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

Youval Nachum  Marvell
Linda Dunbar  Huawei
Ilan Yerushalmi  Marvell
Tal Mizrahi  Marvell

IETF Meeting 84, July 2012
Background - Multi-site Datacenter
Background

- **Challenges in datacenter network scaling:**
  - Large MAC address tables.
  - ARP broadcasts.
  - VM migration.
Proxy ARP (RFC 1027, RFC 1009, RFC 925).

Proxy ARP responds based on IP subnet.
- Assumption: IP subnet implies location.
SARP

- Edge devices: proxy SARP.
- IP subnet does not imply location.
- MAC-W / MAC-E imply location.
SARP Cache

ARP: IP-D

Reply: MAC-E

West Site

East Site

IP-S MAC-S

IP-D MAC-D

SARP Proxy
MAC-E

SARP Proxy

Inter-site Connectivity
SARP – Data Plane

1. IP-S → IP-D, MAC-S → MAC-E

IP-S
MAC-S

West
Site

SARP
Proxy
MAC-W

2. IP-S → IP-D, MAC-W → MAC-E

SARP
Proxy
MAC-E

Inter-site
Connectivity

3. IP-S → IP-D, MAC-W → MAC-D

IP-D
MAC-D

East
Site
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.
SARP – ARP Broadcast Domains

Local SARP cache limits broadcast domain for known IP addresses.
SARP is agnostic to the transport technology, e.g. L2VPN.
SARP with VM Migration

- Gratuitous ARP is used to notify network about migration.
- No need for additional control protocols.
- Transparent to inter-site network and protocols.
Next Steps

History:
- March 2012 – draft 00.
- June 2012 – draft 02.
- Discussion in ARMD mailing list.

Next steps:
- Receive feedbacks from WG.
- WG adoption.
Thanks