Network Working Group
Internet-Draft
Intended status: Informational
Expires: August 18, 2014

A. Yourtchenko
cisco
L. Colitti
Google

February 14, 2014

Reducing Multicast in IPv6 Neighbor Discovery
draft-yourtchenko-colitti-nd-reduce-multicast-00

Abstract

IPv6 Neighbor Discovery protocol makes wide use of multicast traffic, which makes it not energy efficient for the mobile WiFi hosts. This document describes two classes of possible ways to reduce the multicast traffic within IPv6 ND. First, within the boundaries of existing protocols. Second - with what the authors deem to be "minor changes" to the existing protocols.

Table of Contents

1. Introduction

Wireless networks based on the IEEE 802.11 standard (WiFi) are ubiquitous in today's life. The multicast/broadcast behavior in these networks has significantly lower performance than unicast in the majority of the cases.

Also, in the current standard and implementations of the 802.11 protocols from the link-layer media standpoint the multicast is the same as broadcast.

The Neighbor Discovery protocol makes substantial use of multicast packets on the assumption that they provide the same or better efficiency compared to unicast packets.

This misalignment results that the nodes on IPv6 networks with the default configuration perform significantly poorer both from the battery life standpoint and the bandwidth efficiency standpoint.

This document presents two groups of measures which reduce the shortcoming:

- o The measures which are possible without any changes to the existing standards.
- o The measures which require minimal changes to the standards.

Add some text here. You will need to use these references somewhere within the text: [RFC4862] [RFC4861] [RFC6620] [RFC3315]

2. Impact of Multicast Packets in 802.11 Networks

NOTE: much if not all of the subsequent text in this section might need to be transferred to vyncke-6man-mcast-not-efficient-01, which discusses why multicast is not an efficient media in the WiFi environments.

1. Multicast can impact power consumption on hosts if hosts receive multicast packets that are not addressed to them.
2. Excessive use of multicast can reduce the performance of wireless networks.
3. The extra packets are more expensive when they occur with the host not otherwise engaged in using the network.
4. Mobile nodes often have more than one processor and multiple power management states both for the central processing unit and for the WiFi portion (e.g. using only one antenna out of multiple). Often, the battery impact of rejecting a packet in the radio firmware is substantially lower than the impact of passing the packet to the main processor and rejecting it there.

In 802.11 networks, multicast frames towards clients have a greater battery impact than the unicast frames because they are transmitted to all hosts at once, with the AP setting the DTIM bit on the beacon packet to signal to the dozing hosts that the transmission is about to begin.

Thus, if the host were not to wake up right there and then, it would miss the multicast frame. Unicast packets are buffered on the AP and may have a more lenient delivery schedule, which would allow the devices to not have to wake up at every beacon interval (100ms).

The tradeoff between the energy savings and the latency of the multicast delivery may be manipulated by changing the parameter called DTIM interval, which determines how often (every Nth beacon)

the AP can send the indication about the multicast traffic to the clients - with the default values being fairly low, usually in the range of one to three.

Increasing these values increases the latency for the multicast packets, therefore changing the DTIM interval beyond the defaults is usually not recommended.

3. Quantifying the use of Multicast in Neighbor Discovery

Normal operation of Neighbor Discovery uses the following multicast packets.

1. Duplicate Address Detection.
   Expected impact: One packet per IPv6 address (a host may be configured to do 2 or more) every time a host joins the network

2. Router Solicitations.
   Expected impact: One packet every time a host joins the network.

3. Router Advertisements.
   Expected impact:

   * One multicast RAs every [RA interval] seconds

   * One solicited RA per host joining the network (if solicited RAs are sent using multicast)

4. Neighbor solicitations. Expected impact: One every time a host talks to a new on-link destination talked to. The response is cached and typically does not expire unless the ND cache is under pressure and subject to garbage collection. Cache entries are refreshed (and possibly deleted) using unicast NUD packets, so cache refreshes do not cause multicast packets to be sent..

With the exception of periodic RAs (and possibly solicited RAs), none of these packets are addressed to all nodes. RS packets are addressed to all routers, and NS packets are addressed to solicited-node multicast groups. Because solicited-node multicast groups contain the last 24 bits of the IPv6 address, in most networks, each solicited-node group will have at most one member.

4. Multicast-limiting measures with no changes in specifications

4.1. On-device robust multicast filtering

The hosts may implement on-device multicast filtering, such that if devices receive multicast packets that are not addressed to them, they will not send the packets to the main CPU but instead remain in a lower sleep state.

It is worth noting that this may require a less deep sleep state than the one required to monitor the TIM in the beacon frames. Also, filtering the packets on the device does not address the inefficiency in spectrum utilisation caused by excessive multicast frames.

4.2. Unicast Solicited Router Advertisements

[RFC4861] in section 6.2.6 already allows to do so via a MAY verb (if the solicitation's source address is not the unspecified address). This is further weakened by the subsequent qualifier being "but the usual case is to multicast the response to the all-nodes group." As a result of this, a lot of implementations do multicast the solicited RAs, significantly impacting the devices.

To help address this, all router implementations SHOULD have a way to send solicited RAs unicast in the environments which wish to do so.

4.3. Infrastructure-based multicast filtering

Ensure that solicited-node multicasts only go to the specific nodes. This can be implemented either using multicast snooping or by converting multicast packets to unicast packets that are addressed to a subset of the hosts..

The latter can be done in two ways:

- on the 802.11 level alone, preserving the destination within the inner Ethernet frame as multicast

- on the 802.11 and 802.3 levels, as clarified by the [RFC6085]

Some networks track individual device IP addresses for security and tracking reasons, typically by snooping DAD packets or device traffic as described in [RFC6620]

In these networks, the infrastructure is already aware of which IP addresses are mapped to which MAC addresses, and can use this information to selectively unicast neighbor solicitations to the nodes that will be interested in them.

Most wireless networks are infrastructure-based. The 802.11 standard defines that all communications in such networks will happen via the access points. Therefore, the infrastructure has a chance to intelligently filter any multicast packets that are coming from both local (served by the same access point) and remote (located behind the wired infrastructure) hosts or routers, before forwarding them onto the air to their ultimate destination.

4.4. Proxy the Neighbor Discovery protocol on the access point

802.11 standard defines also that all of packets sent from the client to the Access Point (either for the local over-the-air delivery or for forwarding on to the wired side) are acknowledged (even the multicast ones).

With this in mind, in the scenarios like DAD, a proxy ND implementation has inherently a much better chance of working than the "regular" forwarding of the multicast DAD NS (and the return forwarding of the multicast DAD NA in case of DAD collision that was detected).

Therefore, the environments which want to increase the robustness of the DAD, may wish to proxy the ND on behalf of the clients, therefore reducing the overall client-directed multicast traffic (which is unacknowledged) and increasing the robustness against the poor radio conditions.

4.5. Maximized Interval for Periodic RAs

Assuming the solicited RAs are sent unicast, increasing the interval of the periodic RAs is a natural way of further reducing the amount of multicast packets in the air.

The bounding factor is AdvDefaultLifetime, which is limited by the [RFC4861], section 6.1 on the sending side to 9000 seconds.

Thus, to find the "right" value one will have to balance the robustness in the face of higher packet loss on the segment with the energy consumption by the endpoints. Some real-world mid-scale networks (on the order of 10000 hosts within a single /64) successfully used a value of one RA in 1800 seconds.

However, it is impossible to specify the "best" value - everything will depend on the quality of the local WiFi installation and the radio conditions, with the constraint of 9000 seconds currently specified by the standard.

4.6. Increasing the advertised Reachable value

The NUD with the default settings and active traffic will enter the PROBE state as frequently as every ~30 seconds. [RFC4861] section 7.3.3 defines: "If no response is received after waiting RetransTimer milliseconds after sending the MAX_UNICAST_SOLICIT solicitations, retransmissions cease and the entry SHOULD be deleted. Subsequent traffic to that neighbor will recreate the entry and perform address resolution again."

Short-term connectivity issues at link layer may cause a trigger for the symptoms described in the [RFC7048], therefore triggering the nodes to send multicast neighbor solicitations. However, most of the hosts do not implement at this time the changes suggested there. With the default short timeouts and a wireless environment which forwards multicasts without the filtering, these retransmissions may contribute to further possible failures of NUD in other hosts. In the extreme high density and mobility environments (conferences, stadiums) this may result in avalanche effect and significantly increase the portion of multicast traffic.

Furthermore, an 802.11 segment usually has a single gateway (possibly in a FHRP redundant configuration), therefore making NUD not very useful at all: if that gateway does not function, there is no alternative.

For these kinds of environments it may be useful to significantly increase the REACHABLE_TIME from 30000 milliseconds to 600000 seconds and higher. One possible concern here, however, may be the overflow of the ND table on the gateway, so, again, there is no "best" value suitable for all the networks.

4.7. Clearing the on-link bit in the advertized prefixes

The mobile nodes have generally fairly limited memory, so in the environments where there are thousands of nodes on a single /64, it might be burdensome for them to manage a large neghbor table. Having a lot of hosts with large neighbor tables may mean also a lot of NUD maintenance activity, with the potential for the catastrophic failure of the NUD therefore increasing in the high-density environments.

Clearing the on-link bit in the advertised prefixes causes the hosts to send all the traffic to each other via the default gateway - thus dramatically reducing the size of the neighbor table and the burden of its maintenance on the hosts.

The remaining impact of the link-local addresses still present in the cache can then be mitigated by blocking the direct communications

between the hosts at L2, which is a standard feature in the wireless LAN equipment. This operation effectively turns a wireless LAN segment into a collection of point-to-point links between the hosts and the access point, not dissimilar to the operation of private VLANs in the wired LAN case - making the subnet effectively NBMA.

4.8. Explicit creation of state with DHCPv6 address assignment

Turning the WLAN subnet into an NBMA has a consequence that the DAD may no longer work - which may create a problem with the global addresses. Therefore, it may be necessary to transfer the control over the address assignment to a centralized entity.

Also, the 802.11 protocols operate in the unlicensed bands, which means that the radio conditions may vary greatly. The 802.11 LLC protocol itself does have a fairly robust L2 retransmission mechanism for the acknowledged packets (up to 64 retransmissions). However, there still may be times when the radio conditions are so poor that this robustness is not enough. If the network were to use the snooping to maintain the strict policies (e.g. restrict the source addresses of the traffic), merely snooping the ND may not work, and the data-driven recovery mechanisms might be unacceptable.

In these cases one may consider using DHCPv6 as an address assignment mechanism, which would provide the explicit management of state by the client, and the retransmissions required to create the necessary state on the network side without requiring the node to send the data.

4.9. Client link shutdown within the router lifetime expiry

Some nodes after a longer period of time may decide to completely shut down the radio. This will of course result in the best battery usage, but will incur a tradeoff that waking up the client from the network side will be impossible. However, this mode of operation is the only one not using DHCPv6 which may allow complete avoidance of multicast RA packets: if the client never stays awake for longer than the router lifetime, it will not require the multicast RA processing. This optimization is here for completeness of the discussion - since it changes the connectivity of the client.

5. Multicast-limiting measures with small changes in specifications

5.1. Remove the send-side limit on AdvDefaultLifetime of 9000 Seconds

[RFC4861], section 6.1 limits the AdvDefaultLifetime on the sending side to 9000 seconds, while explicitly requiring the receiving side

to process all the values up to 65535 (maximum allowed by 16-bit unsigned integer that the AdvDefaultLifetime is).

This artificial limit means a hard limit on the maximum router lifetime that can be specified in the configuration. (The authors tried two router implementations: Cisco IOS and radvd. More information welcome).

This artificial restriction prevents from using very long router advertisement intervals that would otherwise be possible - with the difference being more than 7x!

Additionally, allowing the router lifetime of 65535 seconds, coupled with sufficiently long lifetimes for the prefix, would cover the vast majority of the lifetimes of the devices on the WiFi networks. 65535 seconds is 18.2 hours, and the typical mobile devices might not even stay on the same network for such a long period of time. This would allow to increase the robustness of the network in the face of bad radio conditions causing the high loss of the multicast RAs.

5.2. Explicitly Client-Driven Router Advertisements

We can logically extend the "client link shutdown" in the direction of smaller connectivity loss, and imagine that the client, instead of completely shutting the radio down, would flap its radio link somewhere close to router lifetime expiry, therefore, while acting fully within the standards it will be able to maintain the connectivity during all but very short period of time, without any use of periodic RAs.

It may be interesting to explore a modification of the client behavior such that the "flap time" converges to zero, and eventually allowing the client to initiate a unicast Router Solicitation some time shortly before the router lifetime expires. This will have the result of the client being able to maintain the connectivity without the need of processing any periodic RAs. The advantage of doing so is that the RS-RA exchange will happen at the time convenient for the client sleep schedule - thus allowing to maximize the battery life.

6. Acknowledgements

Thanks to the following people for the very useful discussions. In no particular order: Erik Nordmark, Pascal Thubert, Eric Levy-Abegnoli, Ole Troan, Eric Vyncke, Federico Lovison, Jerome Henry.

7. IANA Considerations

None.

8. Security Considerations

Not discussed in -00.

9. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3315] Droms, R., Bound, J., Volz, B., Lemon, T., Perkins, C., and M. Carney, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 3315, July 2003.

[RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, September 2007.

[RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless Address Autoconfiguration", RFC 4862, September 2007.

[RFC6085] Gundavelli, S., Townsley, M., Troan, O., and W. Dec, "Address Mapping of IPv6 Multicast Packets on Ethernet", RFC 6085, January 2011.

[RFC6620] Nordmark, E., Bagnulo, M., and E. Levy-Abegnoli, "FCFS SAVI: First-Come, First-Served Source Address Validation Improvement for Locally Assigned IPv6 Addresses", RFC 6620, May 2012.

[RFC7048] Nordmark, E. and I. Gashinsky, "Neighbor Unreachability Detection Is Too Impatient", RFC 7048, January 2014.

Authors' Addresses

Andrew Yourtchenko
cisco
7a de Kleetlaan
Diegem, 1831
Belgium

Phone: +32 2 704 5494
Email: ayourtch@cisco.com

Lorenzo Colitti
Google

Email: lorenzo@google.com