

Internet Engineering Task Force (IETF)
Request for Comments: 9466
Category: Standards Track
ISSN: 2070-1721

Y. Liu, Ed.
China Mobile
T. Eckert, Ed.
M. McBride
Futurewei
Z. Zhang
ZTE Corporation
October 2023

PIM Assert Message Packing

Abstract

When PIM Sparse Mode (PIM-SM), including PIM Source-Specific Multicast (PIM-SSM), is used in shared LAN networks, there is often more than one upstream router. This can lead to duplicate IP multicast packets being forwarded by these PIM routers. PIM Assert messages are used to elect a single forwarder for each IP multicast traffic flow between these routers.

This document defines a mechanism to send and receive information for multiple IP multicast flows in a single PackedAssert message. This optimization reduces the total number of PIM packets on the LAN and can therefore speed up the election of the single forwarder, reducing the number of duplicate IP multicast packets incurred.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at <https://www.rfc-editor.org/info/rfc9466>.

Copyright Notice

Copyright (c) 2023 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction
 - 1.1. Requirements Language
 - 1.2. Terminology
2. Problem Statement

3.	Specification
3.1.	PIM Packed Assert Capability Hello Option
3.2.	Assert Packing Message Formats
3.3.	PackedAssert Mechanism
3.3.1.	Sending PackedAssert Messages
3.3.1.1.	Handling of Reception-Triggered Assert Records
3.3.1.2.	Handling of Timer Expiry-Triggered Assert Records
3.3.1.3.	Beneficial Delay in Sending PackedAssert Messages
3.3.1.4.	Handling Assert/PackedAssert Message Loss
3.3.1.5.	Optimal Degree of Assert Record Packing
3.3.2.	Receiving PackedAssert Messages
4.	Packet Formats
4.1.	PIM Packed Assert Capability Hello Option
4.2.	Assert Message Format
4.3.	Simple PackedAssert Message Format
4.4.	Aggregated PackedAssert Message Format
4.4.1.	Source Aggregated Assert Record
4.4.2.	RP Aggregated Assert Record
5.	IANA Considerations
6.	Security Considerations
7.	References
7.1.	Normative References
7.2.	Informative References
Appendix A.	Use Case Examples
A.1.	Enterprise Network
A.2.	Video Surveillance
A.3.	Financial Services
A.4.	IPTV Broadcast Video
A.5.	MVPN MDT
A.6.	Special L2 Services
	Acknowledgments
	Authors' Addresses

1. Introduction

When PIM-SM is used in shared LAN networks, there is typically more than one upstream router. When duplicate data packets appear on the LAN from different upstream routers, assert packets are sent from these routers to elect a single forwarder according to [RFC7761]. The PIM Assert messages are sent periodically to keep the Assert state. The PIM Assert message carries information about a single multicast source and group, along with the corresponding Metric and Metric Preference of the route towards the source or PIM Rendezvous Point (RP).

This document defines a mechanism to encode the information of multiple PIM Assert messages into a single PackedAssert message. This allows sending and receiving information for multiple IP multicast flows in a single PackedAssert message without changing the PIM Assert state machinery. It reduces the total number of PIM packets on the LAN and can therefore speed up the election of the single forwarder, reducing the number of duplicate IP multicast packets. This can be particularly helpful when there is traffic for a large number of multicast groups or SSM channels and PIM packet processing performance of the routers is slow.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

1.2. Terminology

The reader is expected to be familiar with the terminology of [RFC7761]. The following lists the abbreviations repeated in this document.

AT: Assert Timer
DR: Designated Router
RP: Rendezvous Point
RPF: Reverse Path Forwarding
RPT: RP Tree
SPT: Shortest Path Tree

2. Problem Statement

PIM Asserts occur in many deployments. See Appendix A for explicit examples and explanations of why it is often not possible to avoid.

PIM Assert state depends mainly on the network topology. As long as there is a Layer 2 (L2) network with more than two PIM routers, there may be multiple upstream routers, which can cause duplicate multicast traffic to be forwarded and assert processing to occur.

As the multicast services become widely deployed, the number of multicast entries increases, and a large number of Assert messages may be sent in a very short period when multicast data packets trigger PIM assert processing in the shared LAN networks. The PIM routers need to process a large number of small PIM assert packets in a very short time. As a result, the device load is very large. The assert packet may not be processed in time or even discarded, thus extending the time of traffic duplication in the network.

The PIM Assert mechanism can only be avoided by designing the network to be without transit subnets with multiple upstream routers. For example, an L2 ring between routers can sometimes be reconfigured to be a ring of point-to-point subnets connected by the routers. However, these Layer 2 (L2) and Layer 3 (L3) topology changes are undesirable when they are only done to enable IP multicast with PIM because they increase the cost of introducing IP multicast with PIM.

These designs are also not feasible when specific L2 technologies are needed. For example, various L2 technologies for rings provide sub-50 msec failover mechanisms, something not possible equally with a ring composed from L3 subnets. Likewise, IEEE Time-Sensitive Networking mechanisms would require an L2 topology that cannot simply be replaced by an L3 topology. L2 sub-topologies can also significantly reduce the cost of deployment.

3. Specification

This document defines three elements in support of PIM assert packing:

1. The PIM Packed Assert Capability Hello Option
2. The encoding of PackedAssert messages
3. How to send and receive PackedAssert messages

3.1. PIM Packed Assert Capability Hello Option

The PIM Packed Assert Capability Hello Option (Section 4.1) is used to announce support for the assert packing mechanisms specified in

this document. PackedAssert messages (Section 3.2) MUST NOT be used unless all PIM routers in the same subnet announce this option.

3.2. Assert Packing Message Formats

The PIM Assert message, as defined in Section 4.9.6 of [RFC7761], describes the parameters of a (*,G) or (S,G) assert using the following information elements: Rendezvous Point Tree flag (R), Source Address, Group Address, Metric, and Metric Preference. This document calls this information an "assert record".

This document introduces two new PIM Assert message encodings through the allocation and use of two flags in the PIM Assert message header [RFC9436]: the Packed (P) and the Aggregated (A) flags.

If P=0, the message is a (non-packed) PIM Assert message as specified in [RFC7761]. See Section 4.2. In this case, the A flag MUST be set to 0 and MUST be ignored on receipt.

If P=1, then the message is called a "PackedAssert message", and the type and hence encoding format of the payload are determined by the A flag.

If A=0, then the message body is a sequence of assert records. This is called a "Simple PackedAssert message". See Section 4.3.

If A=1, then the message body is a sequence of aggregated assert records. This is called an "Aggregated PackedAssert message". See Section 4.4.

Two aggregated assert record types are specified.

The "Source Aggregated Assert Record" (see Section 4.4.1) encodes one (common) Source Address, Metric, and Metric Preference as well as a list of one or more Group Addresses. Source Aggregated Assert Records provide a more compact encoding than the Simple PackedAssert message format when multiple (S,G) flows share the same source S. A single Source Aggregated Assert Record with n Group Addresses represents the information of assert records for (S,G1)...(S,Gn).

The "RP Aggregated Assert Record" (see Section 4.4.2) encodes one common Metric and Metric Preference as well as a list of "Group Records", each of which encodes a Group Address and a list of zero or more Source Addresses with a count. This is called an "RP Aggregated Assert Record", because with standard RPF according to [RFC7761], all the Group Addresses that use the same RP will have the same Metric and Metric Preference.

RP Aggregation Assert Records provide a more compact encoding than the Simple PackedAssert message format for (*,G) flows. The Source Address is optionally used in the assert procedures in [RFC7761] to indicate the source(s) that triggered the assert; otherwise, the Source Address is set to 0 in the assert record.

Both Source Aggregated Assert Records and RP Aggregated Assert Records also include the R flag, which maintains its semantics from [RFC7761] but also distinguishes the encodings. Source Aggregated Assert Records have R=0, as (S,G) assert records do in [RFC7761]. RP Aggregated Assert Records have R=1, as (*,G) assert records do in [RFC7761].

3.3. PackedAssert Mechanism

PackedAsserts do not change the PIM Assert state machine specification [RFC7761]. Instead, sending and receiving of PackedAssert messages, as specified in the following subsections, are

logically new packetization options for assert records in addition to the (non-packed) Assert message [RFC7761]. There is no change to the assert record information elements transmitted or their semantics. They are just transmitted in fewer but larger packets, and a fewer total number of bytes is used to encode the information elements. As a result, PIM routers should be able to send and receive assert records faster and/or with less processing overhead.

3.3.1. Sending PackedAssert Messages

When using assert packing, the regular Assert message encoding [RFC7761] with A=0 and P=0 is still allowed to be sent. Routers are free to choose which PackedAssert message format they send -- simple (Section 4.3) and/or aggregated (Section 4.4).

- * When any PIM routers on the LAN have not signaled support for assert packing, implementations MUST only send Asserts and MUST NOT send PackedAsserts under any condition.
- * The protocol or system conditions for which an implementation sends PackedAsserts instead of Asserts are out of scope for this specification. Protocol conditions include protocol triggers such as data-triggered asserts or Assert Timer (AT) expiry-triggered asserts, and system conditions include high or low load or control plane packet reception rates.
- * Implementations are expected to specify in documentation and/or management interfaces (such as a YANG data model) which PackedAssert message formats they can send and under which conditions they will send them.
- * Implementations SHOULD be able to indicate to the operator (such as through a YANG data model) how many Assert and PackedAssert messages were sent/received and how many assert records were sent/received.
- * A configuration option SHOULD be available to disable PackedAssert operations. PIM-SM implementations [RFC7761] that introduce support for assert packing from day one MAY omit this configuration option.

When a PIM router has an assert record ready to send according to [RFC7761], it calls one of the following functions:

- * send Assert(S,G) / send Assert(*,G) ([RFC7761], Section 4.2)
- * send Assert(S,G) / send AssertCancel(S,G) ([RFC7761], Section 4.6.1)
- * send Assert(*,G) / send AssertCancel(*,G) ([RFC7761], Section 4.6.2)
- * send Assert(S,G) ([RFC7761], Section 4.8.2)

If sending of PackedAsserts is possible on the network, instead of sending an Assert message with an assert record, any of these calls MAY instead result in the PIM implementation remembering the assert record and continuing with further processing for other flows, which may result in additional assert records.

PIM MUST then create PackedAssert messages from the remembered assert records and schedule them for sending according to the considerations in the following subsections.

3.3.1.1. Handling of Reception-Triggered Assert Records

Avoiding additional delay because of assert packing compared to immediately scheduling Assert messages is most critical for assert records that are triggered by reception of data or reception of asserts against which the router is in the "I am Assert Winner" state. In these cases, the router SHOULD send out an Assert or PackedAssert message containing this assert record as soon as possible to minimize the time in which duplicate IP multicast packets can occur.

To avoid additional delay in this case, the router should employ appropriate assert packing and scheduling mechanisms, as explained here.

Asserts/PackedAsserts created from reception-triggered assert records should be scheduled for serialization with a higher priority than those created because of other protocol or system conditions. They should also bypass other PIM messages that can create significant bursts, such as PIM join/prune messages.

When there are no reception-triggered Assert/PackedAssert messages currently being serialized on the interface or scheduled to be sent, the router should immediately generate and schedule an Assert or PackedAssert message without further assert packing.

If one or more reception-triggered Assert/PackedAssert messages are already serializing or are scheduled to be serialized on the outgoing interface, then the router can use the time until the last of those messages has finished serializing for PIM processing of further conditions. This may result in additional reception-triggered assert records and the packing of these assert records without introducing additional delay.

3.3.1.2. Handling of Timer Expiry-Triggered Assert Records

Asserts triggered by expiry of the AT on an assert winner are not time-critical because they can be scheduled in advance and because the `Assert_Override_Interval` parameter [RFC7761] already creates a 3-second window in which such assert records can be sent, received, and processed before an assert loser's state expires and duplicate IP multicast packets could occur.

An example mechanism to allow packing of AT expiry-triggered assert records on assert winners is to round the AT to an appropriate granularity such as 100 msec. This will cause the AT for multiple (S,G) and/or (*,G) states to expire at the same time, thus allowing them to be easily packed without changes to the Assert state machinery.

AssertCancel messages have assert records with an infinite metric and can use assert packing like any other Assert. They are sent on Override Timer (OT) expiry and can be packed, for example, with the same considerations as AT expiry-triggered assert records.

3.3.1.3. Beneficial Delay in Sending PackedAssert Messages

Delay in sending PackedAsserts beyond what was discussed in prior subsections can still be beneficial when it causes the overall number of possible duplicate IP multicast packets to decrease in a situation with a large number of (S,G) and/or (*,G), compared to the situation where an implementation only sends Assert messages.

This delay can be used in implementations because it cannot support the more advanced mechanisms described above, and this longer delay can be achieved by some simpler mechanisms (such as only periodic generation of PackedAsserts) and still achieves an overall reduction in duplicate IP multicast packets compared to sending only Asserts.

3.3.1.4. Handling Assert/PackedAssert Message Loss

When Asserts are sent, a single packet loss will result only in continued or new duplicates from a single IP multicast flow. Loss of a (non-AssertCancel) PackedAssert impacts duplicates for all flows packed into the PackedAssert and may result in the need for resending more than one Assert/PackedAssert, because of the possible inability to pack the assert records in this condition. Therefore, routers SHOULD support mechanisms that allow PackedAsserts and Asserts to be sent with an appropriate Differentiated Services Code Point (DSCP) [RFC2475] such as Expedited Forwarding (EF) to minimize their loss, especially when duplicate IP multicast packets could cause congestion and loss.

Routers MAY support a configurable option for sending PackedAssert messages twice in short order (such as 50 msec apart) to overcome possible loss, but only when the following two conditions are met.

1. The total size of the two PackedAsserts is less than the total size of equivalent Assert messages.
2. The condition of the assert record flows in the PackedAssert is such that the router can expect that their reception by PIM routers will not trigger Assert/PackedAsserts replies. This condition is true, for example, when sending an assert record while becoming or being assert winner (Action A1/A3 in [RFC7761]).

3.3.1.5. Optimal Degree of Assert Record Packing

The optimal target packing size will vary depending on factors including implementation characteristics and the required operating scale. At some point, as the target packing size is varied from the size of a single non-packed Assert to the MTU size, a size can be expected to be found where the router can achieve the required operating scale of (S,G) and (*,G) flows with minimum duplicates. Beyond this size, a further increase in the target packing size would not produce further benefits but might introduce possible negative effects such as the incurrence of more duplicates on loss.

For example, in some router implementations, the total number of packets that a control plane function such as PIM can send/receive per unit of time is a more limiting factor than the total amount of data across these packets. As soon as the packet size is large enough for the maximum possible payload throughput, increasing the packet size any further may still reduce the processing overhead of the router but may increase latency incurred in creating the packet in a way that may increase duplicates compared to smaller packets.

3.3.2. Receiving PackedAssert Messages

Upon reception of a PackedAssert message, the PIM router logically converts its payload into a sequence of assert records that are then processed as if an equivalent sequence of Assert messages were received according to [RFC7761].

4. Packet Formats

This section describes the format of new PIM extensions introduced by this document.

4.1. PIM Packed Assert Capability Hello Option

The PIM Packed Assert Capability Hello Option is a new option for PIM Hello messages according to Section 4.9.2 of [RFC7761].

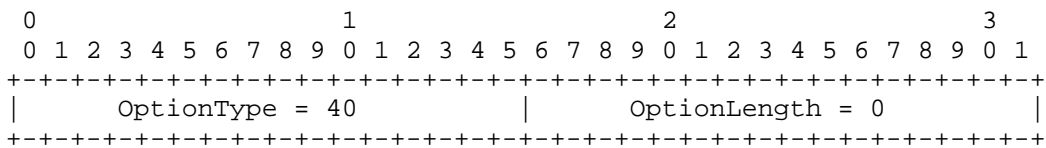


Figure 1: PIM Packed Assert Capability Hello Option

OptionType 40 (Packed Assert Capability):

Indicates support for the ability to receive and process all the PackedAssert encodings defined in this document.

OptionLength 0:

The Packet Assert Capability has no OptionValue.

4.2. Assert Message Format

Figure 2 shows a PIM Assert message as specified in Section 4.9.6 of [RFC7761]. The Encoded-Group and Encoded-Unicast address formats are specified in Section 4.9.1 of [RFC7761] for IPv4 and IPv6.

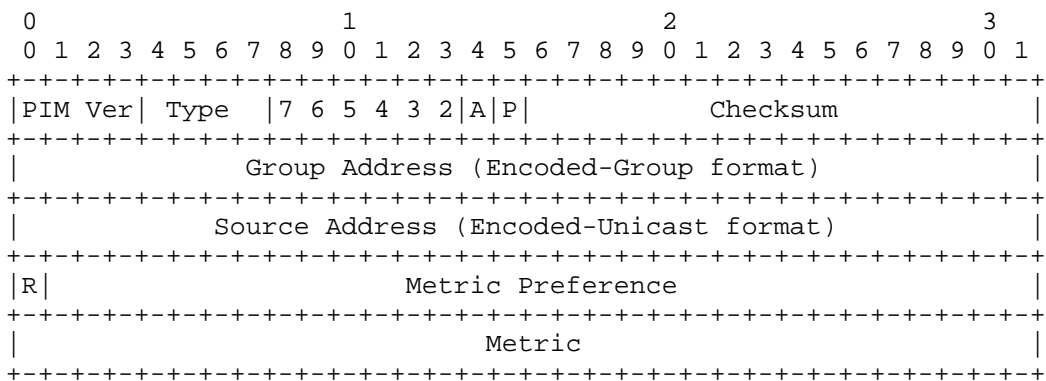
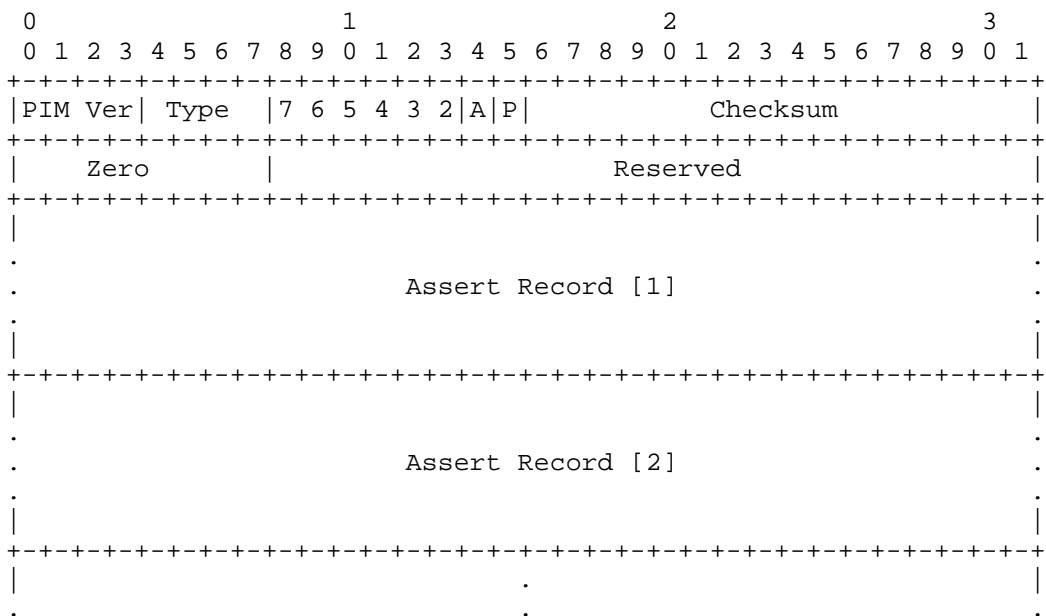


Figure 2: Assert Message Format

This common header shows the "7 6 5 4 3 2" flag bits (as defined in Section 4 of [RFC9436]) and the location of the P and A flags (as described in Section 5). As specified in Section 3.2, both flags in a (non-packed) PIM Assert message are required to be set to 0.

4.3. Simple PackedAssert Message Format



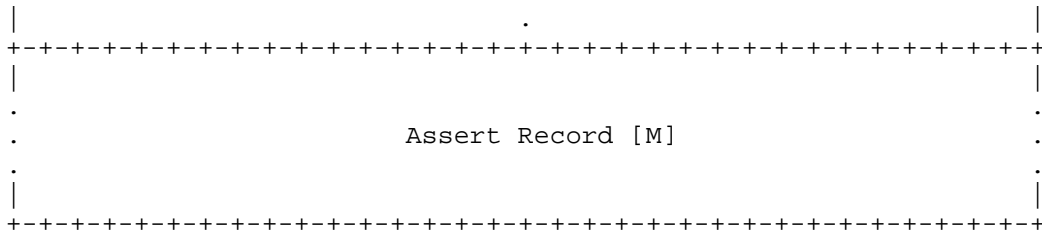


Figure 3: Simple PackedAssert Message Format

PIM Version, Type, Checksum:

As specified in Section 4.9.6 of [RFC7761].

"7 6 5 4 3 2":

Flag bits per Section 4 of [RFC9436].

P:

Packed flag. MUST be 1.

A:

Aggregated flag. MUST be 0.

Zero:

Set to zero on transmission. Serves to make PIM routers that are not capable of assert packing to fail in parsing the message instead possible mis-parsing of the message as an Assert message [RFC7761] if this field was not zero-filled.

Reserved:

Set to zero on transmission. Ignored upon receipt.

M:

The number of Assert Records in the message. Derived from the length of the packet carrying the message.

Assert Record:

Formatted according to Figure 3, which is the same as the PIM Assert message body as specified in Section 4.9.6 of [RFC7761].

The format of each Assert Record is the same as the PIM Assert message body as specified in Section 4.9.6 of [RFC7761].

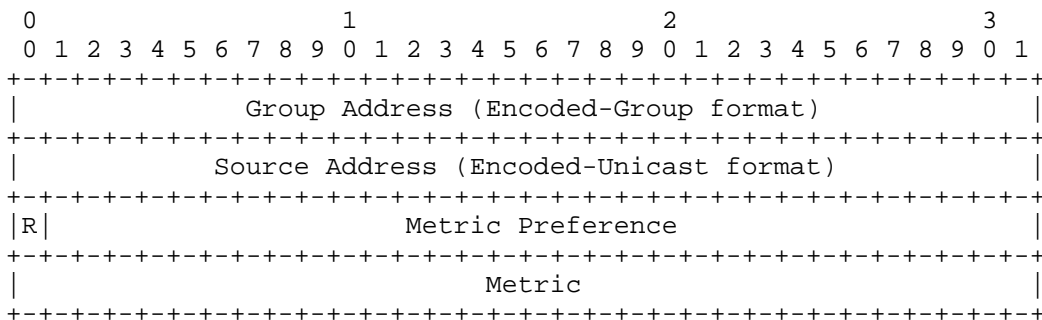
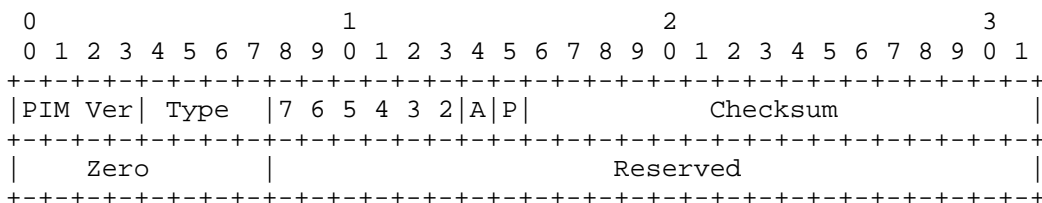


Figure 4: Assert Record

4.4. Aggregated PackedAssert Message Format



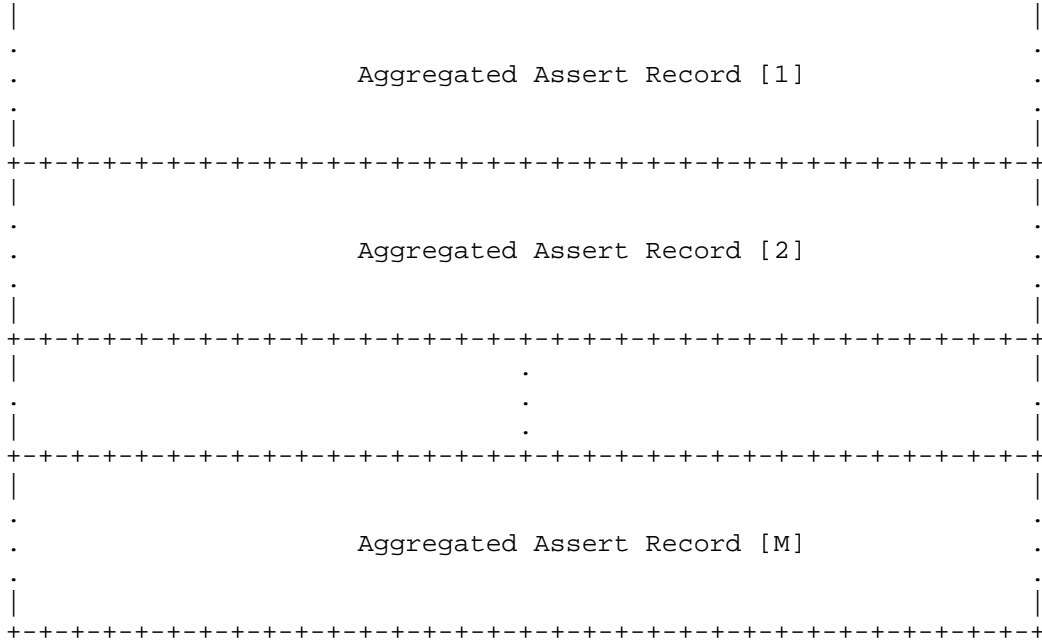


Figure 5: Aggregated PackedAssert Message Format

PIM Version, Type, Reserved, Checksum:
As specified in Section 4.9.6 of [RFC7761].

"7 6 5 4 3 2":
Flag bits per Section 4 of [RFC9436].

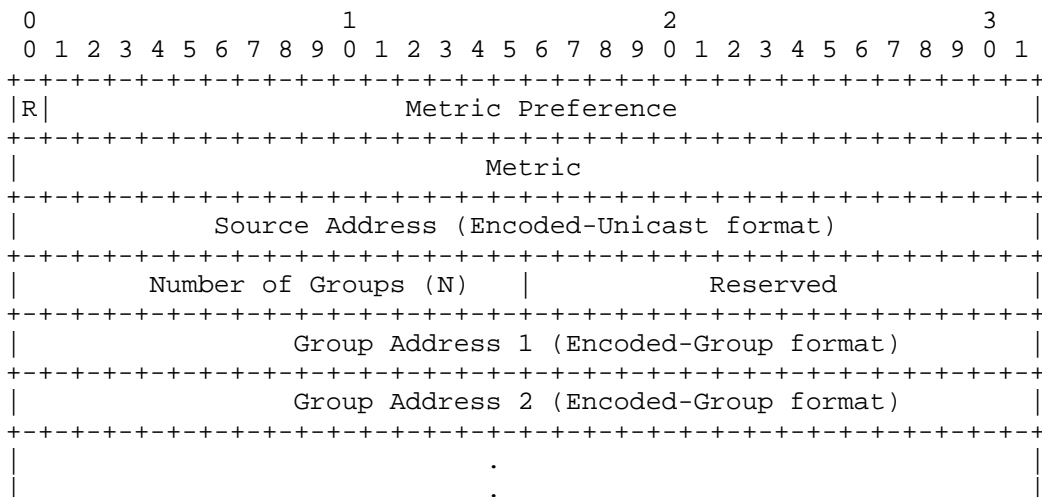
P:
Packed flag. MUST be 1.

A:
Aggregated flag. MUST be 1.

Zero:
Set to zero on transmission. Serves to make PIM routers that are not capable of assert packing to fail in parsing the message instead possible mis-parsing of the message as an Assert message [RFC7761] if this field was not zero-filled.

Aggregated Assert Record:
Formatted according to Figure 5. The number M of Aggregated Assert Records is determined from the packet size.

4.4.1. Source Aggregated Assert Record



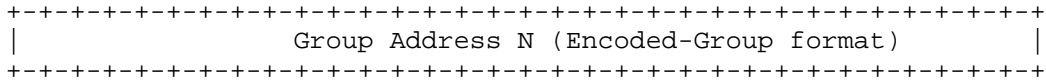


Figure 6: Source Aggregated Assert Record

R:

MUST be 0.

R indicates both that the encoding format of the record is that of a Source Aggregated Assert Record and that all assert records represented by the Source Aggregated Assert Record have R=0 and are therefore (S,G) assert records according to the definition of R in [RFC7761], Section 4.9.6.

Metric Preference, Metric, Source Address:

As specified in Section 4.9.6 of [RFC7761]. Source Address MUST NOT be zero.

Number of Groups:

The number of Group Address fields.

Reserved:

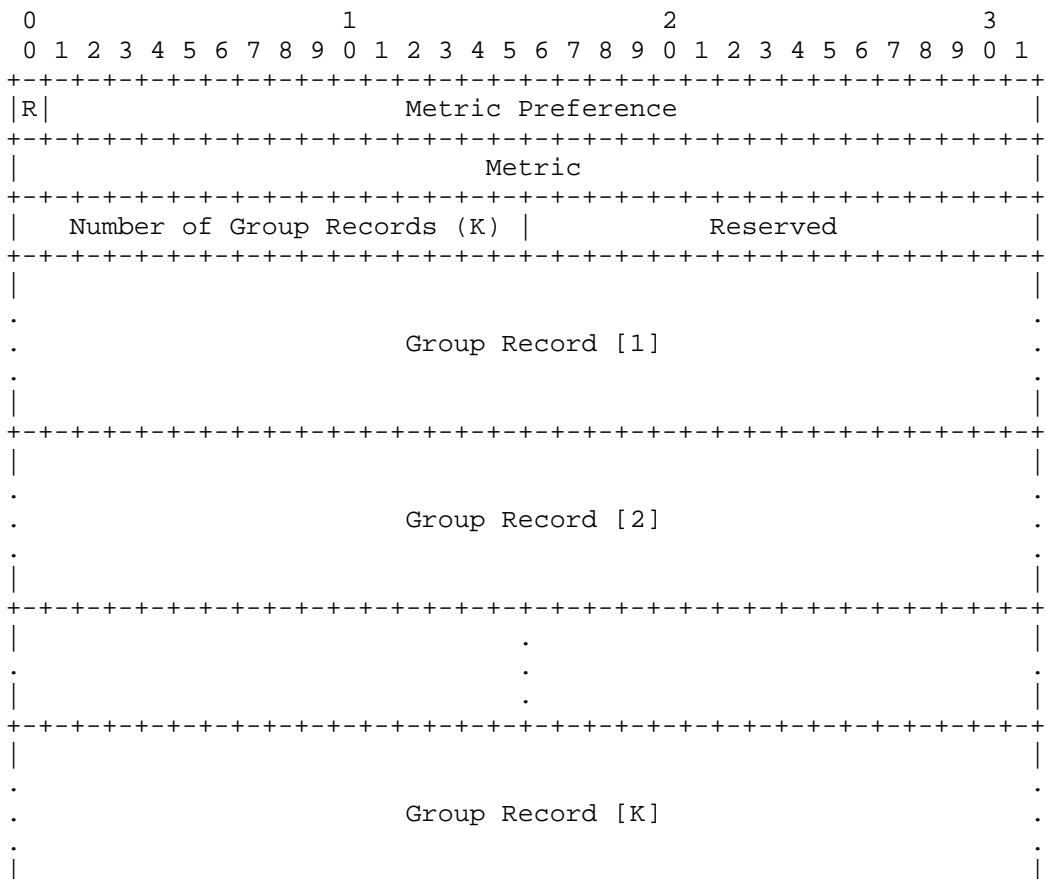
Set to zero on transmission. Ignored upon receipt.

Group Address:

As specified in Section 4.9.6 of [RFC7761].

4.4.2. RP Aggregated Assert Record

An RP Aggregation Assert Record aggregates (*,G) assert records with the same Metric Preference and Metric. Typically, this is the case for all (*,G) using the same RP, but the encoding is not limited to only (*,G) using the same RP because the RP address is not encoded as it is also not present in assert records [RFC7761].



+++++

Figure 7: RP Aggregated Assert Record

R:

MUST be 1.

R indicates both that the encoding format of the record is that of an RP Aggregated Assert Record and that all assert records represented by the RP Aggregated Assert Record have R=1 and are therefore (*,G) assert records according to the definition of R in [RFC7761], Section 4.9.6.

Metric Preference, Metric:

As specified in Section 4.9.6 of [RFC7761].

Number of Group Records (K):

The number of packed Group Records. A record consists of a Group Address and a Source Address list with a number of sources.

Reserved:

Set to zero on transmission. Ignored upon receipt.

The format of each Group Record is:

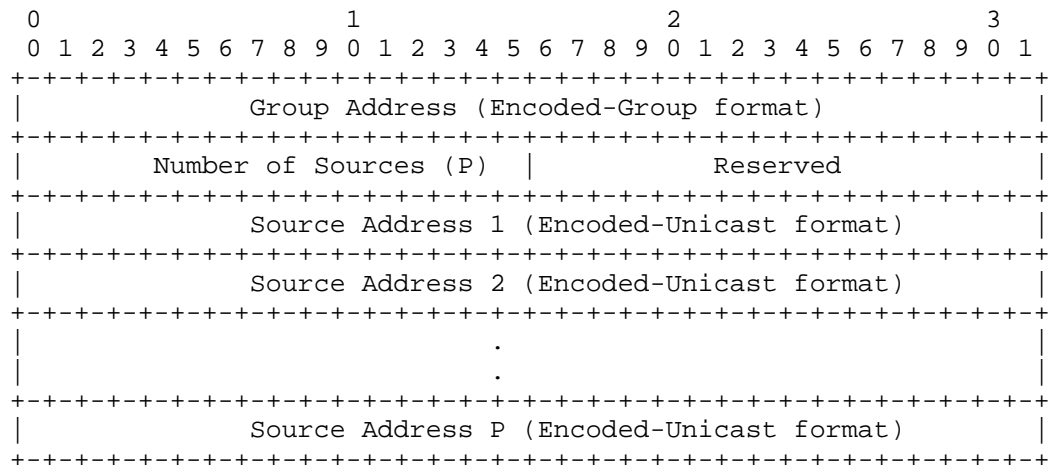


Figure 8: Group Record

Group Address:

As specified in Section 4.9.6 of [RFC7761].

Reserved:

Set to zero on transmission. Ignored upon receipt.

Number of Sources (P):

The Number of Sources corresponds to the number of Source Address fields in the Group Record. If this number is not 0 and one of the (*,G) assert records to be encoded has Source Address 0, then 0 needs to be encoded as one of the Source Address fields.

Reserved:

Set to zero on transmission. Ignored upon receipt.

Source Address:

As specified in Section 4.9.6 of [RFC7761]. But there can be multiple Source Address fields in the Group Record.

5. IANA Considerations

IANA has updated the "PIM Message Types" registry as follows to

include the Packed and Aggregated flag bits for the Assert message type.

Value	Length	Name	Reference
40	0	Packed Assert Capability	RFC 9466

Table 1: PIM-Hello Options

IANA has assigned the following two flag bits for PIM Assert messages in the "PIM Message Types" registry.

Type	Name	Flag Bits	Reference
5	Assert	0: Packed	RFC 9466
		1: Aggregated	RFC 9466
		2-7: Unassigned	[RFC3973] [RFC7761]

Table 2: PIM Message Types

6. Security Considerations

The security considerations of [RFC7761] apply to the extensions defined in this document.

This document packs multiple assert records in a single message. As described in Section 6.1 of [RFC7761], a forged Assert message could cause the legitimate designated forwarder to stop forwarding traffic to the LAN. The effect may be amplified when using a PackedAssert message.

Like other optional extensions of [RFC7761] that are active only when all routers indicate support for them, a single misconfigured or malicious router emitting forged PIM Hello messages can inhibit operations of this extension.

Authentication of PIM messages, such as that explained in Sections 6.2 and 6.3 of [RFC7761], can protect against forged message attacks.

7. References

7.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC7761] Fenner, B., Handley, M., Holbrook, H., Kouvelas, I., Parekh, R., Zhang, Z., and L. Zheng, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", STD 83, RFC 7761, DOI 10.17487/RFC7761, March 2016, <<https://www.rfc-editor.org/info/rfc7761>>.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.
- [RFC9436] Venaas, S. and A. Retana, "PIM Message Type Space

Extension and Reserved Bits", RFC 9436,
DOI 10.17487/RFC9436, August 2023,
<<https://www.rfc-editor.org/info/rfc9436>>.

7.2. Informative References

- [RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, DOI 10.17487/RFC2475, December 1998, <<https://www.rfc-editor.org/info/rfc2475>>.
- [RFC3973] Adams, A., Nicholas, J., and W. Siadak, "Protocol Independent Multicast - Dense Mode (PIM-DM): Protocol Specification (Revised)", RFC 3973, DOI 10.17487/RFC3973, January 2005, <<https://www.rfc-editor.org/info/rfc3973>>.
- [RFC6037] Rosen, E., Ed., Cai, Y., Ed., and IJ. Wijnands, "Cisco Systems' Solution for Multicast in BGP/MPLS IP VPNs", RFC 6037, DOI 10.17487/RFC6037, October 2010, <<https://www.rfc-editor.org/info/rfc6037>>.
- [RFC7431] Karan, A., Filsfils, C., Wijnands, IJ., Ed., and B. Decraene, "Multicast-Only Fast Reroute", RFC 7431, DOI 10.17487/RFC7431, August 2015, <<https://www.rfc-editor.org/info/rfc7431>>.
- [RFC7490] Bryant, S., Filsfils, C., Previdi, S., Shand, M., and N. So, "Remote Loop-Free Alternate (LFA) Fast Reroute (FRR)", RFC 7490, DOI 10.17487/RFC7490, April 2015, <<https://www.rfc-editor.org/info/rfc7490>>.

Appendix A. Use Case Examples

The PIM Assert mechanism can only be avoided by designing the network to be without transit subnets with multiple upstream routers. For example, an L2 ring between routers can sometimes be reconfigured to be a ring of point-to-point subnets connected by the routers. However, these L2/L3 topology changes are undesirable when they are only done to enable IP multicast with PIM because they increase the cost of introducing IP multicast with PIM.

These L3 ring designs are specifically undesirable when particular L2 technologies are needed. For example, various L2 technologies for rings provide sub-50 msec failover mechanisms that will benefit IP unicast and multicast alike without any added complexity to the IP layer (forwarding or routing). If such L2 rings were to be replaced by L3 rings just to avoid PIM asserts, then this would result in the need for a complex choice of a sub-50 msec IP unicast failover solution (such as [RFC7490] with IP repair tunnels) as well as a separate sub-50 msec IP multicast failover solution, (such as [RFC7431] with dedicated ring support). The mere fact that, by running at the IP layer, different solutions for IP unicast and multicast are required makes them more difficult to operate, and they typically require more expensive hardware. This often leads to non-support of the IP multicast part.

Likewise, IEEE Time-Sensitive Networking mechanisms would require an L2 topology that cannot simply be replaced by an L3 topology. L2 sub-topologies can also significantly reduce the cost of deployment.

The following subsections give examples of the type of network and use cases in which subnets with asserts have been observed or are expected to require scaling as provided by this specification.

A.1. Enterprise Network

When an enterprise network is connected through an L2 network, the intra-enterprise runs L3 PIM multicast. The different sites of the enterprise are equivalent to the PIM connection through the shared LAN network. Depending upon the locations and number of groups, there could be many asserts on the first-hop routers.

A.2. Video Surveillance

Video surveillance deployments have migrated from analog-based systems to IP-based systems oftentimes using multicast. In the shared LAN network deployments, when there are many cameras streaming to many groups, there may be issues with many asserts on first-hop routers.

A.3. Financial Services

Financial services extensively rely on IP Multicast to deliver stock market data and its derivatives, and the current multicast solution PIM is usually deployed. As the number of multicast flows grow, many stock data with many groups may result in many PIM asserts on a shared LAN network from the publisher to the subscribers.

A.4. IPTV Broadcast Video

PIM DR deployments are often used in host-side network for IPTV broadcast video services. Host-side access network failure scenarios may benefit from assert packing when many groups are being used. According to [RFC7761], the DR will be elected to forward multicast traffic in the shared access network. When the DR recovers from a failure, the original DR starts to send traffic, and the current DR is still forwarding traffic. In this situation, multicast traffic duplication maybe happen in the shared access network and can trigger the assert progress.

A.5. MVPN MDT

As described in [RFC6037], Multicast Distribution Tree (MDT) is used as tunnels for Multicast VPN (MVPN). The configuration of multicast-enabled VPN Routing and Forwarding (VRF) or changes to an interface that is in a VRF may cause many assert packets to be sent at the same time.

A.6. Special L2 Services

Additionally, future backhaul, or fronthaul, networks may want to connect L3 across an L2 underlay supporting Time-Sensitive Networks (TSNs). The infrastructure may run Deterministic Networking (DetNet) over TSN. These transit L2 LANs would have multiple upstreams and downstreams. This document takes a proactive approach to prevention of possible future assert issues in these types of environments.

Acknowledgments

The authors would like to thank the following individuals: Stig Venaas for the valuable contributions of this document, Alvaro Retana for his thorough and constructive RTG AD review, Ines Robles for her Gen-ART review, Tommy Pauly for his Transport Area review, Robert Sparks for his SecDir review, Shuping Peng for her RtgDir review, John Scudder for his RTG AD review, ric Vyncke for his INT AD review, Eric Kline for his INT AD review, Paul Wouter for his SEC AD review, Zaheduzzaman Sarker for his TSV AD review, Robert Wilton for his OPS AD review, and Martin Duke for his TSV AD review.

Authors' Addresses

Yisong Liu (editor)

China Mobile
China
Email: liuyisong@chinamobile.com

Toerless Eckert (editor)
Futurewei
United States of America
Email: tte@cs.fau.de

Mike McBride
Futurewei
United States of America
Email: michael.mcbride@futurewei.com

Zheng (Sandy) Zhang
ZTE Corporation
China
Email: zhang.zheng@zte.com.cn