Technologies for IPv4/IPv6 coexistance

PLNOG 2

Robert Raszuk

raszuk@juniper.net

Jan'2009

Agenda:

- **Background**
- **Prerequisites: SIIT, NAT/NAT-PT/NAPT-PT, cones..., ALG, DNS ALG, CGN,**
- **Translation solutions (stateless & stateful):**
  - *** IVI, NAT6, NAT64, sNAT-PT; Control: ICE,**
  - *** DNS Application Layer Gateway/DNS64, FTP ALG ...**
- **Tunneling solutions:**
  - *** Dual stack lite (former SNAT), DSTM**
- **Address/Port mapping/allocation & tunneling solutions:**
  - *** A+P, SAM (former APBP), NAT444**
- **Standards**

Background ....

- **Where all the action is taking place in IETF ...**
  - **Behave WG**
  - **Softwire WG**
  - **V6ops WG**
  - **INTAREA WG**
  - **Interim Meetings**

- **Who are the leading customers:**
  - **NTT (interested in NAPT CGNs)**
  - **Comcast (interested in Dual stack lite)**
  - **CERNET/CERNETv2 (is testing IVI ... Eager to try SDK)**

Background ....

- Yesterday's panel indicated a very scary desire ... Most people were debating on how to switch to v6. Bad news for those ...

It is not going to happen soon if at all !!!!

- It it just like trying to jump from one already established v4 train network to another incompatible, lacking coverage & brand new v6 train network without any **hubs** interconnecting them.

- Experience of real early adopters of v6 for end customers (Japan, China) proves that only joining the train networks and allowing folks to run on the combined v4v6 one works. Customers can build dual stack servers to at the end in years to come - move to the v6.

- This presentation talks partially about how to build those interconnect hubs as well as how to squeeze more out of current v4 address space.

Background ....

- **Illustration of v4/v6 in train analogy:**

No more room in v4

Quite empty v6

Background ....

- **People say IPv4 addresses are over ....**

Not true at all !!!

- **What is true is that RIRs will soon have no free pools for allocation**
- **But there are many idle v4 addresses allocated which are going to be sold commercially and to get IPv4 address range will be just a matter of $$$.**
- **The process is already starting, RIRs are publishing documents when IPv4 address space resale is going to be legal. (ie. when no more IPv4 addresses is to be available).**

- **And those interconnect hubs of v4 & v6 must be easy and transparent for any users. Those must include:**
- **DNS ALG**
- **Packet translations**
- **Routing advertisements**

Agenda:

- Background
- Prerequisites: SIIT, NAT/NAT-PT/NAPT-PT,NAT-types, ALG, DNS ALG, CGN,
- Translation solutions (stateless & stateful):
  * IVI, NAT6, NAT64, sNAT-PT,
  * DNS Application Layer Gateway/DNS64, FTP ALG ...
- Tunneling solutions:
  * Dual stack lite (former SNAT), DSTM
- Address/Port mapping/allocation & tunneling solutions:
  * A+P, SAM (former APBP), NAT444
- Standards

Stateless IP/ICMP translation SIIT

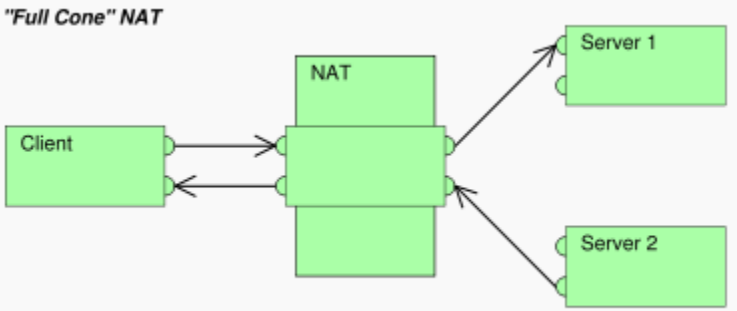
- **Defined in RFC 2765 & draft-baker-behave-v4v6-translation**
- **To make IPv6-only nodes interoperate with IPv4-only nodes**
- **Designed to work for attaching v6 site with v6 only hosts to v4 networks**
- **Not desinged to be used as v6 to v6(withv4) address translator (IVI) due to issue with v6 routing requirement to carry full v4 table embedded in v6 addresses (unless stub)**
- **Translator does not retain any state from one packet to the other even when there is need to insert new checksum in UDP & ICMP headers. Translators are independent and routing does not need to be symmetrical.**
- **This does not work for multicast and is hard for IPSec tunnel mode.**
- **Host applications do not require any modifications.**
- **Solution scales very well as it does not leave any per packet/per session state.**
- **Works with the notion of mapped v4 to v6 addresses, mapping format differs**
- **draft-baker proposes to add stateful mode translation to RFC2765**

General NAT 101

- **NAT – device which modifies src and/or dst addresses**
- **NAPT – device which modifies src/dst address & port numbers (PAT)**
- **NAT/NAPT-PT → modifies src/dst address, port numbers and performs protocol translation v4_to_v6 and/or v6_to_v4**
- **NAT Types: Full cone, Address-Restricted cone, Port-Restricted cone, Symmetric**

Full cone NAT, also known as one-to-one NAT

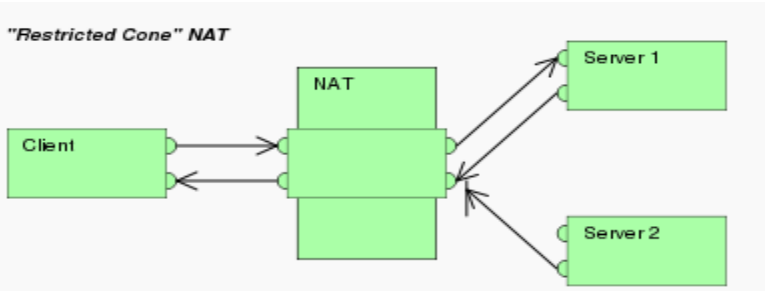
- Once an internal address (iAddr:port1) is mapped to an external address (eAddr:port2), any packets from iAddr:port1 will be sent through eAddr:port2. Any external host can send packets to iAddr:port1 by sending packets to eAddr:port2.



General NAT 101

Address-Restricted cone NAT

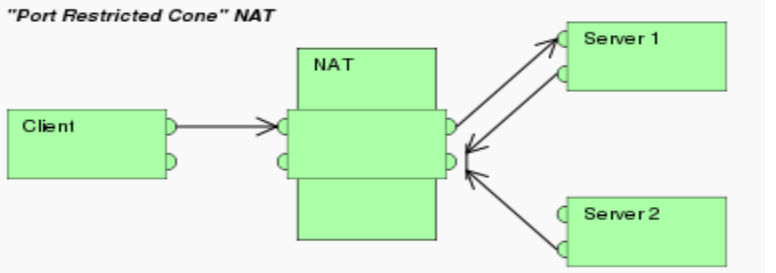
- Once an internal address (iAddr:port1) is mapped to an external address (eAddr:port2), any packets from iAddr:port1 will be sent through eAddr:port2. An external host (hostAddr:any) can send packets to iAddr:port1 by sending packets to eAddr:port2 only if iAddr:port1 had previously sent a packet to hostAddr:any. "any" means the port number doesn't matter.



Port-Restricted cone NAT

Like a **restricted cone NAT**, but the restriction includes port numbers.

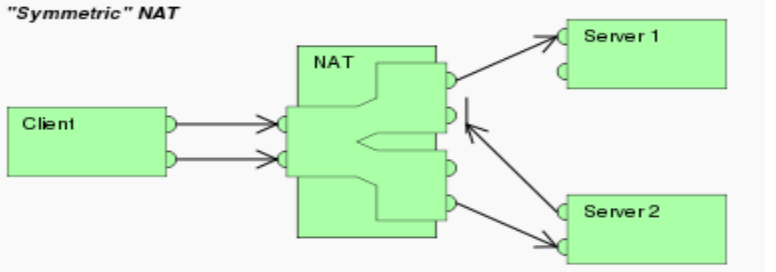
- Once an internal address (iAddr:port1) is mapped to an external address (eAddr:port2), any packets from iAddr:port1 will be sent through eAddr:port2. An external host (hostAddr:port3) can send packets to iAddr:port1 by sending packets to eAddr:port2 only if iAddr:port1 had previously sent a packet to hostAddr:port3.



Symmetric NAT

- Each request from the same internal IP address and port to a specific destination IP address and port is mapped to a unique external source IP address and port.

  If the same internal host sends a packet even with the same source address and port but to a different destination, a different mapping is used.
- Only an external host that receives a packet from an internal host can send a packet back.



General stateful NAT ...

... should conform to number of behavioral requirements:

- **RFC4787 NAT behavioral requirements for unicast UDP**
- **RFC5382 → NAT behavioral requirements for TCP**
- **RFC5135 → Multicast across NAT & NAPT**
- **draft-ietf-behave-nat-icmp-08 → NAT ..... for ICMP**

.... Would be good if it also supports ...

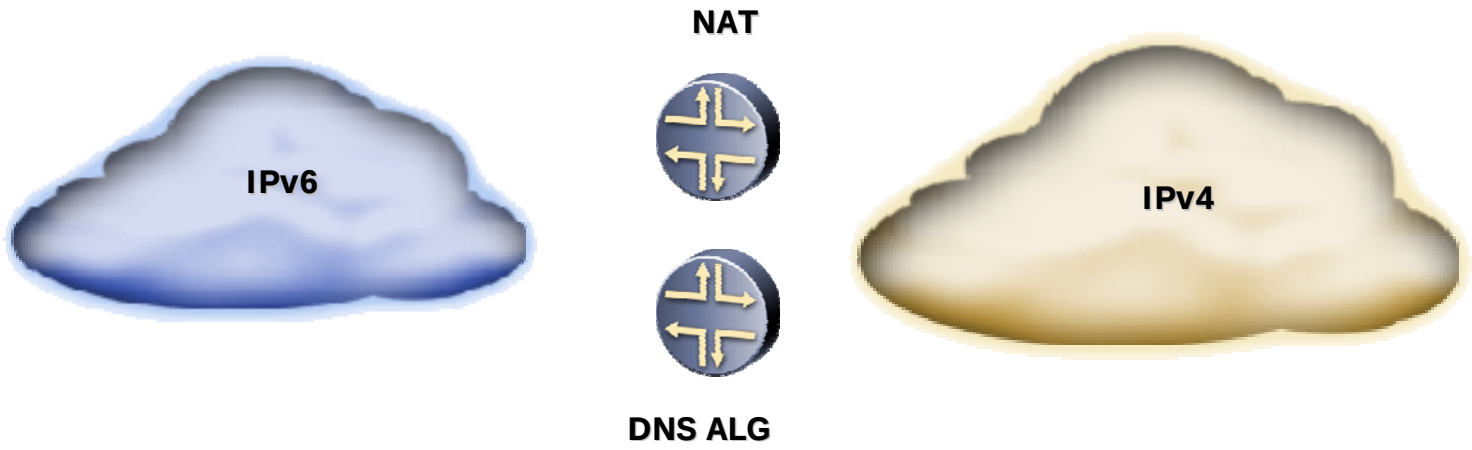
- **Some sort of port mapping protocol example: draft-cheshire-nat-pmp**
- **ALGs for FTP, RTSP/RTP, IPSec pass through, PPTP VPN pass ...**

Agenda:

- Background
- Prerequisites: SIIT, NAT/NAT-PT/NAPT-PT, cones..., ALG, DNS ALG, CGN,
- Translation solutions (stateless & stateful):
  * IVI, NAT6, NAT64, sNAT-PT,
  * DNS Application Layer Gateway/DNS64, FTP ALG ...
- Tunneling solutions:
  * Dual stack lite (former SNAT), DSTM
- Address/Port mapping/allocation & tunneling solutions:
  * A+P, SAM (former APBP), NAT444
- Standards

Translation solutions

- **v4 to v6 translation and v6 to v4 translation ... stateless or statefull translation**



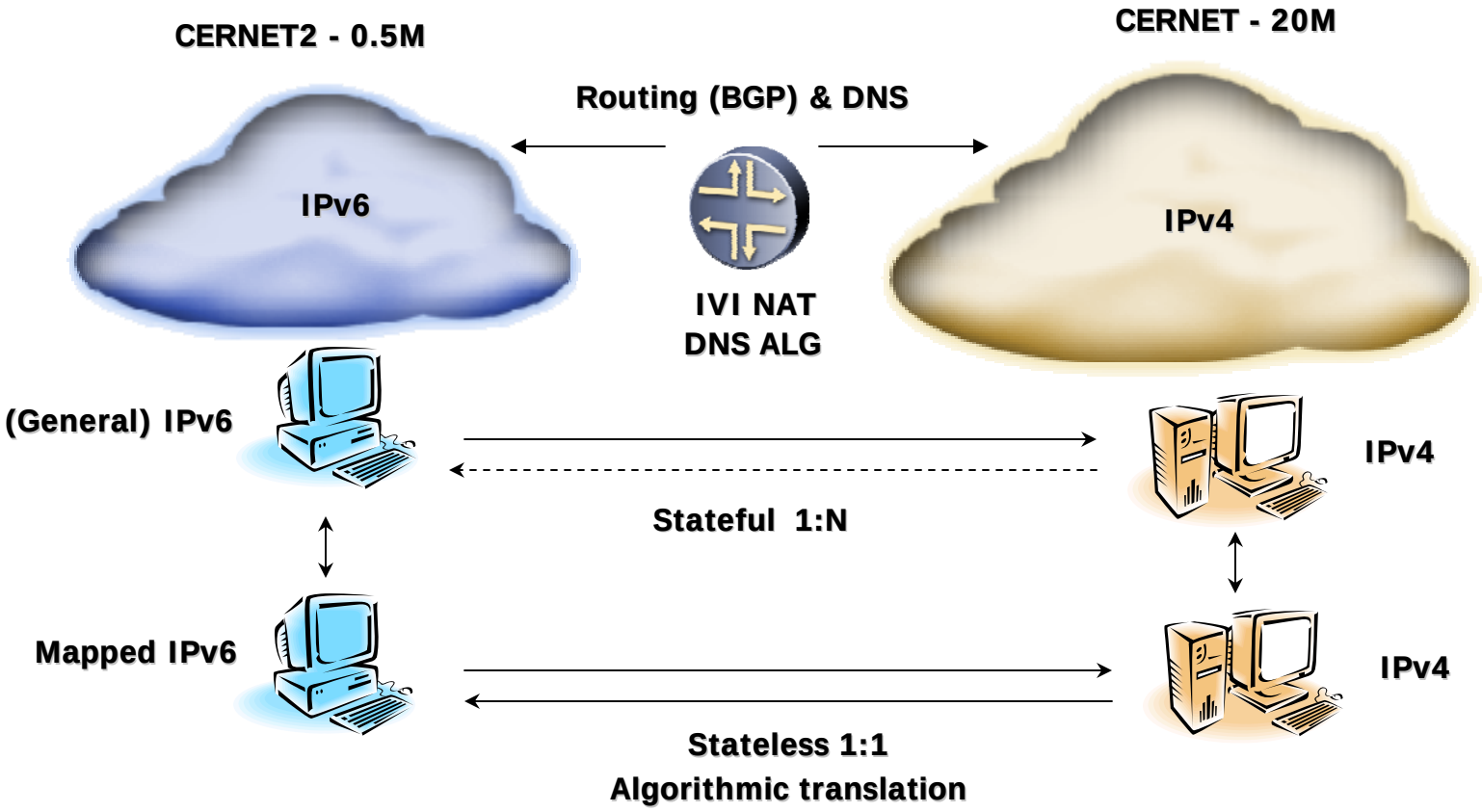
- **state – dynamic per flow or per host state in NAT box**
- **stateless – v4 address embedded in the v6 packet (mapped)**
- **statefull – address/port translation kept in NAT box (unmapped)**

Translation solutions ...

- draft-baker-behave-v4v6-framework → Framework for v4/v6 translation
- draft-bognulo-behave-nat64
- draft-baker-behave-v4v6-translation
- draft-baker-behave-ivi

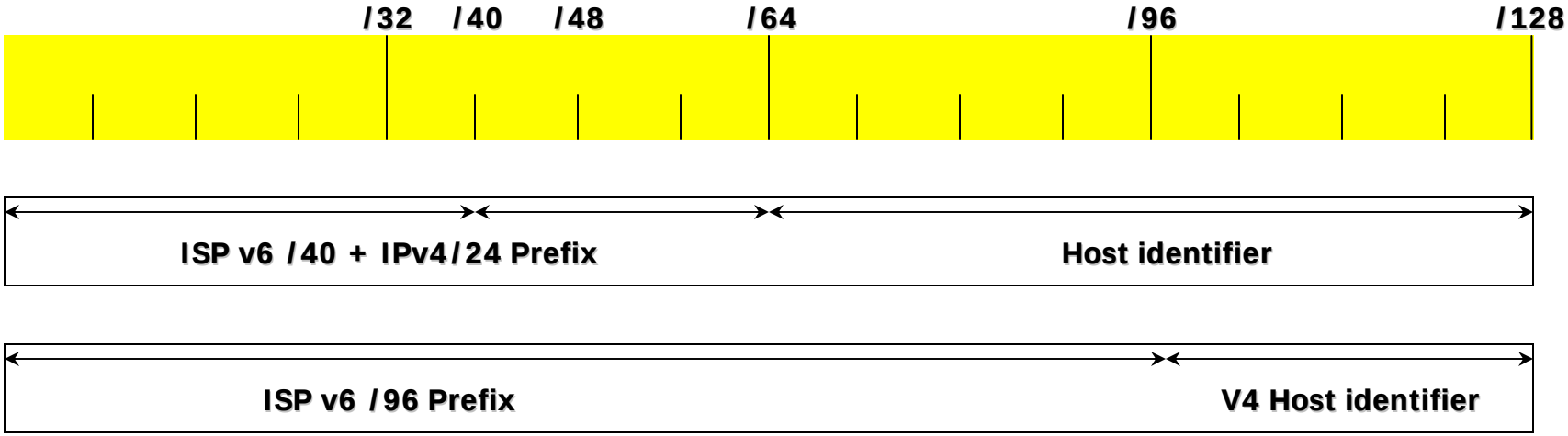
Translation solutions ...

- IPv4 address pool – kept in NAT for stateful translation or assinged as part of one of IPv6 addresses in end hosts



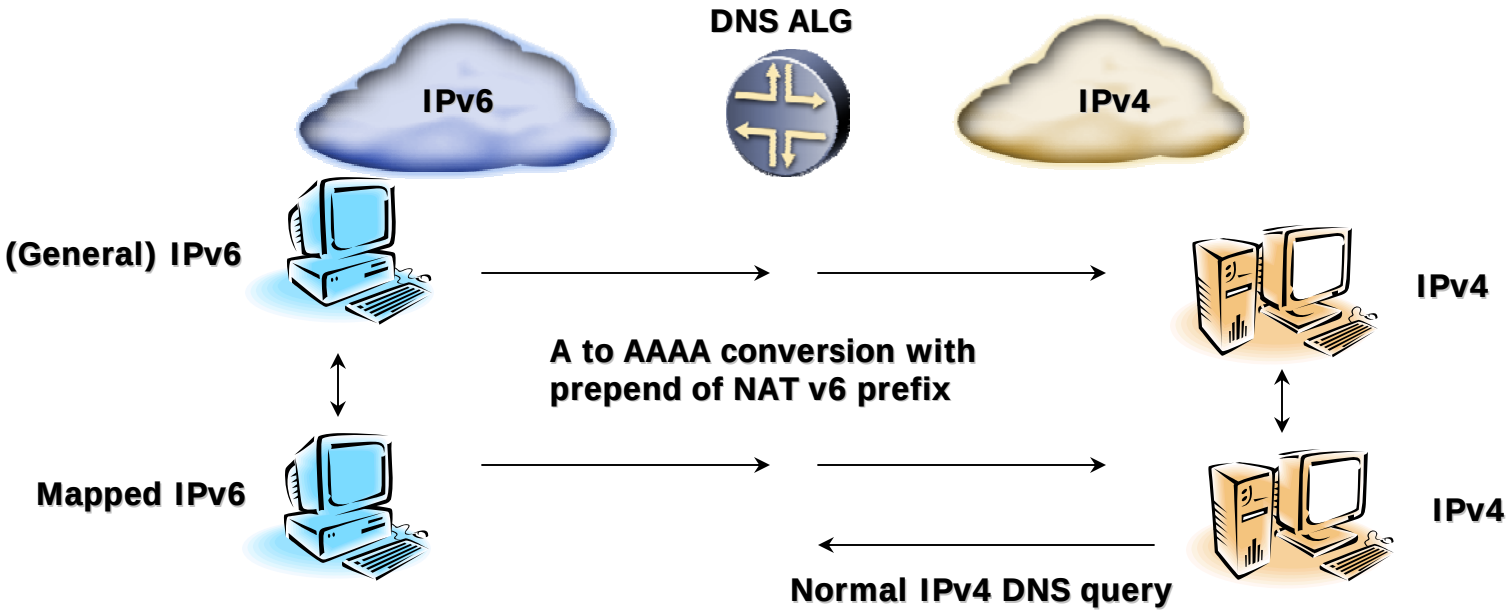
Translation solutions ...

- For v4 mapped v6 addresses prefix /40, /48, /64 or /96
- Prefix 64::/96 appropriate for the CPE & for stub v4 networks
- Putting upper part of v6 prefix in routing locators fits ISP usage



Translation solutions ...

- DNS translation is required in v6 to v4 direction
- Hosts resolve DNS names to addresses
- Mapped v4 portions of v6 addresses are native in IPv4 DNS
- For IPv6 hosts A records are converted to AAAA records by DNS ALG



Translation solutions ...

- Hosts should promote and choose use of mapped addresses when initiating conversation with v4 sites per RFC3484 address selection
- Src & Dst addresses should be as similar as locally possible
- The translation solutions generally do not address v4 address space extension
- The exmaple of IVI implementation available for download as both linux kernel patch as well as as DNS ALG proxy implementation

http://v6s.6test.edu.cn/impl/

Agenda:

- Background
- Prerequisites: SIIT, NAT/NAT-PT/NAPT-PT, cones..., ALG, DNS ALG, CGN,
- Translation solutions (stateless & stateful):
  * IVI, NAT6, NAT64, sNAT-PT,
  * DNS Application Layer Gateway/DNS64, FTP ALG ...
- Tunneling solutions:
  * Dual stack lite (former SNAT), DSTM
- Address/Port mapping/allocation & tunneling solutions:
  * A+P, SAM (former APBP), NAT444
- Standards

What problem are we solving ?

Providers are facing shortage of IPv4 addresses to assign to each customer's CPE or host so that every customer has usable IPv4 connectivity !!!

Private RFC1918 also exhousted in some of large operators.

V4v6 mapped addresses of no help.

Tunneling solutions DS-Lite

- Preserves access to v4 for customers who can not operate over v6 due to legacy equipment
- Allows to share globally unique v4 address by many customers via NAT aggregation
- Provides transport for v4 customers over v6 only core
- Allows for customer static port mapping on NAT via web interface
- Tunnels terminate on CGN NAT44 box
- No need to change DNS .. No need to fabricate AAAA records
- No restrictions on NAT-PT & DNS ALG topologies
- Application ALGs simpler with 32/32 bit translations rather then with 128 to 32 bits within the payload

Tunneling solutions

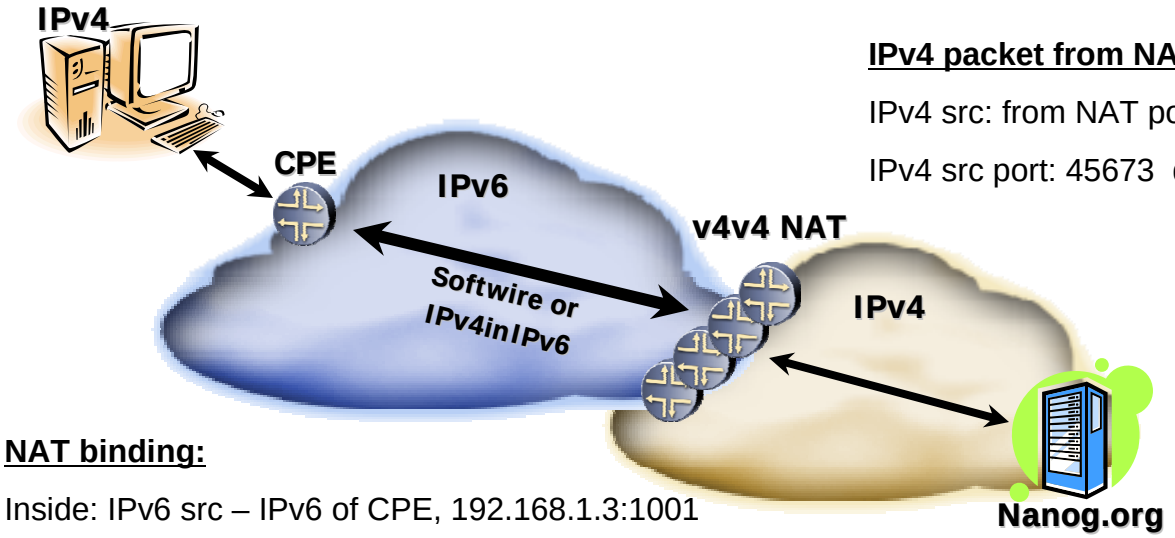
- **Dual-Stack Lite (CPE only assinged IPv6)**
- **Home host no changes (get's IPv4 from CPE as today)**

IPv6 packet from CPE to v4v4 NAT

IPv6 src: IPv6 of CPE dst: IPv6 of v4v4 NAT

IPv4 src: 192.168.1.3 dst: www.nanog.org (resolved via normal DNS) 198.108.95.1

IPv4 src port: 1001 dst port: 80



IPv4 packet from NAT to Internet resource:

IPv4 src: from NAT pool dst: 198.108.95.1

IPv4 src port: 45673 dst port: 80

NAT binding:

Inside: IPv6 src – IPv6 of CPE, 192.168.1.3:1001

Outside: IPv4 from NAT pool:45673

Tunneling solutions

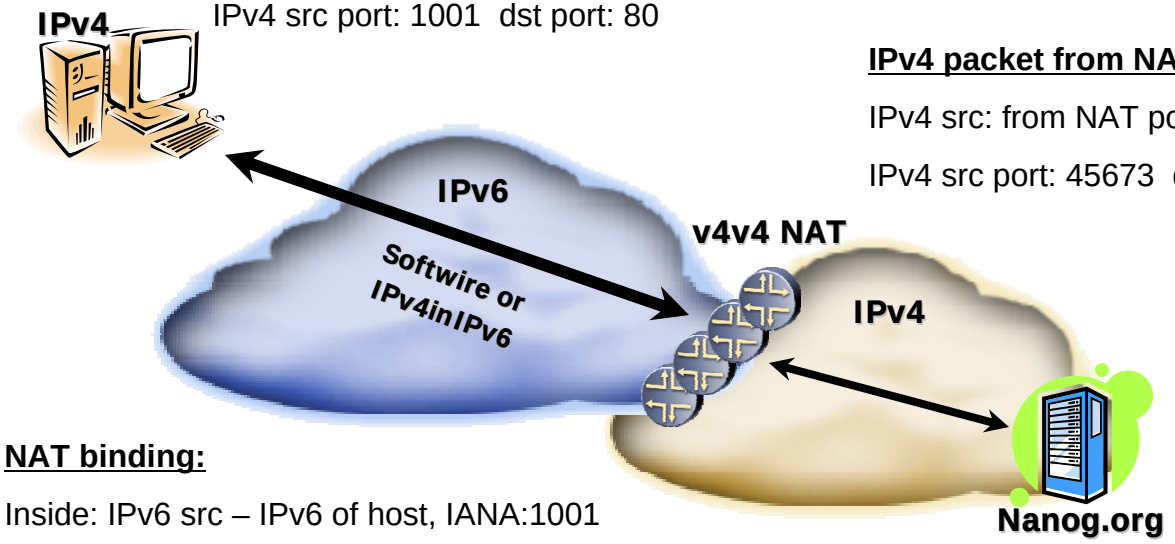
- **Dual-Stack Lite (Host only assinged IPv6, IPv4 on host local only IANA)**
- **Dual stack capable host, but IPv6 only assinged !**
- **All host would have the same IANA IPv4 address**

IPv6 packet from host to v4v4 NAT

IPv6 src: IPv6 of host dst: IPv6 of v4v4 NAT

IPv4 src: well known dst: www.nanog.org (resolved via normal DNS) 198.108.95.1

IPv4 src port: 1001 dst port: 80



IPv4 packet from NAT to Internet resource:

IPv4 src: from NAT pool dst: 198.108.95.1

IPv4 src port: 45673 dst port: 80

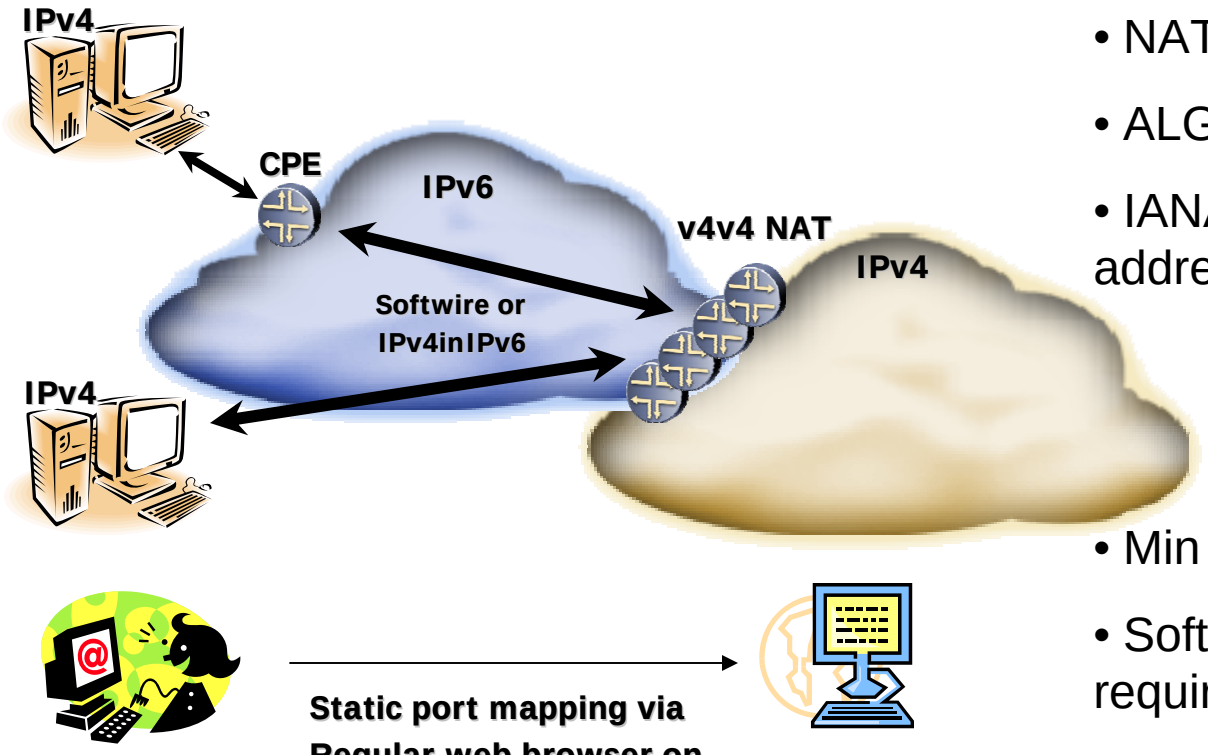
NAT binding:

Inside: IPv6 src – IPv6 of host, IANA:1001

Outside: IPv4 from NAT pool:45673

Tunneling solutions

- **Dual-Stack Lite (CPE only assinged IPv6)**
- **Static port mapping**



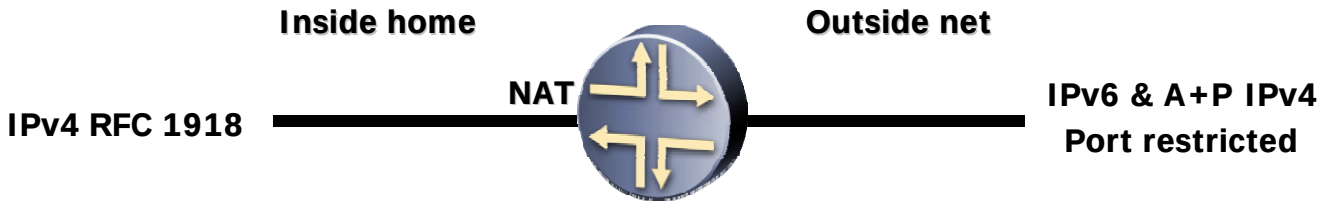
- Merged of DS-lite & S-NAT
- NAT can be very well distributed
- ALG discussion
- IANA to reserve well know IPv4 address block (likely /30)
- Min IP in IP if no control needed
- Softwire L2TP if tunnel control required
- NAT to implement state synchronization & anycasting

Tunneling solutions

- Tunneling between CPE/Host & NAT allows for additional layer of benefits:
  - Allows to place the NAT termination anywhere
  - Allows for placement beyond immediate ISP provider – virtual ISP model – opportunity for new reveneu – already in test/production in Japan for years
  - Use of well known IP-in-IP tunneling protocol
  - Easy horizontal NAT scaling
  - Tunnel termination point can be learned via DHCP

Tunneling solutions

- **DS-Lite & A+P merged ➔ A+P Lite (new name candidate)**
- **What is A+P ?**
  - Divide UDP & TCP port numbers into blocks
  - Assign customers (hosts or CPEs) address + port range to use
  - Share the same address with different port range between many customers
  - Limit routing on address + port to a very small area or eliminate by tunneling in v6 !
- **A+P Gateway:**



Gateway Can either do NAT into assigned port range, pass packets through (for later NAT-ing) and/or port forward A+P ports to end-hosts

Tunneling solutions

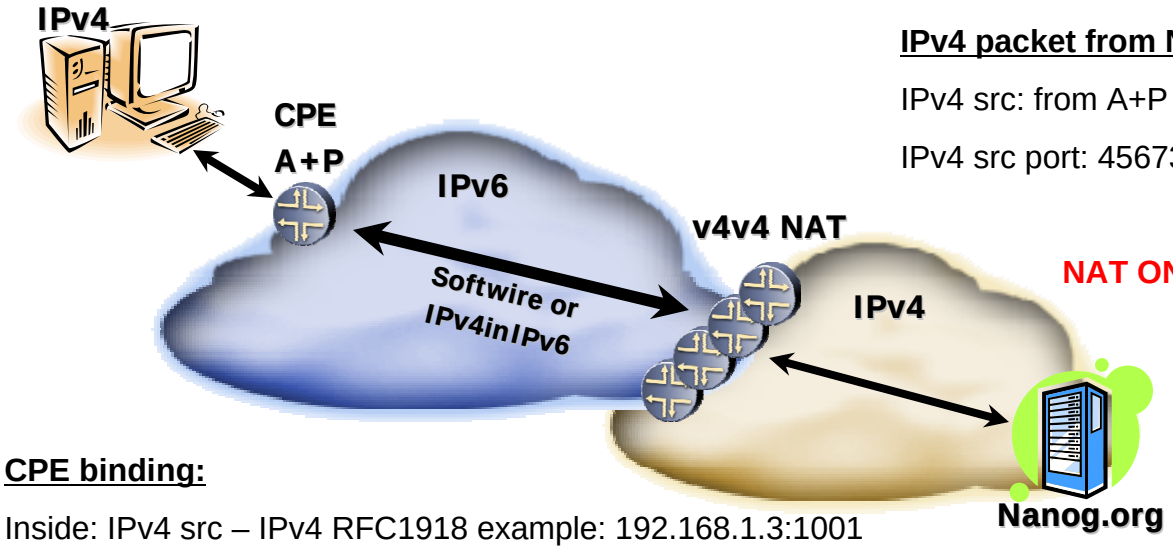
- **Dual-Stack Lite & A+P (example: 133.15.10.24 ports: 45000-46000)**
- **Home host no changes (get's IPv4 from CPE as today ex: rfc1918)**

IPv6 packet from CPE to v4v4 NAT

IPv6 src: IPv6 of CPE (A+P mapped CPE address) dst: IPv6 of v4v4 NAT

IPv4 src: A+P dst: www.nanog.org (resolved via normal DNS) 198.108.95.1

IPv4 src port NATed: 45673 dst port: 80



IPv4 packet from NAT to Internet resource:

IPv4 src: from A+P 133.15.10.24 dst: 198.108.95.1

IPv4 src port: 45673 dst port: 80

NAT ONLY strips/applies v6 header if v4 is A+P !!!

No state in NAT

CPE binding:

Inside: IPv4 src – IPv4 RFC1918 example: 192.168.1.3:1001

Outside: IPv4 from A+P pool: 133.15.10.24:45673

Tunneling solutions

- DS-Lite & A+P merged port allocation/mapping:
- Every customer may provision a fixed set of A+P ports on his CPE which will not to be subject to translation
- Can manually specify reserved and mapped ports (example: via a web site which looks like a home NAT today)
- A larger pool will be allocated on the fly and will pass untouched via provider's NAT
- CPE learns IPv4+Ports assignment via DHCP or other NAT to CPE signalling
- If Provider NAT get's already NATed packets it transparently passes it (removing v6 encap) and conducting BCP38 check (src addresses)
- If packet is not NATed (not from A+P range) NAT operation is performed

Solutions constrains discussion:

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.

Solutions constrains discussion:

5) Efficiency vs. complexity
   Operator has the flexibility to trade off between port multiplexing efficiency (LSN) 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.

Agenda:

- Background
- Prerequisites: SIIT, NAT/NAT-PT/NAPT-PT, cones..., ALG, DNS ALG, CGN,
- Translation solutions (stateless & stateful):
  * IVI, NAT6, NAT64, sNAT-PT,
  * DNS Application Layer Gateway/DNS64, FTP ALG ...
- Tunneling solutions:
  * Dual stack lite (former SNAT), DSTM
- Address/Port mapping/allocation & tunneling solutions:
  * A+P, SAM (former APBP), NAT444
- Standards

Standards review: IPv6

- RFC2460 Main IPv6
- RFC2461 Neighbor Discovery for IPv6
- RFC2462 IPv6 Stateless autoconfiguration
- RFC2463 ICMP for IPv6 specification
- RFC3041 Privacy extension for stateless autoconfig in v6
- RFC3177 IAB/IESG recommendation on address alocation
- RFC3484 Default address selection on v6 nodes & v4v6 dual nodes

Standards review: IPv6

- RFC3879 ULAs – Unique local v6 addresses
- RFC4192 IPv6 gracefull renumbering
- RFC4291 (obs: 2373/3513) IPv6 Address Architecture
- RFC4864 Local Network Protection (LNP) for IPv6
- IPv6 unicast address assignment considerations draft-ietf-v6ops-addcon-10
- IPv6 scoped address architecture (draft-ietf-ipngwg-scoping-arch)
- Mobility support for IPv6 (draft-ietf-mobileip-ipv6)

Standards review: DNS

- **RFC2766 DNS Application Layer Gateway**
- **RFC3493 Socket API changes for TCP/IP to support IPv6**
- **RFC3596 DNS Extensions to Support IP Version 6**
- **RFC4472 Operational issues with IPv6 DNS**
- **draft-bagnulo-behave-dns64-01 → DNS64**
  - DNS extension for NAT from IPv6 clients to IPv4 servers
  - Synthesizes AAAA resource record from A record by prepending /xx IPv6 prefix of a NAT box (ex: /96)
  - Only above prefix is shared between NAT & DNS64
  - DNS64 can reside in NAT, in host or anywhere in the domain

Standards review: v4/v6 Tunneling based

- **RFC2529 Transmission of v6 over v4 without explicit tunnels**
- **RFC3056 Connection of IPv6 domains via IPv4 clouds**
- **RFC4213 Basic Transition Mechanisms for IPv6 Hosts and Routers** (original dual stack recommendation, v6 over v4 tunnels)
- **RFC4380 Teredo → Tunneling IPv6 over UDP via NAT**
- **RFC5214 ISATAP → Intra-site automatic tunnel addressing**
- **draft-tschofenig-v6ops-secure-tunnels → ipsec for v6 over v4**
- **draft-bound-dstm-exp-04 Dual stack v6 dominant transition**
  - DSTM proposes temporary allocation of IPv4 addresses on demand
  - uses DHCP for allocation, no new protocol, single domain.

Other recommended & interesting reading:

- **RFC4966 → Reasons to move NAT to historic status**
- **draft-wing-nat-pt-replacement-comparison**

  Comparison of proposals to replace NAT-PT
- **draft-shirasaki-isp-shared-address**

  4 blocks of /8 of IPv4 addresses would suffice per ISP
- **draft-mrw-behave-nat66-00.txt**

Acknowledgments:

• Those slides would not be possible without work of many indivuals in this area: Miya Kohno, Shin Miyakawa, Fred Baker, Randy Bush, Alain Durand, Xing Li ...

• Participation in IETF, NANOG, RIPE & APRICOT conferences helps a lot

• Discussions with key large service providers and network operators was a great value !

• And kudos for wikipedia too ....

Thanks!