Scaling IPv6 Neighbor Discovery

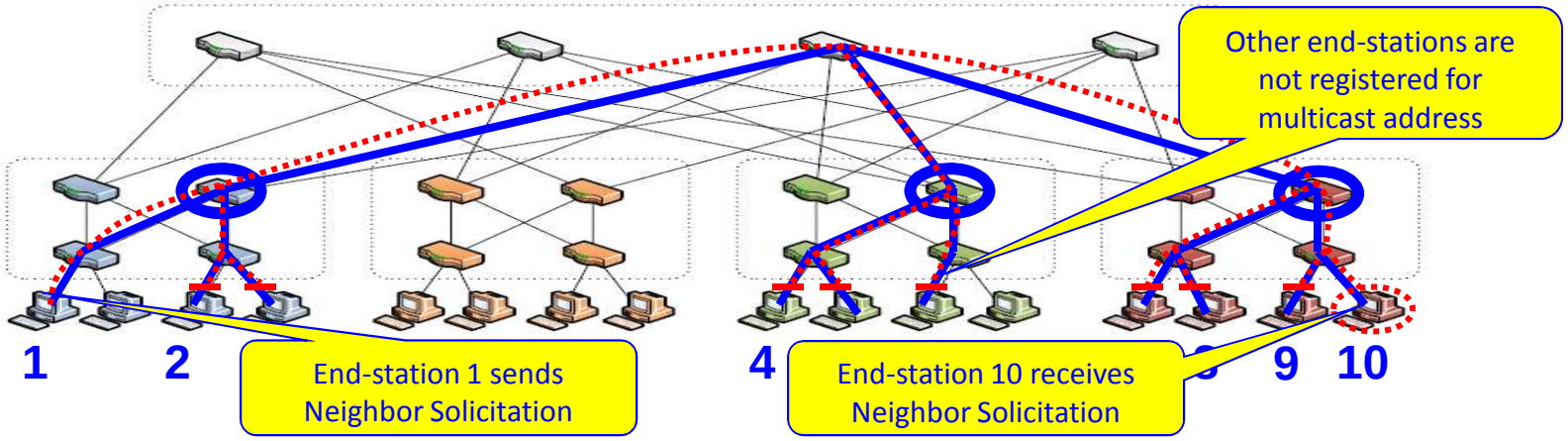
Ben Mack-Crane

(ben.mackcrane@huawei.com)

Overview of Neighbor Discovery Protocol

- IPv6 nodes on the same LAN use Neighbor Discovery (RFC4861) to
  - to find routers and discover link and network parameters,
  - to discover each other's presence,
  - to determine each other's link-layer addresses, and
  - to maintain reachability information about the paths to active neighbors.

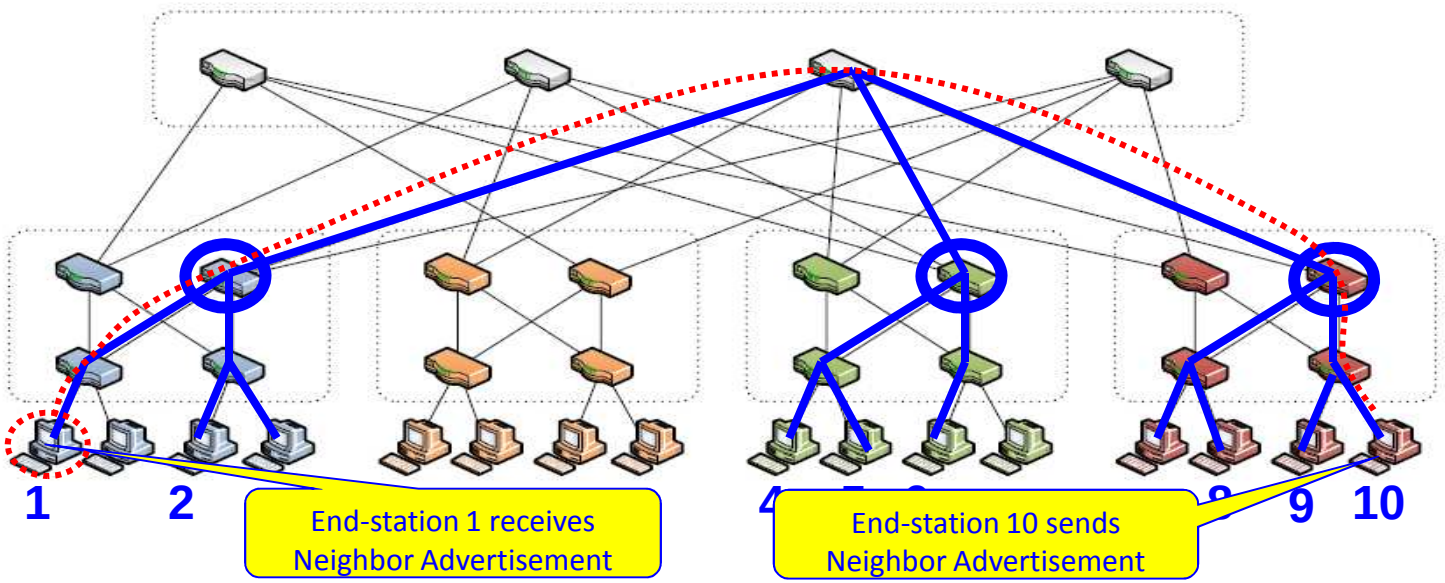
Neighbor Solicitation



End-station 1 wants to resolve the L2 address of end-station 10:

- End-station 1 sends Neighbor Solicitation packet using the *solicited-node multicast address* for end-station 10's IPv6 address;
- The Neighbor Solicitation packet is flooded to all endpoints on the VLAN;
  - When MMRP is not supported, all multicast messages are broadcasted.
- However, only end-station 10 has configured its NIC to receive this multicast address, so no other end-stations must process the Neighbor Solicitation packet;
- Therefore, there shouldn't be significant impact on end-station CPU cycles if Servers are properly designed and no duplicated IPv6 addresses.

Neighbor Advertisement

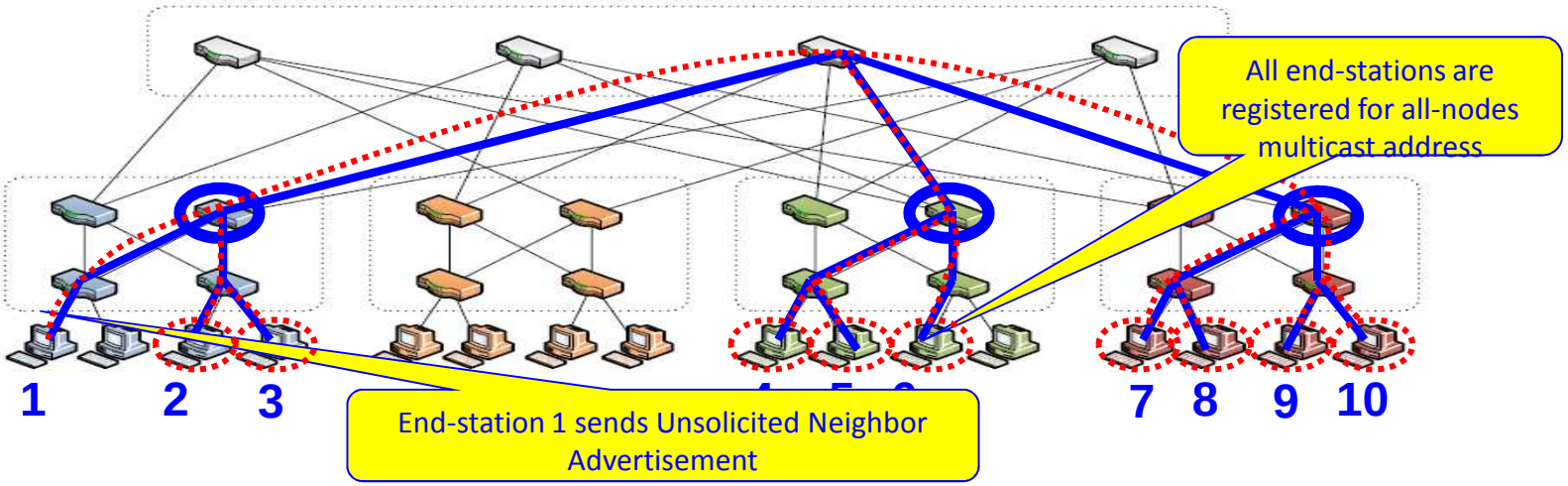


Response to Neighbor Solicitation is unicast:

- End-station 10 sends Neighbor Advertisement packet using end-station 1's unicast address;
- Packet unicast to and processed only by end-station 1.

Differs from ARP in that address resolution does not involve all nodes – only the requesting node and those who register for the solicited-node multicast address.

Unsolicited Neighbor Advertisement



End-station 1 wants to inform all end-stations of a change in L2 address:

- End-station 1 sends an Unsolicited Neighbor Advertisement packet using the *all-nodes multicast address*;
- The Unsolicited Neighbor Advertisement packet is flooded to all endpoints on the VLAN;
- All end-stations in the VLAN process the Unsolicited Neighbor Advertisement;

- Note: this is expected to be a rare event (change of L2 address) and therefore, although all end-stations must process this packet, there would be no significant impact on end-station CPU cycles.

Similar to Gratituous ARP Response

ND Scaling Gap Analysis – Performance

There are three performance scalability concerns:

1) Too many packets are transmitted on links where they are not useful – unnecessary use of bandwidth
2) Too many unnecessary packets are received/processed by nodes – unnecessary node processing
3) Too many packets are transmitted/received/processed to serve a particular purpose (i.e., a more efficient protocol is needed) – inefficient use of bandwidth (new case)

ND Scaling Gap Analysis – Networks

There are a few network scenarios to consider:

1) Edge: A large LAN with a few routers and many 1000's of hosts
2) Core: A large LAN connecting 1000's of routers
3) Network Virtualization: A large number of networks (VLANs) comprising virtual nodes (hosts and routers) and virtual switches (e.g., a number of virtual switches on a single hardware platform)
4) Multi-Site: A large LAN covering multiple, geographically distributed, sites

ND Scaling Gap Analysis – Performance

Neighbor Discovery Messages (basic)

	Who Sends	How Often	DA	Scale	Host Mobility
Router Solicit	hosts	when new (seldom)	all-routers mcast	O(s)	
Router Advert	routers	periodic; when solicited	all-nodes mcast; unicast	O(R)	
Neighbor Solicit	nodes	when no/stale cache entry for Next Hop	solicited-neighbor mcast	O(P)	
Neighbor Advert	nodes	when solicited	unicast	O(P)	
Unsolicited Neighbor Advert	nodes	when L2 address changes (seldom)	all-nodes mcast	O(s)	
Redirect	routers	when needed (Seldom in non-mobile environment, But happens in Cloud Data Center)	unicast	O(s)	

nodes = routers + hosts; R = #routers; H = #hosts; P = #peers/node; s = small number

Impact to hosts is not bad for networks with a few routers and many hosts (each with a few peers). However, the amount of bandwidth consumed by ND depends on where hosts reside.

Problems with IPv6 self addressed hosts

- For user created subnet, the number of hosts in the subnet is up to the user.
  - IPv6 gives user more freedom to create a mega size subnet, potentially millions of virtual hosts.
  - SLAAC: state less address auto configuration & DAD: duplicated address detection
    - When configure IP addresses, use SLAAC and DAD to validate newly configured address.
    - It could blow up the DHCP

ND Scaling Gap Analysis – Networks

1) Scenario 1: Edge: A large Layer 2 network with a few routers and many 1000's of hosts
   - **ND doesn't impose too much burden to servers/hosts.**
   - **However, if each NIC's MAC filter is smaller than the number of MACs' supported, then effectively all the multicast messages will go into servers → ND will impose burden to the server.**
2) Scenario 2: Core: A large LAN connecting 1000's of routers (not big issue in Data Center)
   - **Some messages scale as O(R) or O(P) which will be large in this case**
   - **An alternative to unreachability detection may be preferred here**
3) Scenario 3: Network Virtualization: A large number of networks (VLANs) comprising virtual nodes (hosts and routers) and virtual switches (e.g., a number of virtual switches on a single hardware platform)
   - **This case may share concerns with O(R) scaling since the number of routers/switches is increased by virtualization**
4) Scenario 4: Multi-Site: A large LAN covering multiple, geographically distributed, sites
   - **Neighbor Discovery for Anycast addresses may need to be filtered if it is desired to limit Anycast to a local site**