Scaling the Address Resolution Protocol for Large Data Centers (SARP)
draft-nachum-sarp

Youval Nachum
Linda Dunbar
Ilan Yerushalmi
Tal Mizrahi

Marvell
Huawei
Marvell
Marvell

IETF Meeting 85, November 2012
History of this Draft

- March 2012 – draft 00.
- Discussion in ARMD mailing list.
- July 2012 – IETF 84 – presented in INTAREA WG.
  - Main feedback: need to equally address IPv4 and IPv6.
- October 2012 – draft 03.
Background - Multi-site Datacenter

Inter-site Connectivity

Datacenter Site

Edge Device

Core

EoR

ToR

Server

Server

Server

Server

Edge Device

Datacenter Site

Inter-site Connectivity

Edge Device

Core

EoR

ToR

Server

Server

Server

Server

Edge Device

Core

EoR

ToR

Server

Server

Server

Server

Datacenter Site

Inter-site Connectivity

Datacenter Site

Inter-site Connectivity

Datacenter Site

Inter-site Connectivity
Background

- Challenges in datacenter network scaling:
  - Large MAC address tables.
  - ARP/ND broadcasts.
  - VM migration.
Background – Proxy ARP

- Proxy ARP (RFC 1027, RFC 1009, RFC 925).
- Proxy ARP responds based on IP subnet.
  - Assumption: IP subnet implies location.
Edge devices: proxy SARP.
IP subnet does not imply location.
MAC-W / MAC-E imply location.
SARP Cache

1. ARP/ND: IP-D
West Site

2. Reply: MAC-E

Inter-site Connectivity

SARP Proxy
MAC-E

SARP Proxy
MAC-E

IP-D
MAC-D

East Site

IP-S
MAC-S

West Site
SARP – Data Plane

1. IP-S $\rightarrow$ IP-D, MAC-S $\rightarrow$ MAC-E
2. IP-S $\rightarrow$ IP-D, MAC-W $\rightarrow$ MAC-E
3. IP-S $\rightarrow$ IP-D, MAC-W $\rightarrow$ MAC-D
MAC address table of bridges in the west site:
- Local site addresses, e.g., MAC-S.
- Edge devices, e.g., MAC-E.
- No need for addresses of remote sites.
Local SARP cache limits broadcast domain for known IP addresses.
SARP over Overlay Network

SARP is agnostic to the transport technology, e.g. L2VPN.
• **IPv4**: Gratuitous ARP is used to notify network about migration.
• **IPv6**: unsolicited neighbor advertisement is used.

• No need for additional control protocols.
• Transparent to inter-site network and protocols.
Next Steps

- Receive feedbacks from WG.
- WG adoption.
Thanks