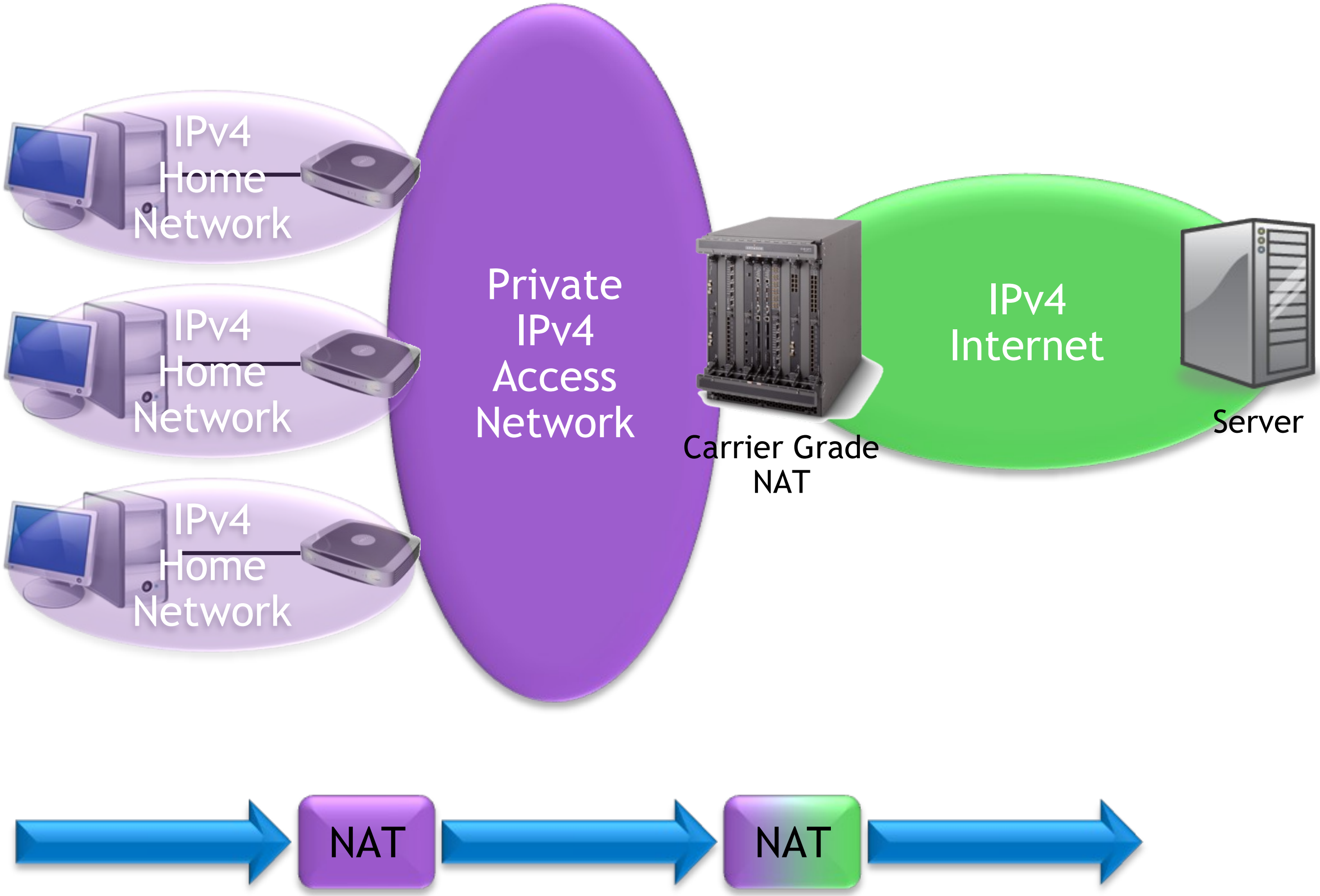
Subscriber Aware NAT

David Miles
david.miles@alcatel-lucent.com

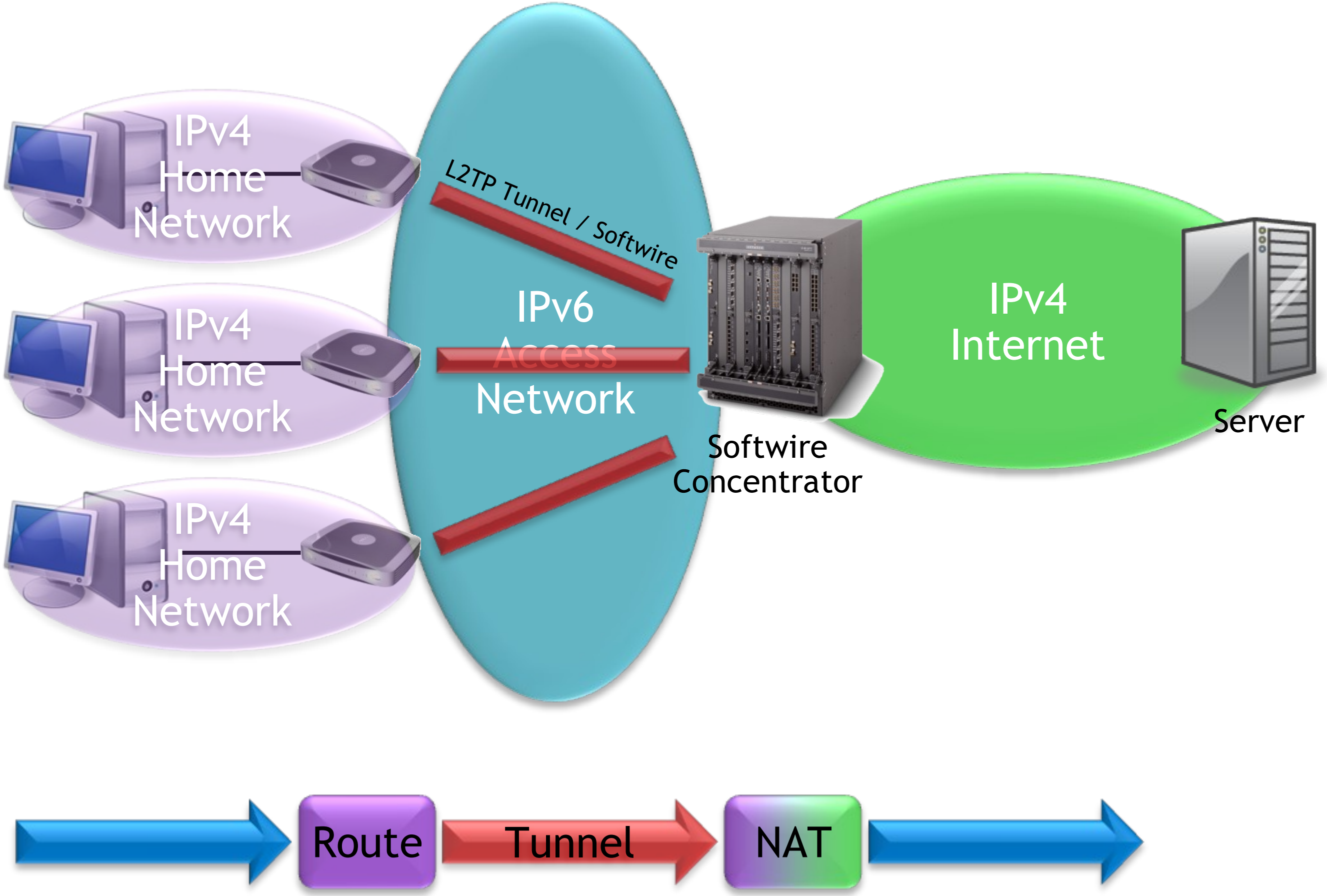
1 IPv4 Continuity

Address Exhaustion

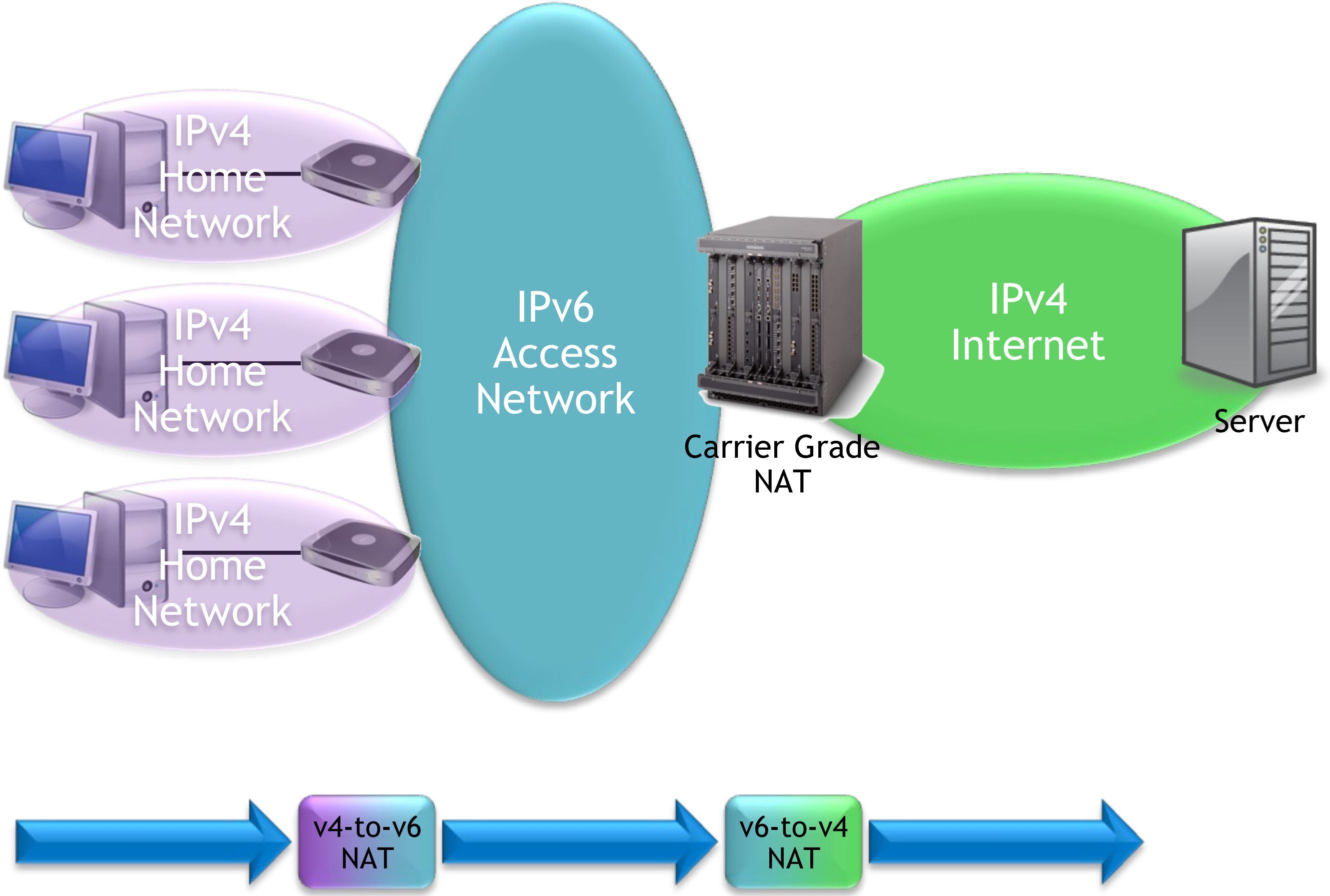
Carrier Grade NAT



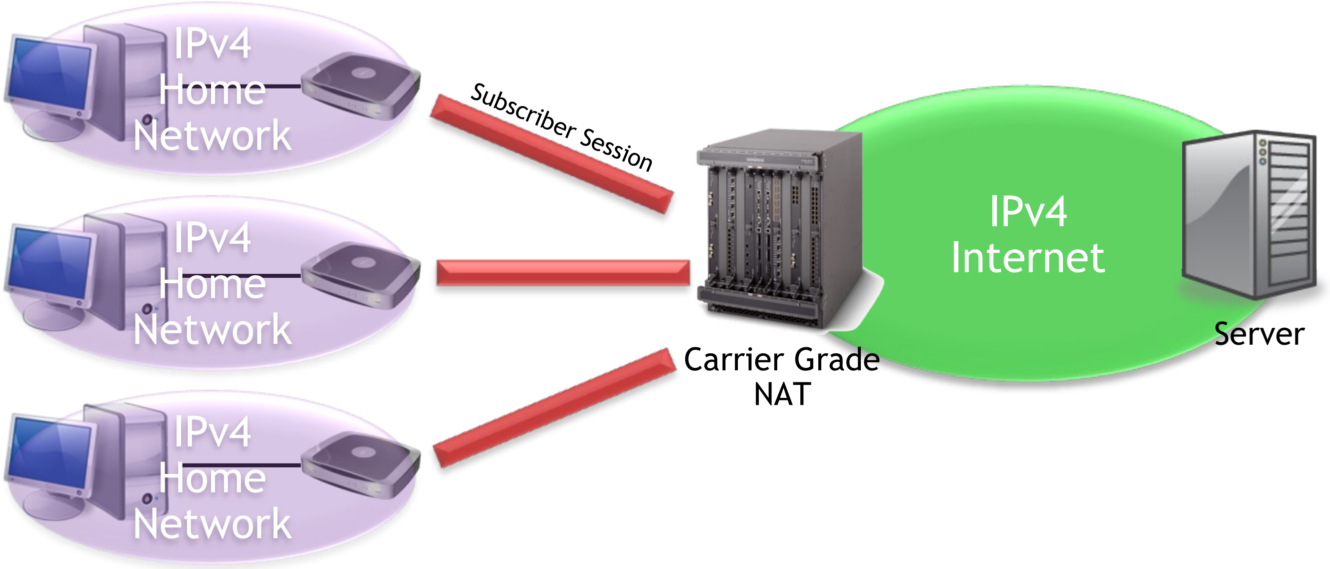
Dual Stack Lite

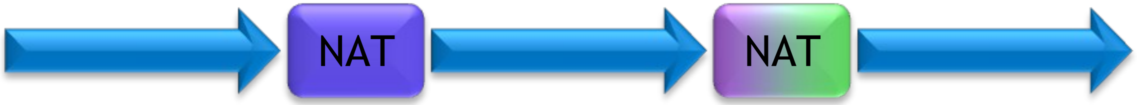


v4-to-v6 Translation



Subscriber-Aware NAT





Options for IPv4 Overloading

Carrier Grade NAT	Dual Stack Lite	IPv4-to-IPv6 Translation	Sub-Aware NAT
No CPE Change	Requires CPE to support Softwires	May require CPE change	No CPE Change
IPv4-to-IPv4 NAT	IPv4-to-IPv4 NAT	IPv4-to-IPv6 NAT	IPv4-to-IPv4 NAT
CGN can be deployed anywhere in the network	Dual-Stack Lite must be deployed in the Softwire Concentrator	IVI can be deployed anywhere in the network	Sub-Aware NAT must be deployed in the BNG
May need a large (/16) assignment re-used in the network	Can use any address	IPv4 addresses translated	Can use any address
Application Servers can sit between subscriber and CGN	All IPv4 traffic must be subject to NAT	All IPv4 traffic must be subject to NAT	All IPv4 traffic must be subject to NAT

Network Address Translation Issues

- Cannot support unsolicited inbound traffic (to broadband subscribers).
- Limited to client-server model
- “Port-forwarding” is not scalable as TCP/UDP use the concept of well-known ports. Ie; 80-HTTP, 443-HTTPS. One port+one IP = one server
- Must consider how to limit per-subscriber sessions so all ports are not consumed. For example, no more than 100 sessions per sub. Also what do we do when all sessions are exhausted? Redirection to captive portal?
- NAT does not address running a server of any kind, including that needed for DSLForum TR-69 (ACS server communicates to the gateway)

Carrier Grade NAT Issues

- Address space (*draft-shirasaki-isp-shared-addr-00*) between CPE and NAT device.
- Use of RFC1918 may collide with the addresses used within the subscriber LAN. A router cannot have the same subnet on two interfaces.
- No address space seems suitable or large enough to cover the number of expected subscribers.
- In the absence of an IANA allocation, proposals exist to “borrow” addresses that have been reserved for other purposes (such as the IETF test network).
- There will inevitably be multiple subscribers with the same address, so separated routing domains may be needed.

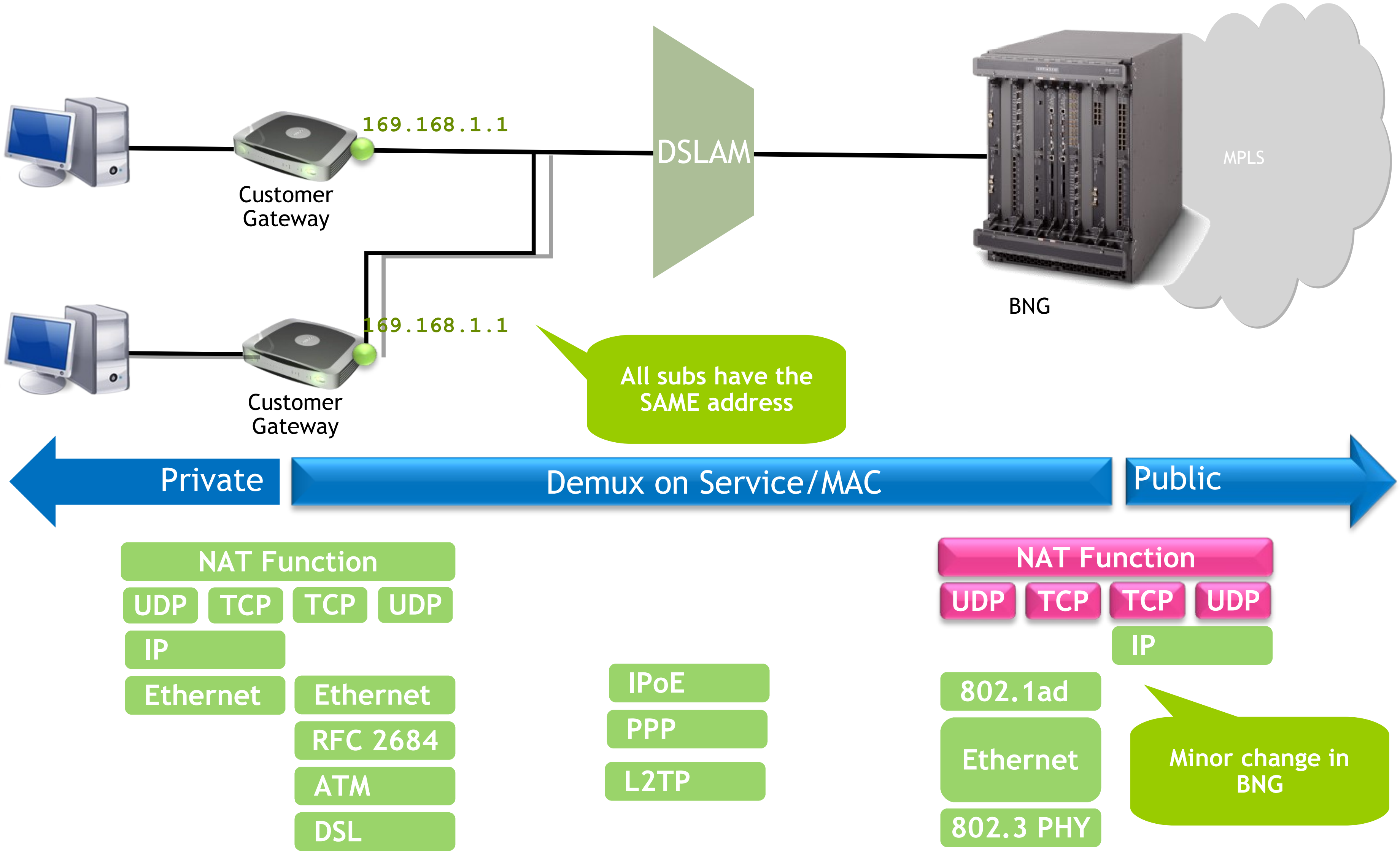
Dual-Stack Lite Issues

- Dual-stack lite is being merged with Softwires SNAT
- Addresses operators who want IPv6-only access networks
- Tunnels IPv4 in a L2TP tunnel, in turn over IPv6 (a Softwire)
- Proposes all endpoints get the same IP address
- But requires CPE change to support the L2TP tunnel
- NAT must be performed in the Softwire Concentrator
- Existing BRAS/BNG cannot apply policy to the tunnelled IPv4 traffic

Subscriber Aware NAT

- Does not require any CPE change
- Subscriber Aware NAT can support existing Windows 3.11, XP, Me, 98, XP, 2000, Vista, etc
- It can support a variety of link-layers and topologies: DSL TR-101, PPPoE, WiMAX, Mobile, Dial-up
- Must be implemented in the BNG/BRAS
- As it is in the subscriber-aware device, per-subscriber policy easily enforced: max-sessions, reserved ports/port-mapping
- Accounting records

Subscriber-Aware NAT



169.168.1.1

Customer Gateway

169.168.1.1

Customer Gateway

DSLAM

BNG

MPLS

All subs have the SAME address

Private

Demux on Service/MAC

Public

NAT Function

UDP TCP TCP UDP

IP

Ethernet

Ethernet

RFC 2684

ATM

DSL

IPoE

PPP

L2TP

NAT Function

UDP TCP TCP UDP

IP

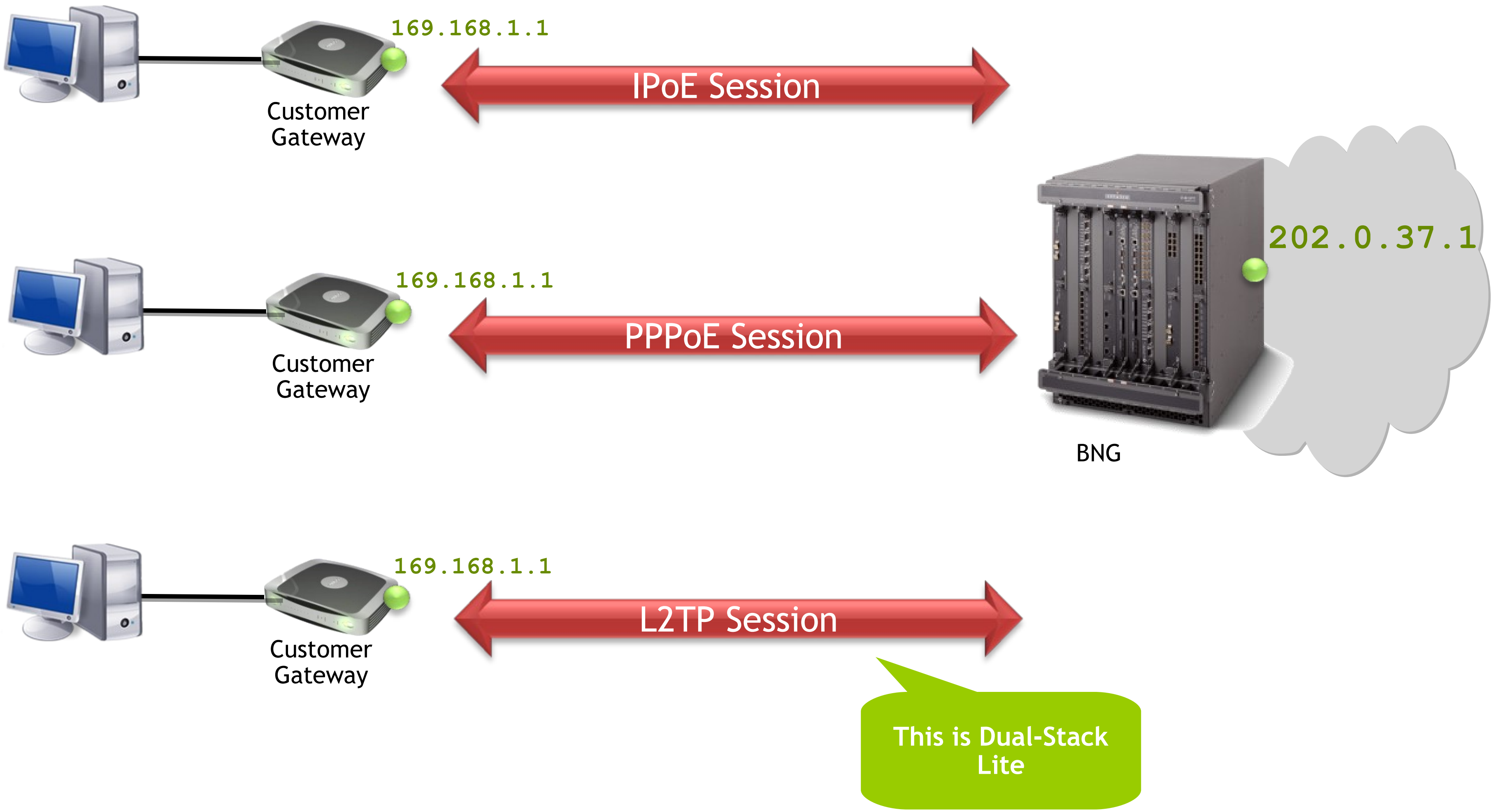
802.1ad

Ethernet

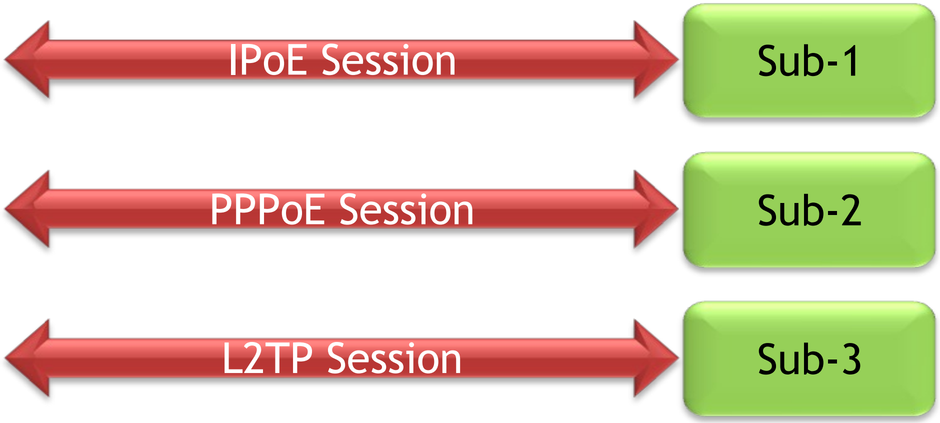
802.3 PHY

Minor change in BNG

Subscriber-Aware NAT



Subscriber-Aware NAT



IPoE Session

Sub-1

PPPoE Session

Sub-2

L2TP Session

Sub-3

Inside IP	Inside Port	Outside IP	Outside Port	Dest IP	Dest Port	Proto
Sub-1	6631	202.0.37.1	8897	88.3.4.2	80	TCP
Sub-2	7765	202.0.37.1	9822	88.3.4.2	80	TCP
Sub-2	7766	202.0.37.1	9893	88.3.4.2	80	TCP

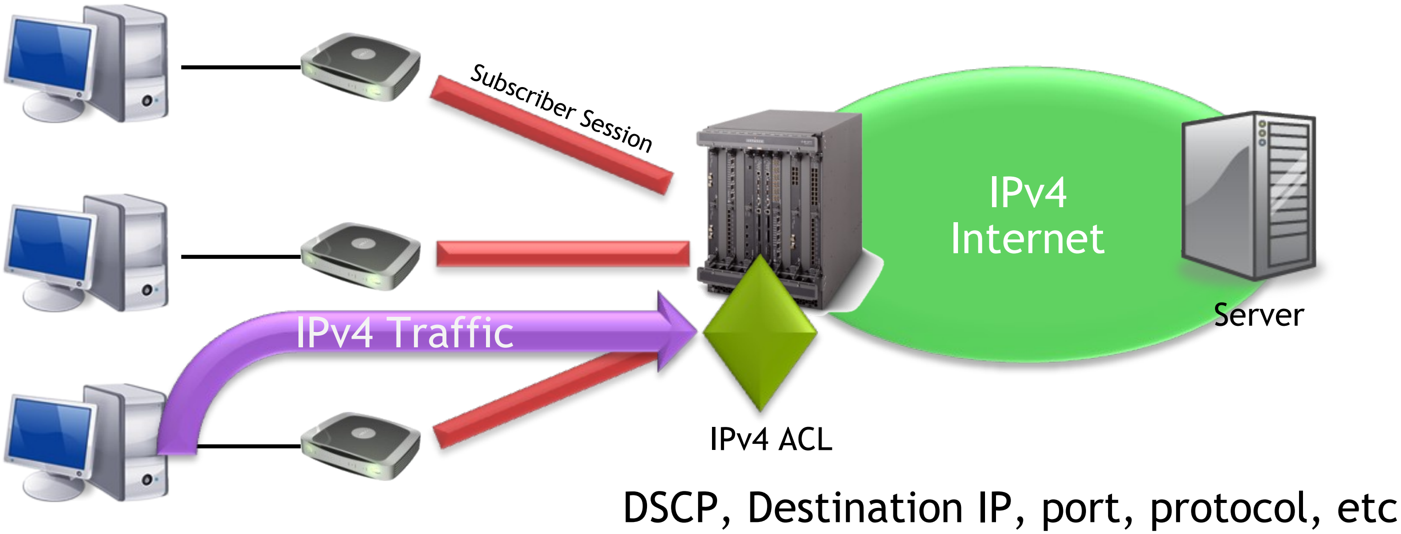
Inside Source IP no longer relevant for NAT

All subscribers have a common IP address (configurable)

2 Subscriber Aware NAT

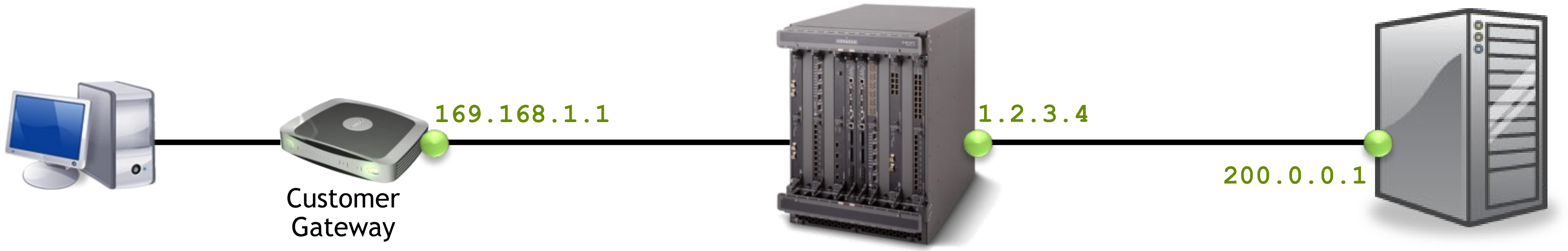
Operations

IPv4 and IPv6 ACL Continue to Work



BRAS/BNG polices still work with IPv4 traffic

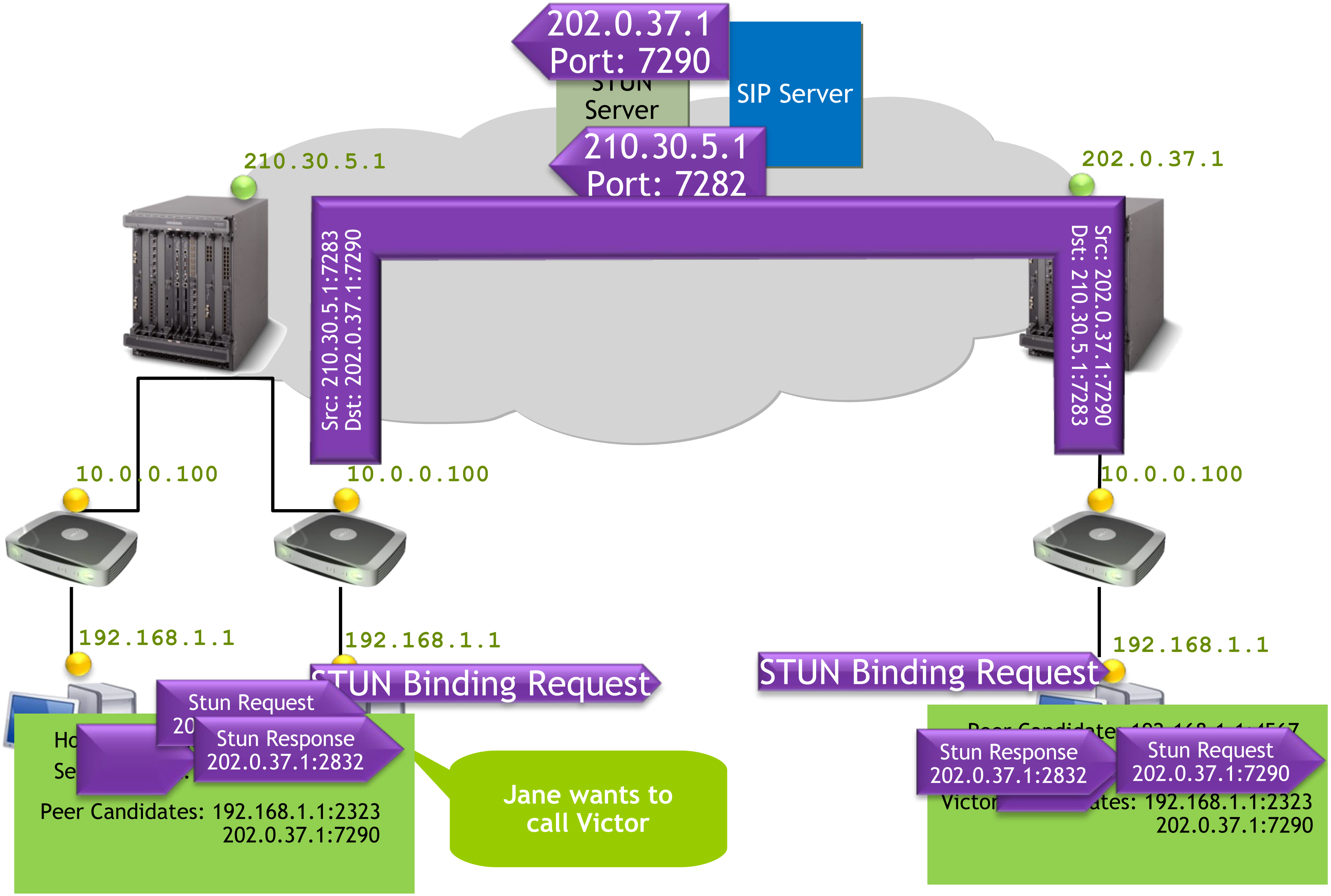
Endpoint Independent Filtering



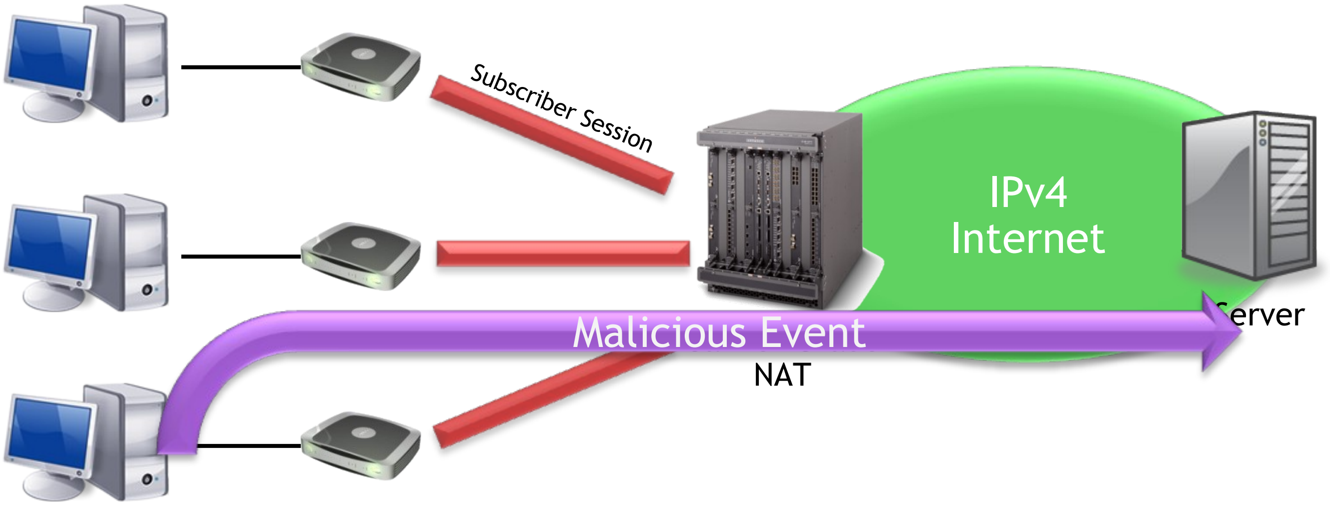
Subscriber	Proto	Inside IP	Port	Outside IP	Port	Dest IP	Port	State	Age
Sub-123	TCP	169.168.1.1	5623	1.2.3.4	10001	200.0.0.1	80	established	7740
Sub-123	TCP	168.168.1.1	5623	1.2.3.4	10001	*	*	listen	7740

This is the basis for STUN

Interactive Connectivity Establishment



Malicious Subscriber Tracking



How do you determine which subscriber accessed the governor's email?

Impact on Law Enforcement

Apache logs:

```
10.0.1.100 - - [21/Feb/2008:17:27:46 +1100] "GET / HTTP/1.0" 200 1456
10.0.1.100 - - [21/Feb/2008:17:27:47 +1100] "GET /apache_pb.gif HTTP/1.0" 200 2326
10.0.1.100 - - [21/Feb/2008:17:30:51 +1100] "GET / HTTP/1.0" 200 1456
10.0.1.100 - - [21/Feb/2008:17:30:51 +1100] "GET /apache_pb.gif HTTP/1.0" 304 -
10.0.1.100 - - [21/Feb/2008:17:31:10 +1100] "GET / HTTP/1.0" 200 1456
10.0.1.100 - - [21/Feb/2008:17:31:10 +1100] "GET /apache_pb.gif HTTP/1.0" 304 -
10.0.1.100 - - [21/Feb/2008:17:31:35 +1100] "GET / HTTP/1.0" 200 1456
10.0.1.100 - - [21/Feb/2008:17:31:35 +1100] "GET /apache_pb.gif HTTP/1.0" 304 -
```

Insufficient detail.

We need to have
Source Port

Apache Mod_Log_Config.c

Currently Apache access logs cannot log source port!
Need to make changes to the source code and recompile Apache

```
/*
 * log_remote_port patch
 */

static const char *log_remote_port(request_rec *r, char *a)
{
        apr_port_t rport;
        apr_sockaddr_port_get(&rport, r->connection->remote_addr);
        return apr_itoa(r->pool, rport);
}
```

http://www.onlamp.com/pub/a/apache/2004/04/22/blackbox_logs.html?page=3

Restricted Port Ranges

- WAND study suggests session setup rate in excess of 2 sessions/subscriber each second.
- A single BNG of 64,000 subscribers could generate over 256,000 create/stop mappings each second
- Not feasible to log this many transactions per second
- Alternative is to restrict each user to a pre-defined port range when a subscriber is substantiated and to provide this in RADIUS accounting

Subscriber	Outside IP	Port Start	Port Stop	Port Mask
Sub-1	1.2.3.4	4,069	8,191	4096/12
Sub-2	1.2.3.4	8,192	12,287	8192/12
Sub-3	1.2.3.4	12,288	16,383	12288/12

Finite Ports

- With any shared resource, one must manage and enforce reasonable-use limits
- When a subscriber exceeds a given number of ports, HTTP-intercept and display a warning. Only intercept when a new mapping is created (avoids impact to a page-load if threshold was exceeded mid-render)
- When port exceeded, new mappings are not created – ICMP messages (code 13) returned. It is not acceptable to destroy old but valid mappings.
- Certain services (by destination IP address, port, or a combination thereof) may be excluded from this threshold for critical services such as email, HTTP to the service provider’s portal, etc

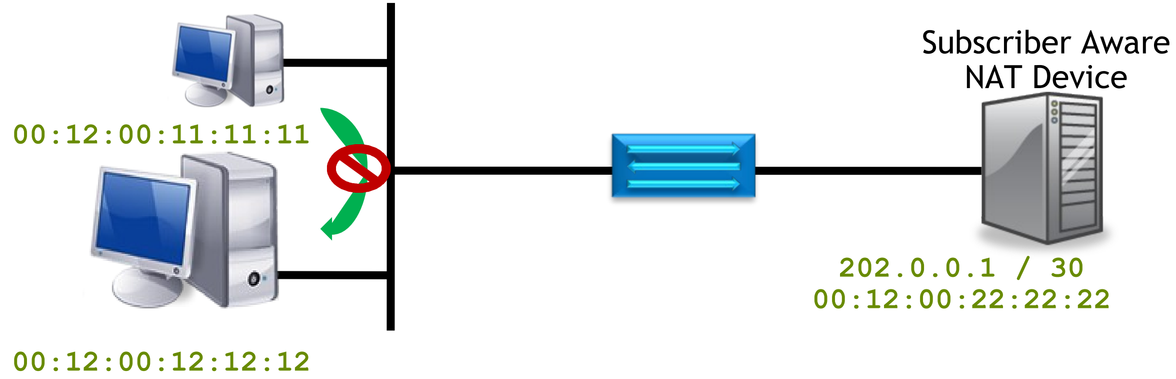
3 Subscriber Aware NAT

Prototype

Subscriber Aware NAT Prototype

- Implements a DHCPv4 Server that offers the *same* IPv4 address to all users
- Responds to ARP requests from clients
- Implements a NAT implementation that allows two different subscribers (MAC addresses in the prototype) to use the *same* IP 5-tuple but be treated as a different NAT session (ie: if the IP src, port, dst, port and protocol were identical)
- Allows hair-pinning of NAT traffic (peer-to-peer via the NAT)
- Does not use ARP for resolving clients IP address, instead using the DHCP lease table for link-layer resolution (thus no ARP down to customers) – this allows customers to have duplicate IP addresses on the same link-layer
- The LAN segment has peer-to-peer disabled. This is to prevent hosts “seeing” their IP address used by another PC (split-horizon).

Subscriber Aware NAT Prototype (Sub-001)



DHCP Discover
chaddr: 00-12-00-12-12-12

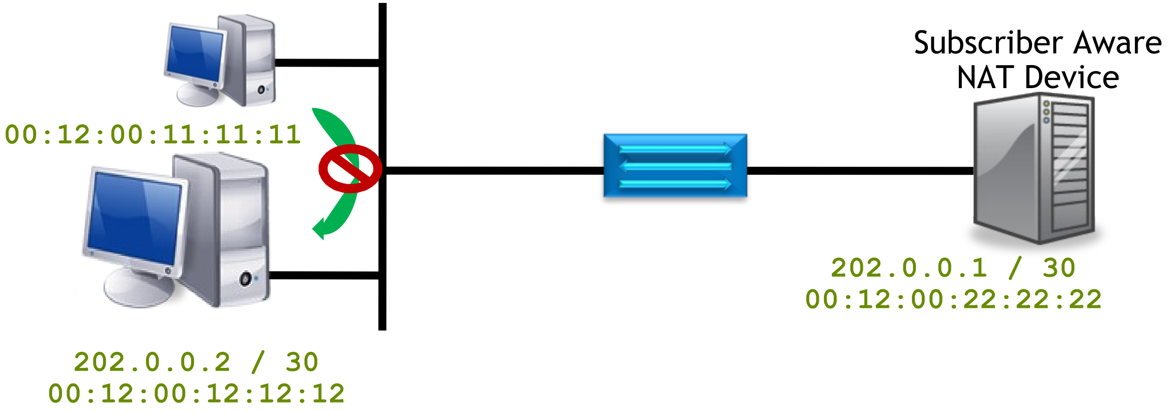
DHCP Offer
yiaddr: 202.0.0.2
subnet mask: 255.255.255.252
router: 202.0.0.1

ARP Request
who is 202.0.0.1?

ARP Response
202.0.0.1 is 00-12-00-22-22-22

TCP Syn:
mac dst: 00-12-00-22-22-22
mac src: 00-12-00-12-12-12
source : 202.0.0.2 : 7762
ip dst: 1.2.3.4 : 80

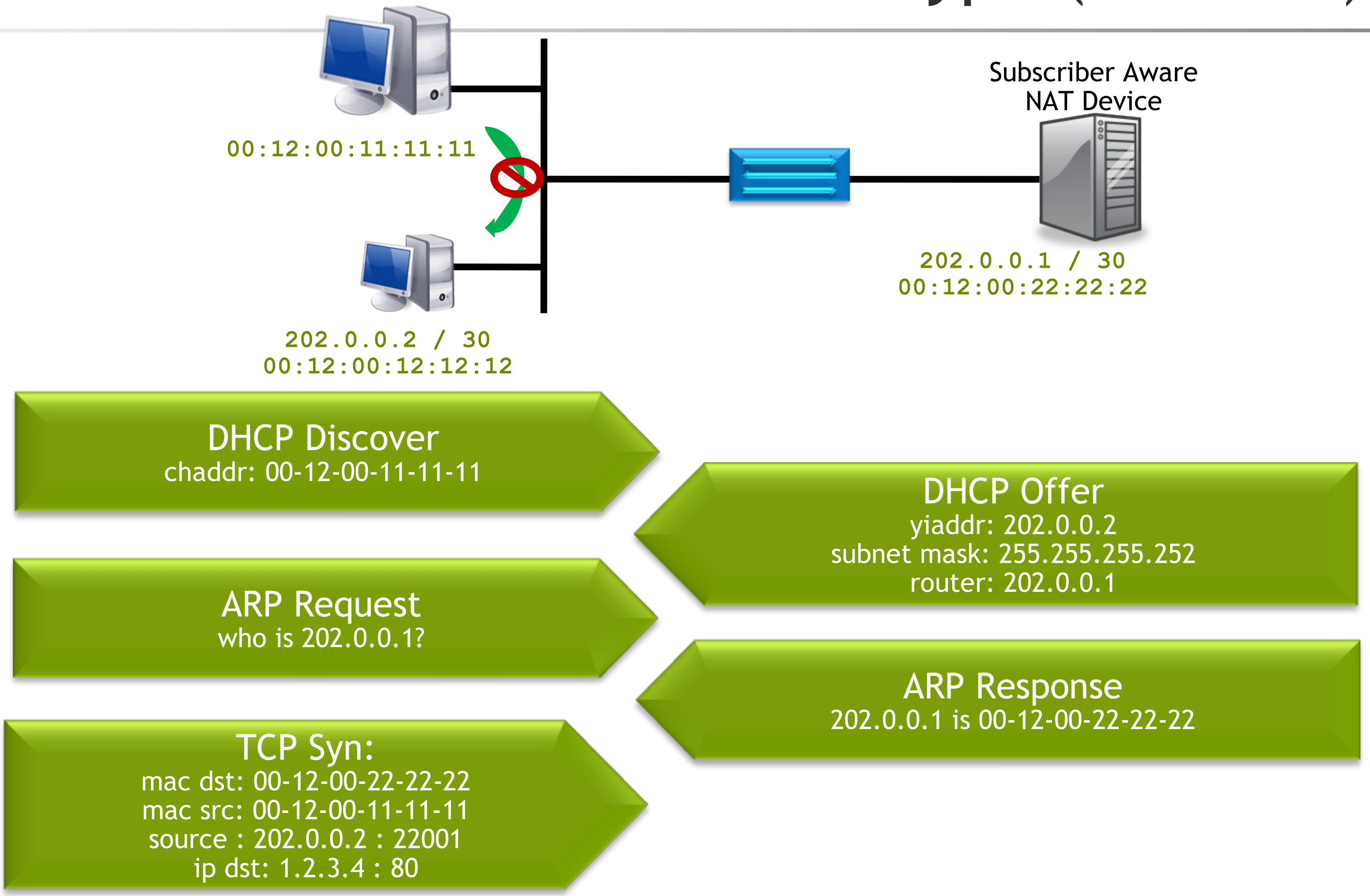
Subscriber Aware NAT Prototype (Sub-001)



Subscriber	Inside		Outside		Destination			State
	IP	Port	IP	Port	IP	Port	Protocol	
Sub-001	202.0.0.2	7762	9.9.9.9	12001	1.2.3.4	80	TCP	SYN SENT

Subscriber	MAC Address	Lease Time
Sub-001	00-12-00-12-12-12	3589s

Subscriber Aware NAT Prototype (Sub-002)



DHCP Discover
chaddr: 00-12-00-11-11-11

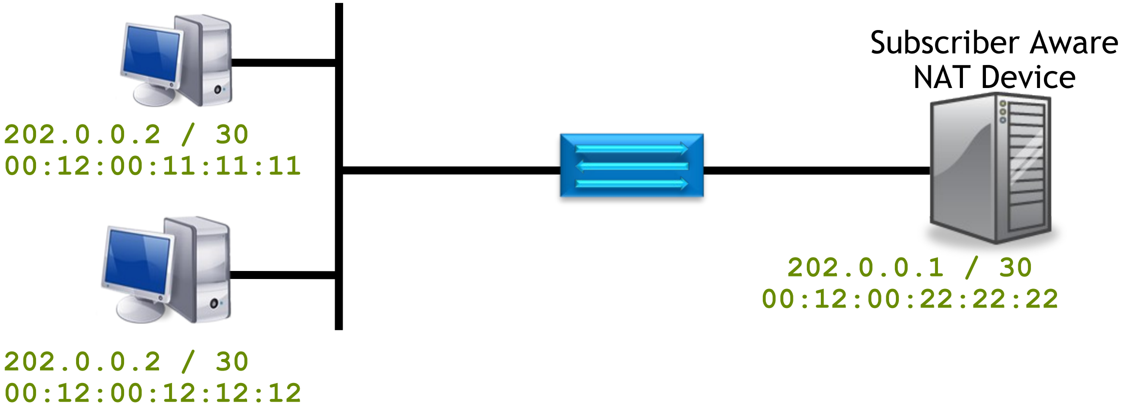
DHCP Offer
yiaddr: 202.0.0.2
subnet mask: 255.255.255.252
router: 202.0.0.1

ARP Request
who is 202.0.0.1?

ARP Response
202.0.0.1 is 00-12-00-22-22-22

TCP Syn:
mac dst: 00-12-00-22-22-22
mac src: 00-12-00-11-11-11
source : 202.0.0.2 : 22001
ip dst: 1.2.3.4 : 80

Subscriber Aware NAT Prototype



Subscriber Aware
NAT Device

202.0.0.2 / 30
00:12:00:11:11:11

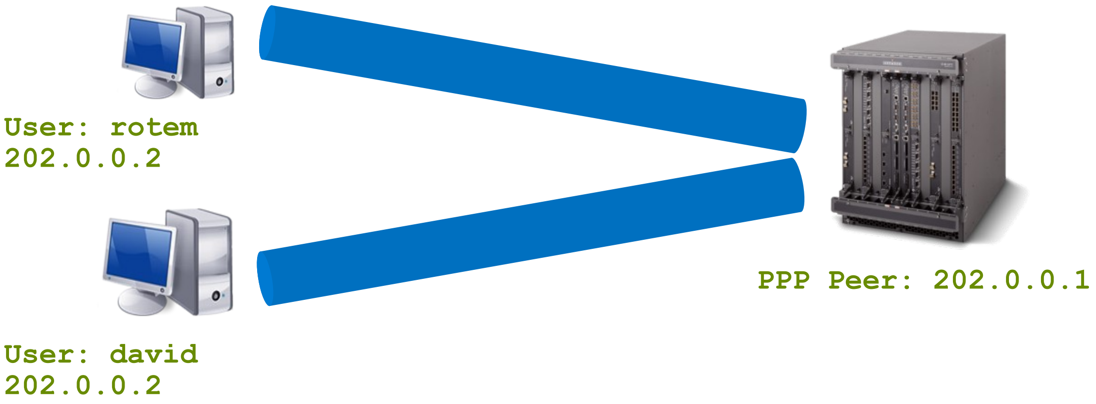
202.0.0.2 / 30
00:12:00:12:12:12

202.0.0.1 / 30
00:12:00:22:22:22

Subscriber	Inside		Outside		Destination			State
	IP	Port	IP	Port	IP	Port	Protocol	
Sub-001	202.0.0.2	5000	9.9.9.9	12001	1.2.3.4	80	TCP	ESTABLISHED
Sub-002	202.0.0.2	5000	9.9.9.9	12002	1.2.3.4	80	TCP	SYN SENT

Subscriber	MAC Address	Lease Time
Sub-001	00-12-00-12-12-12	3213s
Sub-002	00-12-00-11-11-11	3530s

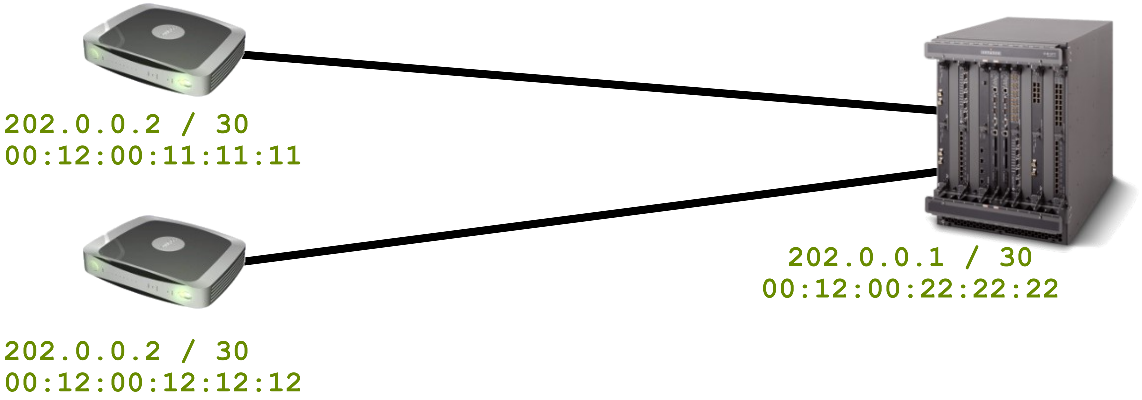
Extend to PPP/PPPoE for Hosts



Subscriber	Inside		Outside		Destination			State
	IP	Port	IP	Port	IP	Port	Protocol	
Rotem	202.0.0.2	7762	9.9.9.9	12001	1.2.3.4	80	TCP	ESTABLISHED
David	202.0.0.2	22001	9.9.9.9	12002	1.2.3.4	80	TCP	SYN SENT

Subscriber	PPP Session	PPP State
Rotem	2/0/2.123-773138	ESTABLISHED
David	2/0/2.124-183941	ESTABLISHED

TR-101 for Routers



Subscriber	Inside		Outside		Destination			State
	IP	Port	IP	Port	IP	Port	Protocol	
Rotem	202.0.0.2	7762	9.9.9.9	12001	1.2.3.4	80	TCP	ESTABLISHED
David	202.0.0.2	22001	9.9.9.9	12002	1.2.3.4	80	TCP	SYN SENT

Subscriber	SAP	MAC Address	Lease Time
Rotem	2/0/2.123	00-12-00-11-11-11	738s
David	2/0/2.124	00-12-00-12-12-12	313s

Thank You