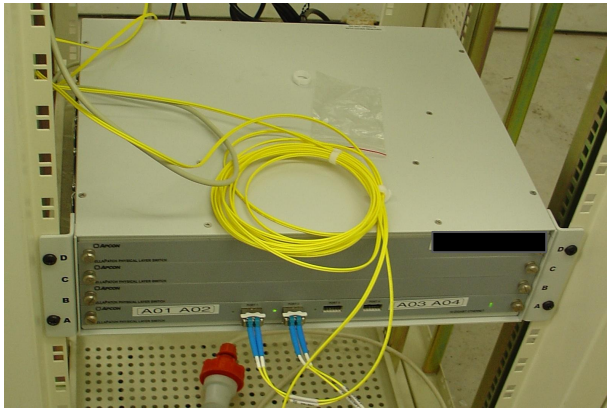
- Lots of money to spend
- Lots of staff (6 person team)
- Resident engineers or vendors falling all over us
- Time (we had pretty tight deadlines)
- **Revenue (at the start)**

We couldn’t have got the business off the ground with an ideal network at the start, so we had to be creative…

Everyone had to know a fair bit about everything for it to work.

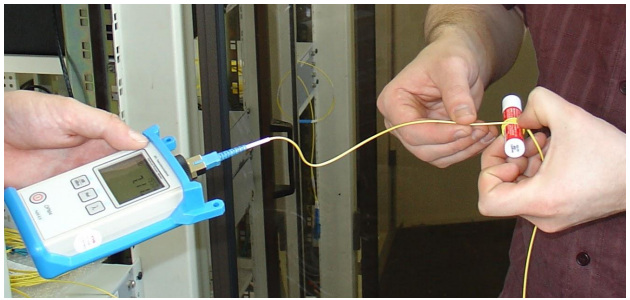
“Transmission network”

- APCON made a ‘fairly reliable’ layer1 repeater
- 10G ZR (80km) optics had just come out
- Sites were determined by driving a car along route
- Find good spots every 70-80km
- Housed APCON or L2 switch at every site with ZRs
- Initially deployed two small DWDM spans
  - This addressed some limited fibre situations
  - Expanded the DWDM across the entire network over time

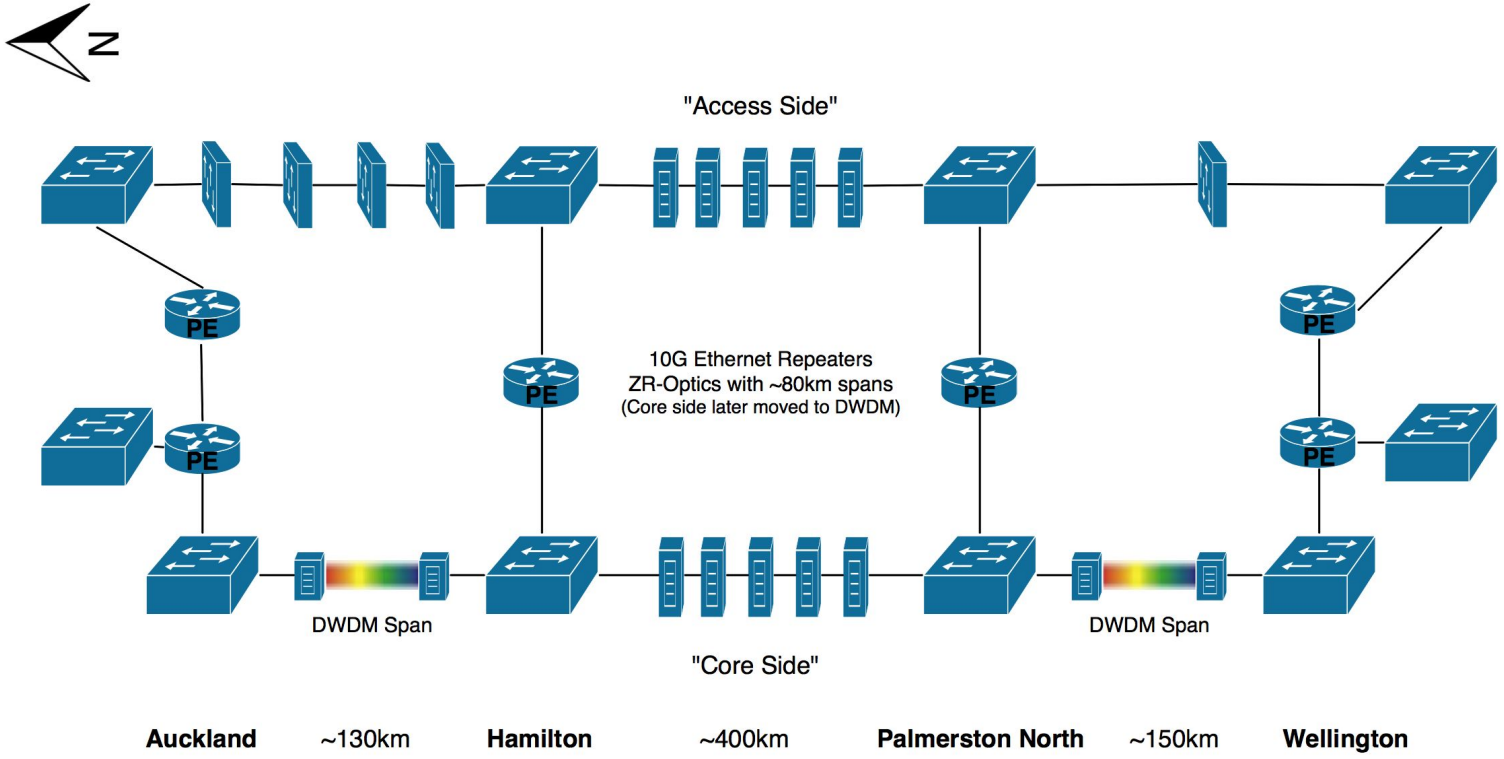


Optical testing

- Initially had minimal optical test equipment (didn’t have an OTDR)
- Had to engage contractors to do testing when critical
- We carried a 6500 around to do this with DOM optics
  - A plethora of low-cost test gear available these days
- One “80km” span was actually 93km - testing is critical
  - Measure optical headroom using variable attenuation. Wind duplex fiber around a pencil. Measure attenuation down one core, device under test on the other. (“Mandrel Wrap”)
  - Grade optics and used the most powerful and sensitive for the long spans



“Transmission network”

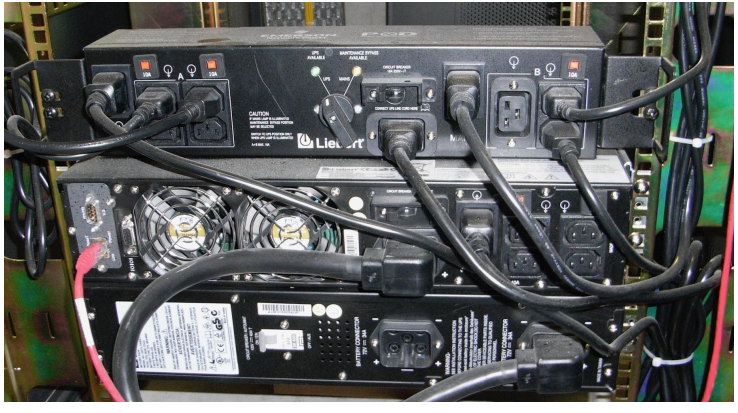


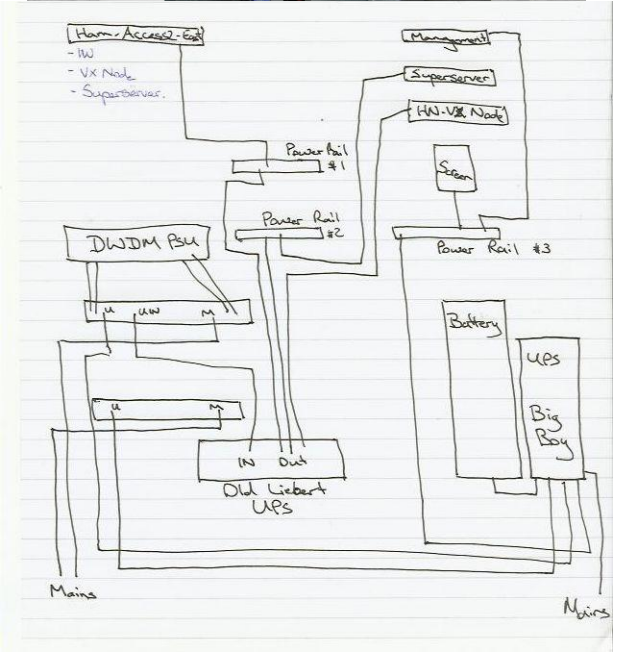
Power

- 3KVA commodity UPSs, or stacks of them, go a long way and don't need an electrician to replace. Racks often had a chassis switch + half a rack of UPS.
  - Key is that everything is behind a transfer switch so that you can replace or service UPSs without impacting the network, allowing for growth without disruption
- You can store up to 500 liters of diesel at any location without licensing (which was good given how dodgy the power is in rural NZ)
- We didn't really use what you would call "Colocation facilities" - more "Brolocation" for these
  - The vault of an old bank that had been brought by a cafe.
  - A concrete bunker that was put on NZ Army land without formal permission
  - Railway control huts
- Pay people for space or land with a 100M internet connection.
- Use air vents, not air conditioning

Power

- You can actually design a power system that switches to a low cost generator pretty cheap
  - Relays and SNMP/terminal servers with dry contacts for control and monitoring of virtually everything.
  - This also helped because we didn't buy the redundant PSUs
  - Rely on dual power supplies to mains and UPS power to avoid UPS/transfer switch SPOF





Colocation

- We didn’t really use what you would call “Colocation facilities” “Brolocation” would be a better description
- Air con is a luxury when you’re on a tight budget - it’s amazing how much gear will survive with just good filtered vents.
  - Relay controlled fans, lights, etc. are a win using SNMP controlled terminal servers with dry contacts
- Pay people for space or land with
  - 100M internet connection (big difference from dialup)
  - Generator power (when you’re remote that’s 3 weeks a year of having a heater)
- Try not to give the master POP key out much (way too many people ended up with access - ended up deploying electronic access control!)

Colocation

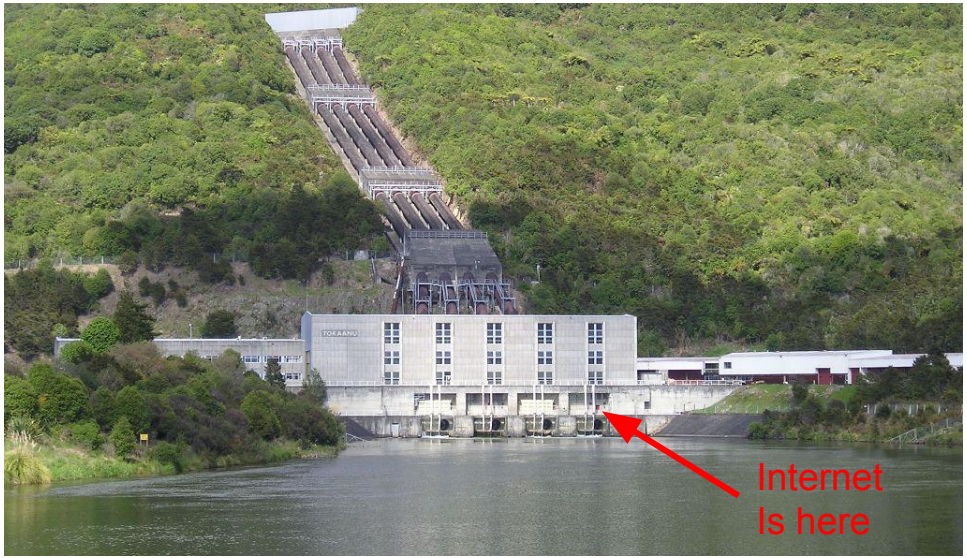
Wellington Railway Station

- Back room in the railway station of our capital city
- POP required seismic ISOBASE
  - None locally available. Imports $$$$.
  - Custom heavy steel plates on bearings with essentially elastic bands holding it together
- Unreliable power - didn’t control the generator
  - Compensated with lots of batteries, monitoring, and detailed knowledge of power draw
- Single PE fed NZ’s primary 911 call center
  - Once moved this PE up 12U while in production without dropping a packet (or the router)

Colocation

Tokaanu

- Two cabinets inside a hydro dam
- Had to get electrical certifications to get entry
- Power was very reliable (that's a pretty big generator!)
- Hard to get to fast if equipment failed



Colocation

Hunterville

- A cafe brought a bank and didn’t know what to do with the vault space
- We put a POP there

Colocation

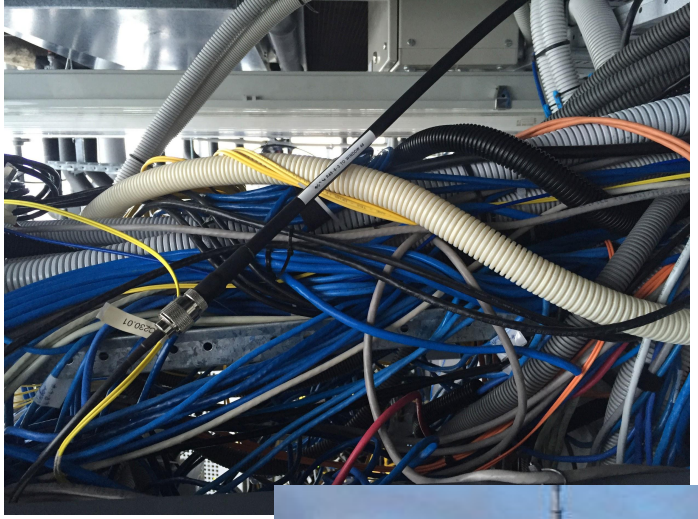
Waiouru

- NZ Army land
- Permission not exactly formal
- Army aren’t very forgiving
- Bunker on the side of the road

“Standard concrete bunker”

- Used standard concrete bunkers that we dropped off every 80km
- Very cheap, easy to just drop on the side of the road

Colocation



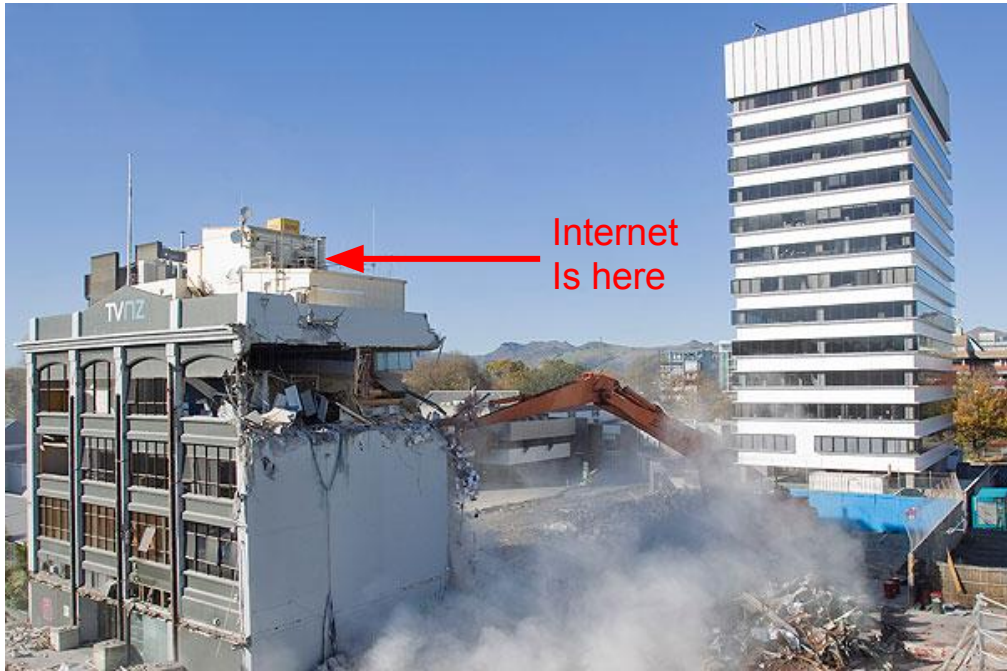
Sky Tower

- Previously the de-facto interconnection point in NZ
  - Primary peering point
  - Most transit to the world passed through the tower
- 48th floor
- Unreliable power
- Gear can reach 50C in summer
- Run your own cross connects
  - What could go wrong?
- Minimal edge equipment kept here

Colocation

Christchurch TV station POP

- Bunker on the roof of a television station
- Easy to get radio shots
- Access was always hard
- During Christchurch Earthquake
  - Building almost collapsed
  - Power failed
  - POP stayed up on generator alone
  - Too risky to top up the generator
  - Race against fuel to move services
- After the earthquake the TV station borrowed our office



Fibre

Many ways to build fast & cheap

- Big rock saw
- Chain trencher
- Bury beside road
- Lay along the railway lines
- Overhead poles
- Bury without conduit
- Cable tie to fences

Daniel Griggs gave a presentation on how we did this at PACNOG 7;
https://nsrc.org/workshops/2010/pacnog7/conference/Fibre_Laying_Presentation.pdf

Network equipment

- Original network built with just 5 P/PEs (Cisco 7600) at intersection points
- This expanded to 40 or so PEs over time
- Most POPs would just have a switch that longlined back to a couple of PEs (often through a break-out in the “APCON transmission network”).
- PEs were often “router on a stick” for most directions out of the POP
  - Build vlans up and down the country for options if things fail
  - Preserve relatively expensive interfaces
- All run as loop-free topologies homing to a couple of PEs (no spanning-tree)
- MPLS core links were generally just another vlan on a transmission network broken out through switches at times
- Main business was L2 wholesale
- Cheaper to build a 500M service as 1G with “soft cap” than buy shaping cards

Network design

We kept it really really simple and it was really stable

- Ran old versions of code
  - We didn’t need to upgrade as we weren’t using or adding features
  - Lots of boxes were doing L2 only
- Simple config
- Traffic was a pretty even mix of L2 point to point services and transit services (network was a wholesale transit) - so we separated the L2 “backbone” - and made the “transit provider” a customer of it.
- It’s amazing how far you can horizontally scale Cisco 7201s as a “transit” provider with L2 underneath
- Most complex thing we did was selling domestic vs international transit (different $ rates)

Service simplification

- No NOC or helpdesk, just the five engineers and one manager.
- Main business was L2 wholesale, so a small number of customers.
- This required a simple set of services.
- We avoided operationally and support complex services, initially
  - No VPLS
  - No L3VPN
- Made failure easier on ourselves by making dual services financially attractive
  - “Want a 10G point to point? Here, take two!”

We simplified our ISP and L2 point to point services by;

- Governing speed by interface speed only (round up to nearest increment)
- Building redundancy with multiple POPs/services

Radio was our friend

Backup paths

- Used radio paths up the country for backup paths when we hadn’t built enough to be redundant.
- This was the only place we ever did shaping/policing
- This was a path for critical traffic only to give a semblance of diversity

BYO Tail

- There wasn’t always a lot of local loop in rural NZ
- Giving people some roof space to provision their own tail on a radio was cheap and easy

Learn to debug and reverse engineer

- Without being a huge spending network, or having resident engineers, you have to get really good at debugging.
- Put servers everywhere
  - Capture TCPdumps
  - Drop a server into a customer service and doing iperfs to prove you can get X from A to B
  - NZ wasn’t used to high speed services, so dealing with big data over high RTT wasn’t well understood by customers
- Get used to not trusting what the router is doing
- Reverse engineer what’s going wrong

Out of band is easy with 96 fibres on every path

- One of the challenges was that almost every ISP used our services in some form.
- There was no way to be sure that if we brought an “out of band” network connection it would stay up if our network failed.
- Big win was that we had a huge fibre resource available to us
- Built a 1G management network in parallel with the production network
  - Initially large layer two strings north to south using ‘dumb’ office switches.
  - Later a proper network with Cisco 1811s running OSPF using a fibre pair on every path
- Whenever we logged into a device it was totally out of band; we could never lock ourselves out.

How did it go?

- We made do with what we had. No-one believed it possible on a low budget.
- We built redundancy in unconventional ways.
  - Detailed understanding of all aspects of the network allows the best trade-offs to be made.
- We incrementally made it better as we sold services and could afford it.
  - Continual improvement resulted in a more conventional network over time.
- We became the primary provider for:
  - National power grid & multiple power plants
  - Police network
  - Government shared services network
  - Almost every ISP bar the two incumbent telcos
- Fundamentally changed the market, taking the price of a 10G wave between major cities from hundreds of thousands of dollars/month to ~$10k/month
- A network that was built on $1M ultimately sold for $116M in 2014

Thank you!

Original FX Engineering team;

- Mike Plant
- Jamie Baddeley
- Neil Fenemor
- Lindsay Druett
- Daniel Griggs
- Jonny Martin
- Tim Hoffman

Founder: Roger De Salis