Aalto University
School of Electrical Engineering

Customer Edge Switching & Realm Gateway Tutorial Session – Day 1

Jesus Llorente Santos
jesus.llorente.santos@aalto.fi

www.re2ee.org

August 20th, 2015

Outline

- Current Internet Model
  - User Location
  - Use of Domain Name System (DNS)
- Issues with Current Internet Model - NATs
- CES to CES communications
- Establishing CES connections
- Application Layer Gateway (ALG)
- Additional Material
  - Introduction to Testbed, System Architecture, OpenFlow…

Current Internet Model

- Internet goes mobile
  - Massive growth of connected users and devices
  - Expect an exponential growth with the arrival of IoT

- Dominant presence of Network Address Translator (NAT)
  - Driven by the IPv4 address exhaustion
  - Allow multiple hosts to connect to the Internet with the same public IP address
  - Separation of private and public networks
    - Reuse same private networks over and over (~18M IPs)
    - 10.0.0.0/8, 172.16.0.0/12, 192.168.0.0/16
  - Requires binding state of IPs and ports when packets traverse the NAT: public-to-private and private-to-public
  - Acts as a first layer of security blocking inbound connections

Current Internet Model

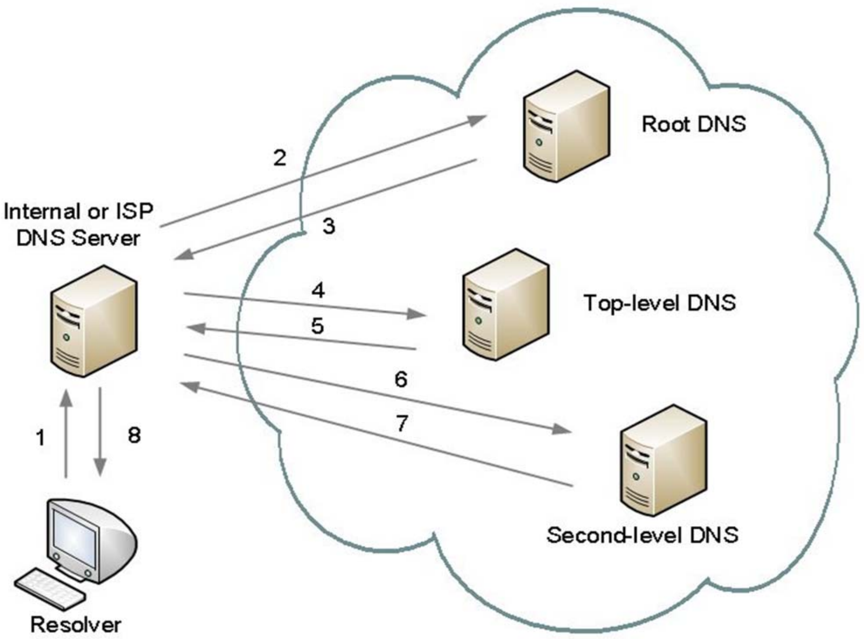
- Location of communicating nodes
  - Users typically located in private networks behind NATs
    - Reduce the amount of public IP addresses needed
    - Need to be able to initiate connections towards public servers
    - Example: computers, laptops, smartphones, etc.
  - Public servers and/or services must be publicly reachable
    - Directly reachable at IP layer via routing
    - Reachable via a proxy or frontend
    - Need to serve requests from connecting users
    - Example: Mail, SSH, HTTP(S), etc.

Current Internet Model

- Identification of hosts and services
  - By IP address
    - Valid on public networks may cause problems across private networks
    - Binds together host identity and routing locator
    - Not always easy to remember: *130.233.224.254*
  - By name
    - Typically following a hierarchical naming scheme, i.e. Fully Qualified Domain Name (FQDN) in DNS
    - Decouples host identity from routing locator
    - Easier to remember: *comnet.aalto.fi*

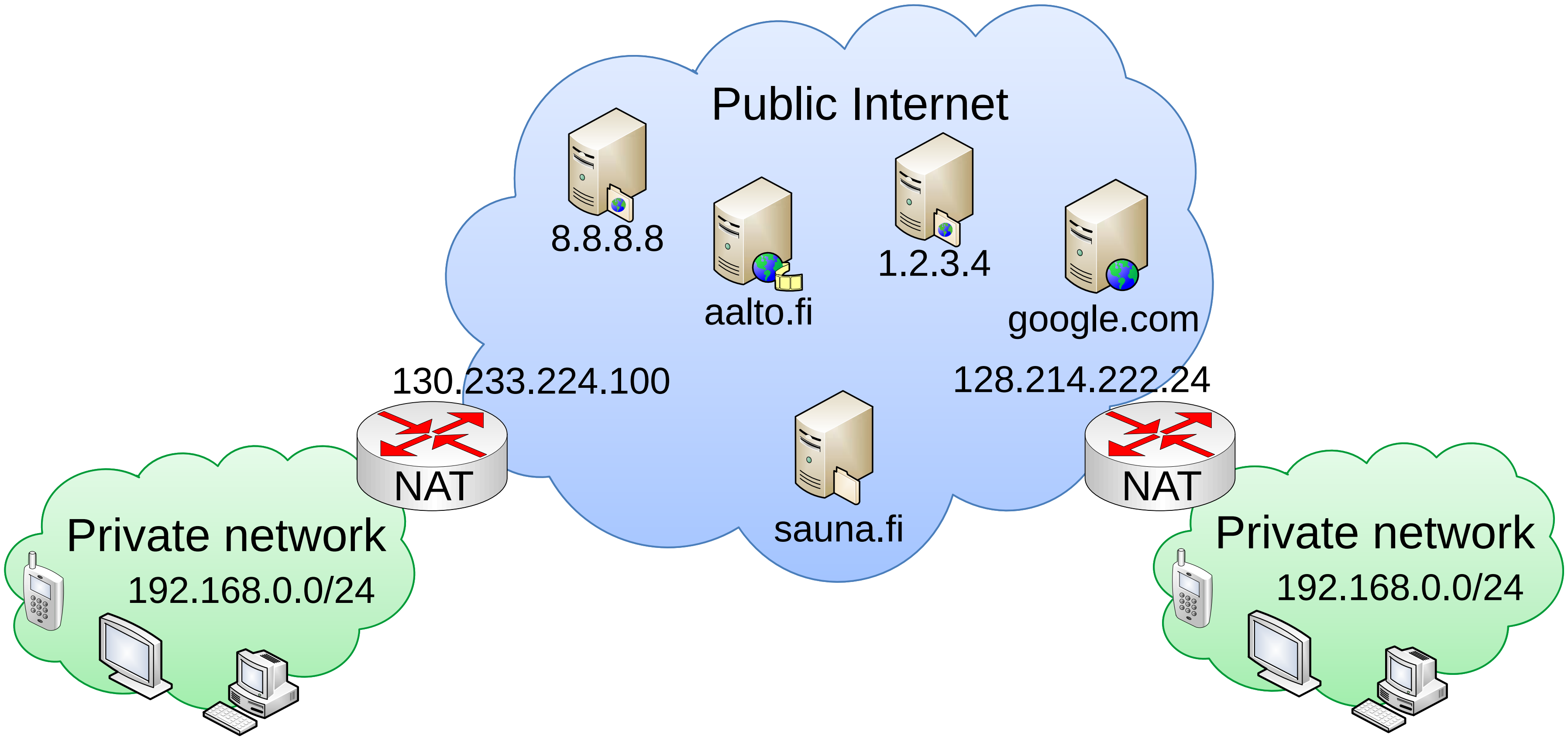
Current Internet Model

- Domain Name System – DNS
  - Resolves FQDN names to IP addresses (most typical function)
  - Transaction based Query/Response
  - Client-Server architecture



Current Internet Model

- Internet Architecture

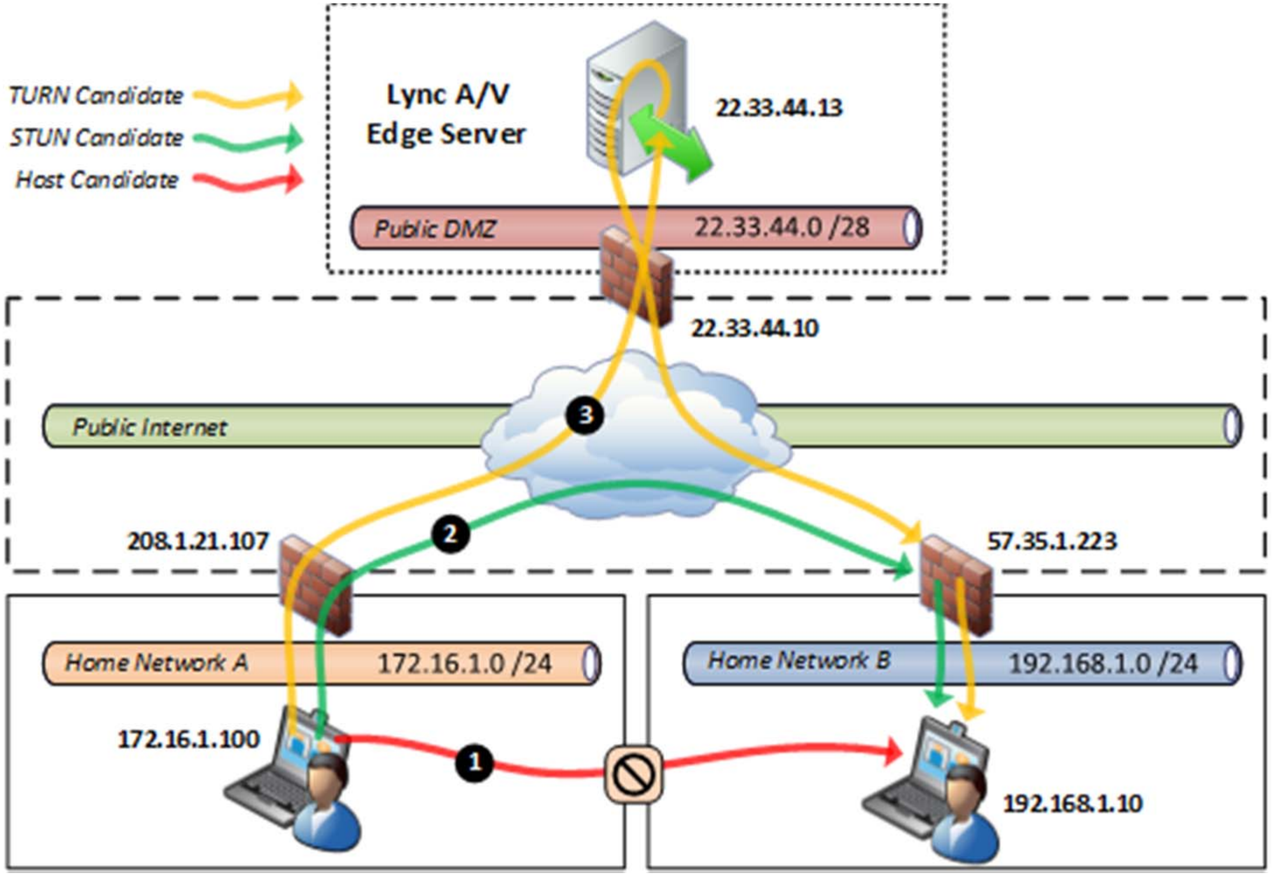


Issues with the current Internet Model

- NAT introduces reachability problem
  - Block inbound connections from reaching the private network
  - NAT-unfriendly protocols are negatively affected by NATs
    - Use of IP address literals or separate control/data connections
    - Require specific Application Layer Gateways e.g. SIP, FTP
  - Traversal of the NAT requires additional protocols
    - STUN/TURN/ICE
    - Results in increased delays during connection setup
    - Requires specific application code and increases network traffic

Issues with the current Internet Model

- More on STUN / TURN / ICE



Issues with the current Internet Model

- Unwanted traffic: Any source can send a packet to any destination address

- Possibility of source address spoofing makes it difficult to attribute evidence of misbehavior to the legitimate source

CES Communications

- CES replaces the existing NAT node of the network
- CES provides name resolution and gateway functionality
- Addressing of the private network is not modified
  - Hosts remain connected with their private addressing
- Does not require changes in either hosts or protocols
- Host identification is always based on names FQDN
  - IP addresses are not used for identification due to their private nature and because they can be repeated across networks

CES Communications

- Provides policy based communications
  - Connection establishment is determined by a set of requirements
  - Reduces unwanted traffic in destination
  - Contributes to mitigate DDoS attacks

- Overcomes the reachability problem of NATs
  - Enables global communications using private IP addresses
  - ALGs are still required for specific protocols that exchange IP address literals as part of the signaling, e.g. SIP, FTP, etc.

- Tunnels end-to-end user data packets across CES edge nodes over any connected network

CES Connection Establishment

There are 3 phases to establishing CES connections

1. Discovery of CES endpoint
   - Triggered by name resolution of a remote host – DNS query
   - Availability of CES service encoded in DNS NAPTR records
   - b.ces. 30 IN NAPTR 10 6 "U" "CETP+cesid“ "!^(.*)$!cesid:1=cesb.ces.?ip=192.0.2.10?alias=IXP!" .
     - Service: CETP+cesid
     - CES Identifier: cesid:1=cesb.ces
     - Endpoint: 192.0.2.10
     - Alias network: IXP

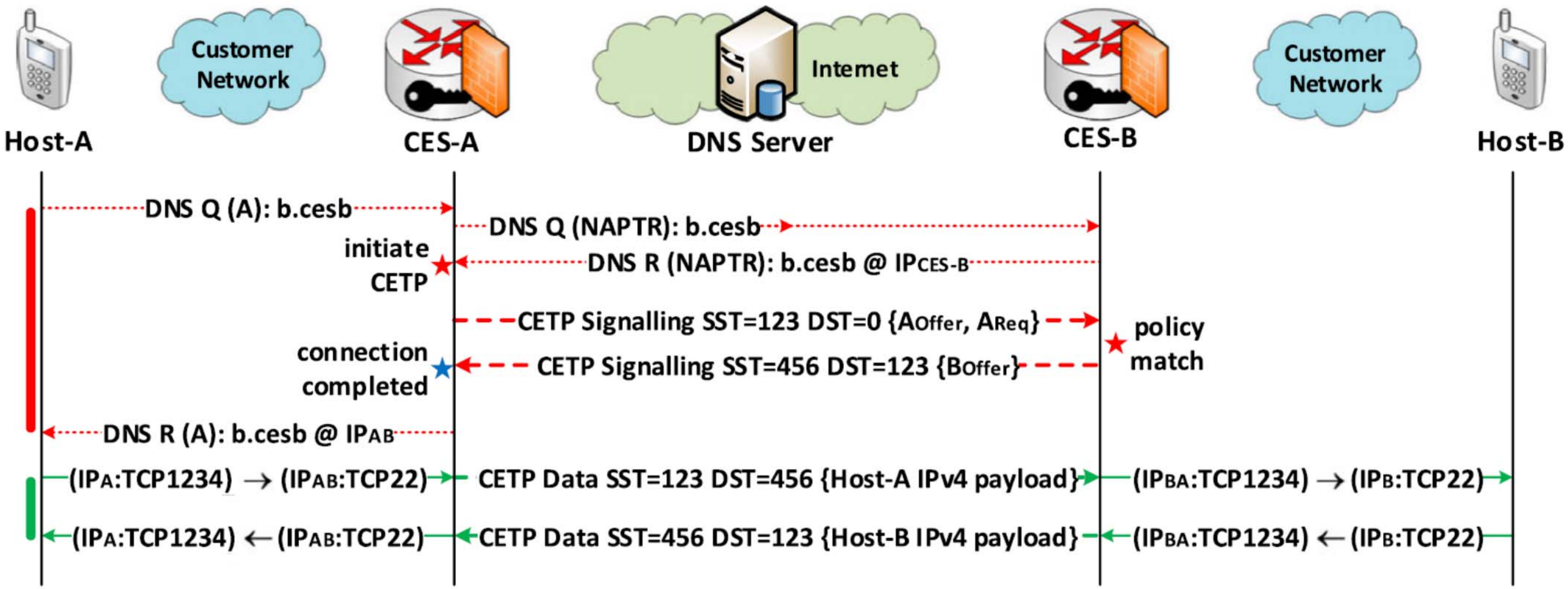
CES Connection Establishment

2. Policy negotiation followed by CES discovery
   - Typically 1 to 3 rounds of signaling exchange
   - Minimizes computation on the inbound CES
   - Mutual exchange between CES nodes of host policy requirements
     - Success: Allocation of IP proxy addresses for end-to-end data forwarding
     - Failure: Notification via DNS response with error code NXDOMAIN
   - Allocation of session tags for connection identification
     - Source Session Tag / Destination Session Tag
     - Currently using 32-bit tags for experimentation
   - First connection suffers additional delay during policy negotiation
   - Following connections have virtually zero delay due to DNS cache

CES Connection Establishment

3. Data forwarding after successful policy negotiation
   - Stateful binding on each CES
     - CES session tags
     - CES routing locators, e.g. Ethernet, IPv4, IPv6, etc.
     - Hosts IDs
     - Hosts FQDNs (useful for PTR reverse queries)
     - Host local IP and allocated proxy IP address
   - CES to CES encapsulated user data with address translation at the edges similar to layer 3 VPN service end to end
     - Proxy IP is allocated from a private pool, e.g. 10.0.0.0/8
     - Proxy IP is a just a local representation of the remote host
     - Proxy IP is meaningless outside the scope of the CES connection

CES Connection Establishment



IP_A: 192.168.0.100
P_{AB}: 10.0.0.1

CES_A: 198.18.0.11/24
CES_B: 198.18.1.12/24

IP_B: 192.168.0.100
P_{BA}: 172.16.0.1

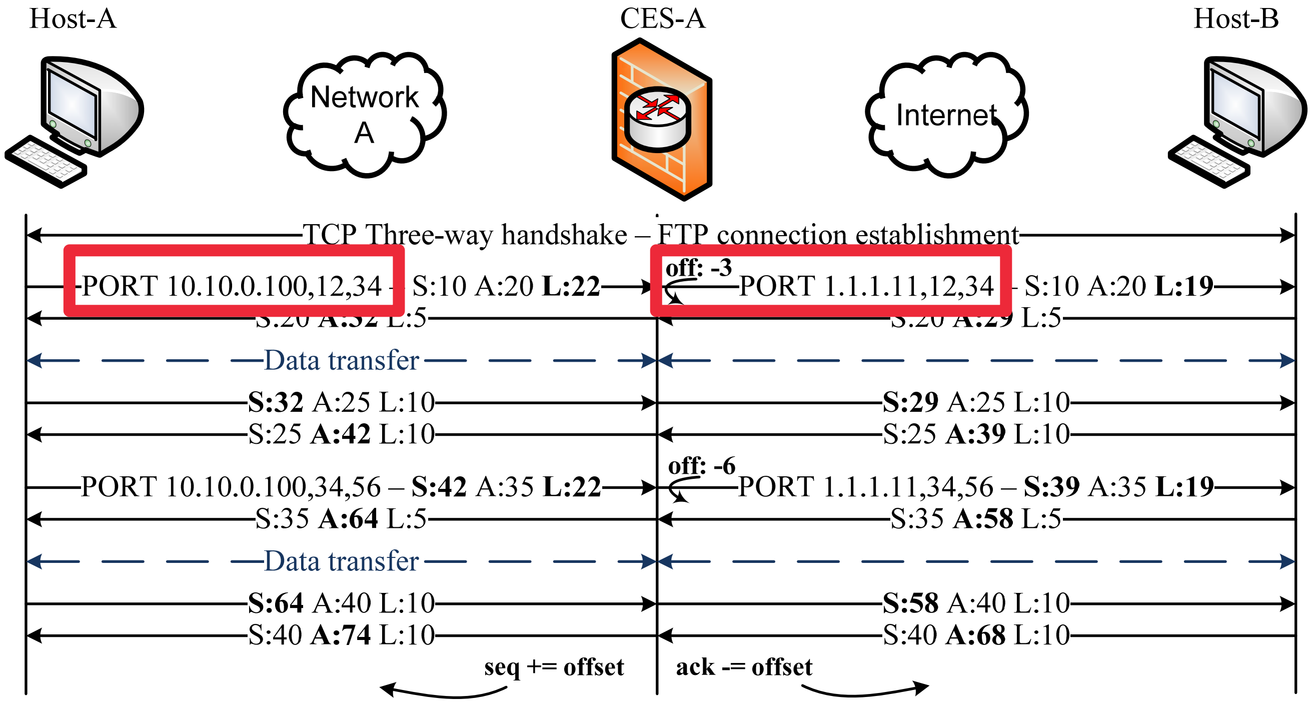
CES Application Layer Gateway ALG

Application Layer Gateways (ALG) developed for the following protocols

- ICMP and ICMP error packets
  - Address transformation at edges
- UDP based SIP – Session Initiation Protocol
  - Replacement of IP address literals by FQDN
- TCP based FTP – File Transfer Protocol
  - Replacement of IP address literals by FQDN
  - Introduces an offset in subsequent TCP segments (SEQ, ACK)
- TCP based RTSP - Real Time Streaming Protocol
  - Replacement of IP address literals by FQDN
  - Introduces an offset in subsequent TCP segments (SEQ, ACK)

CES Application Layer Gateway ALG

FTP Case – Stateful ALG with TCP header rewrite



$$\text{Offset} = \text{Length}_{new} - \text{Length}_{original} + \Delta\text{Offset}$$
$$\text{ACK}_{new} = \text{ACK}_{current} - \text{Offset}$$
$$\text{SEQ}_{new} = \text{SEQ}_{current} + \text{Offset}$$

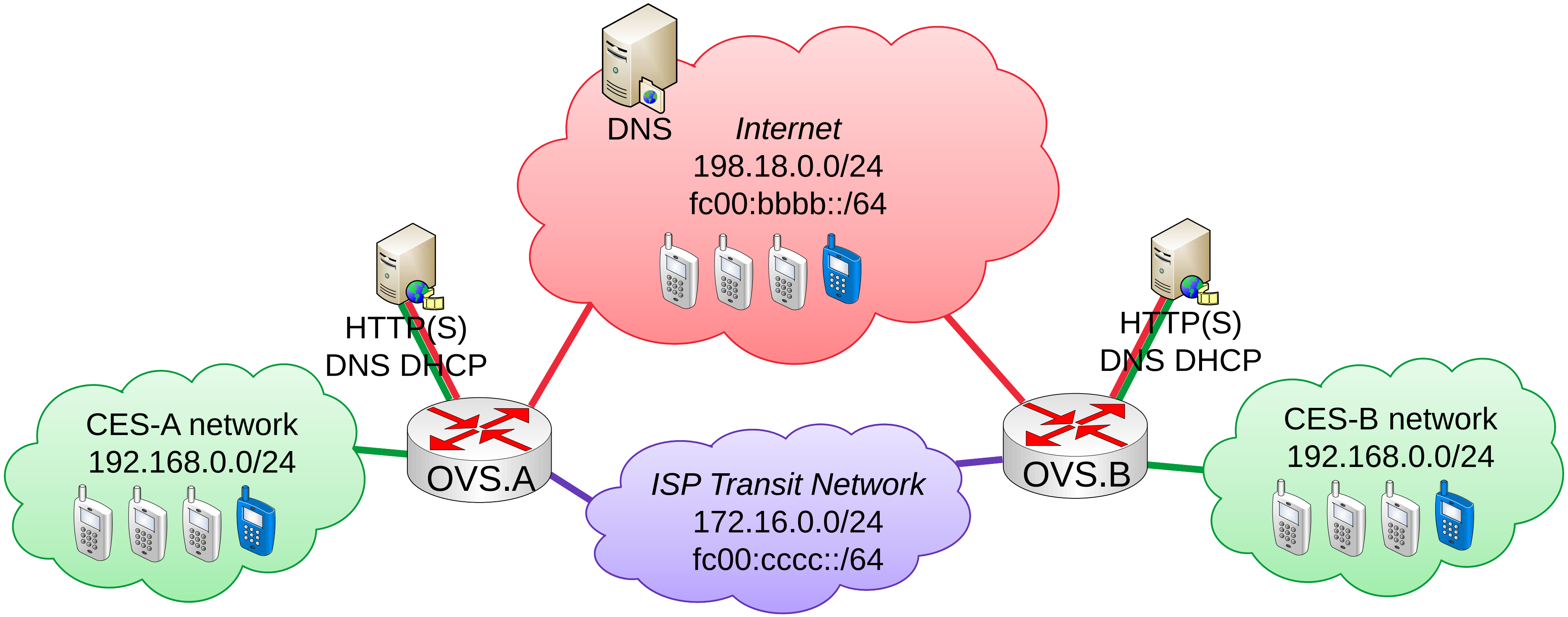
Extra 1: Development Architecture

Current testbed relies on Proxmox VE 3.4

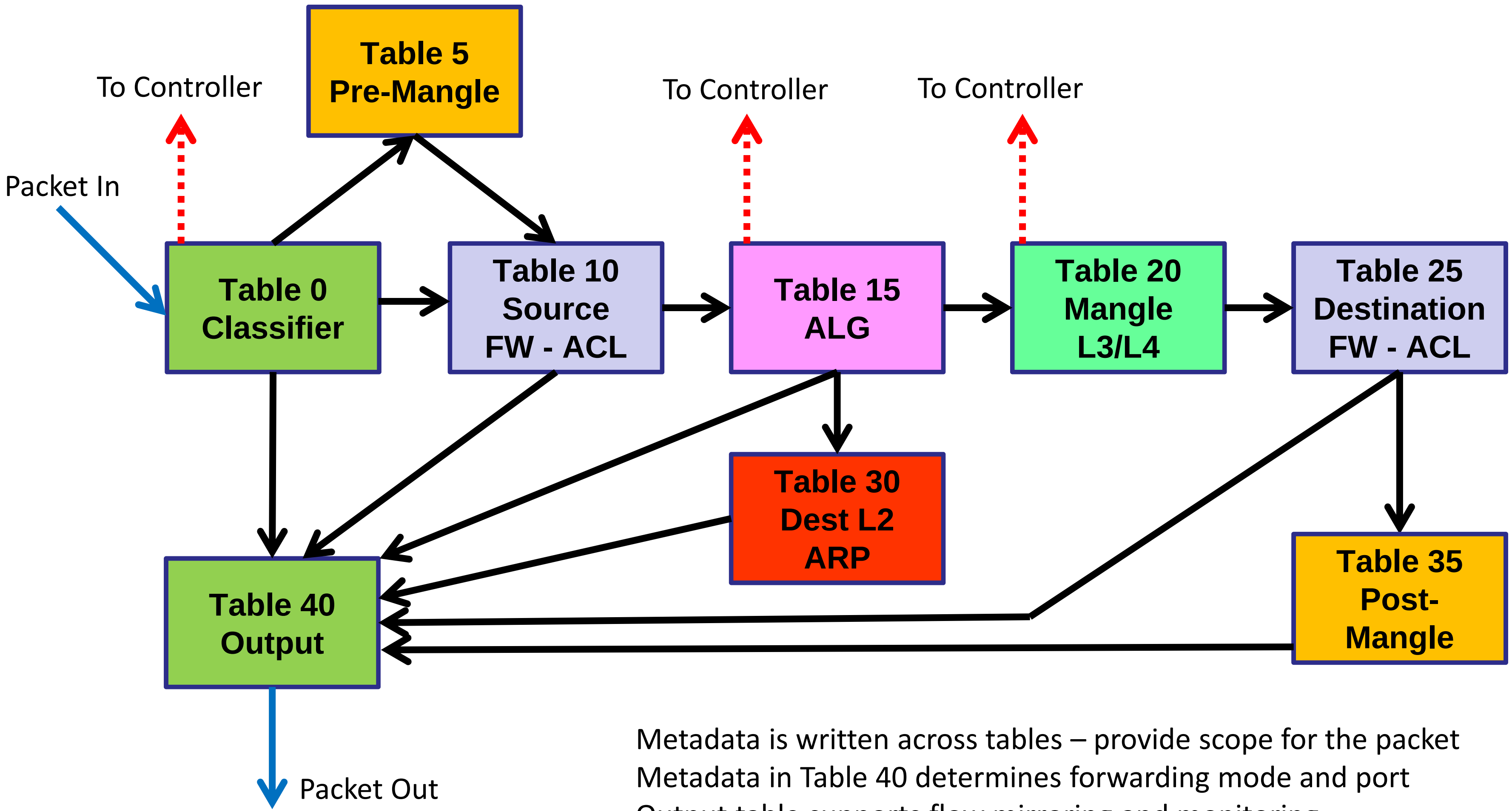
- Supports both KVM and containers with OpenVZ
- Containers are more lightweight compared to full-blown VM
- Available at http://proxmox.com/en/proxmox-ve

- Our whole testbed sits on a single VM running Proxmox
  - All hosts and nodes are virtualized with containers
  - Includes kernel support for OpenvSwitch
  - Networking scenario is made of:
    - Linux bridges
    - OpenvSwitch bridges
    - Virtual Ethernet pairs

Extra 1: Development Architecture



Extra 3: OpenFlow Tables



Metadata is written across tables – provide scope for the packet
Metadata in Table 40 determines forwarding mode and port
Output table supports flow mirroring and monitoring

Thank you!

Q & A ?