Can an ARP Packet Storm Impact BFD sessions on the Nexus 7000 Platform

Contents

Introduction

This document describes an impact of ARP Packet Storm on Control Plane protocols such as BFD, OSPF, and others, running on Nexus 7000 switches.

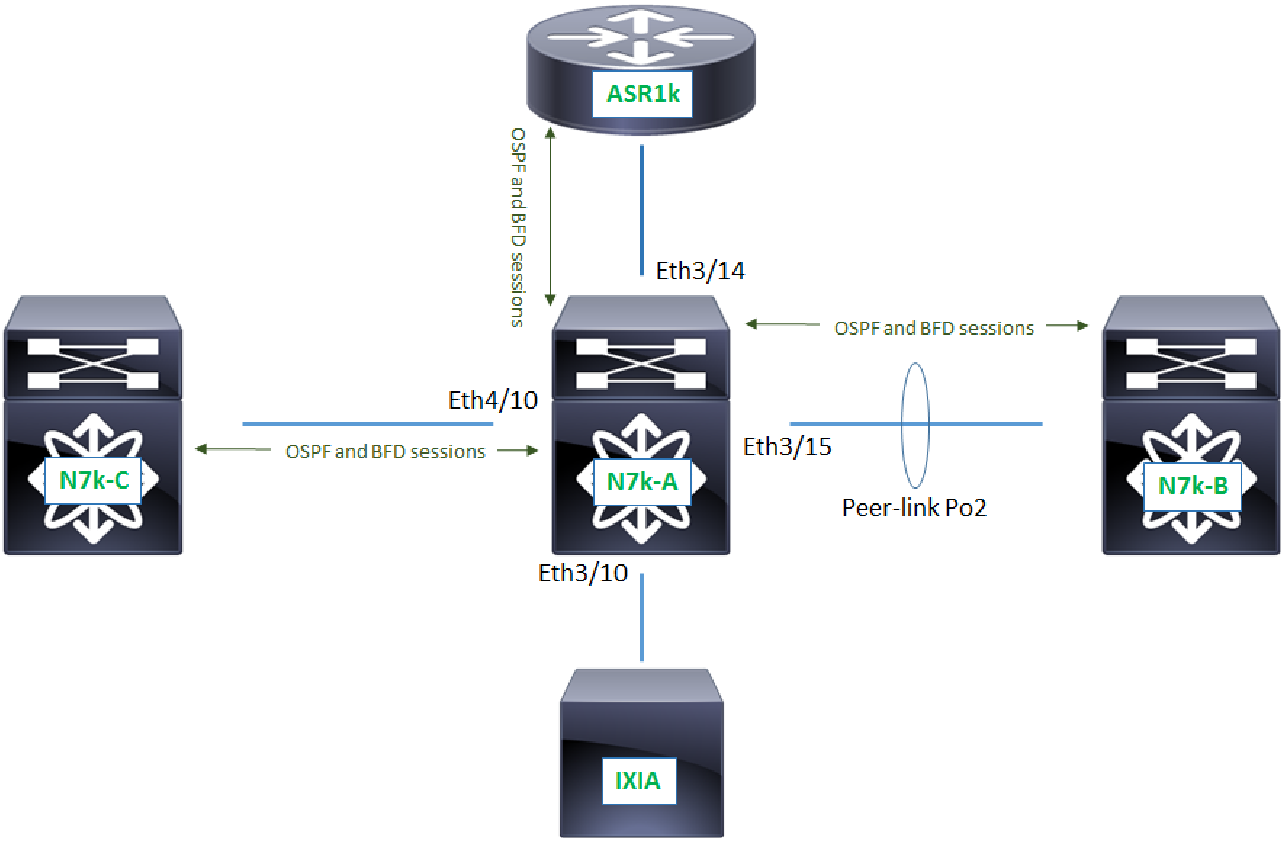
Contributed by Nishad Mohiuddin, Nikolay Kartashev, Cisco TAC Engineers.

Q. Since Cisco NX-OS can distribute BFD operation to compatible modules that support BFD, would an ARP packet storm have any impact on BFD sessions on Nexus 7000 platform?

A. In general, an ARP Packet Storm can have negitive impact on the stability of BFD sessions running on Nexus 7000 switch. Exact symptoms depend on the longetivity and magnitute of the ARP Packet Storm event. Below are test results from Cisco TAC lab network.

Lab setup details

The following lab setup is built to test the impact of amounts of ARP traffic hitting CPU of Nexus 7000 switch.



Here N7k-A is used as Device Under Test (DUT). DUT is a Nexus 7009 switch with the following hardware configuration

```
N7k-A# show module
Mod Ports Module-Type Model Status
--- ----- ----------------------------------- ------------------ ----------
1 0 Supervisor module-1X N7K-SUP1 active *
2 0 Supervisor module-1X N7K-SUP1 ha-standby
3 32 10 Gbps Ethernet Module N7K-M132XP-12 ok
4 32 10 Gbps Ethernet Module N7K-M132XP-12 ok
N7k-A#
```

N7k-A has the following devices connected to it

- N7k-B is a VPC peer, connected to interface Ethernet 3/15
- ASR1k is a Layer-3 neighbor, connected to interface Ethernet 3/14
- N7k-C is a Layer-3 neighbor, connected to interface Ethernet 4/10
- IXIA Traffic Generator is in vlan 6, connected to interface Ethernet 3/10, which is configured as Layer 2 access port

DUT has three BFD sessions, one on linecard in slot 4 towards N7k-C, and two on linecard in slot 3 towards N7k-B and ASR1k

```
N7k-A# show bfd neighbors

OurAddr NeighAddr LD/RD RH/RS Holdown(mult) State Int
10.80.6.173 10.80.6.174 1090519061/4105 Up 4951(3) Up Eth3/14

10.80.1.162 10.80.1.161 1090519054/1090519044 Up 4203(3) Up Eth4/10

10.80.1.61 10.80.1.62 1090519060/1090519059 Up 5921(3) Up Vlan6

N7k-A#
```

DUT also has three OSPF sessions, one on linecard in slot 4 towards N7k-C, and two on linecard in slot 3, towards N7k-B and ASR1k.

```
N7k-A# show ip ospf neighbors
 OSPF Process ID 1
Total number of neighbors: 3
 Neighbor ID Pri State Up Time Address Interface
 10.80.0.2 1 FULL/ - 00:13:26 10.80.1.62 Vlan6
10.80.4.25 1 FULL/DR 00:12:40 10.80.6.174 Eth3/14
10.80.0.3 1 FULL/DR 20:15:07 10.80.1.161 Eth4/10
N7k-A#
```

OSPF is registered with BFD

```
router ospf 1
 bfd
 router-id 10.80.0.1
```

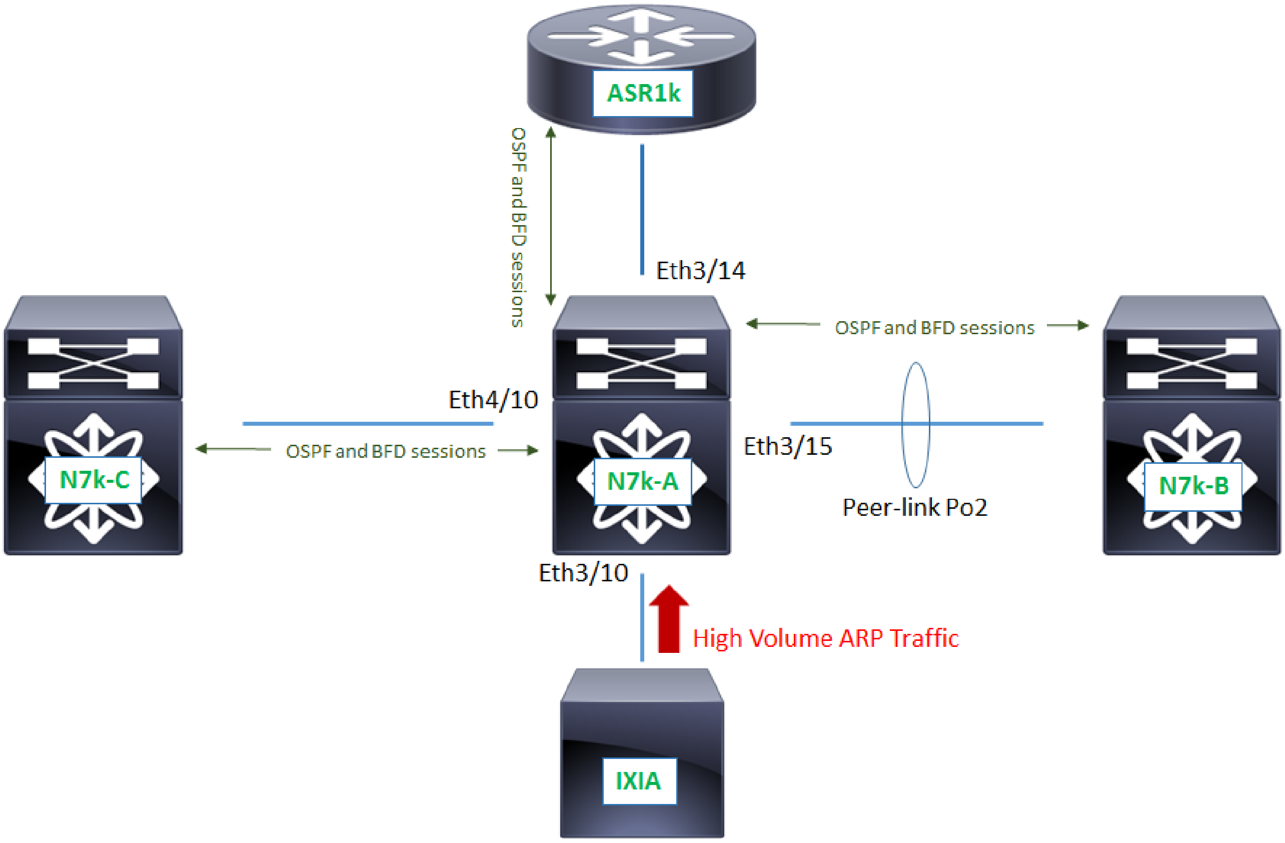
Also, ARP table on N7k-A has entries for all three BFD/OSPF neighbors

```
N7k-A# show ip arp

Address Age MAC Address Interface
10.80.1.62 00:13:30 4055.390f.48c1 Vlan6
10.80.6.174 00:12:46 88f0.774b.0700 Ethernet3/14
10.80.1.161 00:15:13 6c9c.ed44.6841 Ethernet4/10
N7k-A#
```

ARP Storm Begins

IXIA Traffic Generator is used to simulate unstable part of the network, which results in high volume of ARP traffic sent to DUT, as can be seen in the diagram below



The following output shows an increase of input traffic on interface Ethernet 3/10, where the IXIA Traffic Generator is connected. These are broadcast ARP packets received in vlan 6

```
N7k-A# show interface Ethernet3/10 | grep "30 seconds input rate"
 30 seconds input rate 3102999976 bits/sec, 6062053 packets/sec
N7k-A#
```

Since a copy of each broadcast ARP packet is sent to the CPU on N7k-A in this scenario, we see increase of violated bytes on module 3 in CoPP

```
N7k-A# show policy-map interface control-plane class copp-system-p-class-normal
Control Plane

 service-policy input: copp-system-p-policy-strict

 class-map copp-system-p-class-normal (match-any)
 match access-group name copp-system-p-acl-mac-dot1x
 match protocol arp
 set cos 1
 police cir 680 kbps , bc 250 ms
module 3 :
 conformed 2295040 bytes; action: transmit
violated 20569190016 bytes; action: drop

module 4 :
 conformed 128 bytes; action: transmit
violated 0 bytes; action: drop

N7k-A#
```

> **Note**: Note that *there are no violated bytes on module in slot 4*, since the source of broadcast ARP storm is connected to interface on module 3 only

At the point when ARP storm begins, the above outputs are usually the first (and only) signs that indicates an issue on the network. In most cases, these signs go unnoticed or are overlooked by network operators and quickly progress to a situation that leads to major connectivity issues.

ARP Storm Starts impacting Control Plane

By default, the ARP timeout value on the Nexus 7000 platform is configured for 25 minutes or 1500 seconds. The Nexus switch has to periodically refresh the local ARP cache entries in order to keep up-to-date IP-to-MAC resolution of its next hop Layer 3 neighbors.

The following is the output of the ARP cache table on DUT after ARP cache entries expired.

```
N7k-A# show ip arp

Address Age MAC Address Interface
10.80.1.62 00:00:06 INCOMPLETE Vlan6
10.80.6.174 00:00:10 INCOMPLETE Ethernet3/14
10.80.1.161 00:12:59 6c9c.ed44.6841 Ethernet4/10
N7k-A#
```

Notice that ARP cache entries for devices connected to the linecard in slot 3 show **INCOMPLETE** status, whereas the entry for switch N7k-C, which is connected to the linecard in slot 4 is being successfully refreshed as expected.

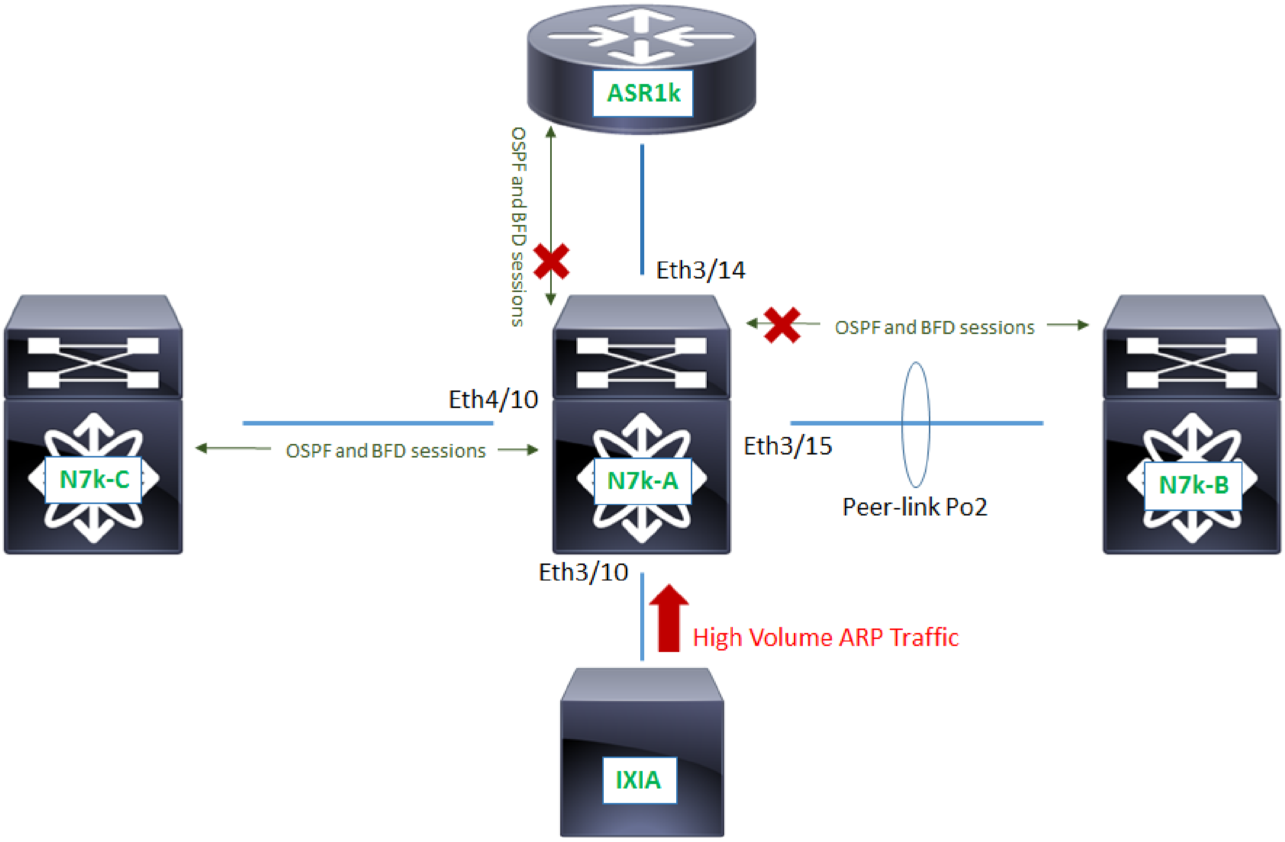
The following DUT log messages indicate the impact on the Control Plane level

```
N7k-A# show logging log
...
2016 Nov 16 22:12:55 N7k-A %BFD-5-SESSION_STATE_DOWN: BFD session 1090519060 to neighbor
10.80.1.62 on interface Vlan6 has gone down. Reason: 0x3.
```

```
2016 Nov 16 22:12:55 N7k-A %OSPF-5-ADJCHANGE: ospf-1 [10600] Nbr 10.80.1.62 on Vlan6 went DOWN
2016 Nov 16 22:12:55 N7k-A %BFD-5-SESSION_REMOVED: BFD session to neighbor 10.80.1.62 on
interface Vlan6 has been removed
2016 Nov 16 22:12:56 N7k-A %OSPF-5-ADJCHANGE: ospf-1 [10600] Nbr 10.80.1.62 on Vlan6 went
EXSTART
2016 Nov 16 22:13:40 N7k-A %OSPF-5-ADJCHANGE: ospf-1 [10600] Nbr 10.80.6.174 on Ethernet3/14
went DOWN
2016 Nov 16 22:13:40 N7k-A %BFD-5-SESSION_STATE_DOWN: BFD session 1090519061 to neighbor
10.80.6.174 on interface Eth3/14 has gone down. Reason: 0x3.
2016 Nov 16 22:13:40 N7k-A %OSPF-5-ADJCHANGE: ospf-1 [10600] Nbr 10.80.6.174 on Ethernet3/14
went EXSTART
2016 Nov 16 22:13:46 N7k-A %BFD-5-SESSION_REMOVED: BFD session to neighbor 10.80.6.174 on
interface Eth3/14 has been removed
2016 Nov 16 22:15:45 N7k-A %OSPF-5-ADJCHANGE: ospf-1 [10600] Nbr 10.80.6.174 on Ethernet3/14
went INIT
...
N7k-A#
```

Notice in this output that OSPF toggles between DOWN to EXSTART state, and then back to INIT state. This occurs because OSPF uses unicast to exchange prefixes during the EXSTART state. Since ARP resolution is incomplete on module in slot 3 at the time of the ARP packet storm, route exchange never completes resulting in the OSPF adjacency to not form.

> **Note**:ARP to IP-to-MAC resolution of next hop relies on unicast as does BFD operation. Given that we can conclude that BFD requires ARP to be resolved for proper operation.



The following outputs confirm the impact of an ARP packet storm on both BFD and OSPF sessions on the module in slot 3. Contrary to this BFD and OSPF session(s) on the module in slot 4 are established and remain stable.

```
N7k-A# show bfd neighbors
```

```
OurAddr NeighAddr LD/RD RH/RS Holdown(mult) State Int
10.80.1.162 10.80.1.161 1090519054/1090519044 Up 5764(3) Up Eth4/10

N7k-A#
```

```
N7k-A# show ip ospf neighbors
OSPF Process ID 1
Total number of neighbors: 3
 Neighbor ID Pri State Up Time Address Interface
 10.80.0.2 1 EXSTART/ - 00:02:54 10.80.1.62 Vlan6
10.80.4.25 1 INIT/DR 00:00:05 10.80.6.174 Eth3/14
10.80.0.3 1 FULL/DR 20:29:28 10.80.1.161 Eth4/10
N7k-A#
```

What happens when an ARP Packet Storm stops?

When an ARP Packet Storm stops, the following recovery occurs automatically and the network begins to converge and enjoys the stable state that it did prior to the ARP broadcast storm.

1. ARP cache entries get resolved on N7k-A
2. BFD sessions on module in slot 3 re-establish
3. OSPF sessions on module in slot 3 re-establish

Conclusion

Even though Cisco NX-OS can distribute BFD operation to compatible modules that support BFD, high volumes of ARP traffic hitting the switch's CPU for a period longer than the time left to refresh local ARP cache entries on Nexus 7000 platform will cause instability in BFD sessions and any client protocols registered with BFD.

This can be attributed to BFD operation which requires ARP resolution of next hop which is unicast. Should the ARP cache entry for the next hop not get refreshed in time, BFD session(s) will fail.