# IP Network Architecture

Lessons learned over time...

kurtis@kurtis.se

woody@pch.net

# Agenda

- IPv4
  - Network topology
  - Addresses and numbers
  - Routing
  - Services
  - Security

# Agenda cont.

- IPv6
  - Network topology
  - Addresses and numbers
  - Services
  - Security

# Who's this Kurtis anyway?

- I built 3 ISPs from scratch
- Used to be Director of Network architecture at KPNQwest, a Tier-1 provider in Europe
- Today freelance consultant working as managing director for Netnod/ Autonomica that operates the Swedish IXP and i.root-servers.net among other things
- Member of the Internet Architecture Board
- Member of the IETF Administrative Oversight Committee
- Chairman of the Swedish Operators Forum
- Working group (co-) chair of IETF shim6, multi6 and v6ops, RIPE NCC Services WG

# Who's this Bill anyway?

- He likes to peer...
- Runs PCH...
- Oh, and he has also started an ISP and built a few networks...
- And he ***claims*** to have a better digital camera...

# Network topology

# Network topology

- Factors driving network topology
  - Geography
  - Cost of capacity
  - Cost of hardware
  - Services
  - Complexity of design and clue-level of available workforce

# Network Topology

Your most useful tools are:

- Documentation
  - Contract times, prices, cancellation penalties, network layout
  - Important that you know that all circuits are in use, otherwise cancel
  - Use tools for network maps
    - Visio, OmniGraffle etc

# Network Topology

- Link monitoring
  - Knowing your capacity usage is critical for planning
  - Trend analysis combined with sales forecasts and sales pipelines are the inputs to your capacity planning process
    - You want your upgrades delivered “just in time”
  - Use tools like RRD or MRTG for graphing usage

# Geography

- Prioritise service delivery by population density, or are there regulatory demands for “Universal Service?”
- Keep local traffic local
  - Capacity might be cheaper on shorter distances
  - Long distances will impact TCP throughput with standard window size. Might be problematic as traffic tends to get more national (http://proj.sunet.se/lsr2)
- Customer base spread will impact capacity planning

# Geography

- Geography dictates backbone link options:
  - Mountainous terrain requires many microwave hops to traverse
  - Running fiber may be very expensive, unless it can be hung on existing long-haul power lines
  - Land-locked countries don't have inexpensive international fiber options
  - Satellite is high-latency, and offers no local-traffic benefit

# Geography

- Geography can also dictate local-loop options:
  - Mountain valleys are well-suited to point-to-multipoint wireless, because antenna locations are obvious and have good LoS.
  - However, wireline will always offer more capacity, and valleys tend to concentrate user populations into small dense areas.

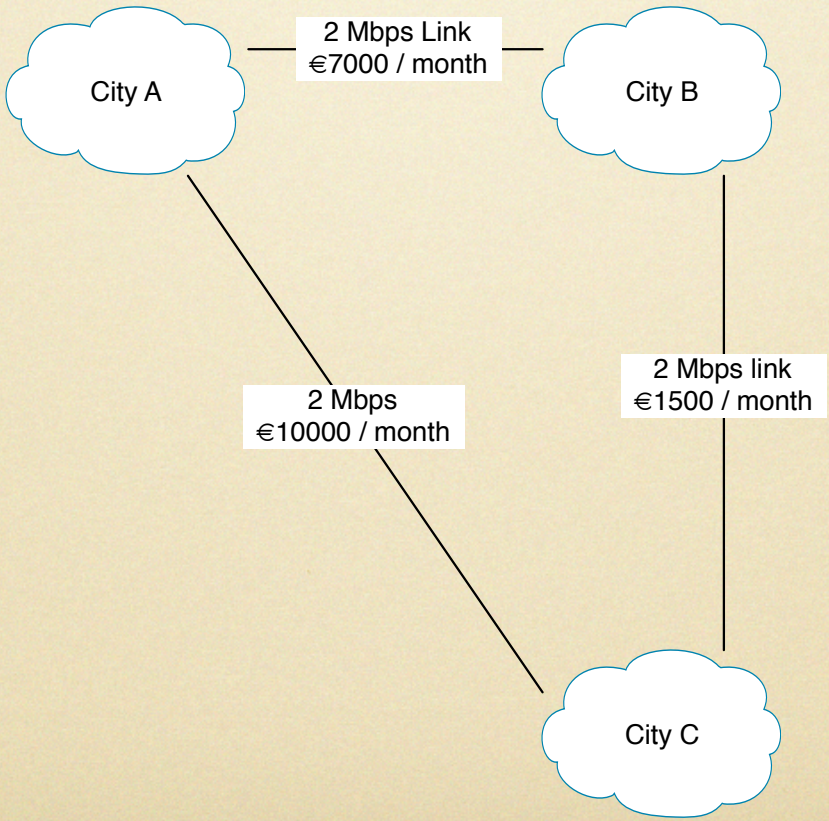
# Cost of capacity

- In more complex topologies, always make sure you have enough backup bandwidth
  - For example in a ring, make sure that if one segment fails you can send traffic the other way around
- "Quality of Service"
  - You can't sell capacity you don't have
  - QoS can't transmit packets, it can only ***drop*** packets. Whose packets get dropped, and how do you retain their business?
  - QoS works on the local loop, but never in the backbone

# Cost of capacity

- Adding distance to packet path might bring price of capacity down
- “To protect or not protect”?
  - Circuits that are protected on the transmission level are more expensive
  - It’s almost always cheaper to get more unprotected circuits (if you can) than a few protected
    - Build redundancy at layer 3, IP routing

# Cost of capacity



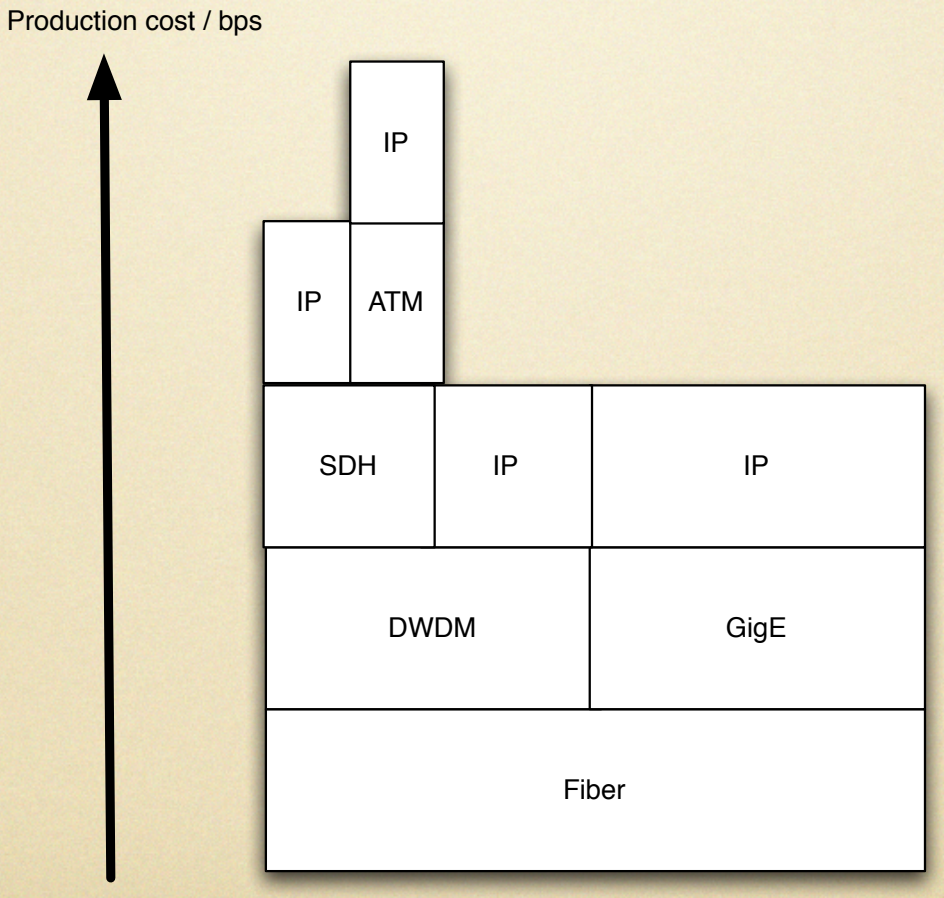
# Cost of capacity

- Build or back-haul?
  - You normally have the option to either build out your POP on location, or back-haul the customer connections to a central location
  - Back-haul gives you two options
    - Do aggregation yourself
    - Buy aggregated back-haul

# Wireless

- Wireless is local-loop and not core infrastructure
  - Wireless is hard to debug and is a shared medium
- Use wireless temporarily to reach new customers
- Migrate to “land-line” infrastructure

# Cost of hardware



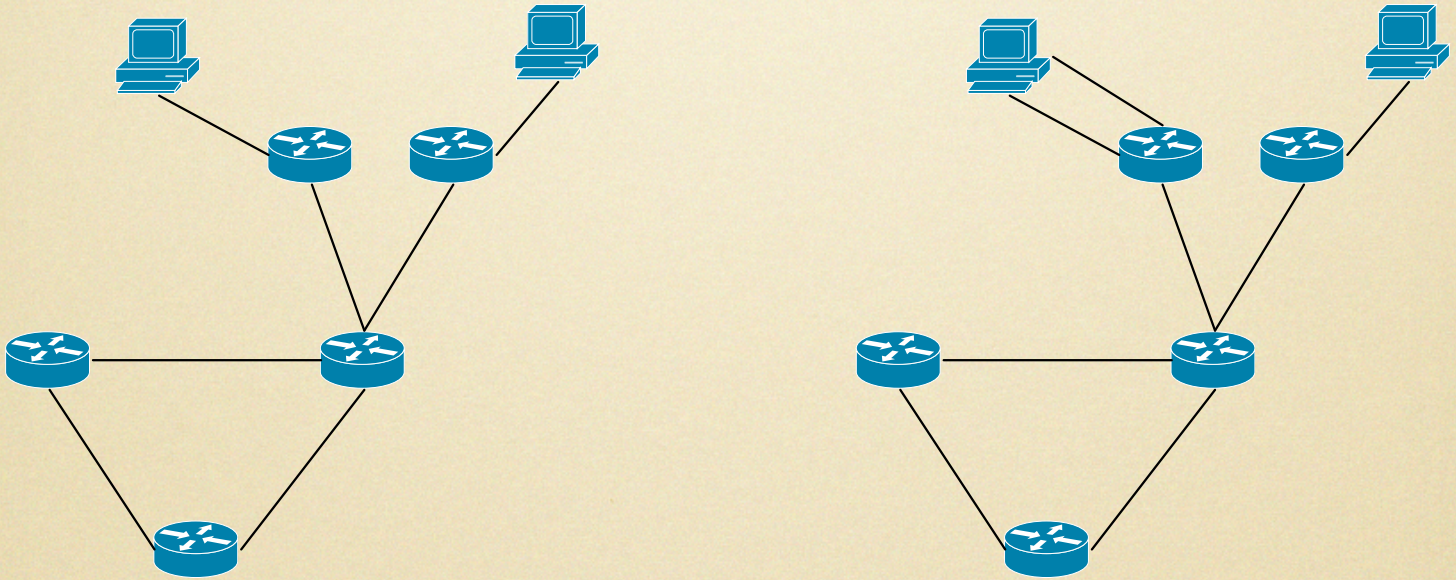
# Cost of hardware

- Your topology will be reflected in your cost for hardware (CAPEX)
  - Needed number of interfaces, types, queuing technology etc.
- But you can also save on your operational expenses (staff etc) with a properly engineered network topology

# Cost of hardware

- Increased network redundancy in your core often means that you
  - Have more time to troubleshoot at failure times
  - Less customer down-time
  - Perhaps do not need 24x7 staff acting on the fault
    - Can wait until day shift comes for example
- Can also allow you to save on OPEX by having cheaper support contracts

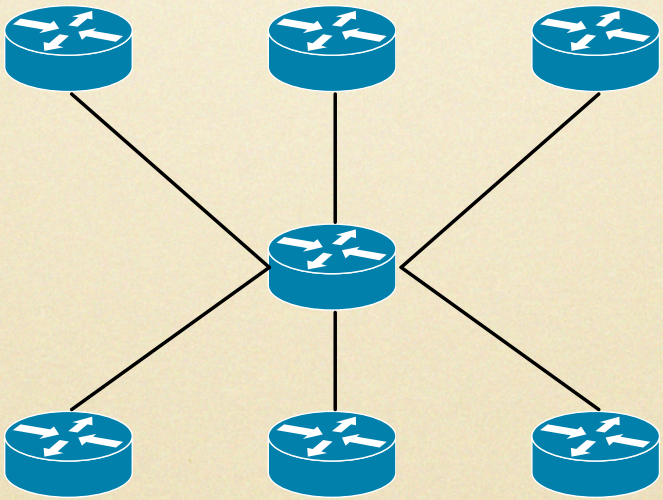
# Network topologies
# Hub-and-spoke



# Network Topology Hub-and-spoke

- Pros
  - Capacity upgrades to a POP is easy
  - Routing will be easy
  - Natural evolution
- Cons
  - No redundancy out to the POPs

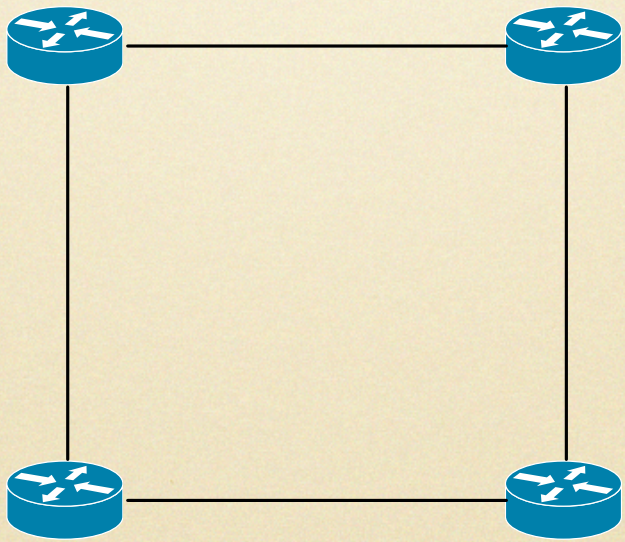
# Network topology
# Star topology



# Network topology
# Star topology

- Pros
  - Central management
  - Easy capacity planning
- Cons
  - Central point of failure
  - Inefficient routing between leaf locations

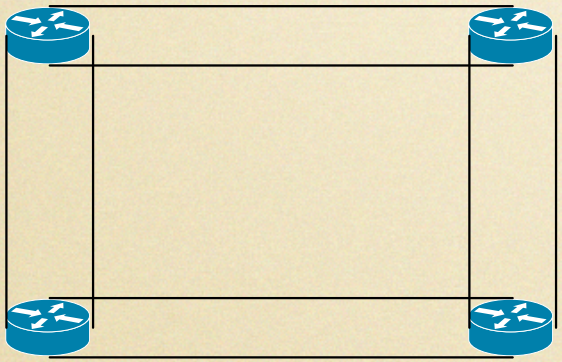
# Network topology
# Ring topology

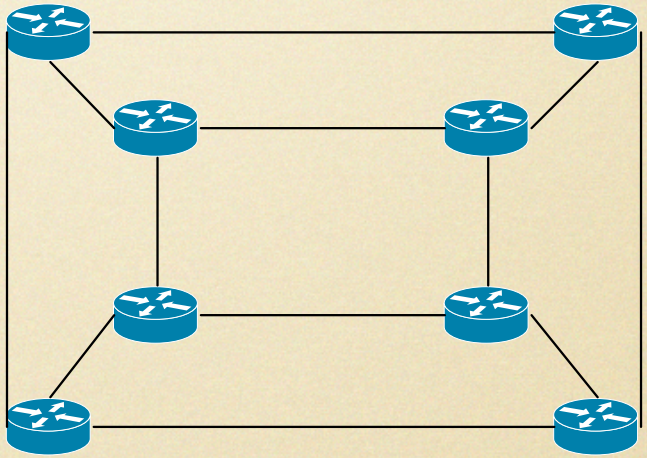


# Network topology
# Ring topology

- Pros
  - Looks cool in diagrams
  - *Can* improve redundancy
- Cons
  - Never actually works very well in practice
  - Capacity planning so you are sure to handle failure modes becomes much harder

# Network topology
# Dual-ring topology





# Network topology
# Dual-ring topology

- Pros
  - Very tolerant to faults
- Cons
  - Most often more complex than worthwhile
  - Routing complexity
  - Capacity planning is even more complex

# Addresses and numbers

# Addresses and numbers

- Address planning
  - Group blocks of addresses to customer locations
    - Will make IGP aggregation easier if needed
    - But the IGP should really only carry infrastructure addresses...(see Philip's excellent tutorial)

# Addresses and numbers

- Apply for your own address block from your RIR/NIR
  - Migrating customers away from upstream address space in the future is a nightmare
- Register sub-allocations with your RIR/NIR
  - Will help others tracing abuse issues
  - Could be a problem with privacy laws
- Use “role” handles
  - Yu do not want to have to update all your RIR objects when one of your staff quits

# Addresses and numbers

- Use “role” addresses
  - Same reason as above
  - You should also have a role address to make sure that mails to contacts for address blocks will always be answered, for example for abuse
- ***Register an abuse contact!***
- Register route objects
  - Others build filters for their incoming routes based on what is registered
  - You should do that as well!
  - Tools like IRRToolset

# Addresses and numbers

- Keep customer records
  - Even for historic data
  - Will help for example in legal investigations
- Keep your records up-to-date

# Routing

# Routing

- Most of what you need to know is in Philip Smith's excellent tutorials
- Do use dynamic routing. Even in small networks it reduces chance of pilot error
- Run BGP to your upstream
- Do not exchange IGP routes with your customers!

# Routing

- Use loopbacks for iBGP
- Filter routes to peers
- Peer as much as possible!
  - Most providers benefit from peering freely
  - Implement the configuration guidelines for peering in Philip Smith's tutorials
  - It will save you money on your transit bill!
  - Take IX locations into account when planning network topology

# Routing

- Do not run the same IGP with customer premises equipment as in your backbone
- You can get better convergence and performance by trimming the routing protocols timers
  - **But be sure you know what you're doing!!!**
  - If values do not match no adjacencies will form

# Services

# Services

- Security
  - Offer basic filtering capabilities
- Mail
  - Do not run an open mail-server!!!
  - Look at grey listing
    - It does not work for everyone
  - Automatic spam filtering (running spamassassin for example) might be an idea

# Services

- DNS
  - Offer DNS services for your customers domains
  - Use two redundant slave-servers that are not on the same subnetwork
    - Might be worth swapping favours with your competitors
      - You will all gain from stability
  - Make sure your resolvers are only accessible by your customers

# Services

- AAA
  - Use an authentication database for your users and logins
  - RADIUS, Kerberos, TACACS+ etc
  - Also use it to authenticate logins on your servers and routers
    - Log all access and parse for anomalies
- Database backends
  - Use a database backend to keep track of users and their permissions
  - Can be integrated with all services, SMTP AUTH, IMAP, router logins, dial-up, Wlan

# Security

# Security

- Most of this covered by Merike, Gaurab and Vicky
- Basic security can be divided into
  - Protecting your customers from others
  - Protecting your own infrastructure
  - Protecting the Internet from your customers

# Security
# Customer protection

- Customer protection
  - Offer customers some sort of basic firewalling, like a standard template ACL for traffic to the customer
    - That for example do not allow Windows filesharing, X-Windows, NFS, etc packets into the customer network
  - Offer SSL transport for SMTP and IMAP/POP

# Customer Protection

- If you wonder why you should provide customers with secure login, I have the passwords of the following accounts in 24h
  - POP: adminc, registry, merike, aslam
  - SNMP: public
  - And a few web-mail accounts

# Customer protection

```
! Customer Incoming ACL
!----------------------
!
! Customer : Kurtis
! Customer Network: 194.15.141.0/24
! Customer Network: 194.15.141.16/30
!
access-list 120 permit ip 194.15.141.64 0.0.0.15 any
access-list 120 permit ip 194.15.141.16 0.0.0.3 any
access-list 120 deny   ip any any log
access-list 121 deny   ip 194.15.141.64 0.0.0.15 any log
```

# Customer protection

```
! Customer Incoming ACL
!---------------------
!
! Customer : Kurtis
! Customer Network: 194.15.141.0/24
! Customer Network: 194.15.141.16/30
!
access-list 121 deny   ip 194.15.141.64 0.0.0.15 any log
access-list 121 permit icmp any any echo
access-list 121 permit icmp any any echo-reply
access-list 121 permit icmp any any administratively-prohibited
access-list 121 permit icmp any any unreachable
access-list 121 permit icmp any any parameter-problem
access-list 121 permit icmp any any packet-too-big
access-list 121 permit icmp any any time-exceeded
access-list 121 permit icmp any any source-quench
access-list 121 permit tcp any any established
access-list 121 permit tcp any eq ftp-data any gt 1023
access-list 121 permit udp any any gt 32768
access-list 121 permit tcp any any eq domain
access-list 121 permit udp any any eq domain
access-list 121 permit udp any eq domain any gt 1023
access-list 121 permit ip any 224.0.0.0 0.255.255.255
access-list 121 permit pim any any
access-list 121 deny   ip any any log-input
```

# Infrastructure protection

- Only allow access to your core routers from networks that your NOC uses
- Only use ssh to access your routers
  - If not supported use jump-hosts that you ssh to
    - Jump host is a trusted secured internal host that can telnet to your routers
- On CPE routers only allow access from the customer network and your own office networks
  - But restrict access methods to telnet or ssh

# Protecting the Internet

- Filter packets coming in from your customers to make sure they can not send spoofed packets
- Filter your outgoing routes
- Filter incoming routes
  - Or As-paths
- Filter out bogons
- Alternatively run RPF

# Security
# Protecting the Internet

- Run netflow on your routers
  - Collect data and analyze it
    - Will help you trace malicious traffic originating inside your network
  - Can also be used for troubleshooting and network planning
  - Use for example flow-tools or cflowd

# IPv6

# IPv6

- Some general remarks
  - If you haven't started playing with IPv6 yet, you should
  - Customer demand is small but there are a lot of developments and it's gaining popularity
  - Launching IPv6 is a good way to improve your knowledge of your IPv4 network
  - Remember: IPv6 is just 96 more bits!

# Network Topology

# Network topology

- Remember, IPv6 is just 96 more bits!
- Your long-term topology doesn't need to be any different than for IPv4
- To avoid routing complexity (or error), the IPv4 and IPv6 topology should match where you have IPv6 services

# Transition

- How do I get from an IPv4 network to an IPv6 network?
  - Transition mechanisms!!!
- You have
  - Customer is IPv6 enabled but not the ISP
  - The ISP has some IPv6 connectivity but not to the PE, and the customer is IPv6 enabled
  - Support for IPv6 at all of the ISP network as well as the customer

# Transition

- The last bullet is the easiest, dual-stack
  - Simply enable both  IPv4 and IPv6
- In the case the end-user supports IPv6 but the ISP doesn't or the ISP just supports it partly some form of tunnelling solution is required
  - Static or dynamic tunnels such as GRE
  - Tunnel solutions such as
    - Teredo
    - 6to4

# Addresses and numbers

# Obtaining IPv6 addresses

- RIR allocation policies are more or less the same
  - Be a LIR, plan to allocate 200 blocks within 2 years, and you will get a /32
- If you are an LIR you should get your allocation and play with it
  - Take the opportunity and talk to the APNIC hostmaster center while you are here!

# Addressing IPv6 networks

- Customers / Customer sites are allocated /48s
- Each network / LAN is assigned a /64
- LIRs can be assigned a larger block than a /32 if they can show the need
  - Based on the HD ratio
  - Hardly likely that anyone would need it though

# Addressing IPv6 networks

- All networks, even point-to-point links are supposed to have /64 blocks
- The remaining 64-bits are EUI64 coded
  - Either hardcoded (unlikely)
  - Or calculated using prefix advertised with “stateless address autoconfiguration”
  - Or calculated using RFC3041 to protect privacy

# Addressing end-users

- Current policy means that :
  - There is no PI address space
  - No other way for end-uses to get addresses except through their upstream provider
- End-users that really needs PI have not went away
- Reasoning behind current policy is to protect against routing table growth
- Work on solving part of the problem is done in the IETF shim6 WG

# Routing

# Routing and IPv6

- The theory is exactly the same
  - Longest prefix match decides on the forwarding path
- There are some pitfalls with using IS-IS
  - IS-IS does not use IP for transport so reachability information might be passed although the topology does not match

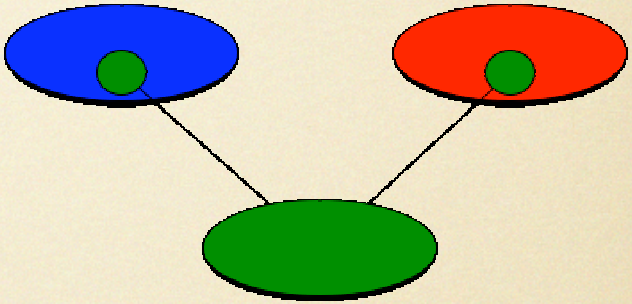
# Routing and IPv6

- OSPF
  - IPv6 support added as OSPFv3
  - Works per interface, also means changes in configurations
- IS-IS
  - Two new TLVs, interface and reachability

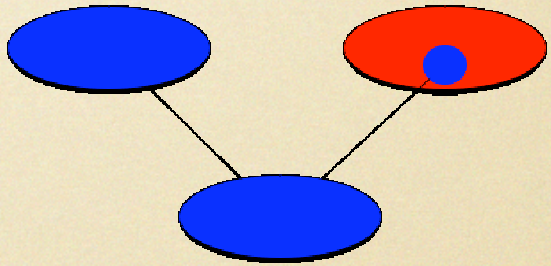
# Routing and IPv6

- In BGP IPv6 information is distributed using BPG Multiprotocol support
- Otherwise more or less the same
  - Remember that router_id is a 32-bit value
- Not BGP specific but multihoming might be more challenging

# Multihoming in IPv6



- AS+PI
  - Get an AS
  - Get PI space
  - Advertise and use BGP



- More specific PA
  - Advertise the more specific route from the ISP

# Multihoming in IPv6

- If customer network does not qualify for a /32 allocation from RIR/NIR
  - Use the second model on previous page
  - Use two PA blocks with two addresses on each server
    - Can create problems when one upstream fails

# Services

# IPv6 Services

- Most operating systems and server software support IPv6 today
- Enabling it is normally is as easy as just configuring it
  - Might require a recompile on some systems
- You will need to tell the world that your hosts support IPv6
  - Insert AAAA records into your DNS zone

# IPv6 DNS entry

- Example DNS entry for a web-server

```
www         A           195.43.225.69
            AAAA        2001:670:87:3001::6
```

# Security

# Security

- MOST IMPORTANT!!
  - ACLs that you have written for IPv4 are... well... written for ***IPv4***. Make sure you have ***IPv6*** ACLs as well!
- Similarly if you use transition mechanisms
  - Watch out for tunnels through your filters and firewalls
  - Watch out for packet spoofing

# Summary

# Summary

- Avoid any form of complexity
  - Use “Occam’s razor” liberally
- Most of the problems or choices you will face have already been solved by someone else
  - There are plenty  of helpful people out there on mailinglist like SANOG