Mboned
Internet-Draft
Intended status: Standards Track
Expires: April 21, 2019

J. Holland
Akamai Technologies, Inc.

October 18, 2018

DNS Reverse IP AMT Discovery
draft-jholland-mboned-driad-amt-discovery-00

Abstract

This document defines a new DNS resource record (RR) used to advertise addresses for Automatic Multicast Tunneling (AMT) relays capable of receiving multicast traffic from the owner of the RR. The new AMTRELAY RR makes possible a source-specific method for AMT gateways to discover appropriate AMT relays, in order to ingest traffic for source-specific multicast channels into multicast-capable receiving networks when no multicast connectivity is directly available between the sending and receiving networks.

Table of Contents

1. Introduction

AMT (Automatic Multicast Tunneling) is defined in [RFC7450], and provides a method to transport multicast traffic in a unicast tunnel, in order to traverse non-multicast capable network segments.

Section 4.1.5 of [RFC7450] explains that relay selection might need to be source dependent, since a relay must be able to receive

multicast traffic from the desired source in order to forward it. It suggests DNS-based queries as a possible approach. This document defines a DNS-based solution, as suggested there. This solution also addresses the relay discovery issues outlined in [RFC8313], in the "Disadvantages" lists in Sections 3.3 and 3.4.

The goal is to enable multicast connectivity between separate multicast-enabled networks when neither the sending nor the receiving network is connected to a multicast-enabled backbone, without requiring any peering arrangement between the networks.

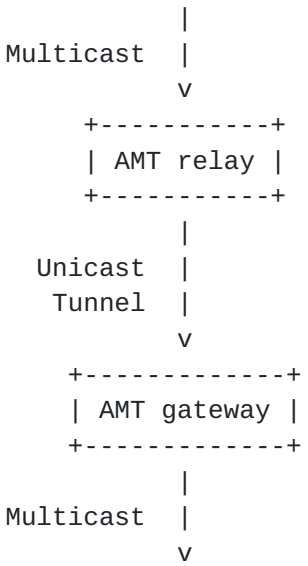
1.1. Background and Terminology

The reader is assumed to be familiar with the basic DNS concepts described in [RFC1034], [RFC1035], and the subsequent documents that update them, particularly [RFC2181].

The reader is also assumed to be familiar with the concepts and terminology regarding source-specific multicast as described in [RFC4607] and the usage of group management protocols for source-specific multicast as described in [RFC4604].

The reader should also be familiar with AMT, particularly the terminology listed in Section 3.2 and Section 3.3 of [RFC7450].

It's especially helpful to recall that once an AMT tunnel is established, the relay receives native multicast traffic and encapsulates it into the unicast tunnel, and the gateway receives the unicast tunnel traffic, unencapsulates it, and forwards it as native multicast:



1.2. Requirements Notation

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 [RFC2119] and [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Relay Discovery Operation

2.1. Overview

The AMTRELAY resource record (RR) is used to publish the address or host name of an AMT relay that can forward multicast traffic from a particular source host. The owner of the RR is the sender of native multicast traffic, and the RR provides the address or hostname of an AMT relay that can receive traffic from it.

The primary use case for the AMTRELAY RR is when a router that can act as an AMT gateway gets a signal indicating that a client in its receiving network has joined a new source-specific multicast channel, (hereafter called an (S,G), as defined in [RFC4607]), for example by receiving a PIM-SM (S,G) join message as described in Section 4.5.2 of [RFC7761].

When the source of a newly joined (S,G) is not reachable via a multicast-enabled next hop, the AMT gateway can connect to an AMT relay and propagate the join signal to that relay. The goal for source-specific relay discovery in this situation is to ensure that the AMT relay chosen is able to receive multicast traffic from the given source. More detailed example use cases are provided in Section 2.2 and Section 2.3, and other applicable examples appear in [RFC8313], Sections 3.3, 3.4, and 3.5.

Often an AMT gateway will only have access to the source and group IP addresses of the desired traffic, and will not know any other name for the source of the traffic. Because of this, typically the best way of looking up AMTRELAY RRs will be by using the source IP address as an index into one of the reverse mapping trees (in-addr.arpa for IPv4, as described in Section 3.5 of [RFC1035], or ip6.arpa for IPv6, as described in Section 2.5 of [RFC3596]).

Therefore, it is RECOMMENDED that AMTRELAY RRs be added to reverse IP zones as appropriate. AMTRELAY records MAY also appear in other zones, but the primary intended use case requires a reverse IP mapping for the source from an (S,G) in order to be useful to most AMT gateways.

When the reverse IP mapping has no AMTRELAY RR but does have a PTR record, the lookup is done in the fashion usual for PTR records. The IP address' octets (IPv4) or nibbles (IPv6) are reversed and looked up with the appropriate suffix. Any CNAMEs or DNAMEs found MUST be followed, and the AMTRELAY RR is queried with the resulting domain name.

When AMTRELAY RRs as defined in this document are available, it is RECOMMENDED that AMT gateways give the AMTRELAY RR precedence over AMT discovery using the anycast IPs defined in Section 7 of [RFC7450].

2.2. Example Receiving Networks

2.2.1. Tier 3 ISP

One example of a receiving network is an ISP that offers multicast ingest services to its subscribers, illustrated in Figure 1.

In the example network below, subscribers can join (S,G)s with MLDv2 or IGMPv3 as described in [RFC4604], and the AMT gateway in this ISP can receive and forward multicast traffic from one of the example sending networks in Section 2.3 by discovering the appropriate AMT relays with a DNS lookup for the AMTRELAY RR with the reverse IP of the source in the (S,G).

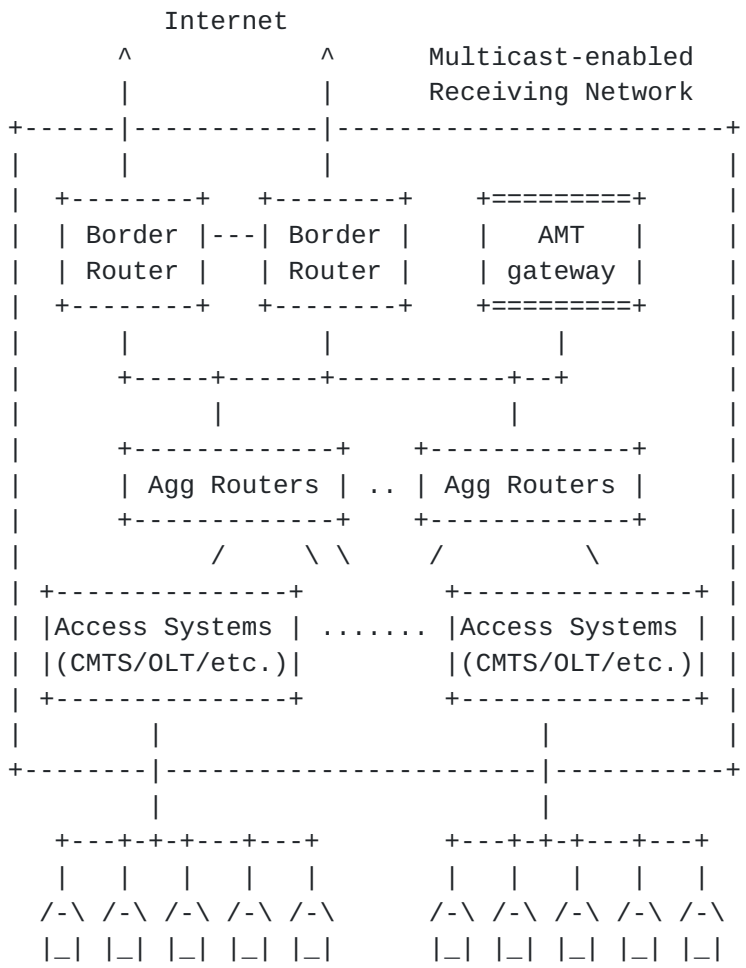


Figure 1: Receiving ISP Example

2.2.2. Small Office

Another example receiving network is a small branch office that regularly accesses some multicast content, illustrated in Figure 2.

This office has desktop devices that need to receive some multicast traffic, so an AMT gateway runs on a LAN with these devices, to pull traffic in through a non-multicast next-hop.

The office also hosts some mobile devices that have AMT gateway instances embedded in apps, in order to receive multicast traffic over their non-multicast wireless LAN.

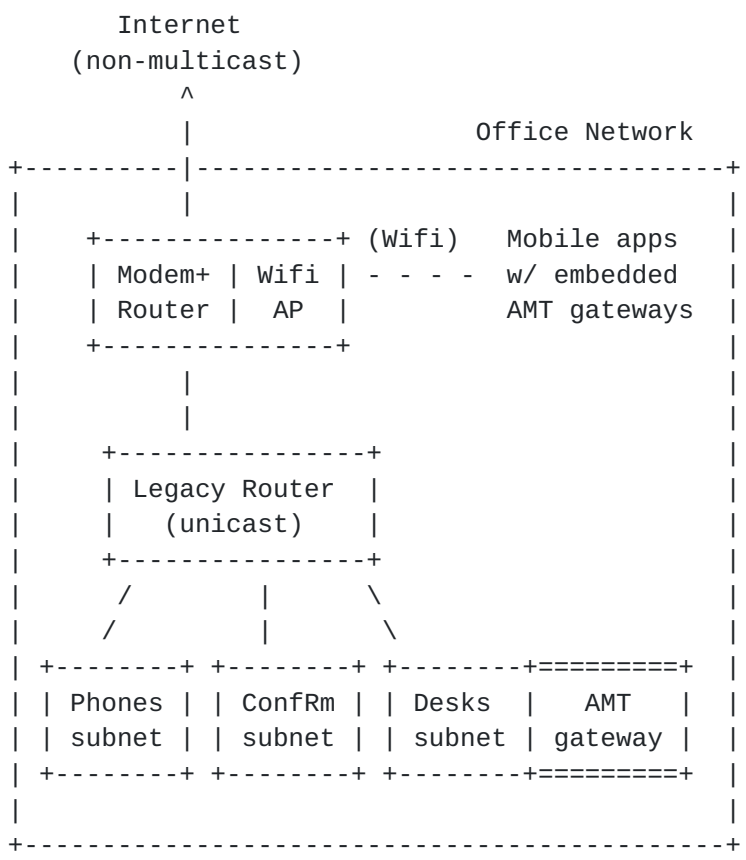


Figure 2: Small Office (no multicast up)

By adding an AMT relay to this office network as in Figure 3, it's possible to make use of multicast services from the example multicast-capable ISP in Section 2.2.1, provided that the AMT gateways contact the local AMT relay instead of an AMT relay upstream of the multicast-capable ISP, and the uplink router performs IGMP/MLD Proxying, as described in [RFC4605].

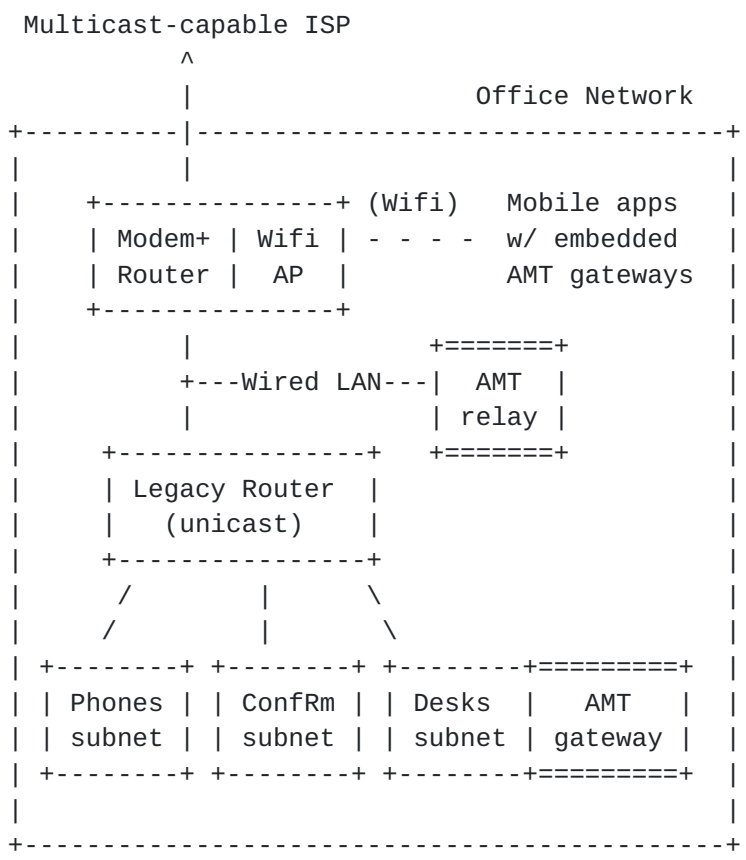


Figure 3: Small Office Example

For this reason, it's RECOMMENDED to provide an AMTRELAY RR referencing _amt._udp.home.arpa for sources, with a more-preferred precedence than the known relays close to source relays like those described in Section 2.3.

<TBD>

.home.arpa is pretty close to what's needed, but since this use case is not a residential home network, should this be another different special-use domain name?

https://tools.ietf.org/html/rfc8375
https://www.iana.org/assignments/
    locally-served-dns-zones/locally-served-dns-zones.xhtml
    special-use-domain-names/special-use-domain-names.xhtml

e.g. _amt._udp.home.arpa
e.g. _amt._udp.most-local.arpa =>
  .local if it's there,
  .home.arpa if it's not,
  .isp.arpa if it's not
(most-local because if somebody bothered to deploy a relay, they did so in a spot where it can do a next-hop receive of multicast, as long as no upstream gateway finds this relay and creates a loop.)

(Can/should "most-local.arpa" be done with the well-known anycast ip? Not sure...)

<\TBD>

2.3. Example Sending Networks

2.3.1. Sender-controlled Relays

When a sender network is also operating AMT relays to distribute multicast traffic, as in Figure 4, each address could appear as an AMTRELAY RR for the reverse IP of the sender, or one or more domain names could appear in AMTRELAY RRs, and the AMT relay addresses can be discovered by finding an A or AAAA record from those domain names.

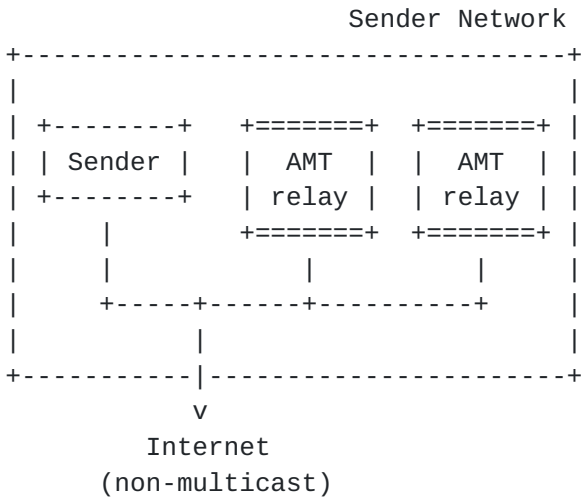


Figure 4: Small Office Example

2.3.2. Provider-controlled Relays

When an ISP offers a service to transmit outbound multicast traffic through a forwarding network, they might also offer AMT relays in order to reach receivers without multicast connectivity to the forwarding network, as in Figure 5. In this case it's RECOMMENDED that a domain name for the AMT relays also be provided for use with the discovery process defined in this document.

When the sender wishes to use the relays provided by the ISP for forwarding multicast traffic, an AMTRELAY RR should be configured to use the domain name provided by the ISP, to allow for address reassignment of the relays without forcing the sender to reconfigure the corresponding AMTRELAY RRs.

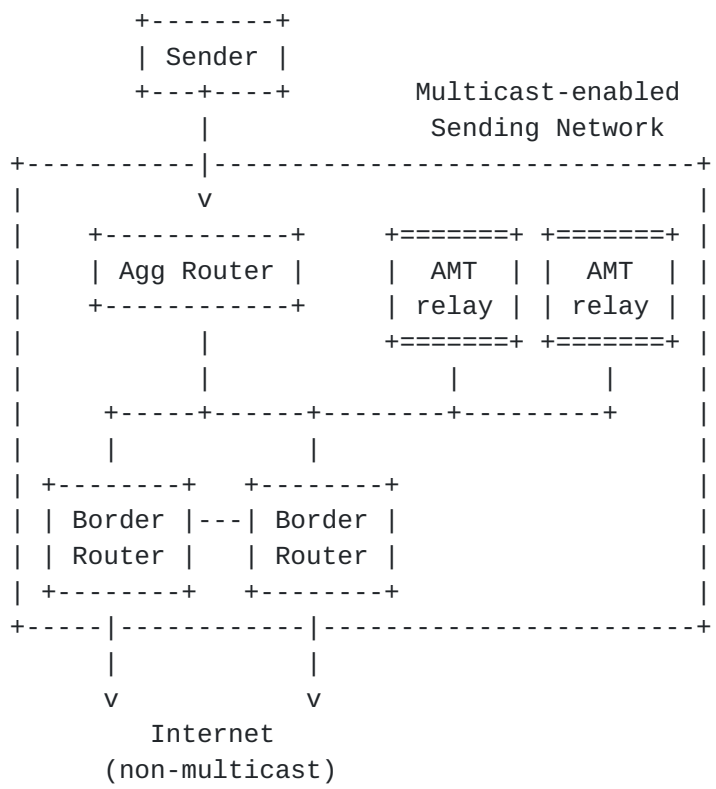


Figure 5: Sending ISP Example

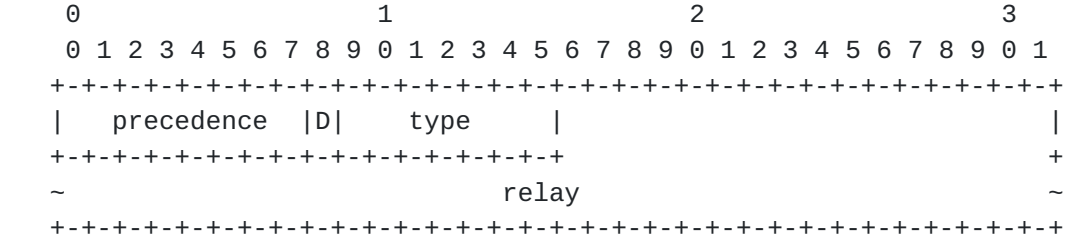
3. AMTRELAY Resource Record Definition

3.1. AMTRELAY RRType

The AMTRELAY RRType has the mnemonic AMTRELAY and type code 68 (decimal).

3.2. AMTRELAY RData Format

The AMTRELAY RData consists of a 8-bit precedence field, a 1-bit "Discovery Optional" field, a 7-bit type field, and a variable length relay field.



3.2.1. RData Format - Precedence

This is an 8-bit precedence for this record. It is interpreted in the same way as the PREFERENCE field described in Section 3.3.9 of [RFC1035].

Relays listed in AMTRELAY records with a lower value for precedence are to be attempted first.

Where there is a tie in precedence, the default choice of relay MUST be non-deterministic, to support load balancing. The AMT gateway operator MAY override this default choice with explicit configuration when it's necessary for administrative purposes.

For example, one network might prefer to tunnel IPv6 multicast traffic over IPv6 AMT and IPv4 multicast traffic over IPv4 AMT to avoid routeability problems in IPv6 from affecting IPv4 traffic and vice versa, while another network might prefer to tunnel both kinds of traffic over IPv6 to reduce the IPv4 space used by its AMT gateways. In this example scenario or other cases where there is an administrative preference that requires explicit configuration, a receiving network MAY make systematically different precedence choices among records with the same precedence value.

3.2.2. RData Format - Discovery Optional (D-bit)

The D bit is a "Discovery Optional" flag.

If the D bit is set to 0, a gateway using this RR MUST perform AMT relay discovery as described in Section 4.2.1.1 of [RFC7450], rather than directly sending an AMT request message to the relay.

That is, the gateway MUST receive an AMT relay advertisement message (Section 5.1.2 of [RFC7450]) for an address before sending an AMT request message (Section 5.1.3 for [RFC7450]) to that address. Before receiving the relay advertisement message, this record has only indicated that the address can be used for AMT relay discovery, not for a request message. This is necessary for devices that are not fully functional AMT relays, but rather load balancers or brokers, as mentioned in Section 4.2.1.1 of [RFC7450].

If the D bit is set to 1, the gateway MAY send an AMT request message directly to the discovered relay address without first sending an AMT discovery message.

This bit should be set according to advice from the AMT relay operator. The D bit MUST be set to zero when no information is available from the AMT relay operator about its suitability.

3.2.3. RData Format - Type

The type field indicates the format of the information that is stored in the relay field.

The following values are defined:

- type = 0:
  The relay field is empty (0 bytes).

- type = 1:
  The relay field contains a 4-octet IPv4 address.

- type = 2:
  The relay field contains a 16-octet IPv6 address.

- type = 3:
  The relay field contains a wire-encoded domain name. The wire-encoded format is self-describing, so the length is implicit. The domain name MUST NOT be compressed. (See Section 3.3 of [RFC1035] and Section 4 of [RFC3597].)

3.2.4. RData Format - Relay

The relay field is the address or domain name of the AMT relay. It is formatted according to the type field.

When the type field is 0, the length of the relay field is 0, and it indicates that no AMT relay should be used for multicast traffic from this source.

When the type field is 1, the length of the relay field is 4 octets, and a 32-bit IPv4 address is present. This is an IPv4 address as described in Section 3.4.1 of [RFC1035]. This is a 32-bit number in network byte order.

When the type field is 2, the length of the relay field is 16 octets, and a 128-bit IPv6 address is present. This is an IPv6 address as described in Section 2.2 of [RFC3596]. This is a 128-bit number in network byte order.

When the type field is 3, the relay field is a normal wire-encoded domain name, as described in Section 3.3 of [RFC1035]. Compression MUST NOT be used, for the reasons given in Section 4 of [RFC3597].

3.3. AMTRELAY Record Presentation Format

3.3.1. Representation of AMTRELAY RRs

AMTRELAY RRs may appear in a zone data master file. The precedence, D-bit, relay type, and relay fields are REQUIRED.

If the relay type field is 0, the relay field MUST be ".".

The presentation for the record is as follows:

```
  IN AMTRELAY precedence D-bit type relay
```

3.3.2. Examples

For zone files in resolvers that don't support the value natively, it's possible as a transition path to use the format for unknown RR types, as described in [RFC3597].

```
  IN AMTRELAY 128 0 3 amtrelays.example.com.
```

or (see Appendix B):

```
  IN TYPE68  \# ( 24 ; length
           80 ; precedence
           83 ; D=1, relay type=3
           616d7472656c6179732e6578616d706c652e636f6d2e ) ; relay
```

As described in Section 2.2.2, a record for _amt._udp.home.arpa SHOULD also be present with a more preferred precedence:

```
  IN AMTRELAY 16 0 3 _amt._udp.home.arpa.
```

or (see Appendix B):

```
  IN TYPE68  \# ( 22 ; length
           10 ; precedence
           03 ; D=0, relay type=3
           5f616d742e5f7564702e686f6d652e617270612e ) ; relay
```

4. IANA Considerations

This document updates the IANA Registry for DNS Resource Record Types by assigning type 68 to the AMTRELAY record.

[To be removed (TBD):
Dear IANA, we request 68, since 68 is unassigned and easier to remember than other valid numbers, because the AMT UDP port number is 2268.

Registry URI:
https://www.iana.org/assignments/
dns-parameters/dns-parameters.xhtml#dns-parameters-4
]

This document creates a new IANA registry specific to the AMTRELAY for the relay type field.

Values 0, 1, 2, and 3 are defined in Section 3.2.3. Relay type numbers 4 through 255 can be assigned with a policy of Specification Required (see [RFC8126]).

[TBD: should the relay type registry try to combine with the gateway type from [RFC4025], Section 2.3 and 2.5? They are semantically very similar.
https://www.ietf.org/assignments/
ipseckey-rr-parameters/ipseckey-rr-parameters.xml
]

5. Security Considerations

[TBD: these 3 are just the first few most obvious issues, with just sketches of the problem. Explain better, and look for trickier issues.]

5.1. DNSSEC

If AMT is used to ingest multicast traffic, spoofing this record can enable spoofed multicast traffic.

Depending on service model, spoofing the relay may also be an attempt to steal services or induce extra charges.

5.2. Local Override

The local relays, while important for overall network performance, can't be secured by DNSSEC.

5.3. Congestion

Multicast traffic, particularly interdomain multicast traffic, carries some congestion risks, as described in Section 4 of [RFC8085]. Network operators are advised to take precautions including monitoring of application traffic behavior, traffic

authentication, and rate-limiting of multicast traffic, in order to ensure network health.

6. Acknowledgements

This specification was inspired by the previous work of Doug Nortz, Robert Sayko, David Segelstein, and Percy Tarapore, presented in the MBONED working group at IETF 93.

Thanks also to Jeff Goldsmith for his helpful review and feedback.

7. References

7.1. Normative References

[RFC1034] Mockapetris, P., "Domain names - concepts and facilities", STD 13, RFC 1034, DOI 10.17487/RFC1034, November 1987, <https://www.rfc-editor.org/info/rfc1034>.

[RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, RFC 1035, DOI 10.17487/RFC1035, November 1987, <https://www.rfc-editor.org/info/rfc1035>.

[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>.

[RFC2181] Elz, R. and R. Bush, "Clarifications to the DNS Specification", RFC 2181, DOI 10.17487/RFC2181, July 1997, <https://www.rfc-editor.org/info/rfc2181>.

[RFC3596] Thomson, S., Huitema, C., Ksinant, V., and M. Souissi, "DNS Extensions to Support IP Version 6", STD 88, RFC 3596, DOI 10.17487/RFC3596, October 2003, <https://www.rfc-editor.org/info/rfc3596>.

[RFC3597] Gustafsson, A., "Handling of Unknown DNS Resource Record (RR) Types", RFC 3597, DOI 10.17487/RFC3597, September 2003, <https://www.rfc-editor.org/info/rfc3597>.

[RFC4604] Holbrook, H., Cain, B., and B. Haberman, "Using Internet Group Management Protocol Version 3 (IGMPv3) and Multicast Listener Discovery Protocol Version 2 (MLDv2) for Source-Specific Multicast", RFC 4604, DOI 10.17487/RFC4604, August 2006, <https://www.rfc-editor.org/info/rfc4604>.

[RFC4607] Holbrook, H. and B. Cain, "Source-Specific Multicast for IP", RFC 4607, DOI 10.17487/RFC4607, August 2006, <https://www.rfc-editor.org/info/rfc4607>.

[RFC7450] Bumgardner, G., "Automatic Multicast Tunneling", RFC 7450, DOI 10.17487/RFC7450, February 2015, <https://www.rfc-editor.org/info/rfc7450>.

[RFC8085] Eggert, L., Fairhurst, G., and G. Shepherd, "UDP Usage Guidelines", BCP 145, RFC 8085, DOI 10.17487/RFC8085, March 2017, <https://www.rfc-editor.org/info/rfc8085>.

[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>.

7.2. Informative References

[RFC4025] Richardson, M., "A Method for Storing IPsec Keying Material in DNS", RFC 4025, DOI 10.17487/RFC4025, March 2005, <https://www.rfc-editor.org/info/rfc4025>.

[RFC4605] Fenner, B., He, H., Haberman, B., and H. Sandick, "Internet Group Management Protocol (IGMP) / Multicast Listener Discovery (MLD)-Based Multicast Forwarding ("IGMP/MLD Proxying")", RFC 4605, DOI 10.17487/RFC4605, August 2006, <https://www.rfc-editor.org/info/rfc4605>.

[RFC5507] IAB, Faltstrom, P., Ed., Austein, R., Ed., and P. Koch, Ed., "Design Choices When Expanding the DNS", RFC 5507, DOI 10.17487/RFC5507, April 2009, <https://www.rfc-editor.org/info/rfc5507>.

[RFC6895] Eastlake 3rd, D., "Domain Name System (DNS) IANA Considerations", BCP 42, RFC 6895, DOI 10.17487/RFC6895, April 2013, <https://www.rfc-editor.org/info/rfc6895>.

[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>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

[RFC8313] Tarapore, P., Ed., Sayko, R., Shepherd, G., Eckert, T., Ed., and R. Krishnan, "Use of Multicast across Inter-domain Peering Points", BCP 213, RFC 8313, DOI 10.17487/RFC8313, January 2018, <https://www.rfc-editor.org/info/rfc8313>.

Appendix A. Appendix A

This is the template for requesting a new RRType recommended in Appendix A of [RFC6895].

A. Submission Date:

B.1 Submission Type:
[x] New RRTYPE [ ] Modification to RRTYPE
B.2 Kind of RR:
[x] Data RR [ ] Meta-RR

C. Contact Information for submitter (will be publicly posted):
Name: Jake Holland
Email Address: jakeholland.net@gmail.com
International telephone number: +1-626-486-3706
Other contact handles: none

D. Motivation for the new RRTYPE application.
It provides a bootstrap so that AMT (RFC 7450) gateways can find the specific AMT relays that can receive multicast traffic from a known source, in order to signal multicast group membership and receive multicast traffic over a unicast tunnel using AMT.

E. Description of the proposed RR type.
This description can be provided in-line in the template, as an attachment, or with a publicly available URL.
https://datatracker.ietf.org/doc/draft-jholland-mboned-driad-amt-discovery

F. What existing RRTYPE or RRTYPEs come closest to filling that need and why are they unsatisfactory?
Some similar concepts appear in IPSECKEY, as described in Section 1.2 of [RFC4025]. The IPSECKEY RRType is unsatisfactory because it refers to IPSec Keys instead of to AMT relays, but the motivating considerations for using reverse IP and for providing a precedence are similar--an AMT gateway often has access to a source address for a multicast (S,G), but does not have access to a domain name or a good relay address, without administrative configuration.

Defining a format for a TXT record could serve the need for AMT

relay discovery semantics, but Section 5 of [RFC5507] provides a compelling argument for requesting a new RRType instead.

G. What mnemonic is requested for the new RRTYPE (optional)?
AMTRELAY

H. Does the requested RRTYPE make use of any existing IANA registry or require the creation of a new IANA subregistry in DNS Parameters?
No.

I. Does the proposal require/expect any changes in DNS servers/resolvers that prevent the new type from being processed as an unknown RRTYPE (see RFC3597)?
No.

J. Comments:
None.

Appendix B. Appendix B

In a DNS resolver that understands the AMTRELAY type, the zone file might contain this line:

```
IN AMTRELAY 128 0 3 amtrelays.example.com.
```

In order to translate this example to appear as an unknown RRType as defined in [RFC3597], one could run the following program:

<CODE BEGINS>

```
$ cat translate.py
#!/usr/bin/python3
import sys
name=sys.argv[1]
print(len(name))
print(''.join('%02x'%ord(x) for x in name))

$ ./translate.py amtrelays.example.com.
22
616d7472656c6179732e6578616d706c652e636f6d2e
```

<CODE ENDS>

The length and the hex string for the domain name "amtrelays.example.com" are the outputs of this program, yielding a length of 22 and the above hex string.

22 is the length of the domain name, so to this we add 2 (1 for the precedence field and 1 for the combined D-bit and relay type fields) to get the length of the unknown RData.

This results in a zone file line for an unknown resolver of:

```
  IN TYPE68  \# ( 24 ; length
          80 ; precedence
          03 ; relay type=domain
          616d7472656c6179732e6578616d706c652e636f6d2e ) ; relay
```

Author's Address

Jake Holland
Akamai Technologies, Inc.
150 Broadway
Cambridge, MA 02144
United States of America

Email: jakeholland.net@gmail.com