Port Range Proposals

Minneapolis / 2008.11.20

Gabor Bajko <Gabor.Bajko@nokia.com>
Randy Bush <randy@psg.com>
Rémi Després <remi.despres@free.fr>
Pierre Levis <pierre.levis@orange-ftgroup.com>
Olaf Maennel <olaf@maennel.net>
Teemu Savolainen <teemu.savolainen@nokia.com>

Problem Statement

Providers will not have enough IPv4 space to give one IPv4 address to each CPE or terminal so that every consumer has usable IPv4 connectivity.

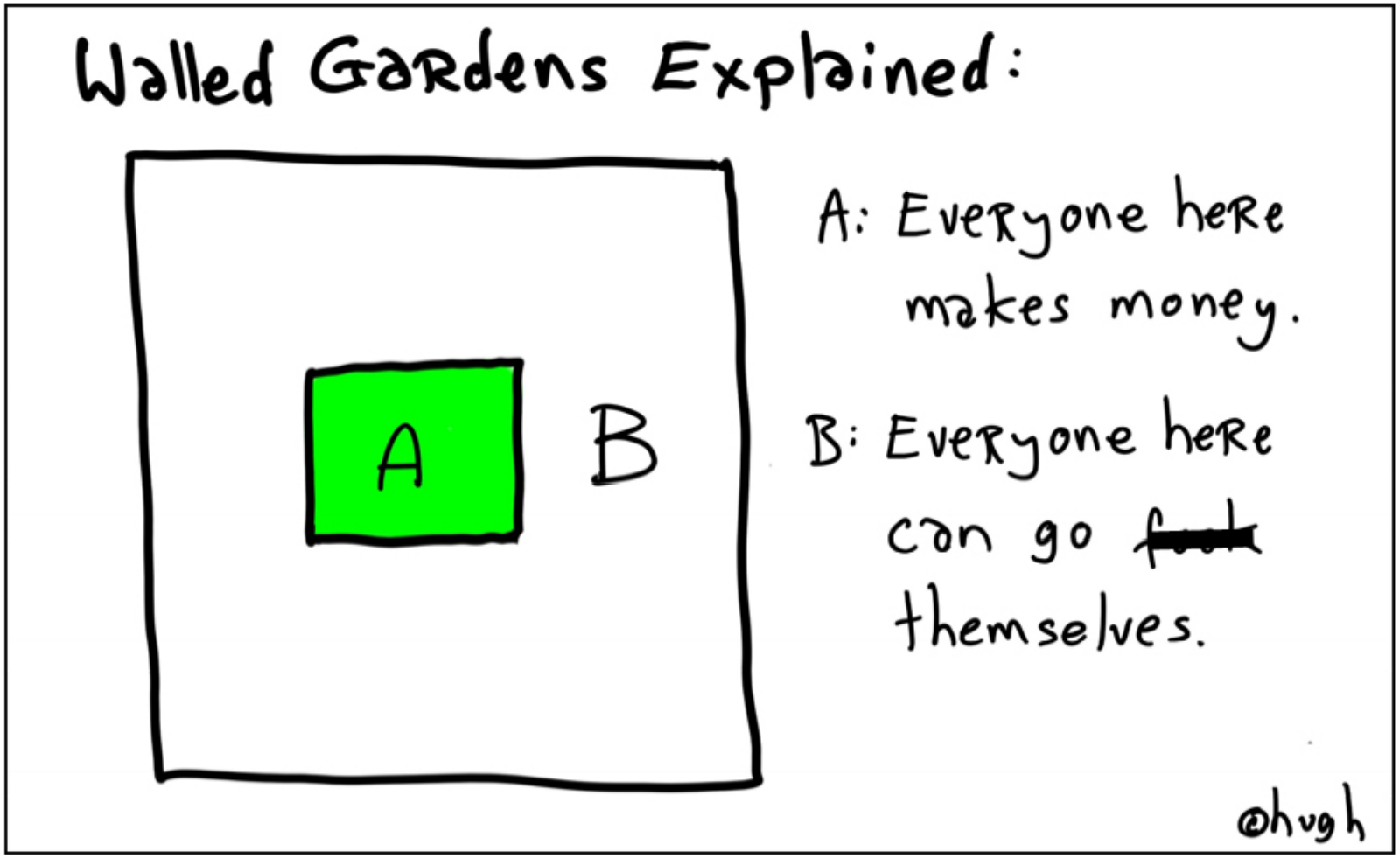
Carrier Grade NAT

- NAT in the core of the provider's network
- Customer has 4to6 NAT and the core re-NATs 6to4 for v4 destinations

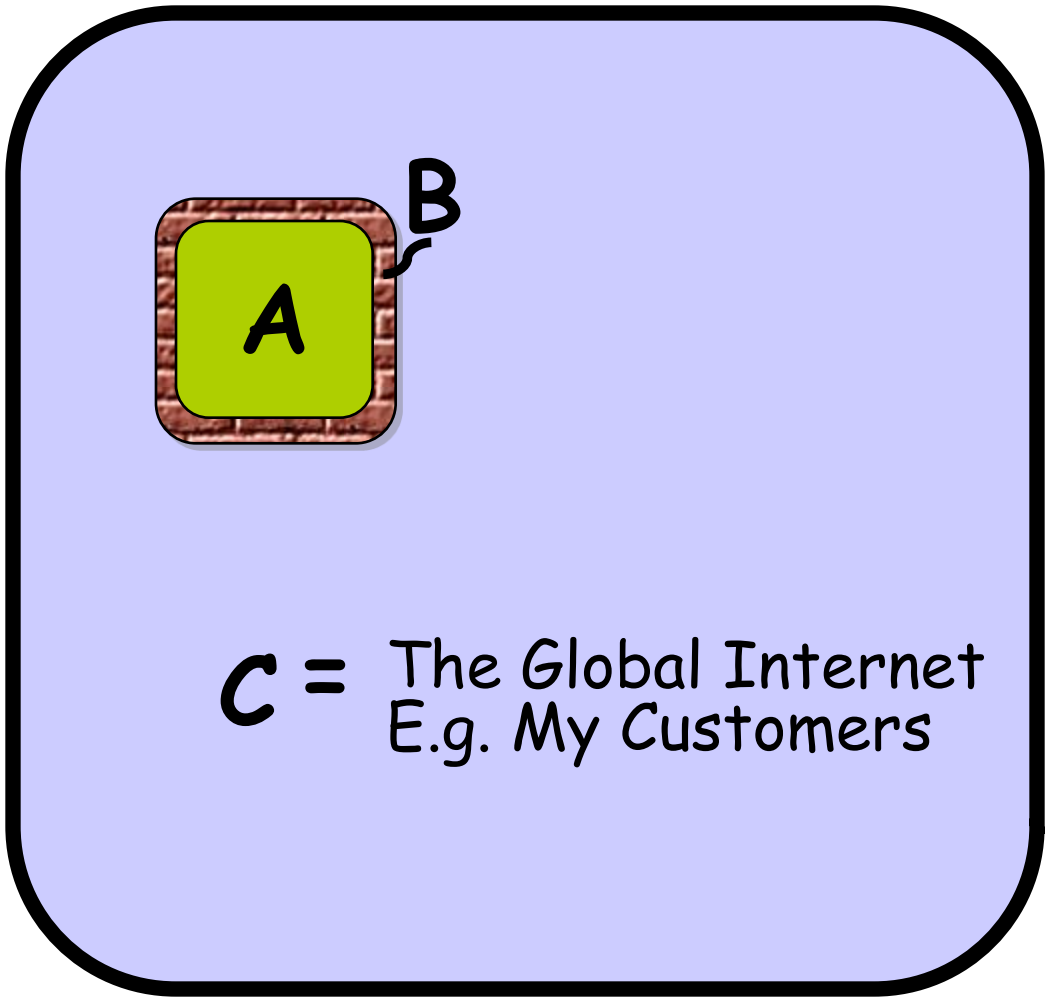
CGN Breaks the Net

- Not only does this cause problems for the carrier, but also for the whole net, as these captive customers can not try or use new disruptive technology
- NAT in middle of net has the problems of a smart core
- Walled gardens here we go!

I Googled "Walled Garden"



Walled Garden Re-Explained



A: Isolated, exploited, & restricted

B: Everyone here makes money

C: Everyone here can go fsck themselves

This Need Not Be Inevitable

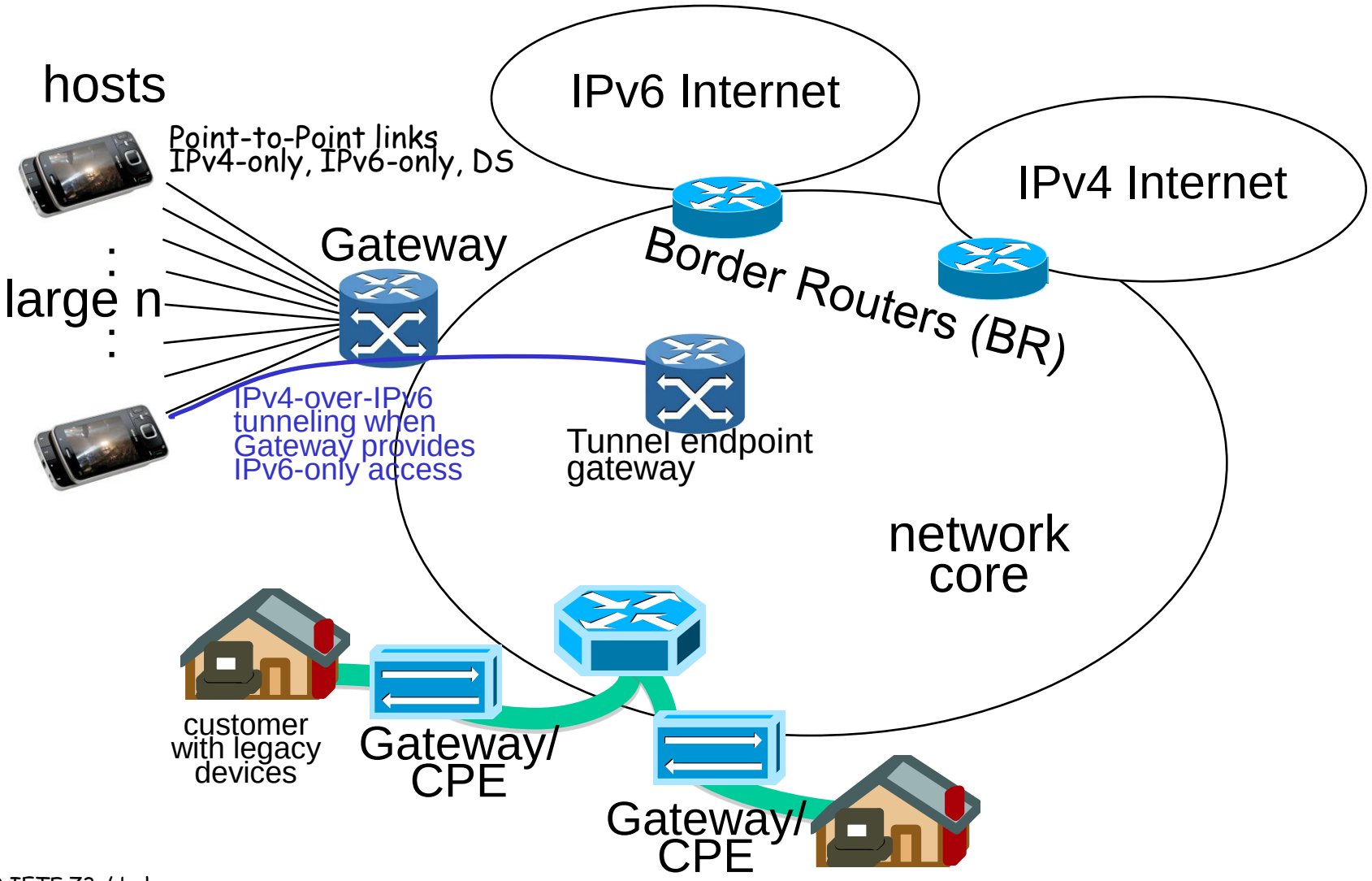
Move the NAT to the Gateway/CPE

As Alain Says

"It is expected that the home gateway is either software upgradable, replaceable or provided by the service provider as part of a new contract."

Constraints for possible solutions

Terminology



Constraints (I)

1) **Incremental deployability and backward compatibility.**

   The approaches shall be transparent to unaware users. **Devices or existing applications** shall be able to work without modification. Emergence of new applications shall not be limited.

2) **End-to-end is under customer control**

   Customers shall have the possibility to send/receive packets unmodified and deploy new application protocols at will.

3) **End-to-end transparency through multiple intermediate devices.**

   Multiple gateways should be able to operate in sequence along one data path without interfering with each other.

4) **Highly-scalable and state-less core.**

   No state should be kept inside the ISP's network.

Constraints (II)

5) **Efficiency vs. complexity**
   Operator has the flexibility to trade off between port multiplexing efficiency (CGN) and scalability + end-to-end transparency (port range).

6) **Automatic configuration/administration**.
   There should be no need for customers to call the ISP and tell them that they are operating their own gateway devices.

7) **"Double-NAT" shall be avoided.**
   Based on constraint 3 multiple gateway devices might be present in a path, and once one has done some translation, those packets should not be re-translated.

8) **Legal traceability**
   ISPs must be able to provide the identity of a customer from the knowledge of the IPv4 public address and the port. This should have the lowest impact possible on the storage and the IS

9) **IPv6 deployment should be encouraged.**

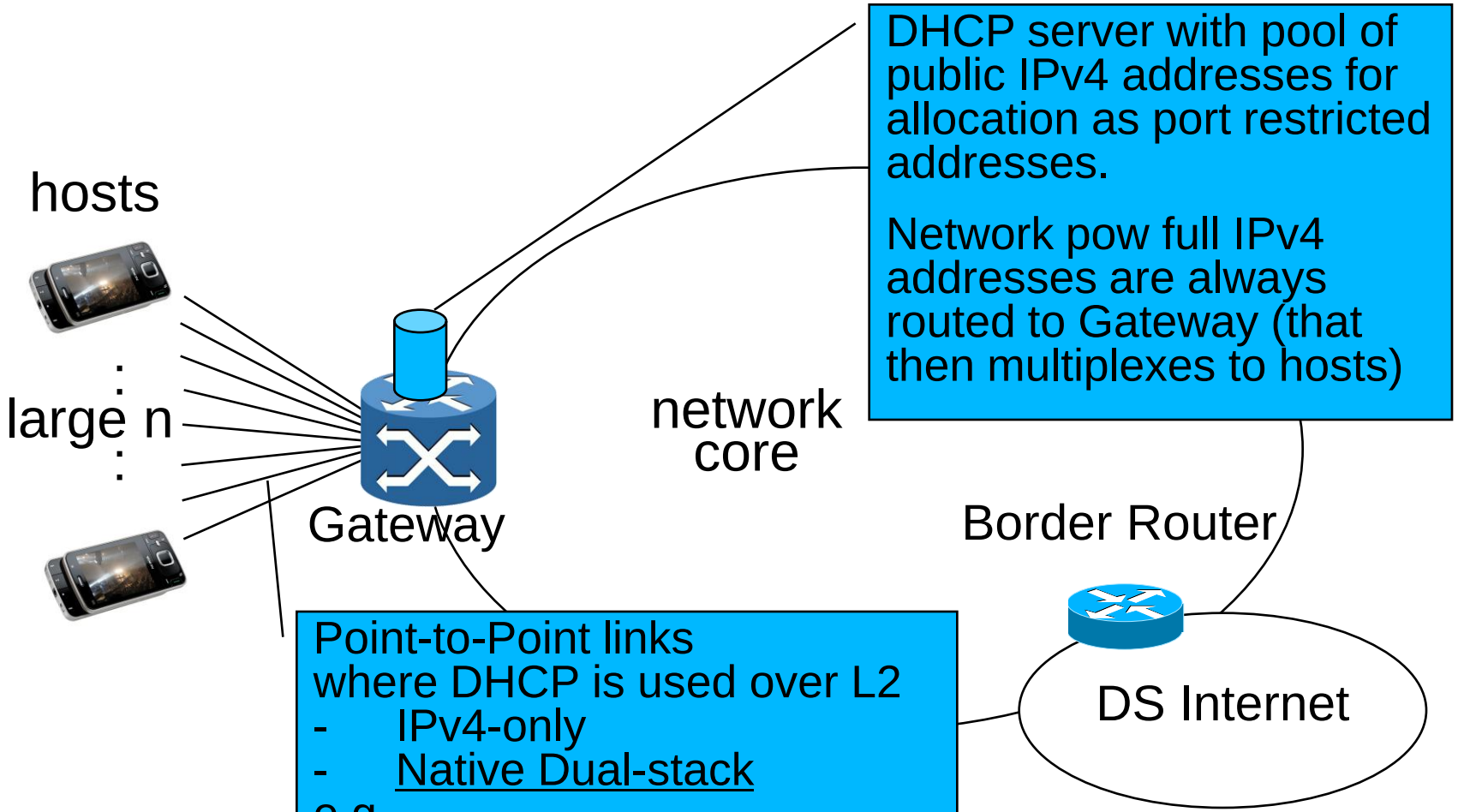
Proposals in short

draft-bajko-v6ops-port-restricted-ipaddr-assign

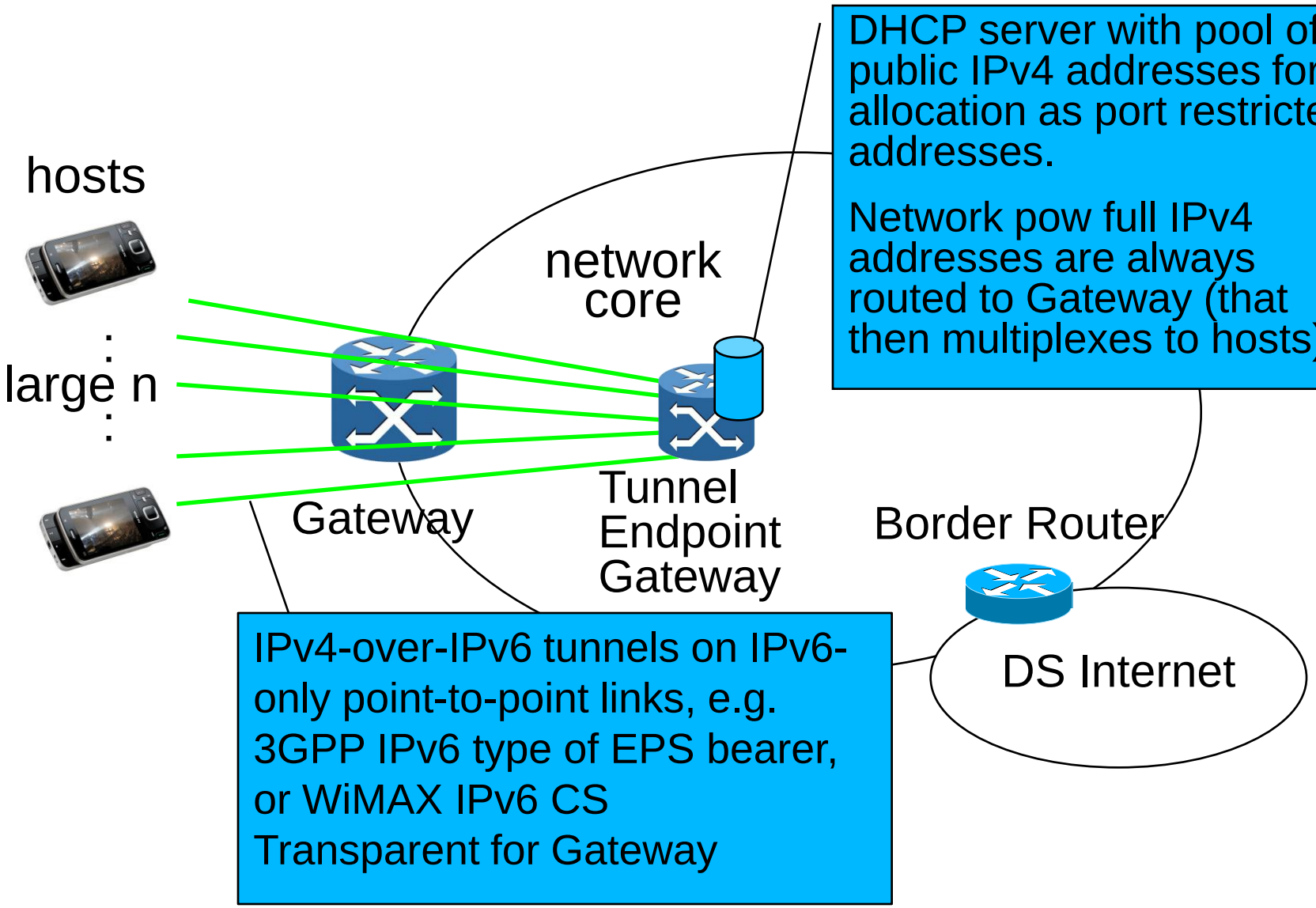
draft-bajko-v6ops-port-restricted-ipaddr-assign

- For **tightly controlled networks**
  - Where hosts can be modified and modifications mandated
  - Cellular networks are the particular example
- Mainly for **point-to-point** links
  - **Physical access links** (L2): e.g. 3GPP IPv4 EPS bearer, WiMAX Forum IPv4 CS
  - **IPv4-over-IPv6 tunneled access links** (L3): e.g. IPv6 clouds, IPv6 PPP, IPv6 EPS bearer, IPv6 CS
- To allow **NAT-less communication**
  - To save on **BATTERY** and complexity

Physical point-to-point links – with or w/o IPv6



Tunneled point-to-point links – over IPv6



About gateway functionality

- Gateway has a pool of public IPv4 addresses
- Gateway can also be acting as a NAT for legacy hosts (CGN)
- Gateway can allocate port-restricted IPv4 addresses and multiplex by ports
- Same stands for both first hop Gateway and Tunnel Endpoint Gateway

Gateway multiplexing tables

- For physical link scenario

Point-to-point link	Public address + port range
Link 1	129.0.0.1 / 5000-5999
Link 2	129.0.0.1 / 6000-6999

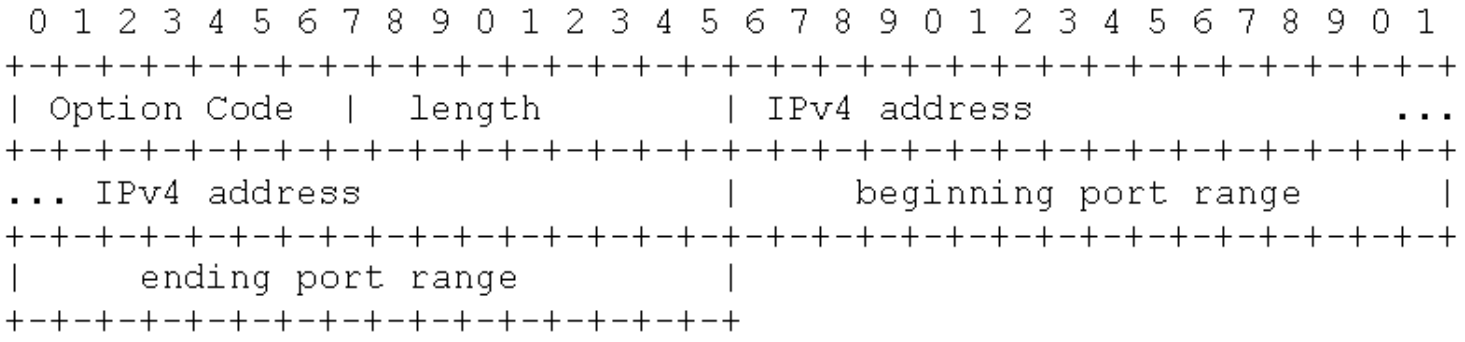
- For tunneled link scenario

Point-to-point tunnel	Public address + port range
Tunnel 1	129.0.0.1 / 5000-5999
Tunnel 2	129.0.0.1 / 6000-6999

  - Very similar to multiplexing done in NATs, except only encapsulation here

DHCP option use and contents

- In case IPv4 connectivity is needed, host requests IPv4 address with OPTION-IPv4-RPR to indicate capability for port-restricted IP addresses
- On ***presence*** of OPTION-IPv4-RPR DHCP server offers OPTION-IPv4-OPR and ‘yiaddr’ of ‘0.0.0.0’
- On ***absence*** of OPTION-IPv4 RPR server allocates full public or private IP address



NAT in a host

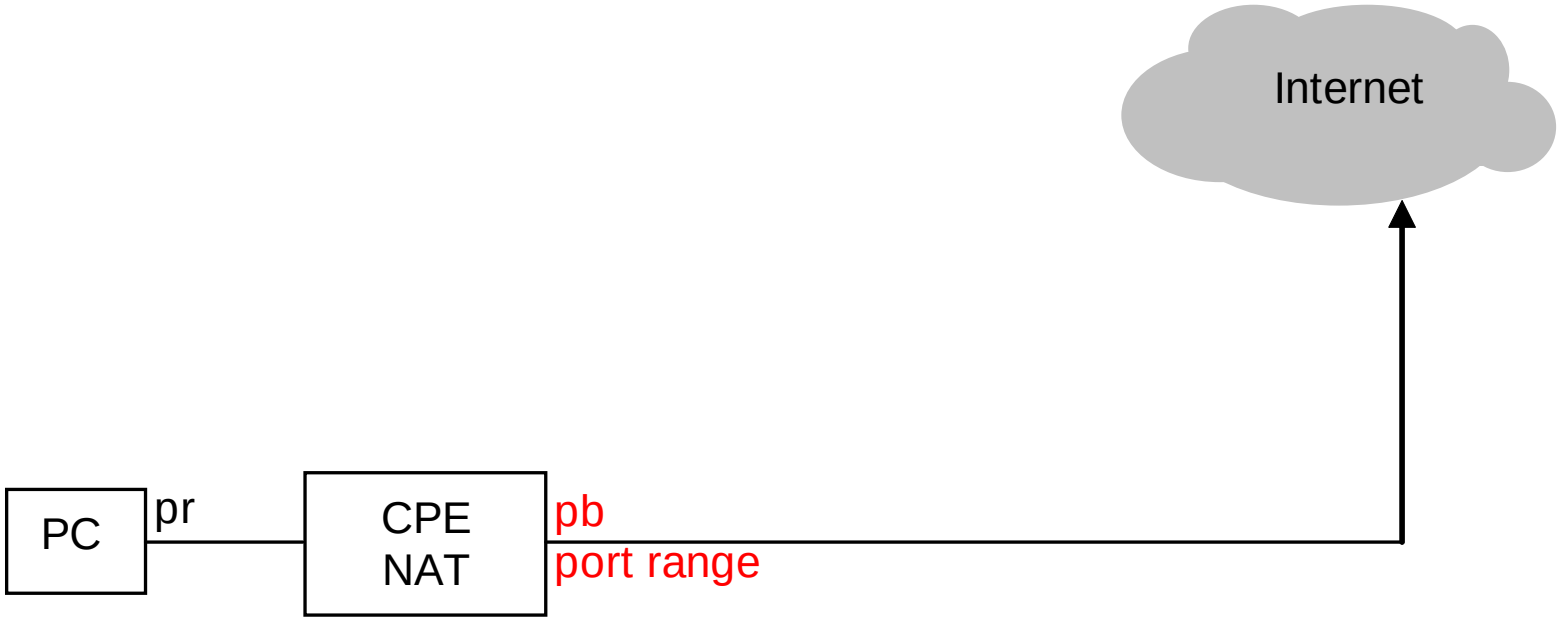
- Hides port-restricted IPv4 addresses from the users and applications
- Distributes NAT functionality to very edges
- Allows host local optimizations for NAT traversal
- Allows NAT control protocols

draft-boucadair-port-range
draft-boucadair-dhc-port-range

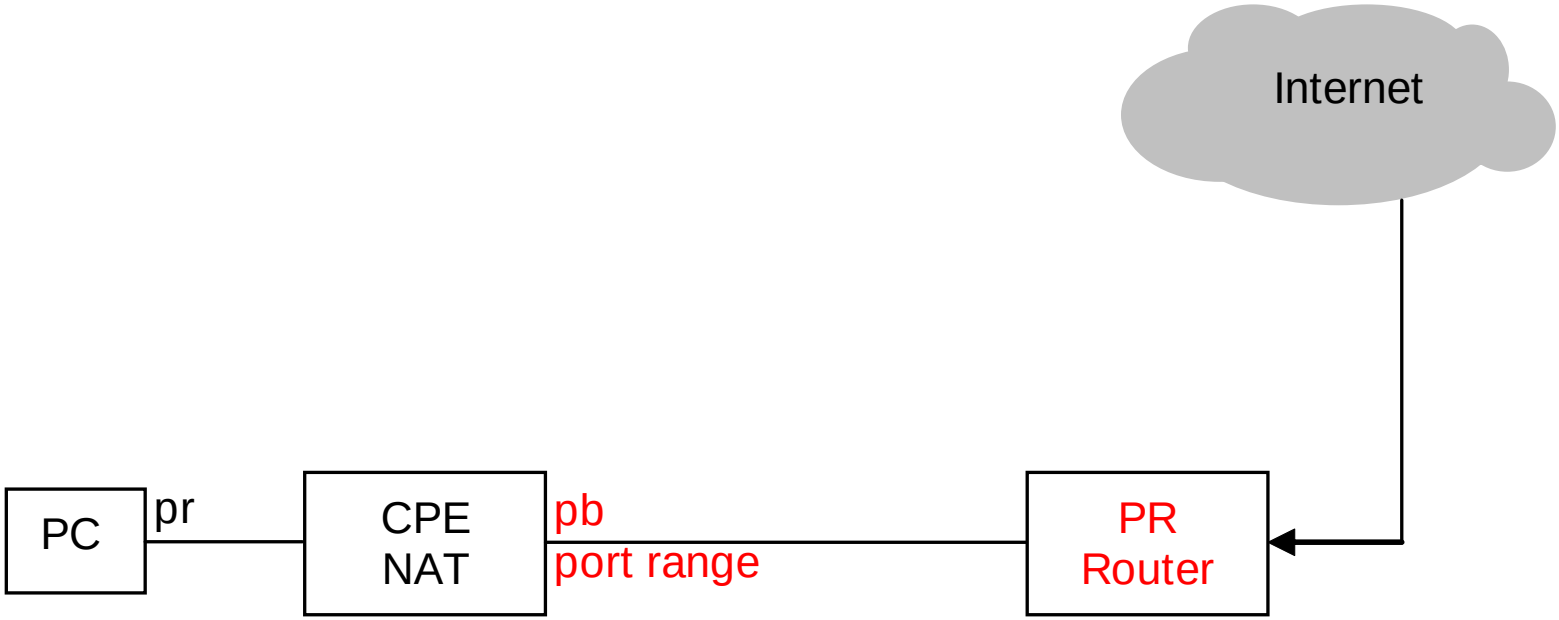
draft-boucadair-port-range
draft-boucadair-dhc-port-range

- Solution Space:
  - Fixed broadband network
  - Residential customers
  - CPEs provided by the ISP

Functional Architecture (1/2)



Functional Architecture (2/2)



Some constraints

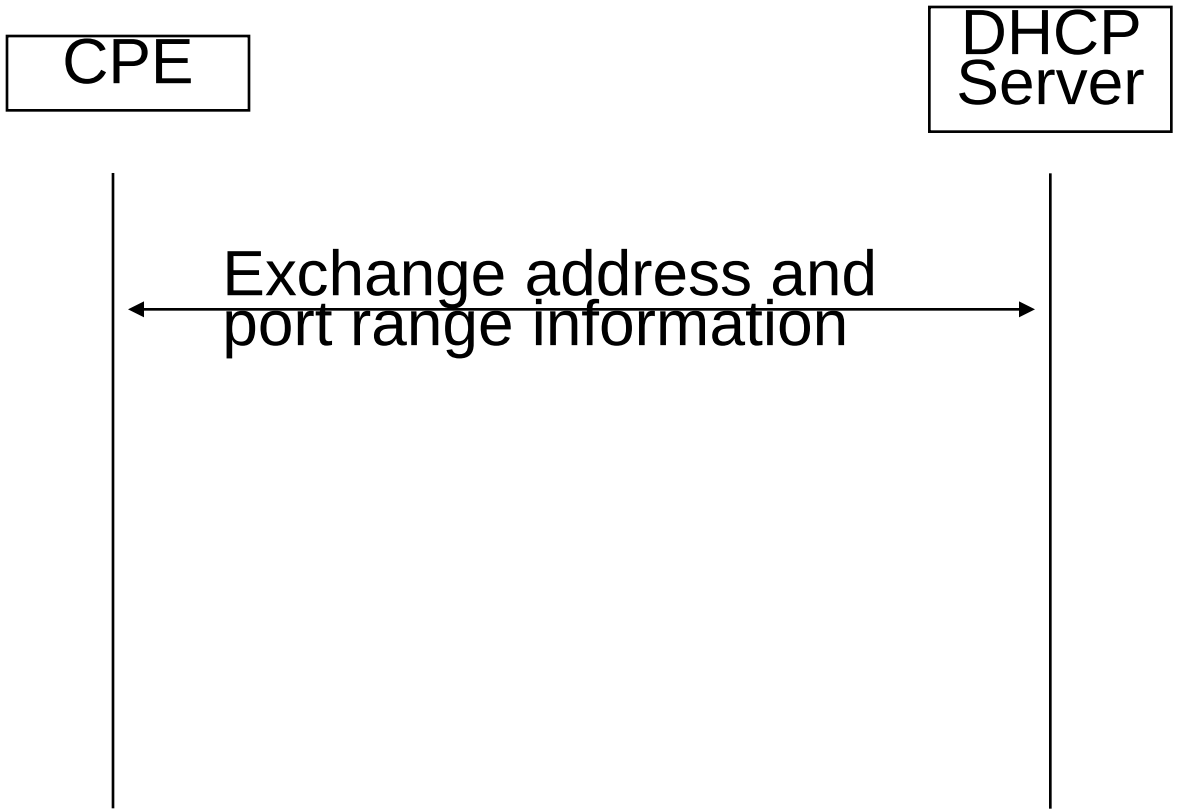
- The PRR must have a route to reach each CPE it covers
- Packets from a customer to another customer must pass through the PRR that handles the destination subnet
- Communications between two CPEs attached to the same PRR must go up to this PRR
- There is no intermediate routers between the PRR and the CPEs

Some architectural choices

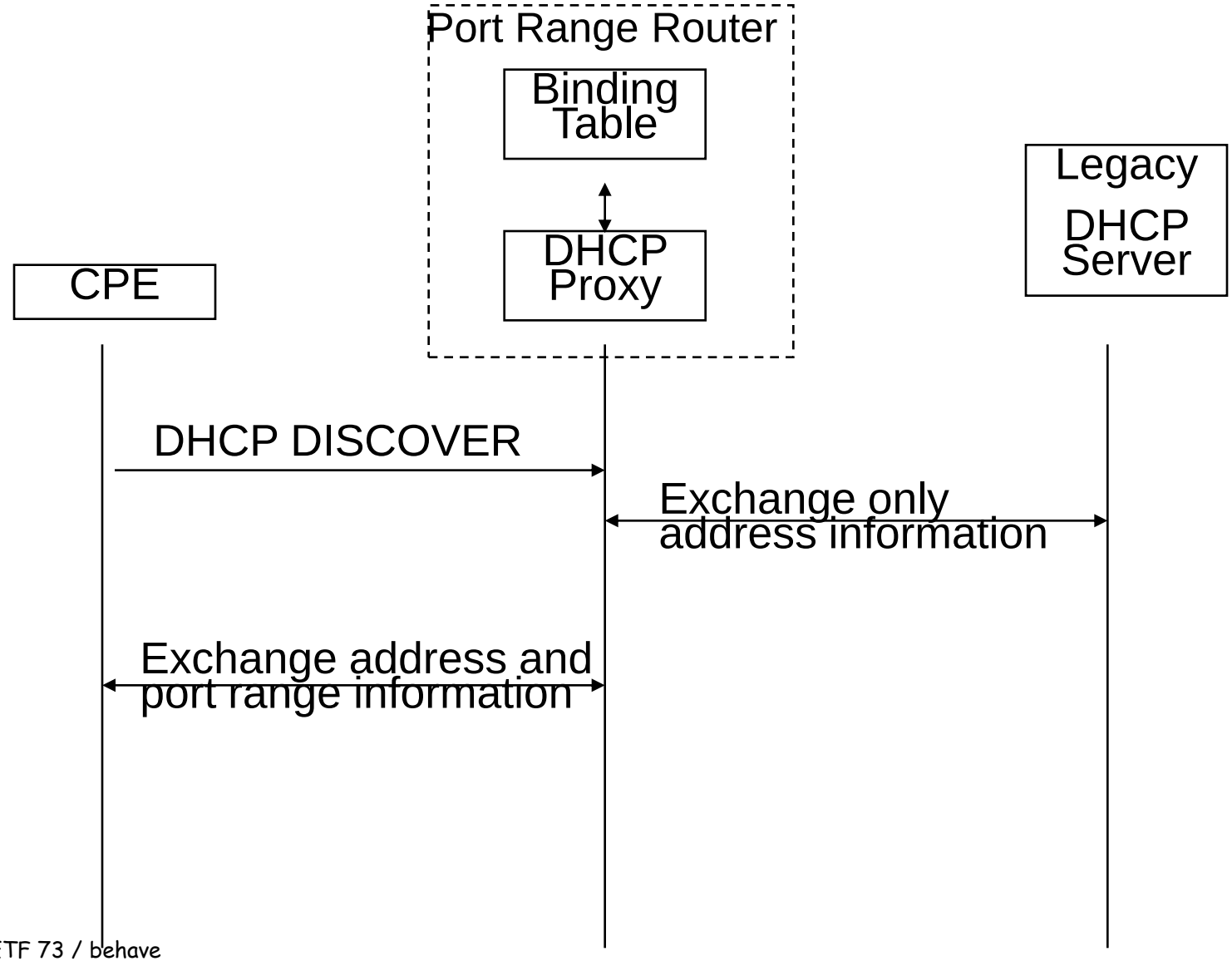
- The choices depend on the ISP requirements and engineering context
- Where to put the PRR?
  - Close to the user vs. close to the core
  - Distributed vs. centralized
- How to route from PRR to CPEs?
  - Point to point relationship (ex L2TP)
  - Private address to CPE, and v4 in v4
  - Private address to CPE, and MAC destination address on L2 access
  - IPv6 address to CPE, and v4 in v6
  - …

Address+Port allocation

Alt1: make your IS port range aware



Alt2: hide port range from your IS



DHCP Option (1/2)

- Port range allocation only (no address)
  - Addresses allocated as today
- Use the notion of Port Mask (similar to Subnet Mask)
- Port Range: a set of port values, may be non-contiguous
- Information carried:
  - Value
  - Mask

DHCP Option (2/2)

- Ex (contiguous):
  - Value: 1000000000000000
  - Mask: 1100000000000000
  - Port Range = 32768-49151
- Ex (non-contiguous):
  - Value: 0000000000000000
  - Mask: 0000001100000000
  - Port Range = 0-255, … ,64512-64767 (64 ranges)
- Other examples are given in the draft

Do we need port masks?

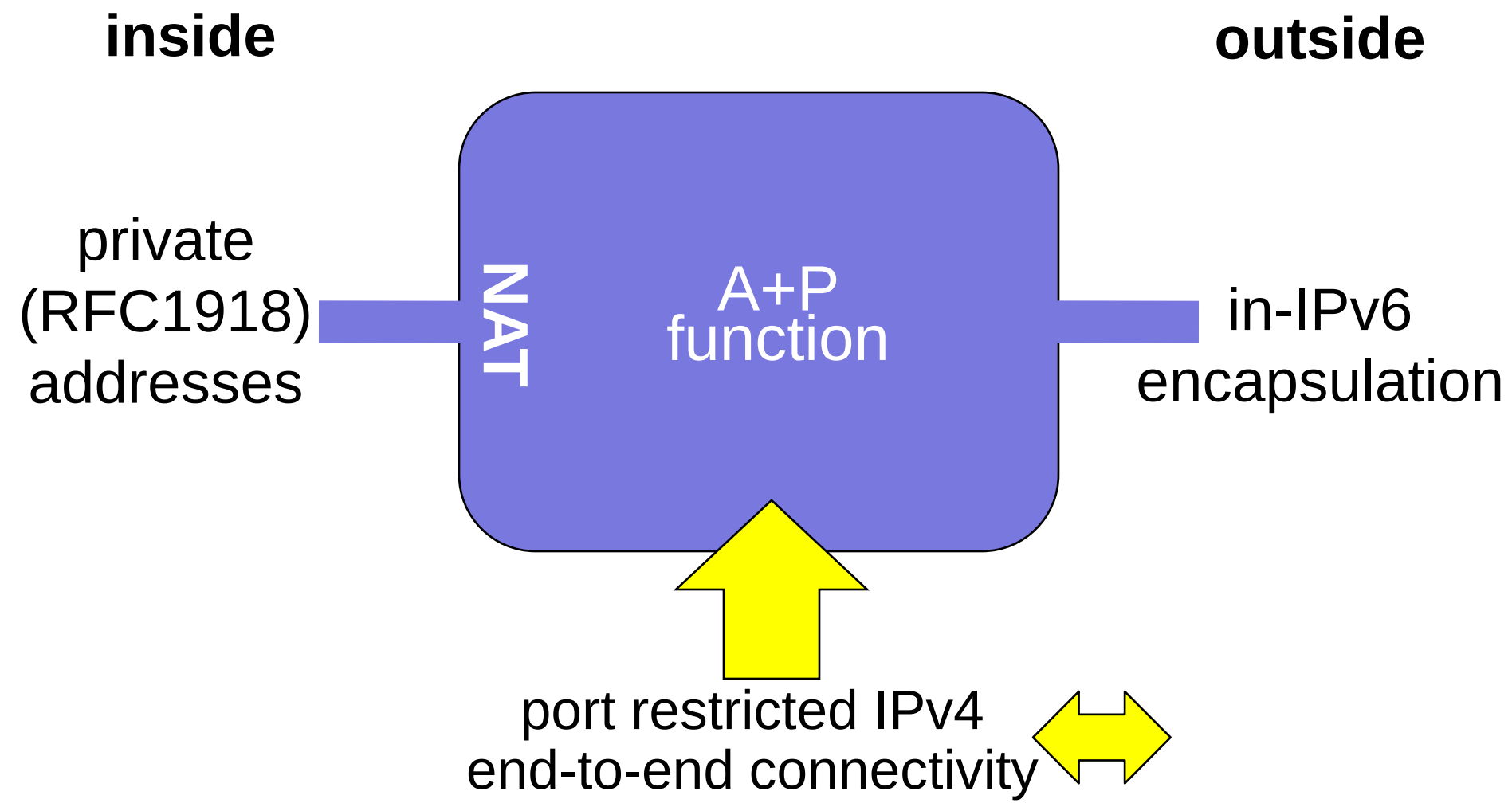
- Brings flexibility
- Non-contiguous values never used for subnets
- But subnet is not port range
  - Subnets are hierarchical, port ranges are not
- Masks restrict to power of two lengths
  - Subnets too
- Port range value will be computed by software, masks are easier to handle than range intervals

draft-ymbk-aplusp

A+P in One Slide

- Similar approach to DS-light (Durand)
- DS-light translate in the core, A+P encaps/decaps in the core, translates at the edge. No state in core.
- Mechanism required that **customer can control their fate**

A+P gateway


inside
outside
private
(RFC1918)
addresses
NAT
A+P
function
in-IPv6
encapsulation
port restricted IPv4
end-to-end connectivity

Encap from CPE

- WKP = well known prefix, 4666::0/64
- Source of v6 packet is WKP+A+P
- Dest address of v6 packet
  - WKP+v4dest
- Border (BR) makes global v4 packet
  - source = A+P
  - dest = v4dest

IPv6 Encap Toward CPE

- BR receives IPv4 packet w/ src/dest
- Encapsulates in IPv6 packet
  - src = WKP+src
  - dest = WKP+dest
- But note that dest is A+P
- It routes normally within ISP core

Note That

- Normal IPv6 backbone routing is used
- Routing out from gateway is based on real destination, not pre-configured tunnel
- Only A+P-gateway (e.g., CPE) and Border Routers are hacked
- No new equipment is introduced
- BRs do not have state or scaling issues

draft-despres-sam

SAMs
Stateless Address Mappings

. v4-v6 Coexistence => various vX/vY encapsulations

. A+P, which extends the global IPv4 space, has to be supported

. A generic mechanism => less specification, less code, less validations, less training…

. SAMs are designed for this (presentation in Softwire 4:40 PM)

Comparison of proposals

Comparison

- Based on current documents
- Most differences come from the addressed architectures
- Authors feel that convergence is worth trying

Comparison matrix (1)

	PR-IP	PRRs	A+P - BP	SAMs
A+P implemented where ?				
Host & gateway	X		X	X
CPE & gateway	X	X	X	X
Host behind CPE & CPE				X
Host behind CPE & CPE & gateway			X	X
A+P tunnelled on what ?				
Point-to-Point link	X	X		
Private IPv4	X	X		X
IPv6	X	X		X
IPv6 - Specific address prefix			X	X
A+P mapping table stored where ?				
gateway	X			
DHCP & gateway		X		
n/a, derived from IPv6 addr			X	
n/a, derived from local addr				X

Comparison matrix (2)

	PR-IP	PRRs	A+P - BP	SAMs
A+P values reserved when ?				
At DHCP request time	X	X		
Statically, independently of usage			X	X
Which kind of gateway address ?				
Non applicable (layer 2)	X			
Unicast	X	X	X	
Anycast				X
Which software handles the port restriction ?				
In the host, the socket-interface module	X			X
The NAT44 of the CPE	X	X		X
In the CPE, a specific module in front of the NAT44			X	

Comparison matrix (3)

	PR-IP	PRRs	A+P - BP	SAMs
Where are IPv4-fragments handled ?				
One dedicated box			X	
Gateway	X	X		X
Which assumptions on routing ?				
None. Single entry point only	X	X		
None. Any order, any entry point			X	
To the same gateway most of the time				X

Discussion Questions?