Problem statement on address resolution in VM migration

draft-liyz-armd-vm-migration-ps-01.doc

Yizhou Li (liyizhou@huawei.com)
Dimensions to be considered

- Network Topology
  - Layer 2 domain restricts VM movement
  - Where is the default gateway (router)?

- Protocols used at layer 2
  - STP based
  - Routing-like protocol, TRILL/SPB
  - Support different scale of network in nature, i.e. #of VMs
ARP difficulty in VM migration (1)

- No ARP “leave” message
  - Magnitude of seconds to complete the migration
  - Impact: black hole at transition time, depending on aging. ARP request storm as no response made at transition time
  - How to forget the address?
    - Keepalive?
    - De-registration?
    - Through management plane? vCenter knows vm is migrating.

- Variety of ARP message type after VM migration
  - ARP request/reply, reverse ARP. Experiments show different implementations on updating ARP table entry on receipt of ARP family message (next slide)
  - Impact: need to guarantee the correct updating on ARP table at gateways
Examples for ARP entry updating in VM migration

<table>
<thead>
<tr>
<th>#</th>
<th>packet sent after VM migration</th>
<th>Is VM's interface updated to ift2 on GW?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>std gratuitous ARP</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>broadcast ARP reply</td>
<td>N</td>
</tr>
<tr>
<td>3</td>
<td>RARP</td>
<td>N</td>
</tr>
<tr>
<td>4</td>
<td>ARP request with GW host as target IP</td>
<td>Y</td>
</tr>
<tr>
<td>5</td>
<td>ARP request with other host as target IP</td>
<td>N</td>
</tr>
<tr>
<td>6</td>
<td>unicast ARP reply with GW as destination</td>
<td>Y</td>
</tr>
<tr>
<td>7</td>
<td>unicast ARP reply with other host as destination</td>
<td>N</td>
</tr>
</tbody>
</table>
ARP difficulty in VM migration

- ARP message unreliable delivery
  - In case of congestion, consecutive ARP messages can be lost. Sending 3 ARP messages does not help much. Need to send more?
  - ARP table is not updated by data frame
  - Impact: delayed cache refreshment, worsen black hole issue, data frames are delivered to old location for traffic from gateway

- Duplicate address detection
  - Gratuitous ARP has function of DAD. No way to perform DAD at transition time
  - ND in IPv6 has a better way to differentiate DAD & address adv.
    - DAD: NS with unspecified address as src.
    - Address resolution: NS with unicast address as src
    - Address advertisement: unsolicited NA with all-nodes multicast address as dest

- Impact: address occupation/contention
Security concerns

- Some existing mechanism may not be applicable
  - MAC locking: locking a MAC address to a specific physical port of the switch.
  - DHCP snooping: binding IP/MAC by snooping DHCP ACK to port of switch.
  - ARP is a signal that a migrated VM starts to receive frame at the new location?
Next step?

- Be merged to general problem statement draft