2010-7-25

Security Level:

Scalable Address Resolution for Data Center and Cloud Computing

Problem Statements

Linda Dunbar (ldunbar@huawei.com)
Sue Hares (shares@huawei.com)

www.huawei.com

HUAWEI TECHNOLOGIES CO., LTD.

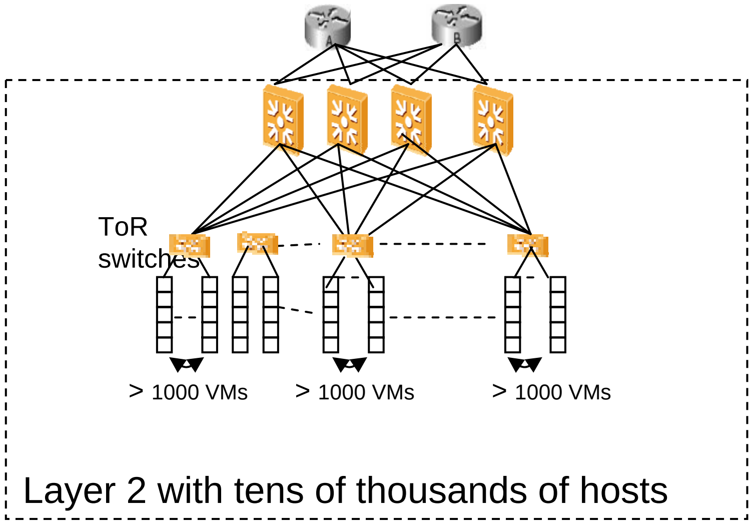
HUAWEI

Goal

- **Give a brief introduction of problem statements to be presented to the ARP222 bar BOF (http://trac.tools.ietf.org/bof/trac/wiki/BarBofsIETF78Arp222)**
- **ARP222:**
  - Address Resolution Protocol for Layer 2 to Anything to Layer 2
- **Explain how ARP222 is relevant to L2VPN and TRILL**

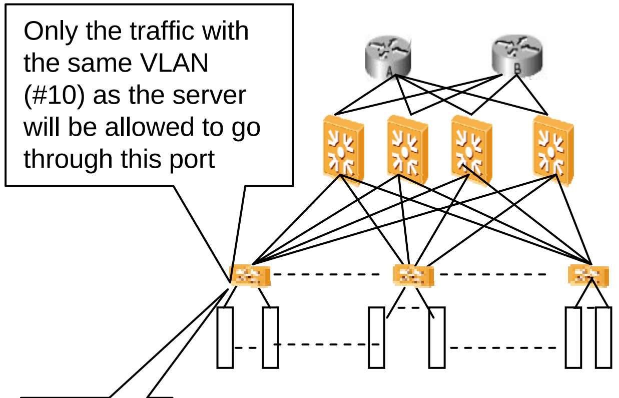
The amount of ARP requests in a data center can be too much burden

- **As Virtual Machines being added to a Data center, the number of hosts within one data center can easily go beyond 20~30K.**
  - Handling up to 1000~2000 ARP requests per second is almost the high limit for any hosts or ARP server. With more than 20K hosts in one Layer 2 domain, the amount of ARP broadcast messages, plus other broadcast messages such as DHCP, can create too much burden to be handled by hosts (or dedicated ARP server).
  - Servers/VM can send a lot of ARP. For Microsoft Windows (versions XP and server 2003), the default ARP cache policy is to discard entries that have not been used in at least two minutes, and for cache entries that are in use, to retransmit an ARP request every 10 minutes



Why VLAN (or smaller subnet) alone is not enough

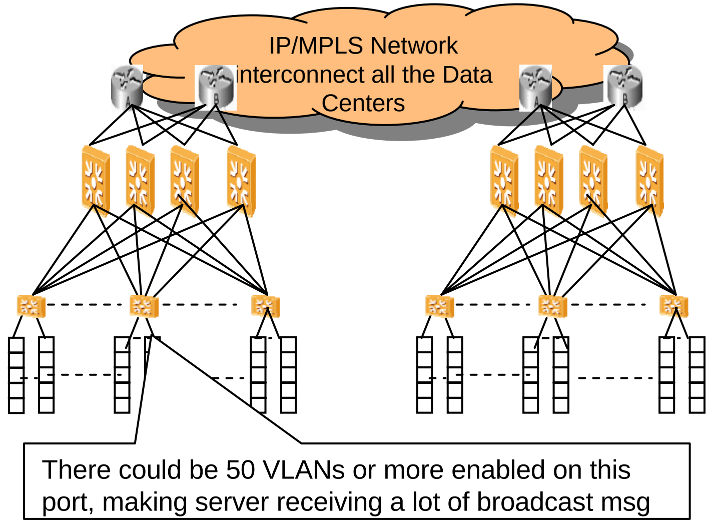
- **Subnet (VLAN) can partition one Layer 2 network into many virtual Layer 2. All the broadcast messages are confined within one subnet (VLAN).**
- **Subnet (VLAN) has worked well when each server serving one single host. The server will not receive broadcast messages from hosts in other subnets (VLANs).**
- **When one physical server is supporting >100 Virtual Machines, i.e. >100 hosts, most likely the virtual hosts on one server are on different subnets (VLANs). If there are 50 subnets (VLANs) enabled on the switch port to the server, the server has to handle all the ARP broadcast messages on all 50 subnets (VLANs). The amount of ARP to be processed by each server is still too much.**
- **When virtual hosts are added or deleted from a server, the switch port to the server may end up enabling more VLANs than the number of subnets actually active on the server, causing more ARP to be sent to the server**
- **For Cloud Computing Service (which is explained in later slides), the number of virtual hosts and virtual subnets can be very high. It might not be possible to limit the number of virtual hosts in each subnet.**



Only the traffic with the same VLAN (#10) as the server will be allowed to go through this port

VLAN #10
Host A

Without VM, each server only sees traffic on one VLAN



IP/MPLS Network interconnect all the Data Centers

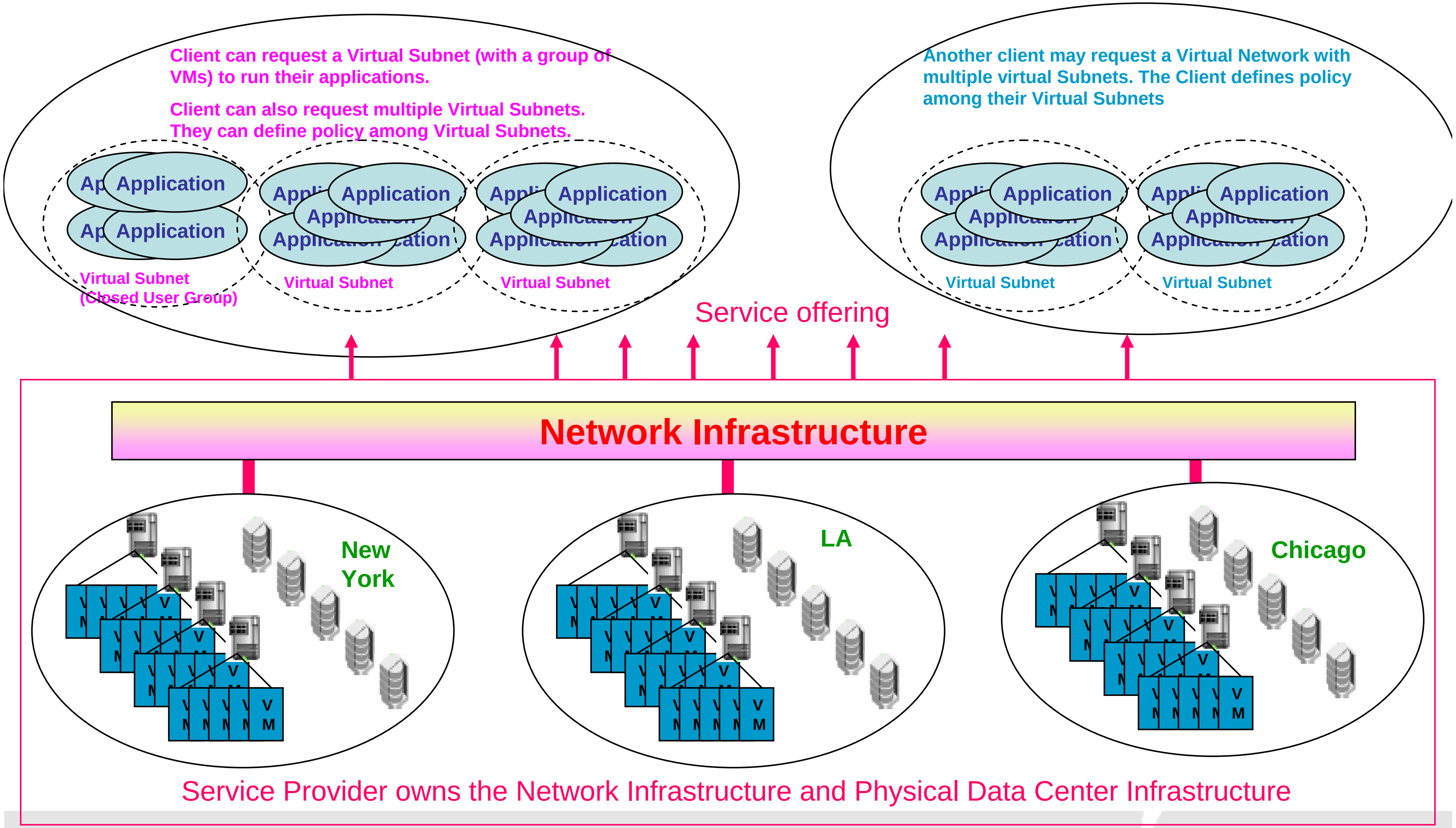
There could be 50 VLANs or more enabled on this port, making server receiving a lot of broadcast msg

Current Status of Bridge’s FDB size:

- Typical bridges support in the range of 16 to 32K MAC Addresses, with some supporting 64K;
- With external memory (TCAM), bridges can support up to 512K to 1M MAC addresses;
- Failure to support sufficient MAC addresses results in increased flooding / poor performance;
- Operators may need to upgrade to higher capacity (ie., more FDB entries) bridges *OR*
- If high-capacity bridges are already in use (e.g., in the core), performance may be compromised by excessive flooding.

Infrastructure As A Service

- Service Model



Client can request a Virtual Subnet (with a group of VMs) to run their applications.

Client can also request multiple Virtual Subnets. They can define policy among Virtual Subnets.

Another client may request a Virtual Network with multiple virtual Subnets. The Client defines policy among their Virtual Subnets

Service Provider owns the Network Infrastructure and Physical Data Center Infrastructure

Virtual Subnet

- **Cloud Computing service needs Network to provide Virtual Subnets and Virtual hosts**
  - Virtual hosts within one virtual subnet can span across different sites due to customer requirement or resource allocation
  - Some virtual subnets have to be connected by private networks (layer 2 or/and layer 3)

Identity to physical IP/MAC address resolution

- Virtual hosts come and go. But each service provider's IP/MAC address space is no unlimited.
- Therefore, it is important to have a scalable Address Resolution protocol to map virtual host's identity to physical IP/MAC addresses.

Virtual Host’s identity

- **Virtual host identity has be associated with its Virtual Network**
- **Similar to IP to Ethernet MAC ARP (IPv4’s RFC826 and IPv6’s RFC4861), Cloud Computing service needs “Virtual Host Identity” to IP/MAC address resolution**

Proposal to IETF

- **Create a new IETF working group (http://trac.tools.ietf.org/bof/trac/wiki/BarBofsIETF78Arp222)**
  - To develop scalable address resolution protocols (IPv4 & IPv6) for Cloud Computing Service with large amount of virtual subnets & virtual hosts, and for data center with large amount of virtual hosts, including
    - Proper identity for virtual subnets and virtual hosts
    - Scalable address/location resolution: identity-to-Address (IP/MAC-VLAN) mapping for customer VMs
  - To develop network solutions to allow virtual hosts see other virtual hosts as if they are on one subnet, but the network interconnect them can be anything (Layer 2 or/and Layer 3).
    - Address isolation (creating and managing silos)
    - Small forwarding tables for Layer 2 switches
  - To develop solutions to scope the broadcast messages, including ARP and DHCP, so that broadcast storm are confined to smaller silos.
  - To develop solutions of handling of multicast messages among virtual hosts in one Virtual Subnet which spans across multiple locations.

How ARP222 is related to TRILL and L2VPN?

- **L2VPN uses MPLS/VPLS to form VPNs to interconnect Ethernet networks.**
  - The provider edge device of L2VPN will have ports facing Ethernet links which potentially may see a lot of virtual hosts and virtual subnets.
- **TRILL targets at campus network, which will have lots of hosts with Ethernet interfaces when servers are virtualized.**
  - Edge RBridge will have some ports facing end stations (virtual hosts) with Ethernet links.
- **The goals of ARP222 include**
  - reducing broadcast messages traversing across Layer 2 silos, which in turn reduces the amount broadcast messages through TRILL ports or L2VPN network ports.
  - reducing the amount of VM addresses to be learnt by L2VPN edge devices and RBridge edge switches.