IPv6 with IPv4 infrastucture: ISATAP, BGP/IGP tunneling

Gérard GASTAUD
BND CTO Office, standards

- **ISATAP**
  - Draft-ietf-ngtrans-isatap.03
- **BGP, IGP tunnel**
  - draft-ietf-ngtrans-bgp-tunnel.04, WG LC
  - draft-many-ngtrans-connect-ipv6-igp-01

ISATAP

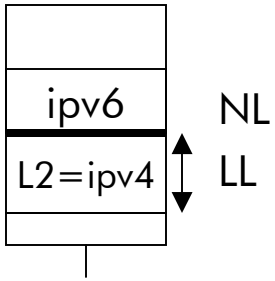
Intra-Site Automatic Tunnel Addressing Protocol

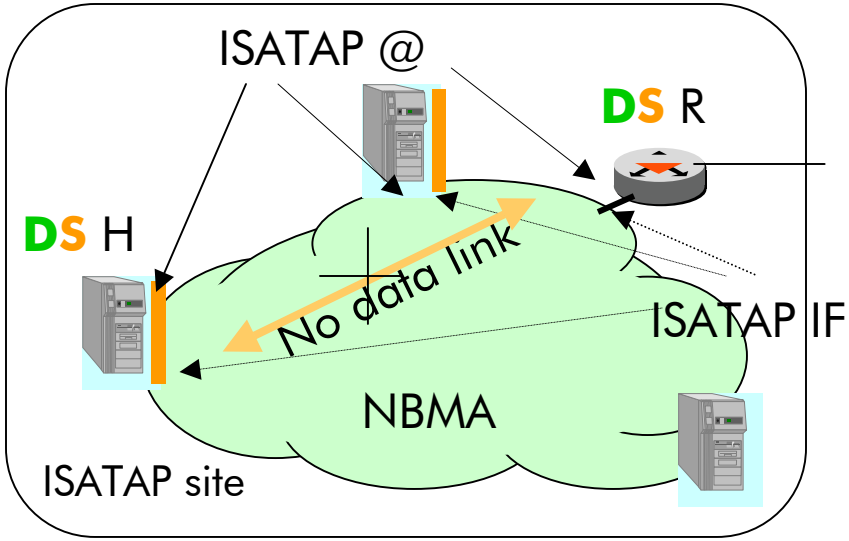
ISATAP (Ice-a-tap) overview

- It **connects isolated dual stack nodes within IPv4 sites via automatic tunnelling** (IPv6 in IPv4)
  - uses an **ISATAP aggregatable global unicast @** embedding a node's IPv4 @
  - treats site's **IPv4 infrastructure as an NBMA Link Layer** for IPv6 using automatic IPv6-in-IPv4 tunneling (i.e., no configured tunnel state)
- ISATAP enables **incremental deployment** of IPv6 Hs within IPv4 sites with **no aggregation scaling issues at border** gateways
- **ISATAP scope: a site intranet**
  - during the IPv4 to IPv6 co-existence phase, sites will deploy IPv6 by increments within their IPv4 interior routing domains
  - ISATAP requires **no** special IPv4 services within the site (e.g. multicast)
- ISATAP supports stateless @ auto-configuration & manual configuration
  - ISATAP supports networks that use non-globally unique IPv4 @ (e.g., private @ but the virtual ISATAP link cannot span a NAT)
- It is compatible with other NGTRANS mechanisms (e.g., 6TO4)

ISATAP terminology

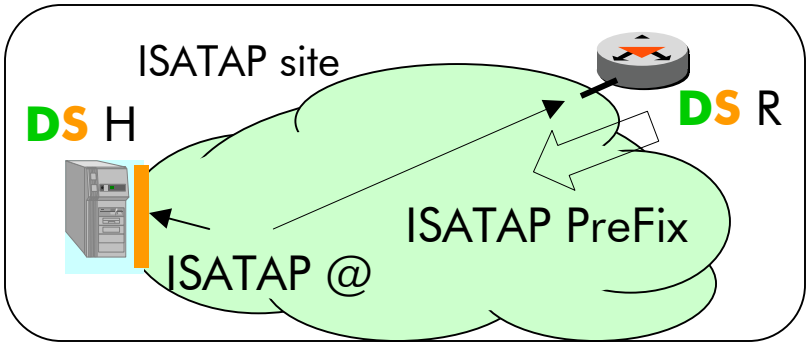
- **LINK**
  - communication medium over which nodes can communicate at link layer e.g. Ethernet, PPP link, IP tunnel
- **Underlying link**
  - a Link Layer supporting IPv4 (for ISATAP), and may be IPv6
- **ISATAP link**
  - one or more underlying links used for IPv4 tunneling. The IPv4 NL @ses of the underlying links are used as LL @ses on the ISATAP link
- **ISATAP interface**
  - Point of attachment to an ISATP link



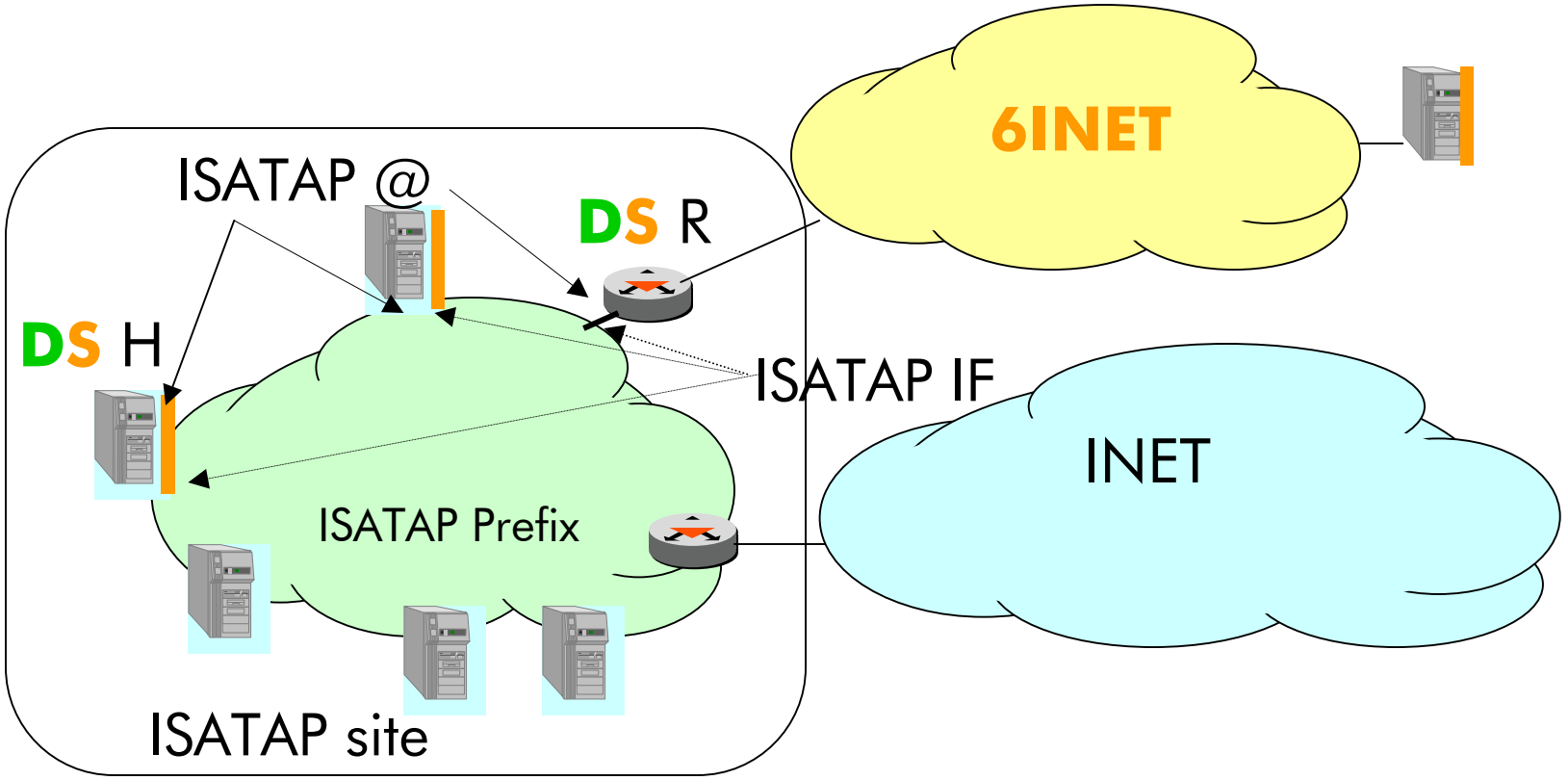


ISATAP terminology

- **ISATAP prefix (PF)**
  - Any local/site/**globally aggregatable 64-bit** IPv6 **PF** (e.g from a native IPv6 ISP) **reserved** by a local network administrator specifically **for ISATAP purposes**
  - It is used to **configure ISATAP @ ONLY** (& not native IPv6 @)
- **ISATAP @**: v6 @
  - with **ISATAP PF and an IPv4 @ embedded** in the interface identifier PF::0:5EFE:a.b.c.d (a.b.c.d IPv4@ of ISATAP link, u=local)
- **ISATAP R** : IPv6 R supporting an ISATAP interface over which it forwards Ps not addressed to itself
  - **normally an interior R** within an heterogeneous v6/v4 network
- **ISATAP H:** An IPv6 H which has an ISATAP interface, & is not a R



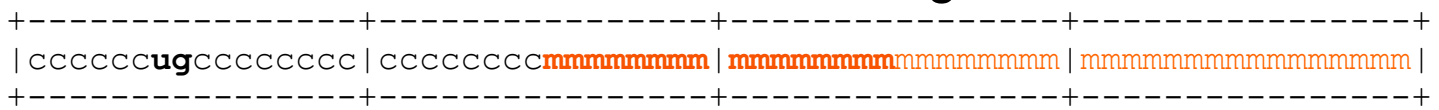
scope



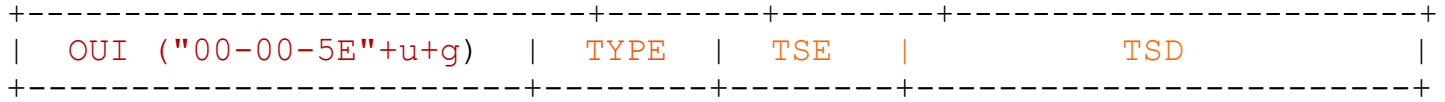
- H does not share a common multiple access data link with DS R, which has access to 6INET (directly or not) but cannot multicast RS because network is NMBA
  - H can't receive RAdv, and R can't receive RSol via Neighbor Discovery

ISATAP address (@)

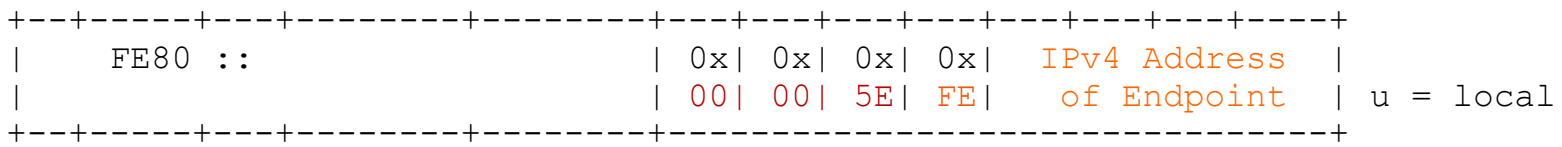
- EUI format as used for IPv6 addressing



- **OUI** (Organizationally-Unique Identifier)= cc..cc (24 bits)= company ID, e.g. IANA ("00-00-5E")
  - g=group/individual, u= global/local scope (0 = local for ISATAP)
- mm..mm= **extension identifier**, assigned within OUI **= TYPE-TSE-TSD**, its format allows future expansion
  - TYPE indicates how (TSE, TSD) are interpreted (1 octet)
  - TSE Type-Specific Extension (1 octet)
  - TSD Type-Specific Data (3 octets)
  - **TYPE= FF, TSE= FE -> legacy EUI-48,**
  - **FE, TSE+TSD = embedded v4 @**



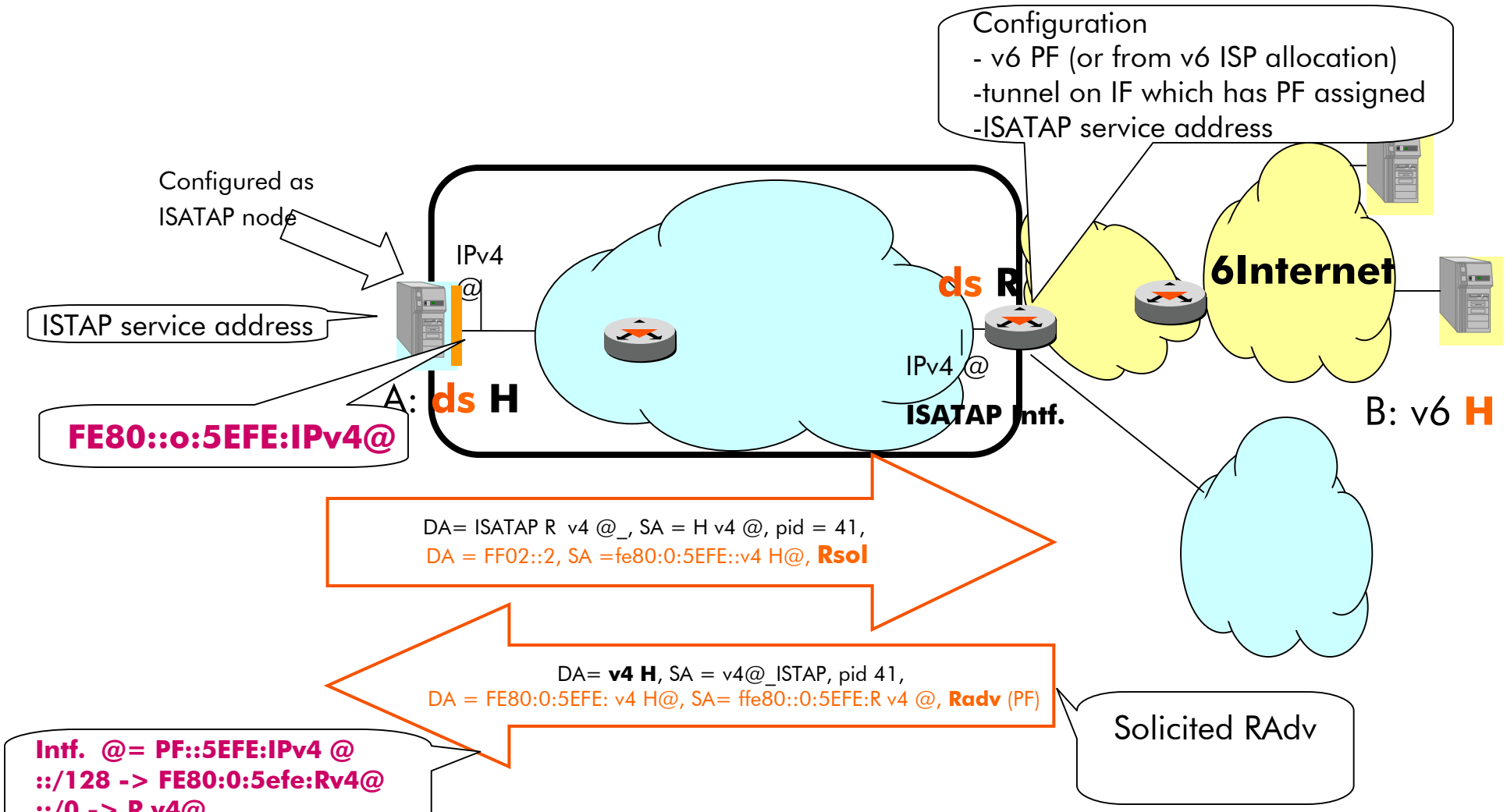
e.g. Link Loacal ISATP @



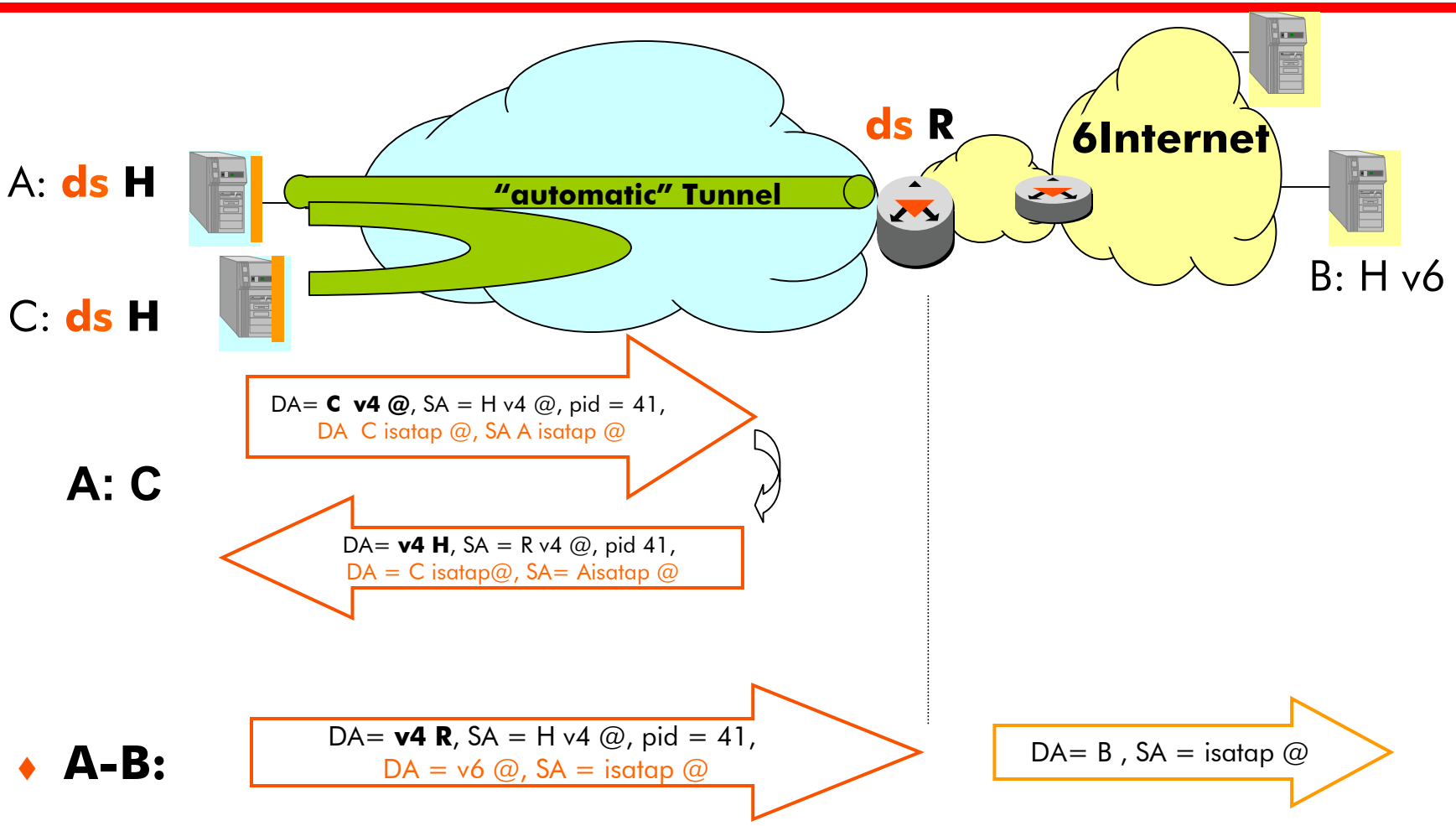
operation

- The DS H
  - If it does not receives native ND (or is configured an ISATAP Int.)
    - builds a Link Local v6 @ for its ISATAP interface (FE80::0:5EFF:v4@H)
    - Gets R v4@ of gateway to 6INETvia DNS well-known service ("ISATAP" service)
    - Builds a potential router list (PRL)with these addresses
    - Then sends a Nsol encapsulated in IPv4 (SA H v4@),DA R v4@) to each of these GWs
  - Upon receiving R's RAdv with the ISATAP PF (e.g. given by ISP), H
    - builds a global unicast @: (PF::0:5EFF:v@h)
    - sets up a **default route** for IPv6 **to this R as Next Hop**
    - Is then ready to send packets to other ISATAP H or to 6Inet
- **Sending rule**
  - When sending a P on an ISATAP interface, check that the IPv6 next hop is an ISATAP @ else discard and send ICMP destination unreachable indication
- **Receiving rule**
  - silently discard any received Ps that do not satisfy ONE OF the checks
    - **source is on-link (**SA: prefix configured on ISATAP interface and an ISATAP-EUI)
    - **previous hop is an on-link ISATAP router (v4 SA in** Potential Routers List)
  - H checks that a received RAdv has a SA found in the PRL (else silent discard)
- No special @ selection rule

ISATAP illustrated

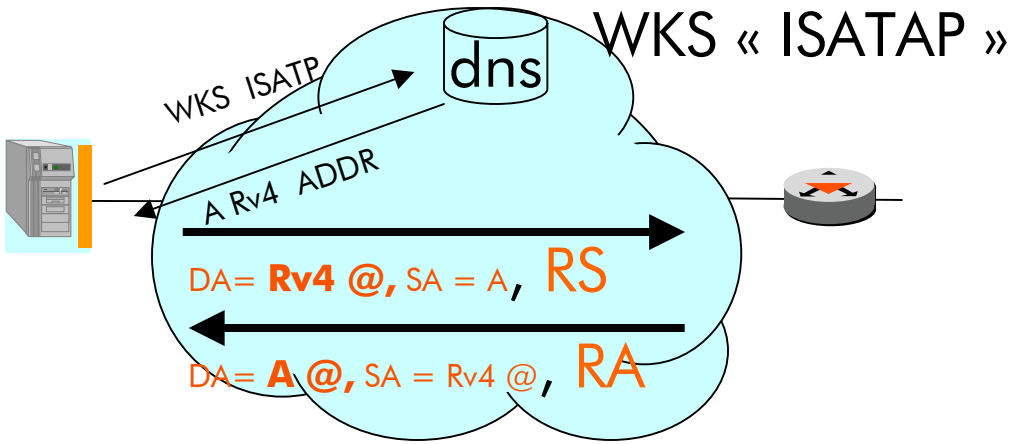


ISATAP: tunneling



- **Correspondent nodes** that
  - **share a common ISATAP PF** communicate using their **ISATAP interfaces**
  - do **not share a common ISATAP PF** communicate via **standard IPv6 routing**

H: How to find ISATAP GW/R?



- The H ask for the WKS name "ISATAP" to the DNS which returns a list of ISATAP gateways (A RR, not SRV RR): query is done over IPv4
  - IPv4 @ of ISTAP routers' advertising interfaces
  - H adds to the PRL the @(ses) returned
  - And then sends a encapsulated unicast Rsol to this address
- DNS should be updated with ISATAP v4 @ (automatically or manually) when an ISATP R is configured in the site

ISATAP vs ND

- **Redirect, Neighbor Unreachability Detection, and next-hop determination = ND**
- **Address resolution and mechanisms for RS/ RA** = **ISATAP** in accordance with ND (ISATAP link is NBMA)
- **Address Resolution (IP @ -> LL @)**
  - IPv6 @ are resolved statically to LL @ (IPv4) by a computation, i.e., last 4 octets treated as an IPv4 @
- **R & PF discovery**
  - Prefix List and Default Router List as specified in ND
  - a NEW data structure **"Potential Router List"** (PRL) per ISATAP link
    - initialized when a node enables an ISATAP link, with IPv4 addresses discovered through name service lookups for the Well-Known Service name "ISATAP"
      - Then periodically updated (ResolveInterval) to detect additions/deletions for the PRL

Data Structures for ISATAP

- PRL (new)
  - Entry:
    - **<IPv4 @ of a R's ISATAP interface, timer for polling>**
    - IPv4 @ (likely to be an "advertising interface") is used to construct the ISATAP LL @ for that interface
    - ResolveInterval (a new configuration variable for ISATAP link)
      - **time between DNS WKS resolutions** (default 1hour)
  - Initialization of the PRL through static IPv4 address assignments
    - Static initialization possible, but periodic WKS above is preferred
  - PRL = context for R discovery and trust basis for R validation
- Neighbor cache, prefix list, default R list, destination cache
  - As in ND

H and R specifics

- R : advertising ISATAP interfaces as for ND except
  - periodic RAdv not required: interval timer is not used
  - RSol unicast in response to R: SA = Link Local ISATAP @
  - RAdv consistency of ND not supported by default
    - Since by default RA are not received by ISATAP Rs
- H
  - Initialize PRL when ISATP link is configured
    - a new entry is polled after a short delay as per ND
  - Periodically, refreshes the PRL via WKS look-up
  - Periodically poll PRL entries the poll timer of which expires, resulting in refreshing the prefix/default router list
    - unicast RSol (DA = fe80::0:5FEFE: v4ADD(PRL entry)
  - Polling timer for entries
    - first set to MinRouterSolicitInterval (between sending RS, = 15 minutes)
    - When a RAdv from the R's in the PRL entry, set to max of 0,5 *(max R lifetime or valid LFT of PFs on the link) or MinRouterSolicitInterval

Deployment and administration

- **Deployment**
  - Underlying link supports both v4 for ISATAP & native v6
    - ISATAP enabled if native v6 does not receive RAdv
    - Later if H receives native RA, polling for PRL is disabled and ISATAP addresses will be gradually deleted (per ND)
- **Administration**
  - ISATAP link: set of Router interfaces and set of nodes which have those addresses in their PRL (administrative, not physical boundaries)
  - ISATAP interfaces assigned
    - one per LL @ (@ of each ISATAP interface added to the PRL)
    - or single interface for multiple LL @ses
      - interface accepts ISATAP Ps addressed to any of the IPv4 LL@ses, chooses one as its primary @, used as SA and represented in the Potential Routers List
  - MinRouterSolicitInterval to control case of M Rs and Ns
  - ISATAP WKS update for the site

Security

- Possible attacks against IPv6 and IPv4
- Many security considerations in 6o4 apply to ISATAP
- IPv4 site security management
  - border gateways SHOULD implement filtering (to detect spoofed IPv4 SA) and ip-protocol-41 filtering
    - IPv4 SA filtering makes ISATP checks effective
    - Unless filtering for pid-41 is not correctly implemented, IPv6 SA is possible, but this can be eliminated if both IPv4 SA filtering, and ISATAP validity checks are implemented
- ND: nodes trust RAdv received from on-link Rs, (255 in 'hop-limit')
  - HL is not decremented when pid 41 Ps traverse multiple IPv4 hops => ISATAP links require a different trust model
    - ONLY those RAdv received from a member of the PRL are trusted; all others are silently discarded (predicated on IPv4 SA filtering
- ISATAP @ format: no privacy extensions for stateless @ auto-configuration
  - ISATAP interface identifier derived from the node's IPv4 @=> not the same level of privacy concerns as IPv6 @ derived from the MAC @

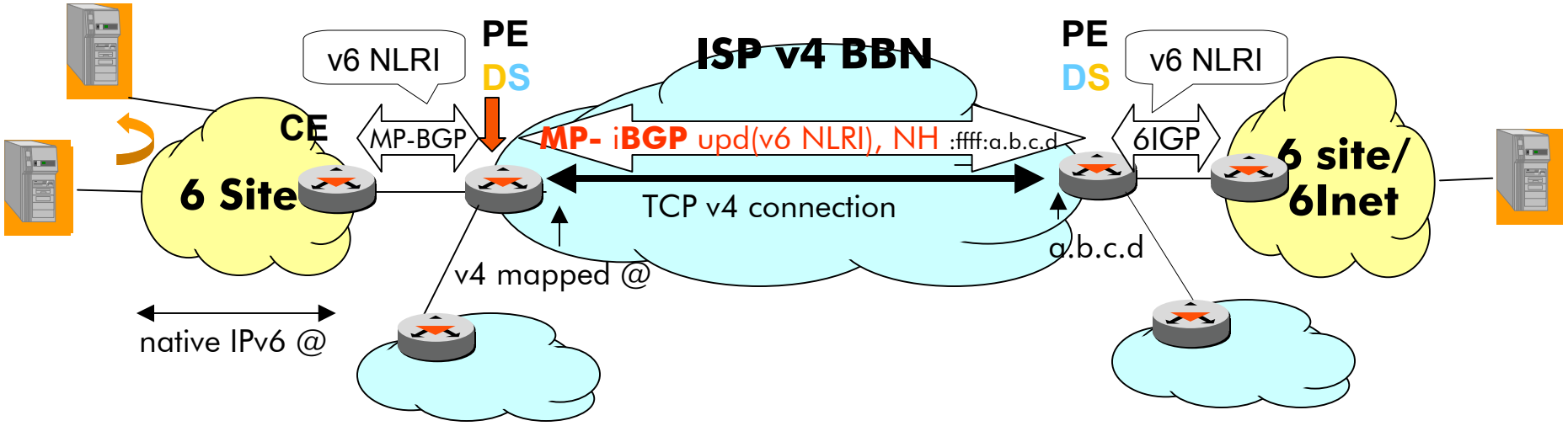
Interaction: ISATAP- other mechanisms

- ISATAP-: DSTM, NAT-PT, 6to4
  - concurrent use of a) DNS, b) of both mechanisms by a H/R, c) both mechanisms by 2 different applications
- **ISATAP-DSTM**
  - a) no issue, b) no issue, c) no issue (proper DNS configuration)
- **ISATAP-6to4**
  - A) no issue, b) no issue, no issue (need a route between ISATAP R and 6to4 R
- **ISATAP-NAT-PT**
  - a)
    - v6 ISATAP -> v4 only: DNS via DNS-ALG
    - v4 H-> ISATAP v6: via ISATAP v4 @, or via NAT-PT but then v4 -> v4 fails: => need several DNS SRVs, ISATAP nodes in separate sub-domain, communicating with primary DNS via NAT-PT
  - b) no issue, c) no issue no issue ISATAP-> IPv4 via NAT-PT, v4 -> ISATAP via proper DNS configuration

BGP Tunnel

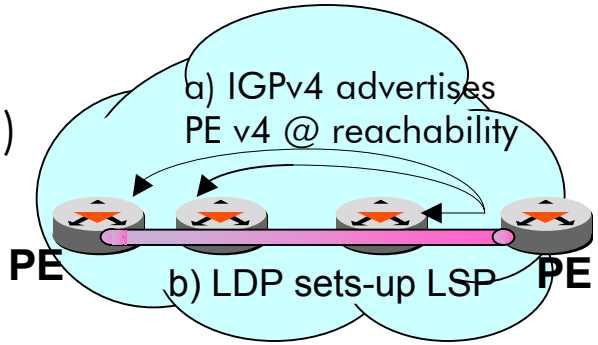
BGP tunneling: MP-BGP over IPv4

- Provider edge routers (PEs)
  - Dual Stack and use **MP-BGP sessions over TCP4** to advertise the sites PreFixes, inside the ISP backbone (BBN)
- the **v6 v4 mapped address** (@) of a PE is used in I MP-BGP update for the NLRI's NH i.e. the iBGP peer
  - ISP BBN may use route reflectors
  - one IPv4 @ per v6 site mono-homing to ISP

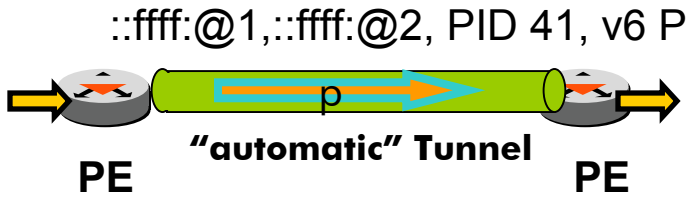


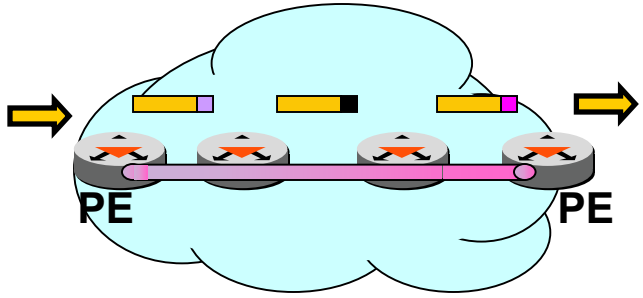
BGP Tunnel

- Automatic Tunnels
  - v6 IPv4 mapped @ ::ffff:IPv4 , for MP-BGP NH for v6 NLRI
  - LSP (v6 Packet **directly** encapsulated in MPLS)
    - IGP establishes the LSP for the FEC of the peer (peer's IPv4 @ is advertised by IGP)
    - 1 label OR
    - 2 labels (MP-BGP update binds a label to v6 NLRI)
      - allows PHP (Penultimate Hop Popping)
        - no IPv6 in penultimate LSR
  - GRE



- NLRI - AFI = IPv6 (2)
  - SAFI
    - IPv4, GRE LSP (mono label) : unicast, multicast or both (1, 2 or 3)
    - LSP (multi label) : label (4) or "VPN" (128)



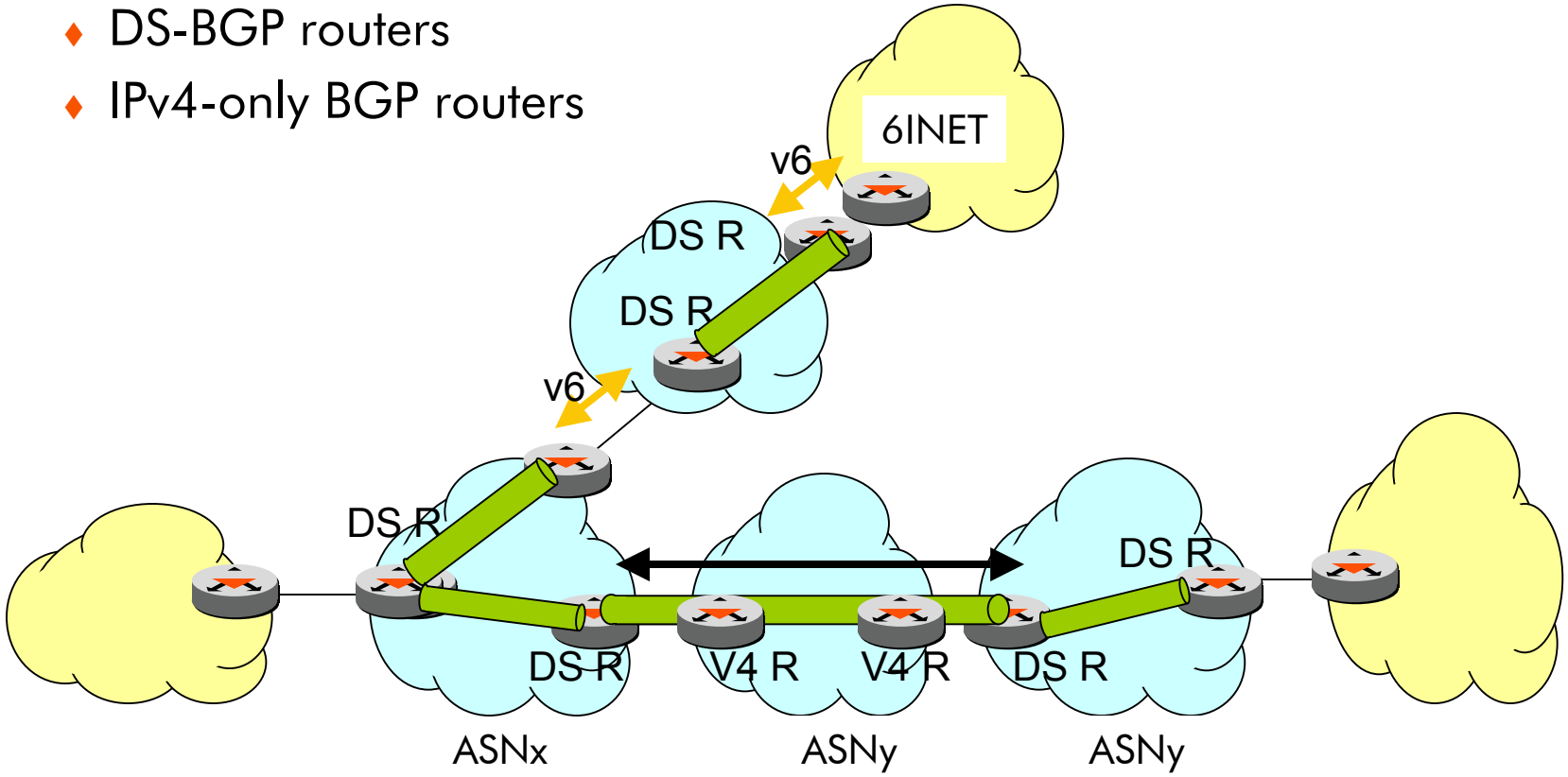


MP-BGP over IPv6 approach

- DS-BGP routers (PE) MUST run MP-BGP over an IPv6 stack (MP-BGP/TCP/IPv6)
- PE sends to its i BGP peer its IPv6 @ as the BGP Next Hop for the v6 NLRI
- transport of MP-BGP messages as well as IPv6 packets over the IPv4 cloud
  - any existing NGTRANS tunneling technique (6TO4, ISATAP, etc.)
  - IPv6 @ of BGP Next Hop matches the actual NGTRANS tunneling technique used in the BBN
  - ingress DS-BGP Router MUST tunnel IPv6 data over the IPv4 cloud towards the Egress DS-BGP Router using the tunneling technique in the BBN applied to the IPv6 address advertised as the BGP Next Hop for the corresponding IPv6 prefix

Inter domain

- border routers between the IPv4 domains are
  - DS-BGP routers
  - IPv4-only BGP routers



tunneling

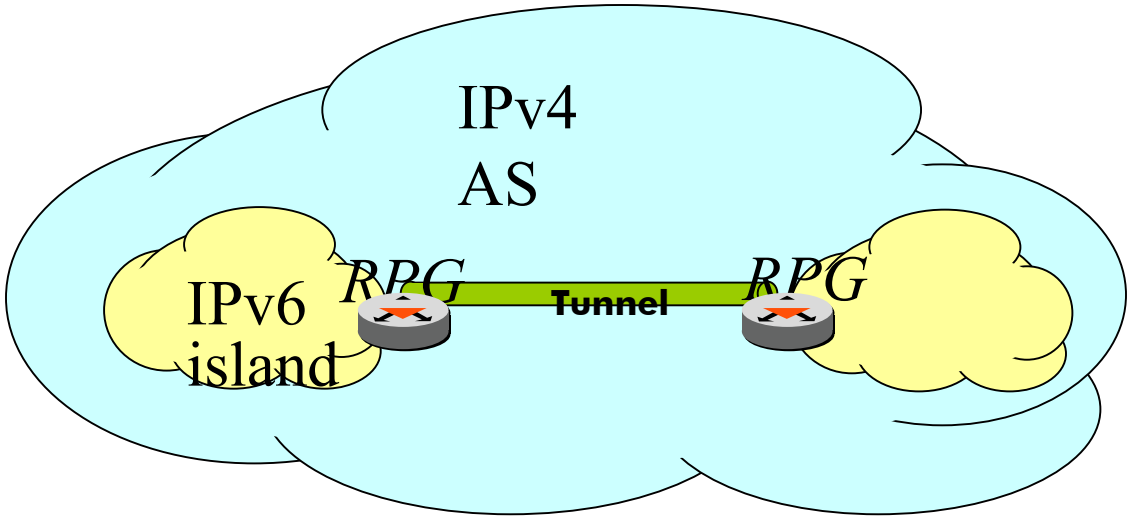
- a. egress sends a labeled BGP route => ingress uses by default an LSP (advertised label (bottom))
  - BGP extended community for tunnel types can indicate other possibilities by order of preference
- b. egress sends an unlabeled BGP route => by default, ingress uses IP in IP e.g. 6 in 4 pid 41
  - BGP extended community for tunnel types can indicate other possibilities by order of preference
- Use of BGP extended communities
  - the egress BGP could use a « tunnel type extended community » to signal the type of tunneling it supports/wishes, thus allowing automatic configuration at both ends

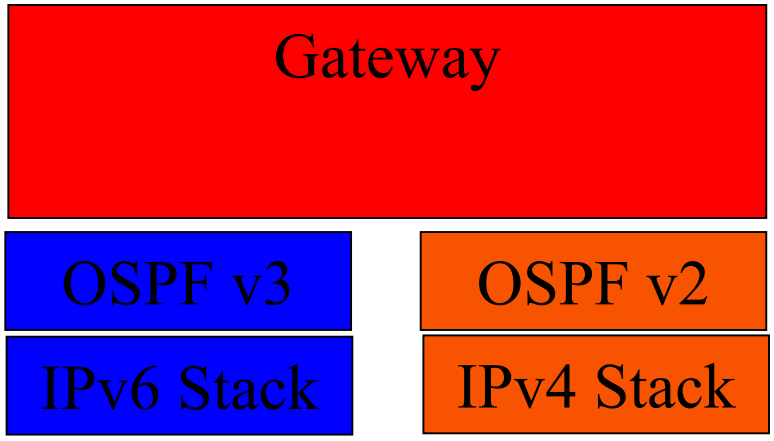
BGP tunnel applicability/benefits

- **For ISPs**
  - That use BGP, possibly offer VPN services
  - No need to upgrade nor change configuration in infrastructure
  - Upgrade only on the edge (upgrade of PE to DS or add separate v6 PEs)
  - IPv6 supported simultaneously with existing services
    - MPLS, MPLS v4_VPNs, QoS, ATM, v4 Internet, …)
  - BGP tunnel allows IPv6 to be deployed over existing infrastructure with **minimal operational impact/cost/risk**
    - It can be easily extended to support IPv6 VPN services (a la RFC2457bis)
      - draft-ietf-ppvpn-bgp-ipv6-vpn-01.txt
- **For sites**
  - v6 CE only has a single routing peer (PE) irregardless of how many remote v6 CEs it communicates with
  - No change on a v6 CE when remote CEs are added/removed
    - reachability is automatically learnt
  - No tunnel /"circuit" to be configured
- BGP tunnel: **flexibility/scalability/low cost & risk**

IGP tunnel

- routing protocol gateway (RPG)
  - Dual stack, transfer information from the IPv6 IGP to the IPv4 IGP and reciprocally
  - The IPv4 tunnel end point address, the IPv6 islands prefixes, and associated metrics are transferred via the IPv4 IGP (flooded in the IPv4 AS via opaque LSA)





ALCATEL

QUESTIONS ?