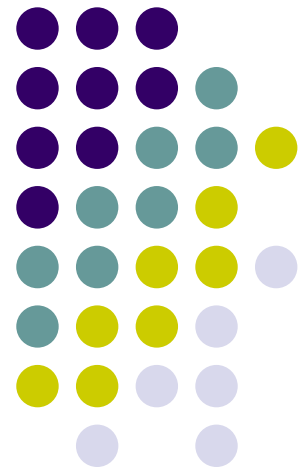


Towards scalable routing for the Internet

Stig Venaas
venaas@uninett.no
UNINETT



The problem



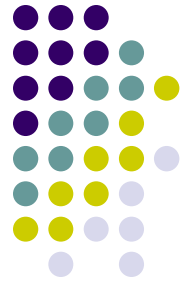
- Recently people have realised that the current Internet routing has to change
- Routing table growth is too large
 - Multihoming is one reason, no provider aggregation
- Increasing number of BGP updates
 - Partly due to lack of aggregation and number of prefixes
 - Also traffic engineering (load balancing with BGP etc)
- The cost of the specialised hardware needed by routers is growing quicker than Moore's Law
 - Also issues with power/heating
- IPv6 is no solution, enterprises are not willing to just use provider assigned addresses and host multi-addressing
 - Also want to change providers without renumbering

The goal



- Find a new way to do routing and/or addressing for the Internet that can scale well into the future
- Multihoming and traffic engineering should be possible
 - Do it near the edge without exposing the Internet core to all details?
- Allow enterprises to have provider independent addressing?
- Needs to work with IPv6 and preferably IPv4
- There should be a simple transition path
 - No flag day
 - Change only parts of the system?
- There is ongoing work in the IETF, in particular in the Routing Research Group in the IRTF
- We will present some of the ideas and proposals

Locators and identifiers



- There is a general agreement that the main problem is IP addresses used as both identifiers and locators
 - An identifier is used to address one specific host
 - Used by transport and application layers
 - A pure identifier should be fixed independent of the location (which network, which provider etc)
 - A multihomed host should still have just one identifier
 - A locator specifies a location, which network, which provider etc
- Locators can be chosen to allow good aggregation

Loc/ID examples



- A typical postal address is e.g. **Jan Modaal**, **Leuk Straatje**, **Amsterdam**, **Netherlands**
- The red part (if unique) would be an identifier
 - The person might move or somehow have two post boxes in two different locations
- The blue part is a locator and can be aggregated (hierarchical)
 - The postal service around the world treats all post to Netherlands the same way, sending it to the same next-hop
 - The postal service in Netherlands can send all Amsterdam post to the same next-hop etc
- When one were looking at IPng (now IPv6) there were proposals like GSE/8+8 for having IP addresses containing prefix and identifier
 - With IPv6 stateless address autoconfig we almost have this
 - **2001:db8:10c:2:1234:56ff:fe78:9abc**
 - However the host treats the entire address as an identifier
 - Many people want the prefix part to be provider independent

Loc/ID alternatives



- Split handled by end hosts
 - The ID may be in lower bits of the address (ref prev slide)
 - The IP address (in packet header) may be locator only
 - Identifiers in e.g. extension header (HIP)
- Split handled by routers
 - First/last-hop routers can rewrite the locator parts of the addresses as needed (e.g. with GSE/8+8)
 - First-hop router may encapsulate the packet with a locator specifying the last-hop router in the outer header
 - Last-hop decapsulates the packet, forwarding the payload to the local host
 - Alternatively do the same at e.g. site border routers where the identifiers are routeable within the site
 - Not quite Id/loc, more hierarchical addressing...

Six/One



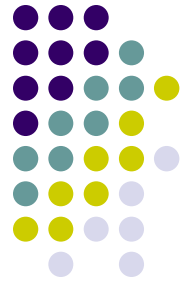
- Host based Loc/ID split with IPv6 (similar to shim6)
- Each host has an address bunch
 - An address from each provider where all share the same last 64 bits (interface identifier)
- Hosts that are to communicate exchange bunches
- They start using one address from each bunch
- Edge routers may rewrite the first 64 bits in order to choose a specific path
- Some context is added to packets
 - Helps the stack write the address back to the original before handing it to the application
 - Transport/application layer uses all 128 bits as an identifier
- Hosts can do later rewriting so edge routers don't have to rewrite every packet

Six/One comments



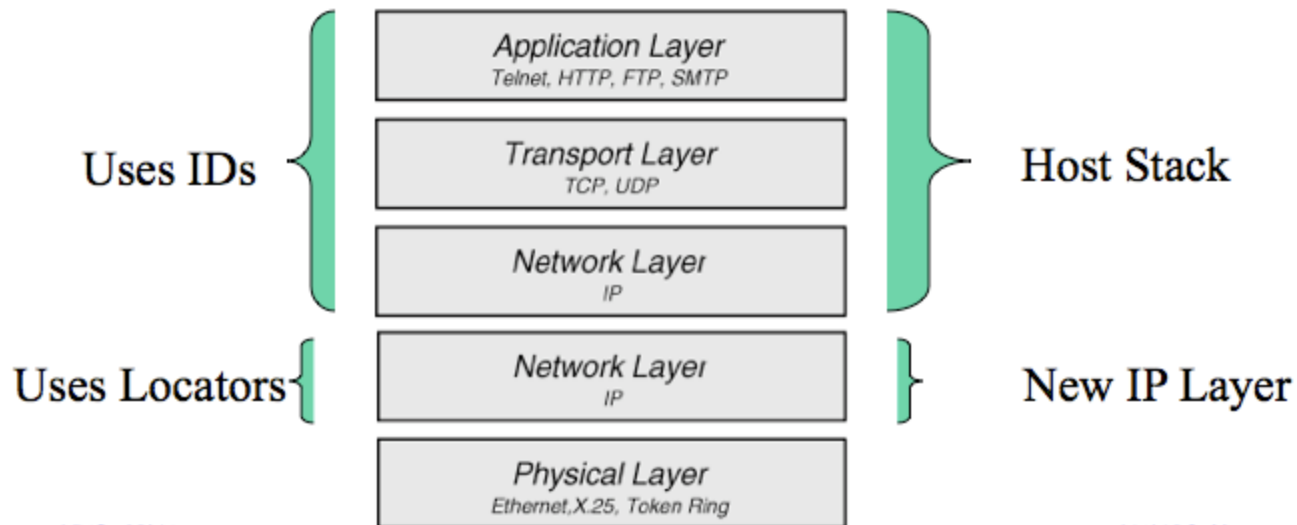
- IPv6 only
- Not a complete Loc/ID split
 - For a host all addresses from a bunch would typically be in the DNS
 - The last 64 bits are the same and uses CGA to prove cryptographically to the peer which addresses are in the bunch
 - One of the addresses will be used as an identifier by applications/transport
 - First 64 bits of each are locators
- Use of CGA avoids the problem of mapping from ID to locator that many other solutions use
 - But means you cannot add new locators during the session since it would require a new identifier
- Only hosts stacks need to be upgraded
 - And edge routers where rewriting is desired
- Easy transition, hosts that use Six/One can communicate with those that don't

Loc/ID split with encapsulation



Locator/ID Split?

- One solution: split the functions -- This is at the heart of the Locator/ID split idea
 - So how might we achieve this?
- Architecturally, we might try to "Jack-up" the existing IP layer

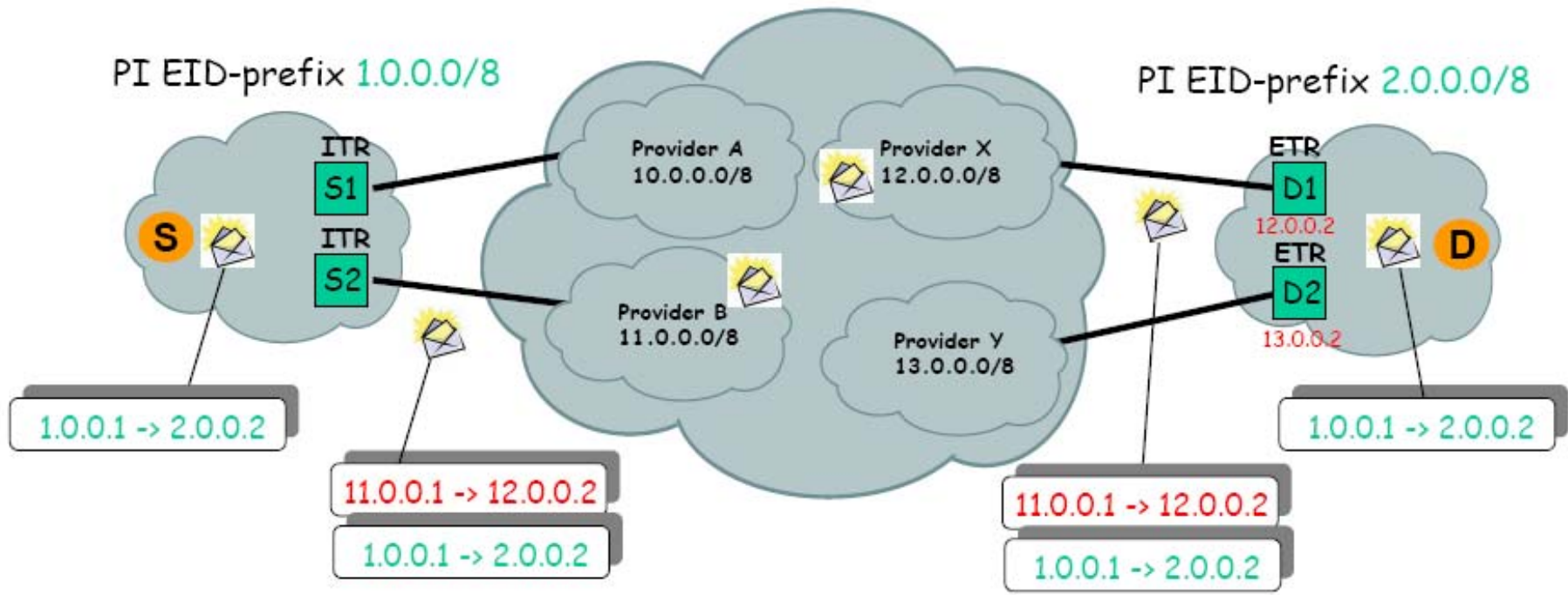


LISP-CONS

NANOG 41

Thanks to Dave Meyer for the slide

LISP



Legend:

EIDs -> Green

Locators -> Red

DNS: D -> 2.0.0.2

Mapping Entry

EID-prefix: 2.0.0.0/8

Locator-set:

12.0.0.2, priority: 1, weight: 50 (D1)

13.0.0.2, priority: 1, weight: 50 (D2)

LISP: A Level of Indirection for Routing

Nanog 41 - ABQ

Slide 12

Thanks to Dino Farinacci for the slide

ID to locator mapping

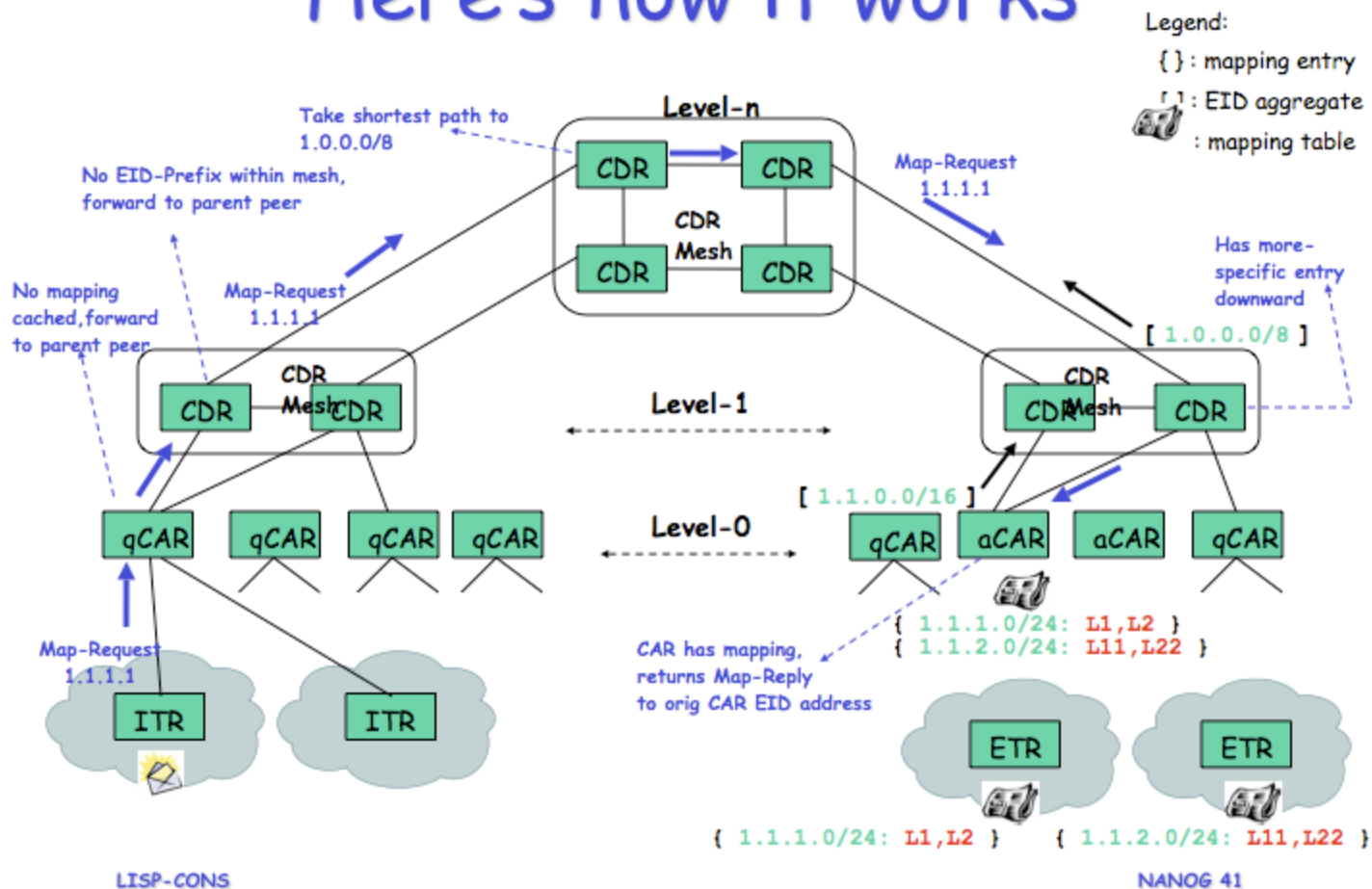


- The hardest part of the Id/loc splitting is how to map identifiers to locators
- Scaling
 - Trade-off of size (amount of state), refresh times and latency
 - Can end system or edge router have full knowledge? How to maintain the information?
 - Can it be requested when needed?
 - Delay or drop data packets until known?
 - Caching may be of some help
 - How quickly may mappings change to provide traffic engineering or some degree of mobility?
- Security
 - How to know that the locator is correct? Can traffic be hijacked?

LISP-CONS ID to locator mapping



Here's how it works



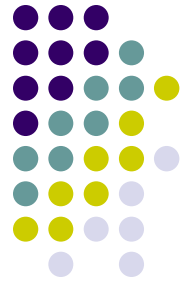
Thanks to Dave Meyer for the slide

Further reading



- Problem statement
 - From IAB routing workshop Amsterdam Oct 2006
 - <http://tools.ietf.org/html/draft-iab-raws-report-02>
- Six/One
 - <http://tools.ietf.org/html/draft-vogt-rrg-six-one-00>
- LISP
 - <http://tools.ietf.org/html/draft-farinacci-lisp-04>
- LISP-CONS
 - <http://tools.ietf.org/html/draft-meyer-lisp-cons-02>

RiNG – Routing in Next Generation



- Routing cluster project
 - Cluster meetings for routing related projects
- Cooperation with other existing initiatives and experts
- Trying to be a bridge between researchers, engineers and operators
 - A lot of on-going research from the pure theoretical to more applied research
- <http://www.ist-ring.eu>
- Reading list on routing
 - <http://wiki.ist-ring.eu/action.php?n=Wiki.Wiki>