



Deploying IPv6

Elite Champions

2011-05-10

Holger Metschulat
Senior System Engineer
hmetschulat@juniper.net



Legal Statement

This presentation sets forth Juniper Networks' current intention and is subject to change at any time without notice. No purchases are contingent upon Juniper Networks delivering any feature or functionality depicted on this roadmap.

The Problem with IPv4

IPv4 created in the 1970s

- Today's Internet unforeseen by most
- 32-bit address space
- ~4.3 billion addresses

Concerns about IPv4 address depletion began in early 1990s

- "Classful" address structure was wasteful
- Huge difference between Class C and Class B
- Class B addresses projected to run dry ~1995

The Advent of IPv6

A larger address pool was required

- Several proposals for “Next-Generation IP” (IPng)
- IPv6 eventually adopted
- 128 bit addresses

Opened an opportunity to incorporate lessons learned

- Improved mobility
- Better multicast
- Integrated security
- Easier extensibility
- More efficient headers

Short-Term Solutions

IPv4 address depletion had to be slowed

- Allowing time for IPv6 development

Classless Inter-Domain Routing (CIDR)

- Eliminated classful IPv4 addressing
- Efficient use of address allocations

Dynamic Address Configuration Protocol (DHCP)

- Enabled sharing of address pool among many hosts

Private IPv4 Addresses (RFC 1918)

- Created globally “reusable” IPv4 addresses

Network Address Translation (NAT)

- Enabled many private hosts to use a few public addresses

IPv6 Gets Sidetracked

Transition to IPv6 intended while IPv4 still plentiful

- All devices run both IPv4 and IPv6 (dual stacks)
- Transition mode intended to span many years
- IPv4 eventually phased out

Short-term solutions proved highly successful

- CIDR proved highly effective
- IP apps of late 1990s made few demands on NAT/Private IP architectures

Need for IPv6 questioned

- Expense, headaches with no tangible benefits
- Everyone wanted to see “IPv6 killer apps”

IPv6 Resurgent

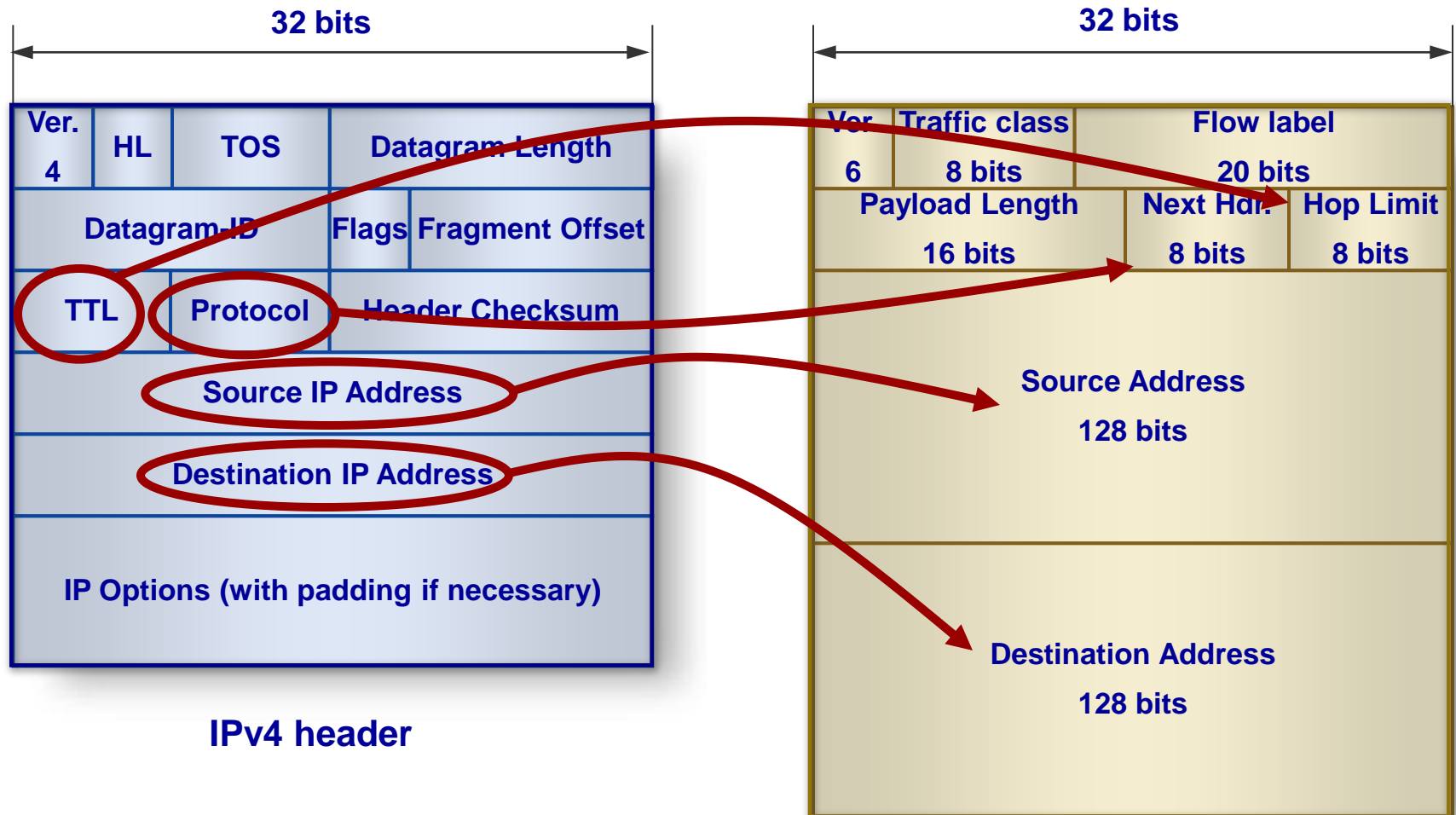
Effectiveness of CIDR declines

- By 2000, previously allocated IPv4 blocks used up
- Demand for new IPv4 allocations increases

21st century address demands explode

- Internet everywhere
- Legacy applications move to IP (voice, video)
- Millions of new IP-enabled mobile handsets
- Expanding economies in populous countries
- IP-enabled consumer electronics
- IP-enabled sensor networks

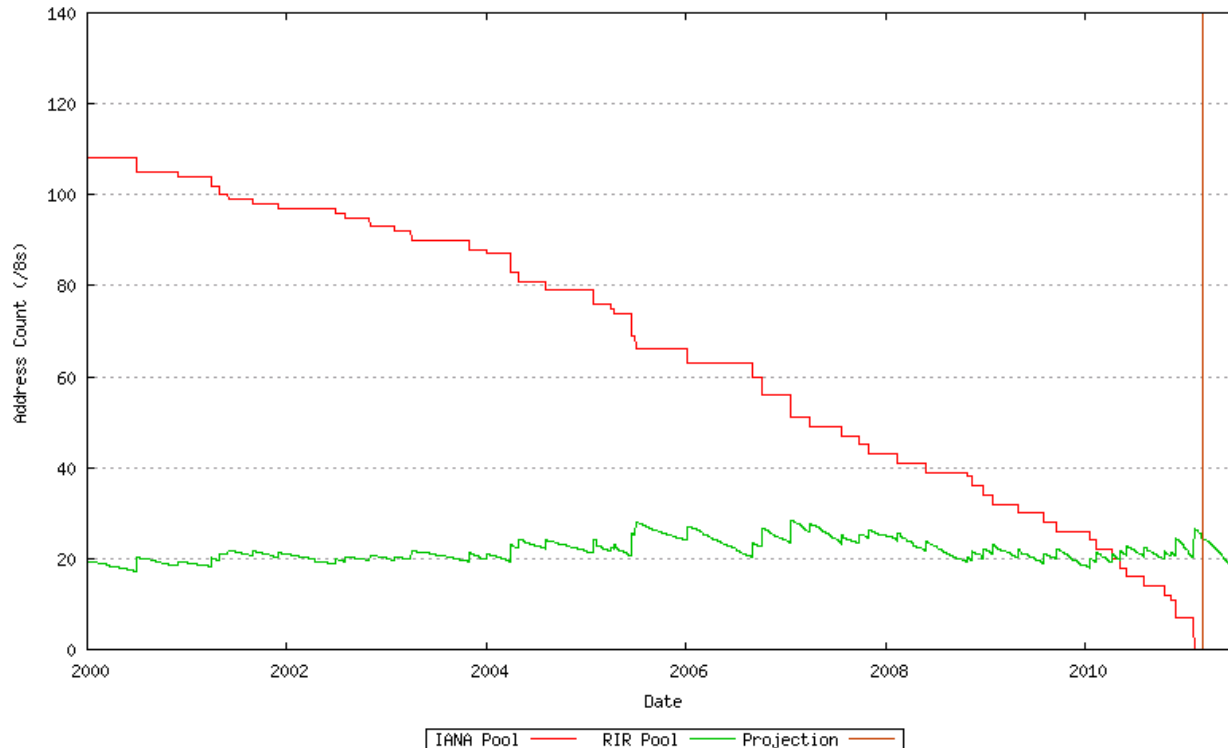
A Comparison of IPv4 and IPv6 Headers



IPv4 header

IPv6 header

The End of the Road Comes into View



IPv4 Exhaustion Counter

▾ Nearest Exhaustion (APNIC)

Reserved / Total Blocks

AfriNIC	2.60 / 4
APNIC	3.8 / 47
ARIN	4.68 / 75
LACNIC	3.01 / 9
RIPE	3.92 / 37

(Remaining /8s)

X-day (Nearest)

Jul 29, 2011

Until X-day (Nearest)

143 days

Num of IPv4 Address

63,490,020

iNetCore via IPv4

Projected RIR and IANA Consumption (nb /8s)

<http://www.potaroo.net/tools/ipv4/index.html>

IPv6 Reality Check: the IPv4 Long Tail

Post IPv4 allocation completion:

- Many hosts & applications in customer residential networks (eg Win 95/98/2000/XP, Playstations, consumer electronic devices) are IPv4-only.
- Most software & servers in enterprise network are IPv4-only
 - They will not function in an IPv6-only environment.
 - Few of those can or will upgrade to IPv6.
- Content servers (web, email,...) are hosted on the Internet by many different parties. It will take time to upgrade those to IPv6.

Current measurement:

0.15% of Alexa top 1-million web sites are available via IPv6

(This number has not changed in the last 12 months)

Source: <http://ipv6monitor.comcast.net>

Is IPv6 Taking Off?















A number of very large ISPs and very large content providers are deploying IPv6 and various transition technologies **now**.

- Still early in the adoption curve.
- However, momentum is building.
- Can't be ignored.

IPv6 does not solve the immediate problem of IPv4 address exhaust.

- Maintaining IPv4 service after IPv4 exhaustion is #1 priority for most players.
- This implies some form or another of IPv4 address sharing: NAT
- Many transition technologies to choose from
 - Impact on routing and network architecture

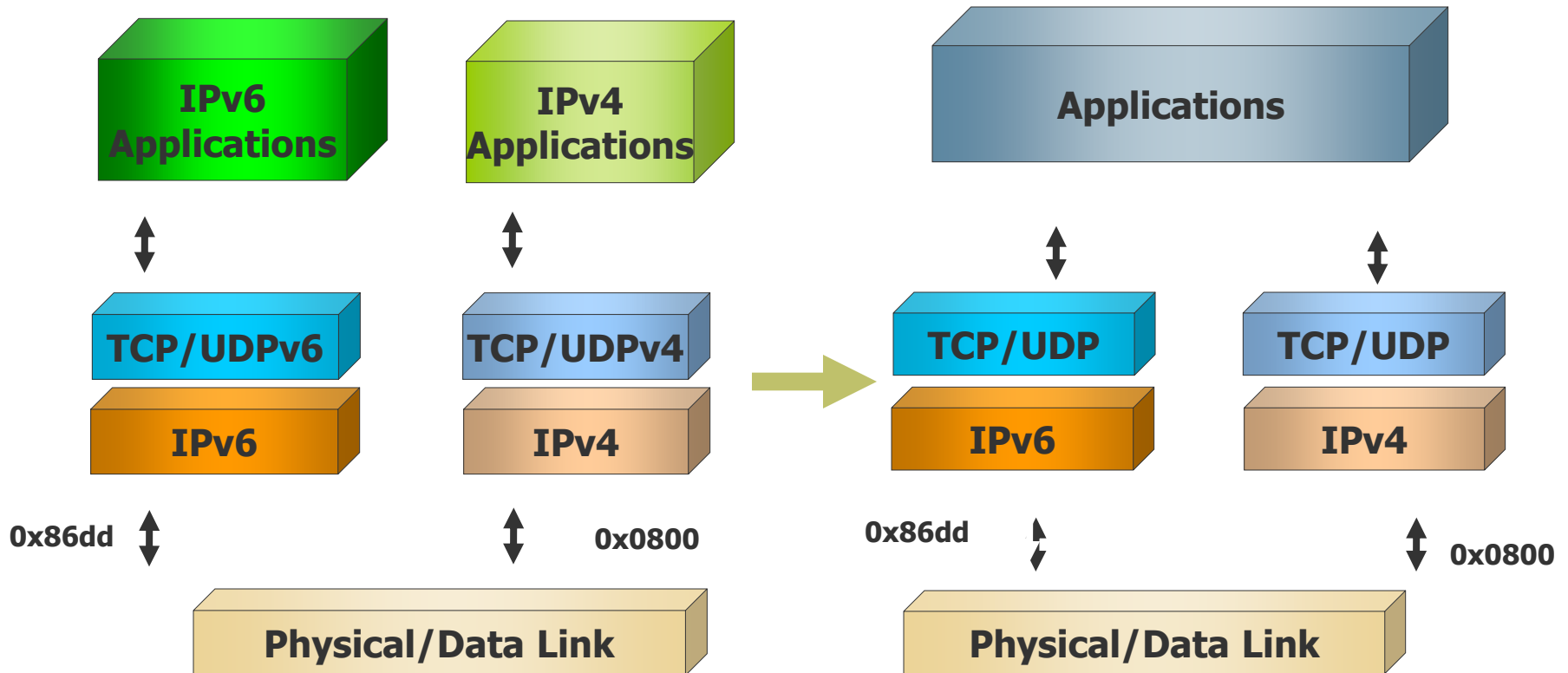
Industry IPv6 score card

Function	Element	Status
Network	Core Router: T	
	Edge Routers: MX, 6PE	
Servers	Linux 2.6+	
	Datacenter equipments, CDN	
End-user clients	Windows 7 (Many XP boxes out there) 	
	MacOS 10.x	
	Game consoles Wii, PS3, Xbox	
Software	Web Browser: Firefox, IE, Safari	
	Skype	
	On-line PC games	
	SSL VPN	
Content	Web content available over IPv6	
CE	CPEs, Mobile Devices	

Number
1 & 2
issues

Dual Stacks

Network, Transport, and Application layers do not necessarily interact without further modification or translation



Observations about Transition techniques

All transition techniques (NAT44, NAT444+6RD, NAT64, DS-Lite) revolve around the notion of sharing IPv4 addresses via some form of NAT.

They all require the exact same amount of IPv4 addresses to be shared in a NAT pool.

- The difference is how packets are transported to the NAT

Sharing addresses among customers introduces issues:

- LEA/Abuse/Logging/Geo-location/Access control

IPv6 is a DRAMA in four acts

Prelude: IPv4 exhaustion happens in 2011.

Act I: NAT solves IPv4 exhaust.

Act II: IPv6 to simplify IPv4 service delivery.

IPv6 networks with IPv4 overlays enable the management of a large number of customers while maintaining an IPv4 service.

Act III: Emergence of IPv6 content.

The decoupling of deploying IPv6 networks from the deployment of IPv6 applications & content solves the chicken and egg problem. IPv6 traffic is a cap& grow strategy around NAT scaling issues.

Act IV: IPv4 dies (very slowly) .

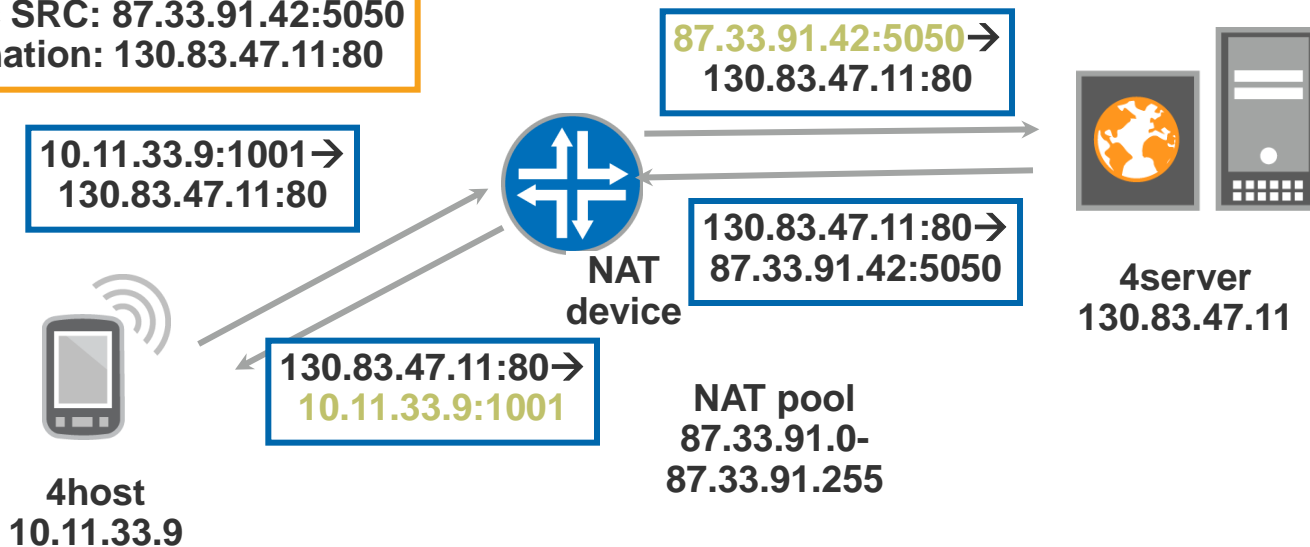
IPv4 & IPv6 co-exist until IPv6 become pervasive.

NAT IPv4 - IPv4

Recall how NAT44 works

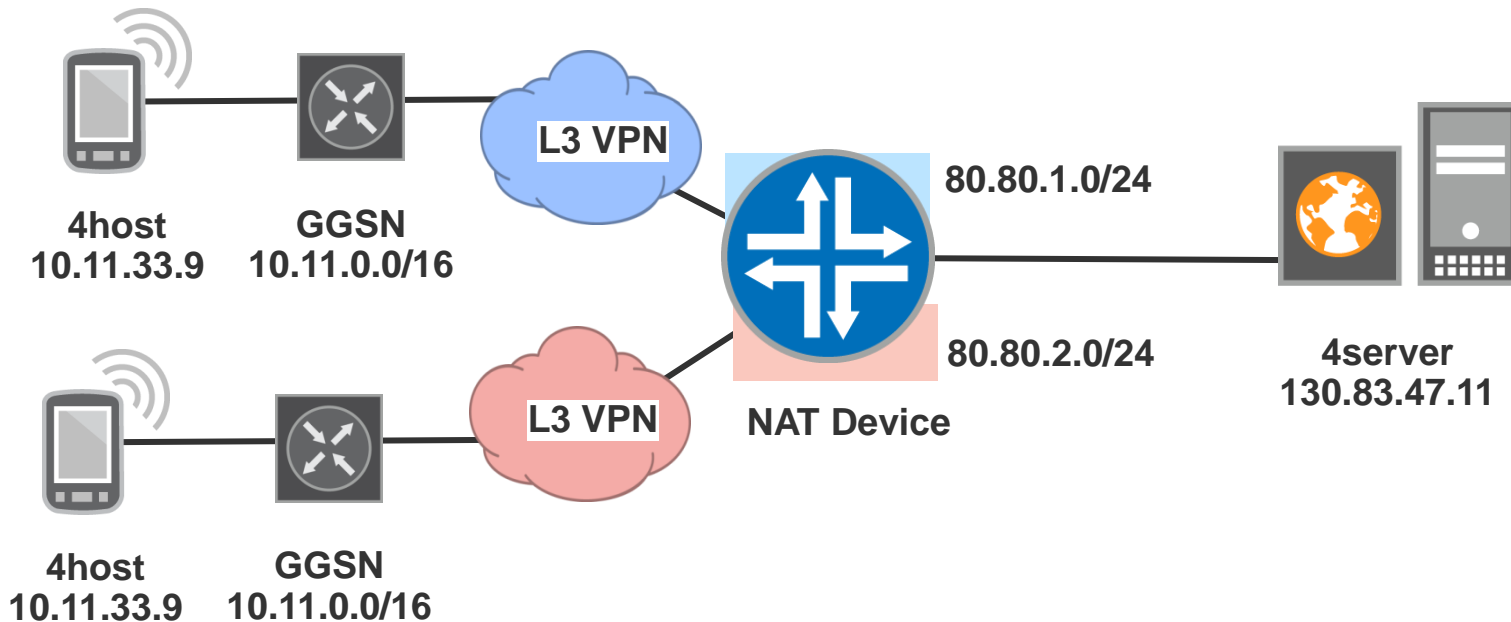
NAT device
session table entry

Private SRC: 10.11.33.9:1001
Public SRC: 87.33.91.42:5050
Destination: 130.83.47.11:80



Virtualising NAT IPv4 - IPv4

Reusing client address ranges or separating access by virtualising the NAT device



IPv4 to IPv6 Transition Mechanisms

Myriad Proposals

Dual Stack

- Host and router

Configured tunnels

- Router to router

Network level translators

Stateless IP/ICMP Translation Algorithm (SIIT)(RFC 2765)

NAT-PT (RFC 2766)

Bump in the Stack (BIS) (RFC 2767)

Transport level translators

Transport Relay Translator (TRT) (RFC 3142)

Application level translators

Bump in the API (BIA)(RFC 3338)

SOCKS64 (RFC 3089)

Application Level Gateways (ALG)

Automatic tunnels

- Tunnel Brokers (RFC 3053)
 - Server-based automatic tunneling
- 6to4 (RFC 3056)
 - Router to router
- DS-Lite
- ISATAP (Intra-Site Automatic Tunnel Addressing Protocol)
 - Host to router, router to host
 - Maybe host to host
- 6over4 (RFC 2529)
 - Host to router, router to host
- Teredo
 - For tunneling through IPv4 NAT
- IPv64
 - For mixed IPv4/IPv6 environments
- DSTM (Dual Stack Transition Mechanism)
 - IPv4 in IPv6 tunnels

Translators

Swap headers of one IP version for headers of the other

Enables interconnection of IPv6-only and IPv4-only devices

- In most cases, IPv6-capable devices are also dual stack capable
- But we do not want Dual Stack everywhere because we need to conserve IP addresses NOW!

Network Address Translation with Protocol Translation (NAT-PT)

- Many translation mechanisms proposed, only NAT-PT has been used

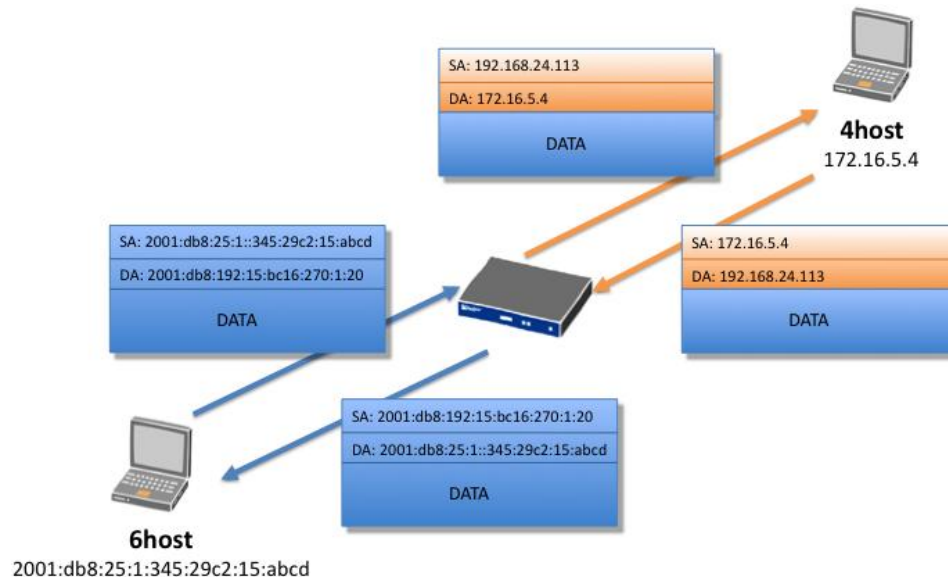
Translators

Headers are swapped

- Non-matching fields must be adjusted

No support required in individual devices

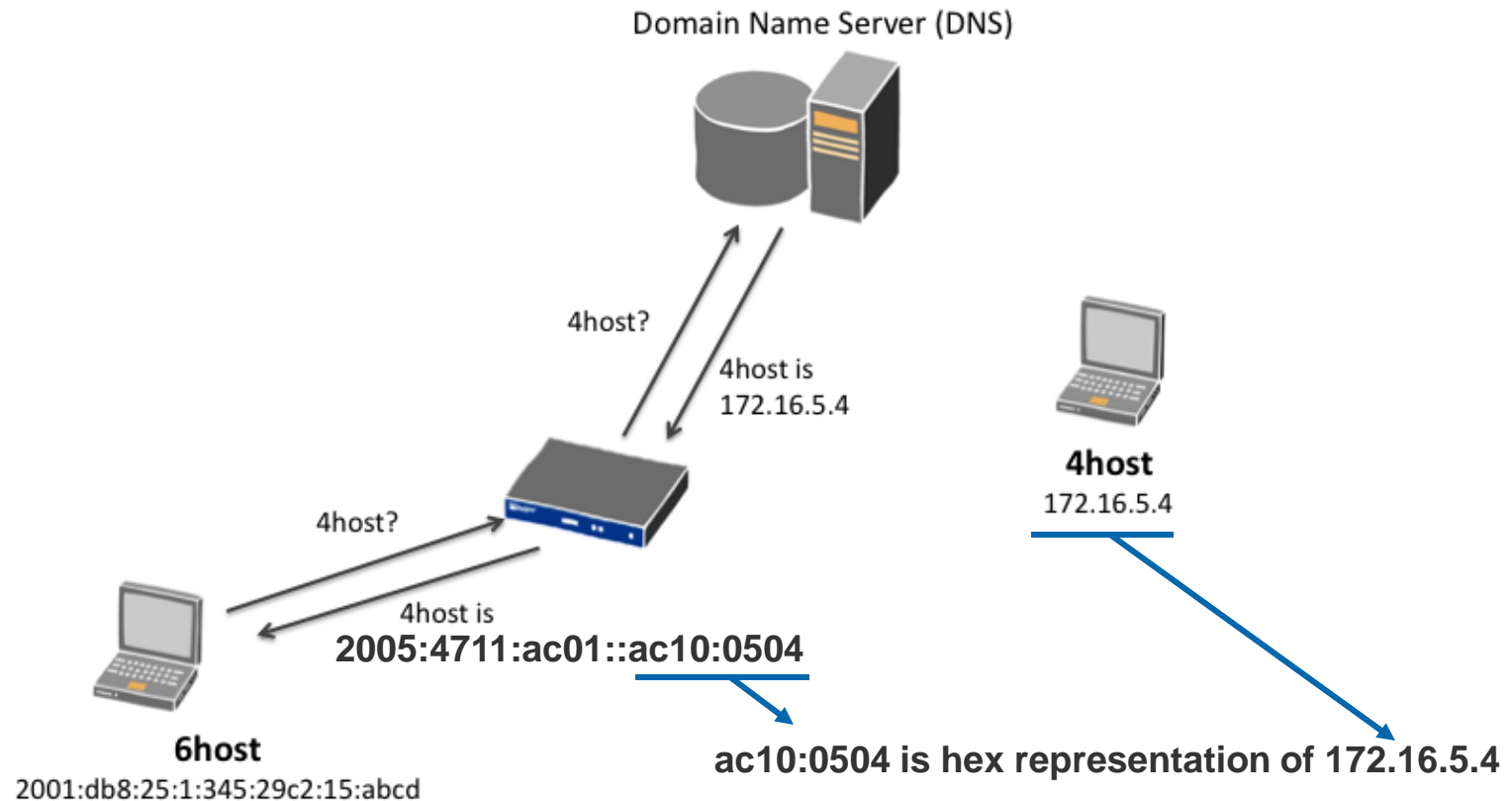
- IPv6 device “thinks” it’s talking to another IPv6 device
- IPv4 device “thinks” it’s talking to another IPv4 device



Translators: NAT-PT or NAT64

DNS records are translated

- DNS Application-Layer Gateway (ALG) in translator



Translators – NAT-PT / NAT64

Pros:

- Meets corner-case requirements

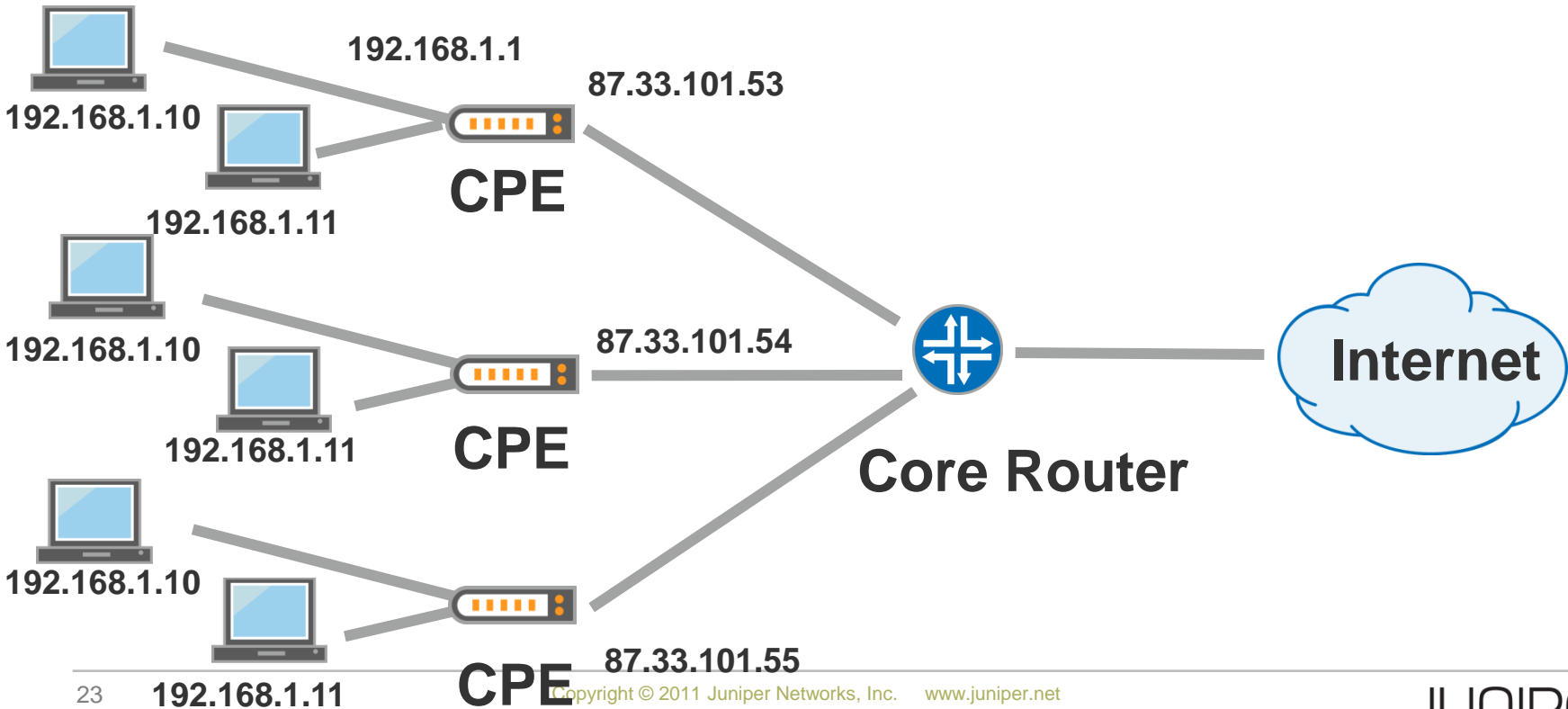
Cons:

- Known scaling problems
- Imposes network design restrictions
 - Traffic flows must be symmetric through same translator
 - DNS must be carefully placed
 - DNS must translate queries and results
- Will not work if customer uses IPv4 instead of DNS (<http://1.2.3.4>)
- Single point of failure; attractive attack target
- No multicast support
- NAT-PT deprecated by IETF
 - Application proxies, SIIT, NAT64, other solutions recommended instead

NAT44

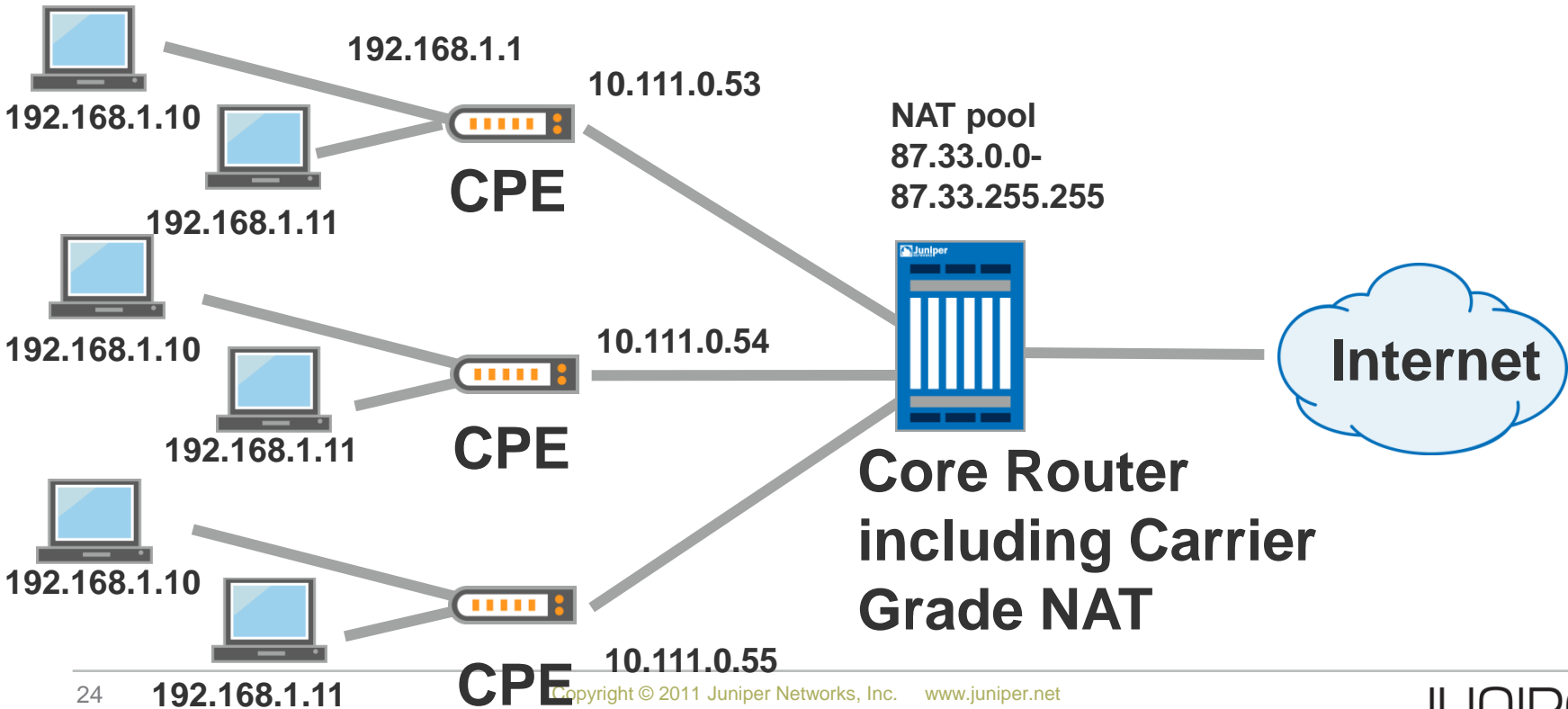
NAT44 today is happening only on the CPE

Every CPE gets one public IPv4 address for the public interface



NAT444

NAT444 introduces another NAT layer between the CPE and a core NAT device (Carrier Grade NAT)



Dual Stack Lite

Although the name sounds like it, this is NOT a stripped-down Dual Stack implementation

End device (or CPE) needs to have a modified IP stack software

Pure IPv6 traffic runs natively – no need to involve a DS lite gateway or modify the IPv6 stack of the CPE

Dual Stack Lite

End Device is Dual Stack (has IPv4 and IPv6 address)

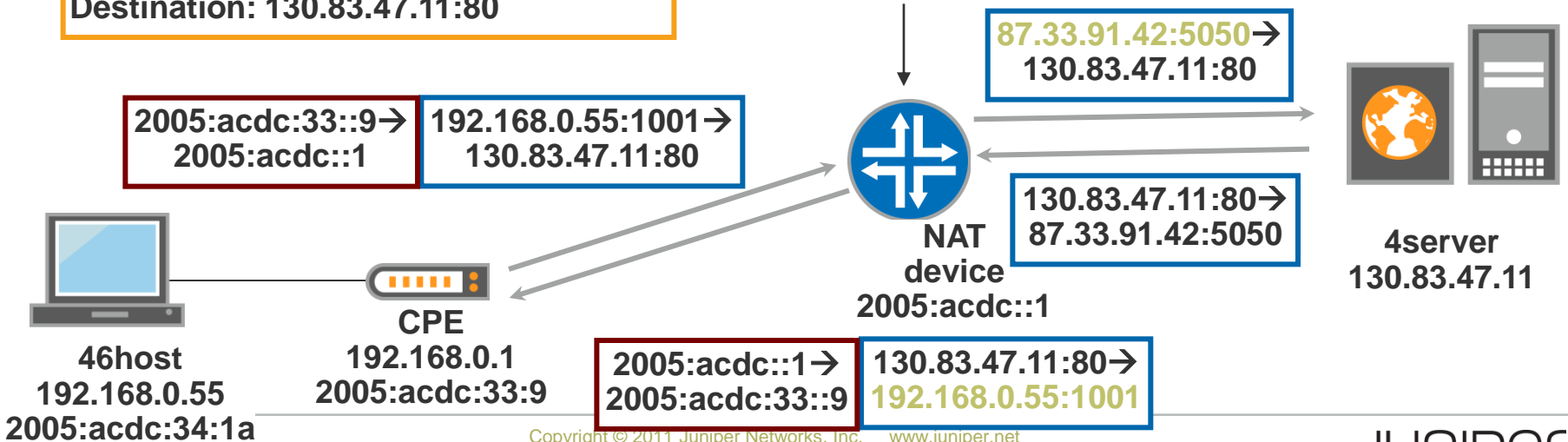
IPv4 traffic is tunneled within an IPv6 packet to the CG NAT device and back

CPE has to do encaps/decaps, no need to do NAT there

NAT device
session table entry

Priv CPE SRC: [2005:acdc:33::9]
Priv Host SRC: 192.168.0.55:1001
Public SRC: 87.33.91.42:5050
Destination: 130.83.47.11:80

NAT pool
87.33.91.0-
87.33.91.255



Dual Stack Lite Findings

End device must support the tunneling of IPv4 in IPv6 packets

End device IPv4 address is irrelevant for the CG NAT translation
 – may be reused amongst clients

No IPv4 in the aggregation network

NAT device
 session table entry

Priv CPE SRC: [2005:acdc:33::9]
 Priv Host SRC: 192.168.0.55:1001
 Public SRC: 87.33.91.42:5050
 Destination: 130.83.47.11:80

NAT pool
 87.33.91.0-
 87.33.91.255



4server
 130.83.47.11

2002:acdc:33::9 → 192.168.0.55:1001 →
 2002:acdc::1 130.83.47.11:80

87.33.91.42:5050 →
 130.83.47.11:80

130.83.47.11:80 →
 87.33.91.42:5050



NAT
 device
 2005:acdc::1



46host
 192.168.0.55
 2005:acdc:34:1a



CPE
 192.168.0.1
 2005:acdc:33:9

2005:acdc::1 → 130.83.47.11:80 →
 2005:acdc:33::9 192.168.0.55:1001

BUT NAT IS EVIL...?

NAT breaks end-to-end communication in some protocols; others (e.g. IPSEC) are not well designed to go through NAT and NAT devices have a hard time with those protocols.

Address Multiplexing as done in NAT may cause restricted end device reachability

- Not good for example for file sharing

End customers that need native IP connectivity and reachability have the possibility to use IPv6 for this

- Most File Sharing systems today support IPv6
- Is this the long-awaited „killer application“?

Have to find a way to position this from a marketing and support perspective

- Two types of subscription models, private IPv4 and public IPv6, or public IPv4 and public IPv6
- The one with private IPv4 is offered at a reduced price or
- Public IPv4 is offered at an extra cost or
- All new customers get private IPv4 addresses, upon complaints or helpdesk calls, customers get moved to public IPv4 addresses immediately

CGN Requirements and Features on Juniper MX

**EIM, APP, EIF
Managing Subscriber Sessions**

Managing subscriber's sessions

Limiting the maximum number of sessions from same subscriber

- This allows to provide fairness of public IP and port availability to different subscribers
- Prevents DDOS attacks: a few subscribers starving all available resources

Port Random-Allocation

- For each IP address in a pool the initial allocated port is assigned randomly and then continuing to allocate ports sequentially from there.
- It lowers the risk of inbound attacks when EIF is enabled

Round Robin address allocation

- For every different internal source address, a different NAT address is allocated in a round robin fashion
- Instead of using all available ports for a specific public IP address, move to the next public IP address whenever there is a new internal host requiring a connection.
- Address Pooling behavior unchanged. I.e. new sessions from the same internal source address will continue to use the same public IPv4

Preserve port range and/or parity

Preserve Range

- RFC4787 defines two port ranges: "Well Known Ports" [0, 1023] and "Registered"/"Dynamic and/or Private" [1024, 65535]
- When the source port of the internal host establishing a new connection falls into one of these ranges the CGN tries to allocate an external source port in the same range. If it fails to find a port, connection fails too.

Preserve Parity

- CGN tries to allocate a even/odd external source port depending on whether the new connection has an internal even/odd source port

Application Level gateways

Although SPs prefer to not have to deal with ALGs some protocols require them to be enabled

- Iphone uses pptp to connect to a VPN
- ICMP, Traceroute, TFTP, RSH, MS-RPC, PPTP, DC, FTP, H.323, SQLnet, RSTP are the currently supported ALGs on MS-DPC

CGN should allow the administrator to selectively enable ALGs on a per protocol basis as there a number of NAT-friendly apps that ALGs can interfere

APP+EIM+EIF

APP, Address Pooling ,“Paired” behavior

- a CGN must use the same external IP address mapping for all sessions associated with the same internal IP address
- It solves the problem of an application opening multiple connections using different source ports.
- Remote servers or peers for the same application reject connections if not all originated by the same IP address. Examples are Instant Messengers

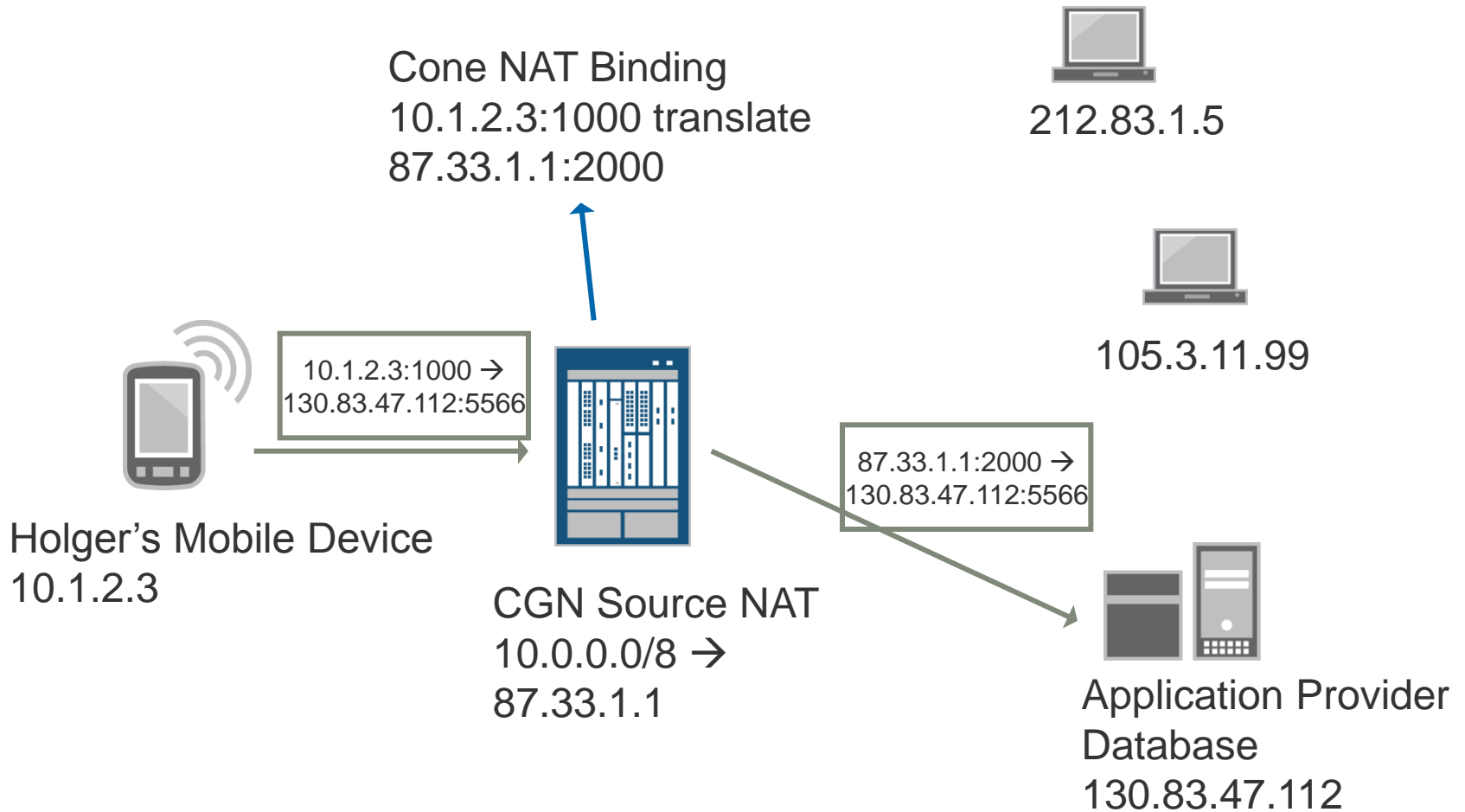
EIM, Endpoint Independent Mapping,

- A CGN must assign the same external address and port for all connections originated from a given internal host if they all use the same internal port
- As a consequence connections originated by same internal IP address, but with different internal port can use a different external IP address
- Enabling EIM allows to have a stable external P address and Port (for a period of time) that external hosts can use to connect. Very important for p2p, gaming and the mobile world
- EIM does not decide who from the external realm can connect to the internal host, that is done by EIF instead.

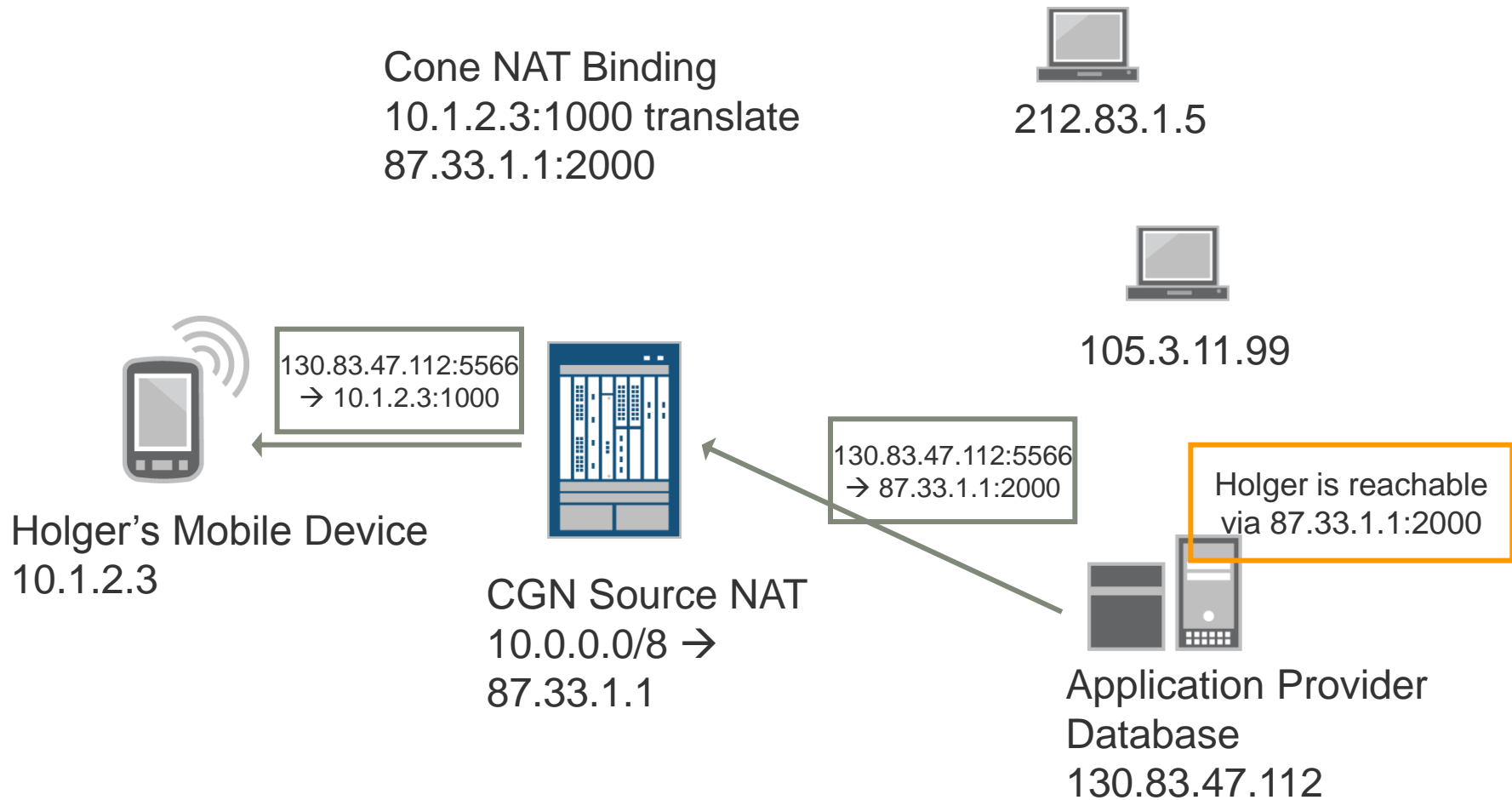
EIF, Endpoint Independent Filtering

- EIM alone does not influence the inbound filtering behavior. Actually the default filtering behavior is Address and Port dependant (APM) which means that only remote Servers or Peers towards which we opened a connection are allowed to reach the internal host using a specific IP and Port.
- EIF filters out only packets only packets not destined to the internal address and port, regardless of the IP address and port of the remote Server or Peer.
- Please consider that differently from Enterprise NATs, the CG-NAT should provide as much transparency as possible to the applications.

ENDPOINT INDEPENDENT MAPPING



ENDPOINT INDEPENDENT MAPPING



ENDPOINT INDEPENDENT MAPPING

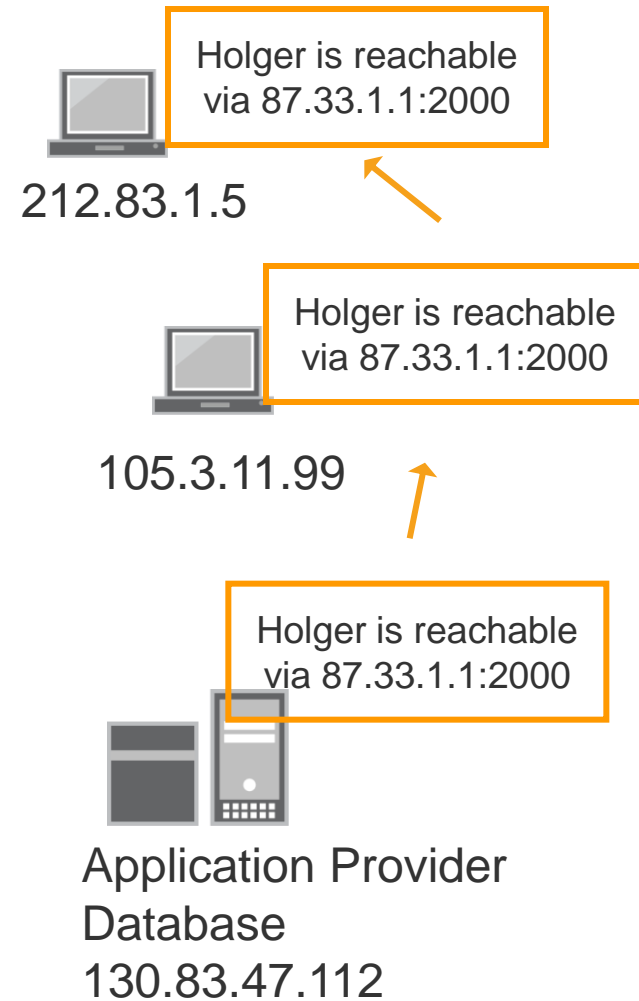
Cone NAT Binding
10.1.2.3:1000 translate
87.33.1.1:2000



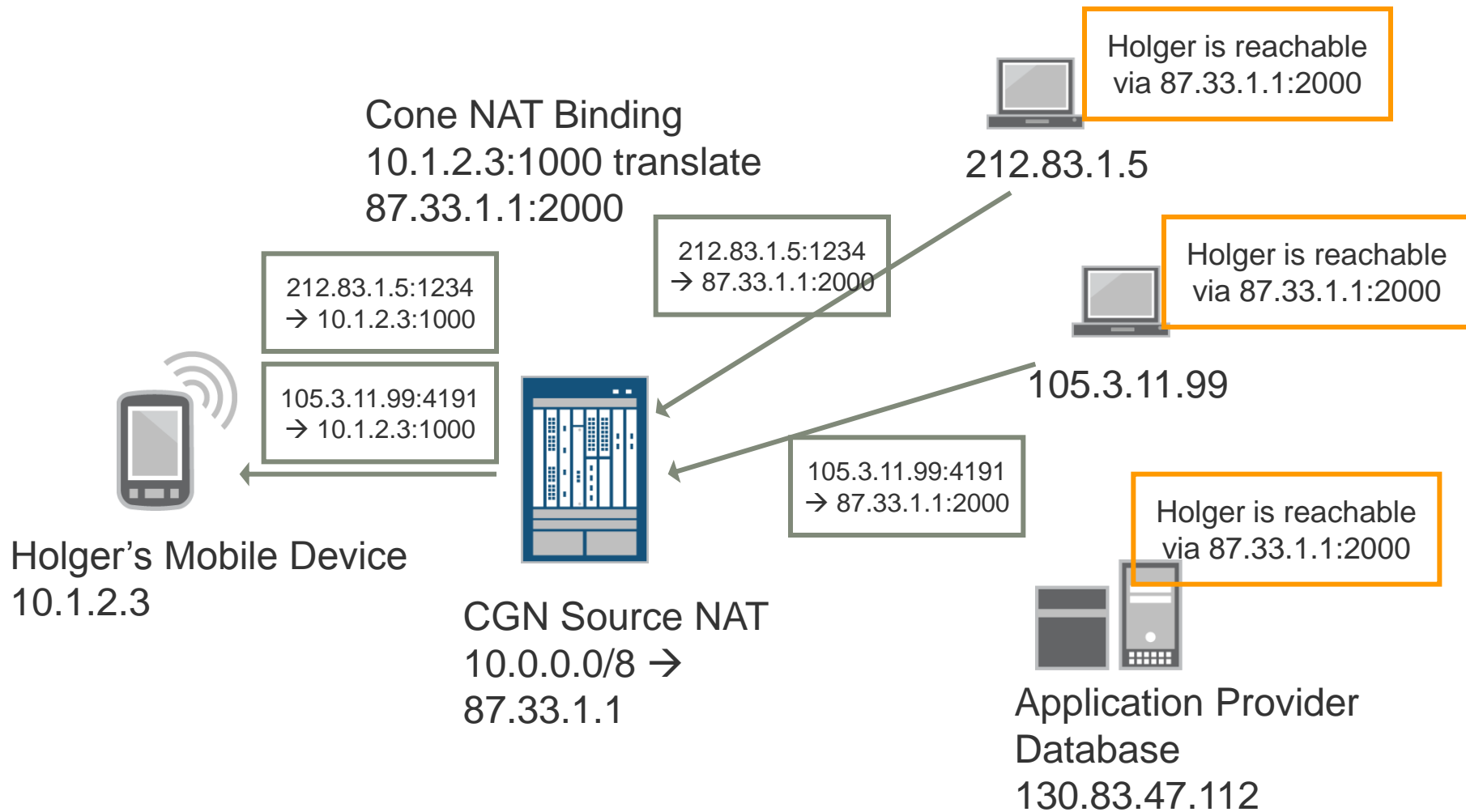
Holger's Mobile Device
10.1.2.3



CGN Source NAT
10.0.0.0/8 →
87.33.1.1



ENDPOINT INDEPENDENT MAPPING



Technology: PCP (New development)

PCP: Port Control Protocol

PCP objectives are to enable applications to receive incoming connections in the presence of an ISP NAT/Firewall.

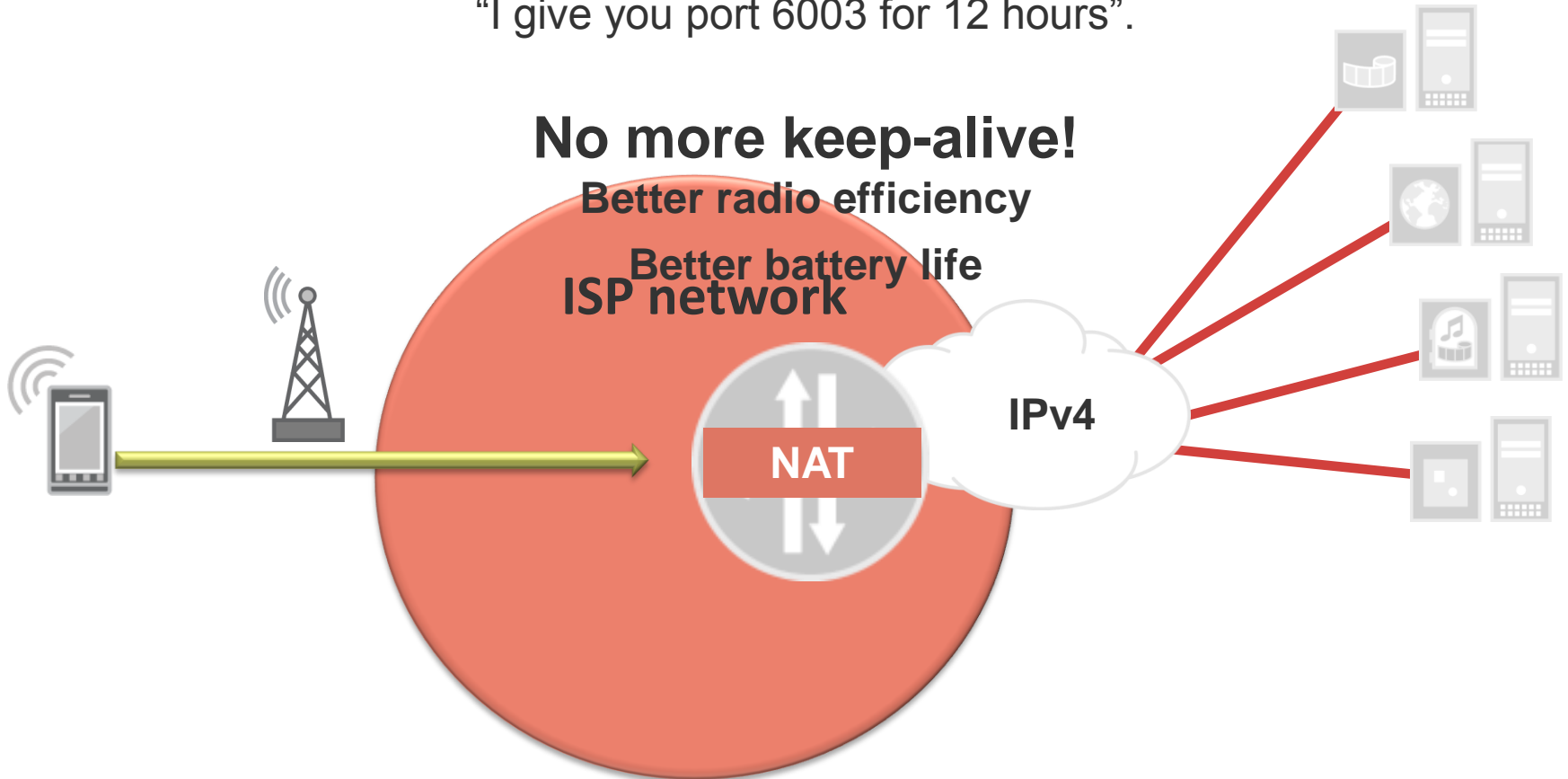
Instead of 'working around' NATs like other NAT traversal techniques like STUN/TURN/ICE, PCP enables an explicit dialog between applications and the NAT.

PCP can be seen as a 'carrier-grade' evolution of UPnP-IGD and NAT-PMP.

The work on PCP is done at IETF in a new working group co-chaired by Alain Durand (Juniper) & Dave Thaler (Microsoft).

PCP in a nutshell

Applications negotiate ports with the ISP NAT to establish external presence. Application asks: "I'd like to get port 5000 for 48 hours", NAT PCP server responds: "I give you port 6003 for 12 hours".



Purpose of Flow analysis

Network planning: oversubscribed or undersubscribed

- Definition of Peak vs. average vs. mean

Protocol trends and usage:

- New applications
- Protocols
- Partnerships

Optimizations: local and network wide

Forecast required storage capacity for logging

- Cost here is a major concern among SPs

Operations and Management: CGN statistics per NPU on JUNIPER MULTISERVICE PIC/DPC

```
user@router> show services stateful-firewall
flow-analysis | no-more
  Services PIC Name:    sp-2/0/0
```

Flow Analysis Statistics:

```
Total Flows Active           :107155
Total TCP Flows Active       :73688
Total UDP Flows Active       :32608
Total Other Flows Active     :859
Created Flows per Second     :1581
Deleted Flows per Second     :1578
Peak Total Flows Active      :292994
Peak Total TCP Flows Active  :232115
Peak Total UDP Flows Active  :70433
Peak Total Other Flows Active:10577
Peak Created Flows per Second:20175
Peak Deleted Flows per Second:20399
Average HTTP Flow Lifetime(ms):0
Packets received             :12675268950
Packets transmitted          :12657387448
Slow path forward            :249264193
Slow path discard            :9253298
```

Flow Rate Data:

```
Number of Samples: 60689
Flow Rate Distribution(sec)
```

Flow Operation :Creation

```
50000+           :0
40000 - 50000   :0
30000 - 40000   :0
20000 - 30000   :1
10000 - 20000   :76
1000 - 10000    :58807
0 - 1000        :1805
```

Flow Operation :Deletion

```
50000+           :0
40000 - 50000   :0
30000 - 40000   :0
20000 - 30000   :1
10000 - 20000   :81
1000 - 10000    :58226
0 - 1000        :2381
```

Flow Lifetime Distribution(sec) :

	TCP	UDP	HTTP
240+	:5503508	985823	4338658
120 - 240	:5922922	552323	
60 - 120	:14187071	951341	
30 - 60	:7432218	3035490	
15 - 30	:9743184	5920873	
5 - 15	:28389196	8161575	
1 - 5	:24445392	21285998	
0 - 1	:118585963	240066552	

Case Study: Mobile

The key issue is license cost:

	Dual-Stack (NAT44)	IPv6-only (NAT64)
License cost 2G & 3G/3GPPr8 (using separate PDP contexts for IPv4 & IPv6)	Two licenses: 1 for IPv4 PDP + 1 for IPv6 PDP	1 for IPv6 PDP
License cost LTE and 3G/3GPPr9 (using a combined PDP context for IPv4&IPv6)	1 for IPv4/IPv6 PDP/bearer	1 for IPv6 PDP/bearer

Preferred

Going IPv6-only + NAT64 works **ONLY** if all applications are converted to IPv6 and there is no connectivity to external devices such as PCs.

Dual-Stack remains the preferred/simplest general solution.

Juniper CGN Solution

Performances
Most deployed features

IP Family transition Services on MS-PIC/MS-DPC

NAT44

- Support CGN requirement
- (draft-ietf-behave-lsn-requirements)

IPv6 Features

- IPv6 NAT and IPv6 Stateful Firewall
- NAT-PT Supported (ICMP ALG)
- NAT-PT DNS ALG (10.4)
- Stateful NAT66 supported
- NAT64 (10.4)

IPv6 Software

- DS-Lite (10.4)
- 6rd/6to4 (11.1-Now)



8 MS-DPC supported by
Single MX Chassis
(1H2011)



Boost of performance in 11.2

Per card (MS-DPC) performance – on average 19Gbps throughput

Metrics	NAPT44(4) PBA ¹	NAT64
Throughput	19Gbps	18Gbps
Total Flows	17M	15M
Peak Flow ² Ramp-up Rate	1.2M Flows/sec	540K Flows/sec
Public Port Pool	4B ports	4B ports
Number of Subscribers	8.5M	7.5M
Ramp-up time (4M Flows)	4sec	8sec

¹Port Block Allocation (PBA): When PBA is configured, ports for a host are allocated in blocks. Subsequent port allocations for the same host come from the previously allocated block.

²Flow = Uni-directional flow through the Router

NAT44(4) MOST deployed features

Address Pooling paired

Round-Robin Allocation across NAT pools

Load-Balancing across Service Cards

TCP/UDP/ICMP configurable timeouts and TCP Keep-Alives

O&M commands and alarms to monitor NAT pool, mapping, session state, etc

O&M commands to monitor total sessions, sessions/sec, sessions lifetime, etc

Application Level Gateways

- For NAT44: ICMP, Traceroute, TFTP, RSH, MS-RPC, PPTP, DC, FTP, H.323, SQLnet, RSTP are the currently supported ALGs on MS-DPC

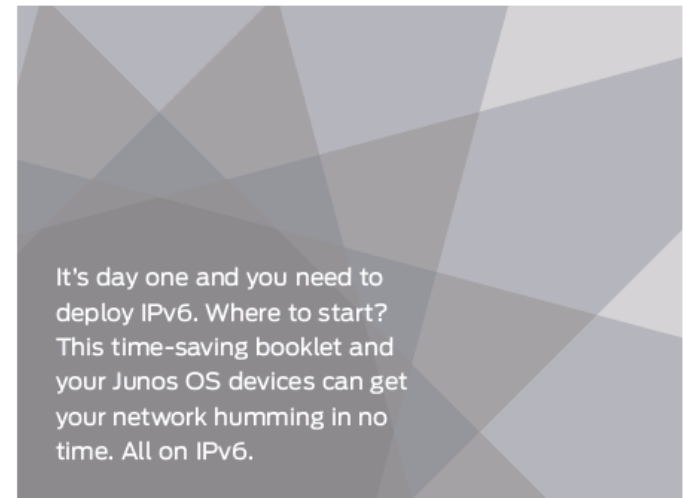
RESOURCES

- Links to standards and drafts in this presentation
- Juniper „Day One“ – Exploring IPv6
<http://www.juniper.net/dayone>

JUNIPER
NETWORKS

Junos® Networking Technologies Series

DAY ONE: EXPLORING IPV6



By Chris Grundemann

Summary

As a matter of fact SPs will need to share limited public IPv4 addresses for many years. CGN serves as a key building block in any transition strategy to IPv6.

CGN must ensure application level transparency

Storage of Logs could significantly impact on costs. Hence strategies to reduce the amount of logs should be considered: PBA, Deterministic NAT

Subscribers who are used to set Port Forwarding on their RG will ask for the same functionality when NAT is either moved into the SP network, i.e. DS-Lite, or added in the SP network, i.e. NAT444 or 6rd+NAT444.

Port Control Protocol as the alternative to Port Forwarding going forward

JUNIPER IPV6 WEBINAR SERIES 2011

Webinar 1

- **Architecting the Network for the IPv6 transition**
- Tuesday May 10th 3pm GMT / Alain Durand

Webinar 2

- **IPv6 Tutorial**
- Tuesday May 17th 3pm GMT/ Raffaele D'Albenzio

Webinar 3

- **Carrier Grade NAT and its evolution**
- Tuesday May 24th 3pm GMT/ Alessandro Salesi

Webinar 4

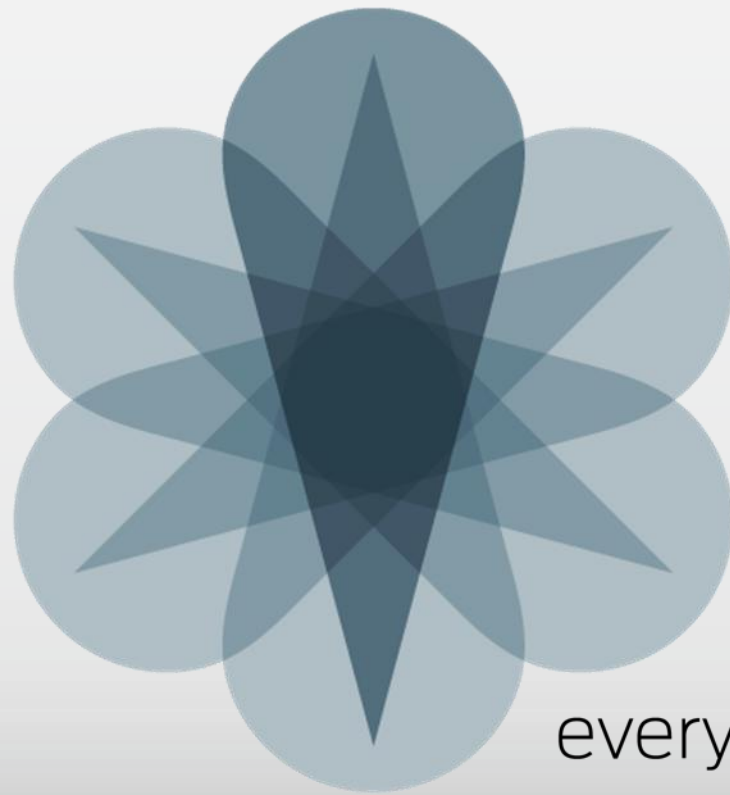
- **SP Case Study on CGNAT and DSLite**
- Tuesday May 31st 3pm GMT/ Michael Melloul

Webinar 5

- **IPv6 in the Mobile**
- Tuesday June 7th 3pm GMT/ TBD

In case you couldn't attend some of the sessions, you can watch the playbacks

Webinar Series at <http://juniper-emea.net/content/ipv6webreg>



everywhere