Cisco Systems

The evolution of NREN Architectures and a case study of "new problems" we need to address

Charles Smith
Senior Consulting Engineer
Academic Research & Technology Initiatives
chas@cisco.com

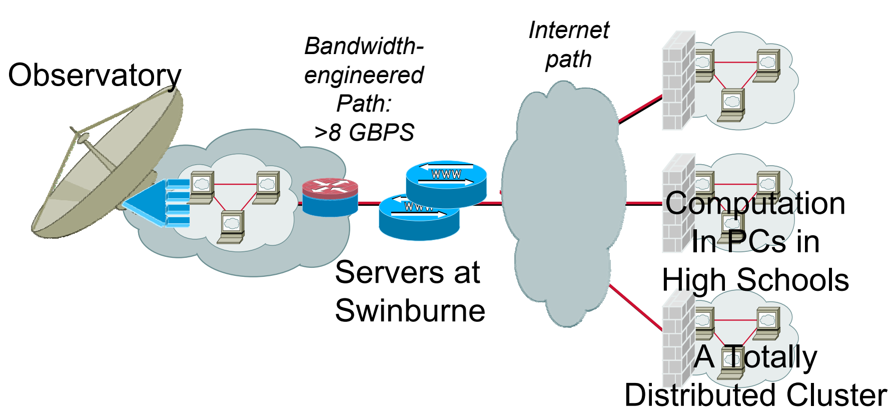
Agenda

- A Case Study in R&E Projects - Radio Astronomy
- Looking at the type of network to address these requirements
- A Short History Lesson in NREN Architecture Evolution
- Where we need to move ?

A Case Study in R&E requirements

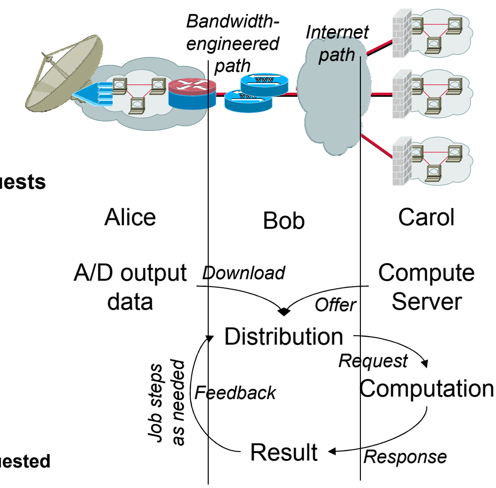
Radio Astronomy

Case Study: Proposed Computation Model



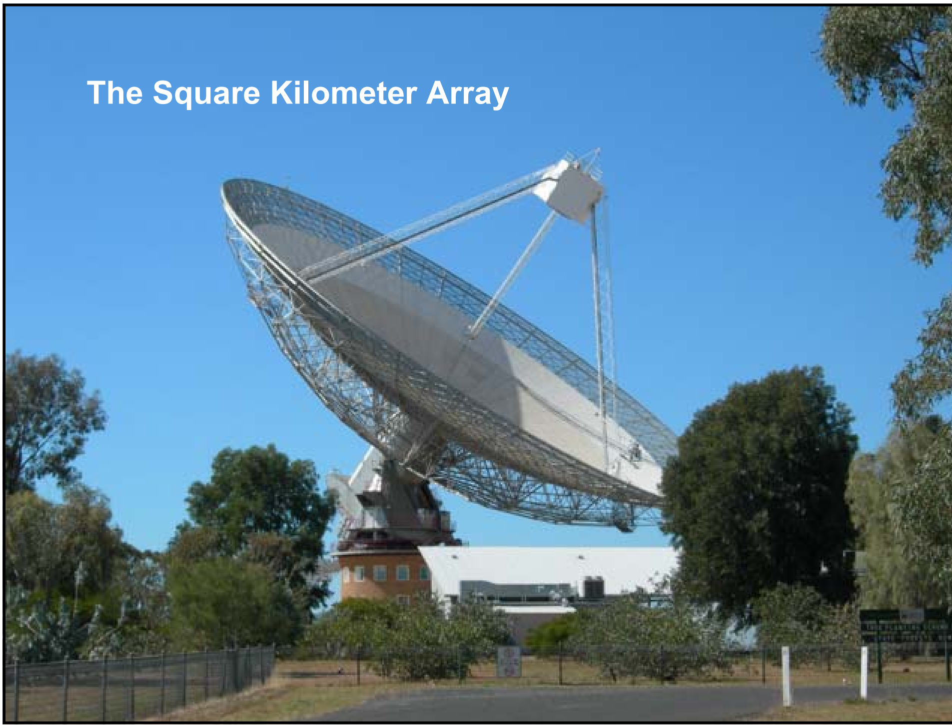
Case Study: Distributed Computation

- Alice exports raw data to Bob
- Carol offers services to Bob
  - Firewall traversal and notification
  - Parameters include:
  - Capabilities of compute server
  - Anonymous identification of server
- Bob replies to Carol's offer with requests for computation
  - Parameters include
  - Type of computation
  - Deadline for completion
  - Characteristics of I/O
  - Input file URLs
  - Output file URLs
- Carol replies
  - Positively, by writing output files as requested and closing connection
  - Negatively by closing connection



Case Study: Issues in this context

- Privacy:
  - You don't want information about your helpers to be inadvertently leaked - bad PR
- Authentication/Authorization:
  - You don't want pseudo-helpers attacking you
- Control of computing:
  - You actually do want to get something done
- Reputation of peer:
  - Even well-meaning peers may not be trustworthy - an offer to help may be circumvented by the janitor who turns off the equipment
- Firewalls and other middleware:
  - You need a firewall traversal model



Radio Astronomers with SKA

- SKA or “Square Kilometer Array”
  - There will be 1,000,000 square meters of actual collecting dish area
  - The array of antennas will span 3,000km in a number of spiral arcs
  - 50% of the antennas will be within 100km of the apex of the spiral arcs
  - The remaining 50% spanning in a logarithmic series of spiral arcs up to 3000km from the apex
  - Scheduled to “go live” by 2014
  - There are 4 competing site locations for SKA these being:
    - South Africa, Australia,China,Argentina
  - The site winner will be chosen in 2007
  - Have a look at http://www.skatelescope.org and http://www.askac.org
- Demonstrators for the SKA are being built now
- This research initiative will fundamentally change how we architect, design, implement operate NREN’s and our GRID’s
- Yes that is a very bold statement but lets just look at some of “the numbers”.

Sensor Collection

Each Sensor Gathers:

8 bits of Sampling

2 bits of Polarisation

2 bits of Nyquist Sampling

Samples at the rate of 256 samples / u second

Note this may move to 512 samples / u second

8 x 2 * 2 * 256 x 10E6 = 8.19E9 bits/s =

7.6 Gigabits / second

That is a 10GbE Port required / Sensor

Antenna Collection

Each Antenna has 100 Sensors:

There is a focal array of 10 x 10 Sensors

7.6 Gigabits / Second x 100 =

763 Gigabits / Second / Antenna

That is a switch with:

- 80 10GbE Ports Input
- 20 40GbE Ports Output

Station Collection

Each Station has 60 antenna:

There are 60 antenna per station

763 Gigabits / Second / Dish x 60 =

45,766 Gigabits / Second / Station

That is a storage requirement of:

- **5,722 Gigabytes / second**
- **5.6 Terabytes / second**
- **20 Petabytes / Hour**

SKA Collection

The SKA has 130 Stations:

There are 130 Stations

45,766 Gigabits / Second / Station x 130 =

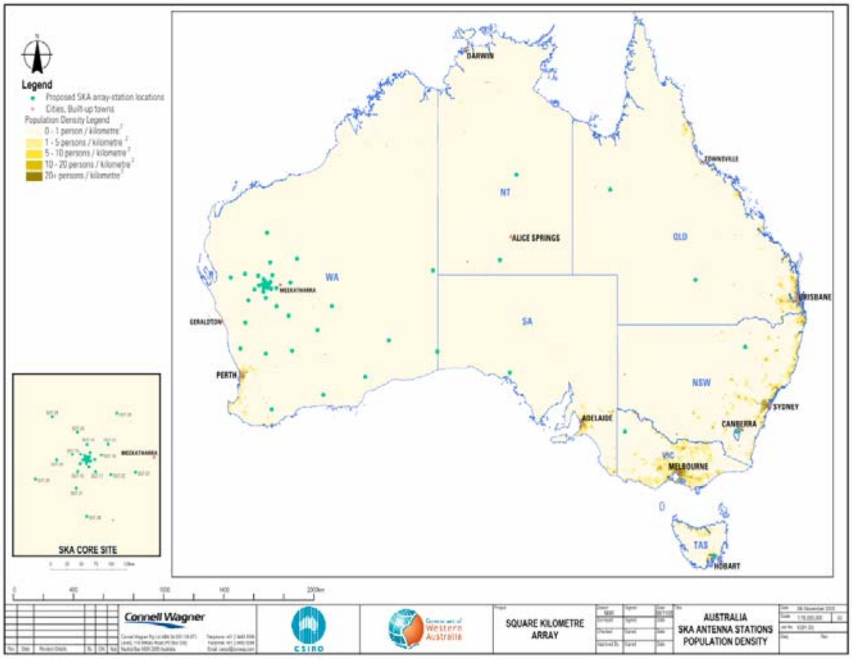
5,950,528 Gigabits / Second

That is a storage requirement of:

- **743,866 Gigabytes / second**
- **726 Terabytes / second**
- **2,554 Petabytes / Hour**

A typical experiment runs for 3 days:

2,554 x 24 x 3 = 183,878 Petabytes of Storage



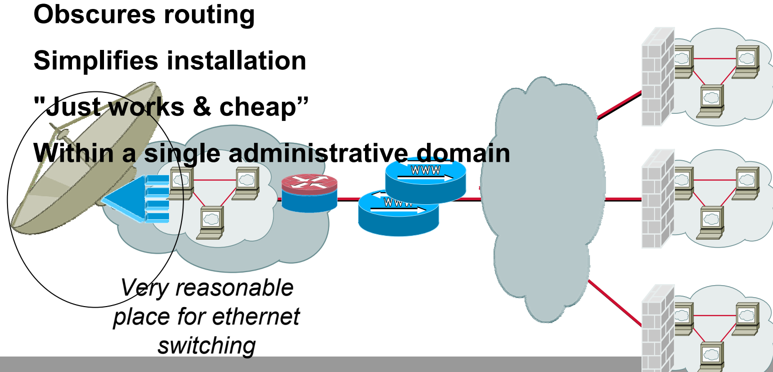
Looking at the type of research network to best address these requirements

Looking at Matthew Baile's Network

- **A GLIF-style solution when needed**
  - **8+ GBPS telemetry path from Parkes Radio Telescope to Swinburne University Cluster (inside single administrative domain)**
- **A general network for distributed compute cluster**
  - **196,000 PCs downloading files and performing computations**
- **Reason each is used?**
  - **It makes sense!**
  - **The cost alone justifies the architecture - Look simply at power both in electricity & CPU. Massive resource not tapped.**

Ethernet switching

- **Intranet Layer**
- **Used when**
  - **Connecting things that one wants to treat as connected**
  - **Obscures routing**
  - **Simplifies installation**
  - **"Just works & cheap"**
  - **Within a single administrative domain**



Lambda switching:

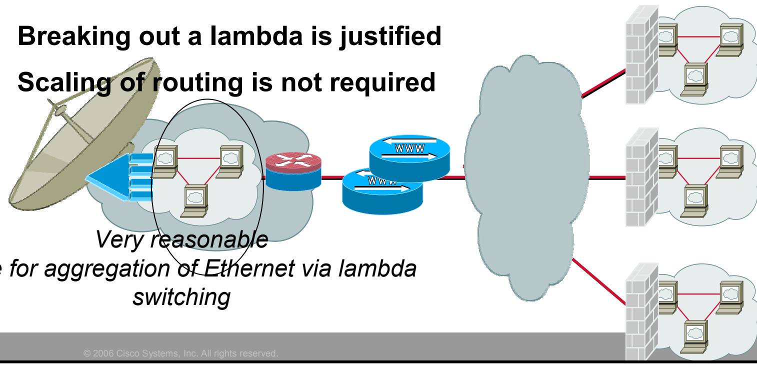
- **Intranet layer**
- **Used when:**

  High capacity is required

  Within an administrative domain

  Breaking out a lambda is justified

  Scaling of routing is not required



IP Routing

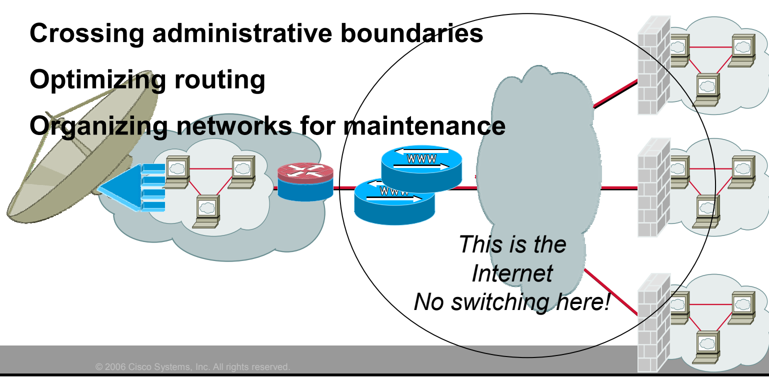
- **Internet Layer**
- **Used when**

  Connecting things that one wishes to manage the connection of

  Crossing administrative boundaries

  Optimizing routing

  Organizing networks for maintenance



Short History Lesson of Evolution of NRENs’

NREN Evolution - Phase 1

- **Individual Service backbone over Circuit Switched, Frame Relay or Cell links**
  - IPv4 Services
  - IP Over HDLC
  - IP Over ATM
  - IP Over X.25
  - IP Over Frame Relay
  - Routers operated by NREN - No one else was buying them then!
  - Links owned & operated by Service Providers
  - Single Autonomous System for research & commodity education
  - Peering between NREN’s established
  - Businesses wanted “in” and links saturated
  - The “Internet Revolution” was born

NREN Evolution - Phase 2

- NREN's establish Individual Services over higher speed Circuit Switched or Dedicated links
  - IPv4 Services
  - IP over PoS STM-4, STM-16
  - IP over ATM STM-1 & STM-4
  - Routers operated by NREN
  - NREN in some cases outsourced to a SP
  - Links owned & operated by Service Providers
  - Commodity Internet operated by Service Providers
  - Single Autonomous System for research & commodity education IP services
  - Peering between NREN's commonplace to reduce transit costs
  - Establishment of common exchange points (PAIX, PNW …)

NREN Evolution - Phase 3

- Multiple Services over High Speed Circuit and Packet links
  - IPv4 Services
  - IPv6 Services
  - IP over PoS STM-16 & STM-64
  - IP over Gigabit Ethernet & 10GbE
  - Backhaul to Core NREN Backbone PoP's over Dark Fibre
  - Routers owned & operated by NREN
  - Outsourcing L3 too expensive & inflexible
  - Links owned & operated by Service Providers
  - Networks split to Multiple Autonomous Systems & Infrastrcutures
    - High speed research
    - Commodity Internet for education
  - e.g. Internet 2

NREN Evolution - Phase 4 - Today

- **Fibre availability realizes the implementation of multi-purpose L1, L2 & L3 backbones over a single physical infrastructure operated by the NREN**
  - IPv4 Services
  - IPv6 Services
  - Storage Services (FC)
  - Distributed Cluster & GRID Services (Ethernet & FC)
  - IP over PoS STM-64
  - IP over Gigabit Ethernet & 10GbE
  - Optical Switched (Gigabit Ethernet) Circuits
  - Lambda Services between researcher vertical communities
  - Multiple Routers owned & operated by multiple NREN's
  - Links owned & operated by NREN communities
  - e.g. TWAREN, GLIF, National Lambda Rail, AARnet, SURFNet, Canarie

NREN Phase 4 - Common Factors to consider

- **Gigabit Ethernet is the required handoff to all sites. High performance Ethernet L3 switches are now moving to the edge.**
- **10 Gigabit Ethernet is preferred for Inter-PoP connectivity vs. OC192 due to lower cost of Router & Switch Interfaces.**
- **Networks in the LAN & WAN are evolving towards peer-to-peer relationships.**
- **IPv6 is now a base level requirement and campus networks will start deploying over the next 2 years**

Where we need to move ?

The Problem Set ?

- Current Drivers of Network Technology in R&E

  Higher education institutions are cooperating to build state-wide and national high-bandwidth research and education networks. Bandwidth requirements are growing exponentially between these sites.

  Need for Bandwidth on demand as – "Scientists need large amounts of dedicated bandwidth, but for relatively short periods of time" Bill Wing, one of the directors of the project and a member of ORNL's Computer Science and Mathematics Division

  Need for national networks to support dedicated bandwidth capabilities between institutions on a time-sharing basis.

  Eliminate "place" as an issue for collaboration.

  Dilemma

    Dedicating very large and expensive resources on a time share basis is good for a small community of researchers but punishes the wider research community (they have no access).

    Lots of dedicated smaller resources punishes those researchers who require high bandwidth intermittently

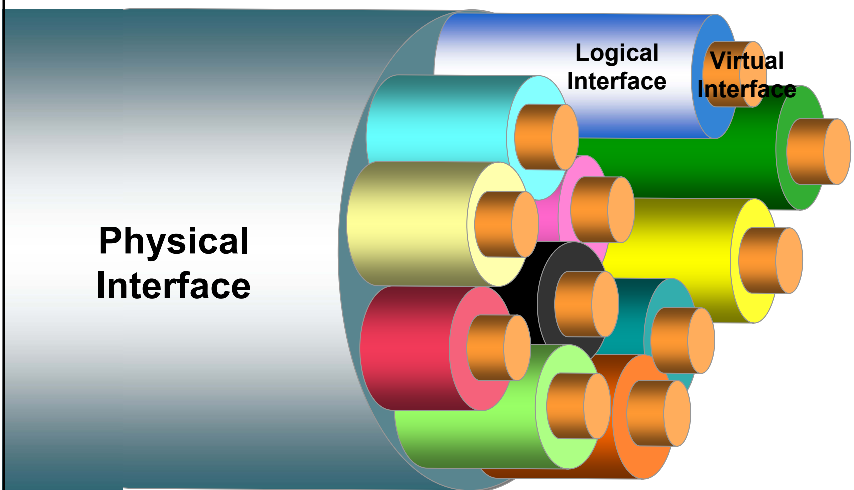
What's Next ?

- NRENs need to accommodate distributed experimental infrastructures based upon Diversified Communities of Interest
- Current GLIF model is reasonable for small communities of interest but impractical for scaling as research communities diversify and grow.
- IP over 10GbE & STM-256 they are big pipes - how best to fill them ?
- Research is a very "spiky" phenomenon, do you split the 10G & 40G up in the "circuit method" into smaller chunks (ie GLIF) or do you retain single dynamic pipes that allow multiple communities the flexibly to share the bandwidth while remaining autonomous of each other.
- The NREN should not be a traditional "service" it needs to evolve as a facilitator at the cross roads of diverse communities of interest.

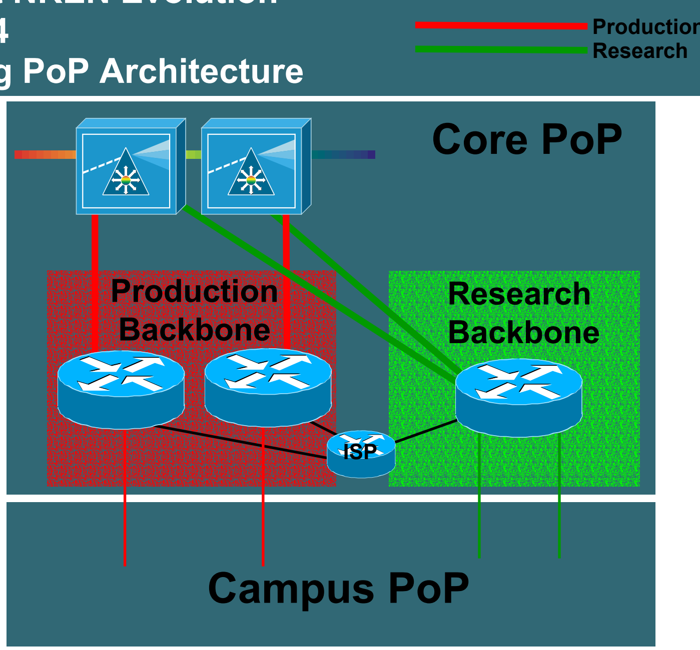
Hierarchical Routing - Is this the answer ?

- Physical Routing is founded on a small number of routers in a point of presence for multiple providers with shared common physical resources i.e the Gigapop or peering exchange will become a single physical router.
- NREN's own and manage Logical Routers and their physical interfaces located within the Physical Router.
- Communities of Interest will own and manage virtual interfaces within the Logical Routers.
- AS's will be based upon these communities of interest rather than geo-political domains.

Interface Virtualization Model

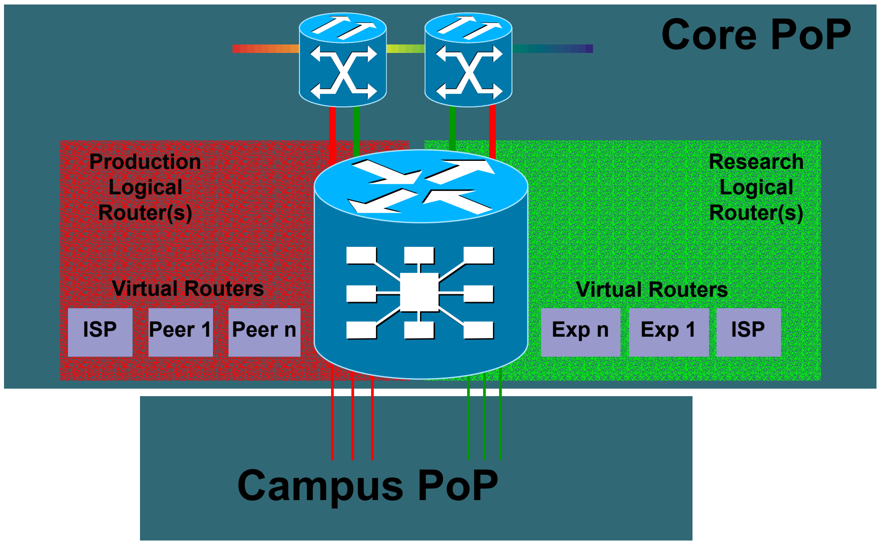


Current NREN Evolution
Phase 4
Existing PoP Architecture



NREN Evolution Phase 5
How would this look ?
Future PoP Architecture

DWDM Wavelength
DWDM Wavelength



Q and A

CISCO SYSTEMS

EMPOWERING THE INTERNET GENERATION