Tony Speakman/Cisco
Lorenzo Vicisano/Cisco
21 January 2003
Expires: July 2003

Reliable Multicast Transport Building Block
Generic Router Assist - Signalling Protocol Specification

Status of this Document

This document is a product of the IETF RMT WG. Comments should be addressed to the authors or to the WG's mailing list at rmt@ietf.org.

To the extent that it applies, this document is in conformance with the requirements of Section 2.1 of RFC3269.

Lexical Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC2119.

Abstract

This draft specifies the signalling protocol to be implemented in both end systems and network elements to activate built-in GRA functionality in network elements. It specifies this protocol specifically in the context of UDP as well as generally in the context of prospective (reliable multicast) transport protocols for IP.

Table of Contents

1. Introduction

Generic Router Assist (GRA), a transport-independent mechanism for providing transport protocols with access to network-element-based functionality, is comprised of several components including at least a signalling protocol at both the network and transport layers within which to encode GRA identifiers and operands; a specification and encoding of GRA functions within the network elements themselves to provide pre-defined GRA functions; a per-transport-session upstream-GRA-neighbour discovery mechanism; implosion control procedures to adapt to the effects of changing receiver populations and changes in multicast routing; and a control protocol for managing the disposition of network-element-based GRA functionality both within individual network elements and within individual transport sessions.

The utility and applicability of GRA are sufficiently wide that it makes sense first to define a base signalling protocol, particularly in the context of UDP, which will permit sufficient development to usefully narrow the requirements on GRA so that a general specification of GRA functionality and control can be undertaken. In the process, this signalling protocol can also be proven and evolved.

To that end, this draft specifies the GRA signalling protocol to be used in the context of UDP or in the context of a prospective (reliable multicast) transport for IP (an RMT) to activate built-in GRA functionality in network elements. For the moment, that built-in functionality will be specified in separate drafts, known as GRA function specifications, each of which will specify specific GRA headers and the network-element-based GRA services associated with those headers required to implement a given function. This draft specifies only the mechanisms by which GRA headers are delivered to built-in GRA functions.

2. Definitions

The definitions of the following terms for the purposes of this document are are provided in Section 11.

Distribution Tree	Direct Path	Suppression
Upstream	Reverse Path	Elimination
Downstream	Direct GRA Packet	Aggregation
GRA Packet	Reverse GRA Packets	Accumulation
GRA Hop		
GRA Neighbours		

3. Applicability

The model assumed here is that GRA may be implemented in some fraction of the network elements in a source-specific multicast distribution tree, that those GRA-capable network elements may discover their upstream GRA neighbours through transport-session-specific announcements flowing down the distribution tree, that there are pre-defined GRA functions in those network elements, and that the source and the receivers in the transport session direct GRA packets into the distribution tree. Those packets are detected by network elements and processed according to established GRA functions the specifications of which determine, amongst other things, the subsequent fate of the packet.

Each GRA function specification defines a related group of one or more GRA headers that may be borne by a subset of the session’s packets together with the GRA services to be applied to those headers to provide some feature to a session. A GRA function specification defines exactly one GRA service for each GRA header, and the services must be related to each other (typically through shared state) such that their joint action results in a single coherent function.

GRA packets are recognized as such during unicast or multicast forwarding, and are processed according to the appropriate GRA function specification.

The GRA header is parsed to identify the session to which the packet belongs and the applicable GRA function for that session, if any. Each GRA packet matches at most one GRA function and at most one GRA service in that GRA function; GRA packets that fail to match any GRA function at a network element are forwarded normally.

The processing defined by the GRA service associated with the GRA function is then carried out, using the values from the GRA header and the state associated with the GRA service. When GRA processing concludes successfully, the packet is either discarded or forwarded, as specified by the selected GRA service. There are two forwarding functions: multicast forwarding on a group of interfaces or unicast forwarding to a network-layer destination.

GRA includes a capability for GRA-capable network elements to return GRA packets from a receiver back to a source on the reverse of the path from the source to the receiver. Any GRA function that specifies the use of this capability must provide a per-transport-session upstream-GRA-neighbour discovery service that permits GRA-capable network elements to associate any given transport session identifier with an upstream GRA neighbour.

Note that GRA functions can only be applied within a single transport session. The coordination of GRA functions across multiple transport sessions is not accommodated here.

The impact of GRA on forwarding path bandwidth will be related directly to the the number of sessions using GRA, to the fraction of GRA packets in each session, and to the state and processing complexity defined in GRA function specifications. The constraints specified in this document on GRA usage in general and on GRA function specifications in particular are intended to mandate the lowest possible upper bounds on these parameters while still providing a useful level of network-element-based functionality.

4. Rationale

The justification for providing GRA functionality lies in the observation that , at least for IP multicast, a distribution tree represents a distributed repository of information about the corresponding multicast session as a whole (such as loss, delay, topology, and membership information) which, when subjected to distributed processing in the network itself, may yield results which can be used by sources, receivers, and network elements themselves to significantly enrich both the efficiency and the intelligence of the operation of the session.

The definition of the GRA signalling protocol specified here is sufficiently general to be applied to any prospective UDP-based application protocol or any RMT for IP that is designed to make use of GRA. Such a protocol may simply insert a GRA header in the transport header of some subset of packets in a session for those packets to be processed by network-element-based GRA functions. For RMTs, GRA functionality is defined entirely in terms of the network layer header and the GRA header and so is completely transport-layer independent. For UDP applications, GRA functionality is defined entirely in terms of the network layer header, the UDP header, and the GRA header and so is completely application-layer independent. Network-element-based GRA functionality can thus be applied with identical semantics across different UDP applications or RMTs.

Given its network-layer, hop-by-hop mechanics, the definition of GRA signalling specified here does not overlap with functionality provided by any other reliable multicast building block.

5. Functionality

The operation of the GRA signalling protocol is specified here in terms of the format of GRA headers and procedures for processing those headers in network elements.

In the context of this document, the direct path is the path taken by a packet from a source to a receiver as determined by IP routing. The reverse path is the path taken by a packet from a receiver to a source as it is forwarded through the sequence of upstream GRA neighbours between the receiver and the source. Direct GRA packets are GRA packets being forwarded by IP routing on the direct path. Reverse GRA packets are GRA packets being forwarded GRA-hop-by-GRA-hop on the reverse path.

5.1. Headers

The GRA header is best conceived of as augmenting the transport header and must be available at any layer in the communications stack at which the transport header (the application header in the case of UDP) is also available.

GRA packets intended for interpretation in network elements must bear a **network layer indication** that the GRA header is present. A new IP option will be created specifically to signal the presence of the GRA header at the network layer.

GRA headers may be used in packets without this network layer indication for the purposes of purely end-to-end GRA-related exchanges within a session.

GRA packets must also bear a **transport layer indication** that the GRA header is present since IP options are not typically available above the network layer. The method for detecting the presence of the GRA header at the transport layer is based upon the G-bit (defined below) being set in the common portion of the GRA header which, when present, must immediately follow the network header.

> **NOTA BENE:** A UDP application or an RMT must specify its own application or transport headers, respectively, such that the value of the bit in this same location in their own headers is always zero.

For UDP, direct GRA packets can be associated with a globally unique transport session based upon the IP address information and the UDP port information. Reverse GRA packets must also be associated with the globally unique transport session to which they refer, and while, at least for UDP, port information is typically available in the transport header, the required IP address information is not available from the IP header since it will be addressed to a GRA neighbour.

Consequently, for reverse GRA packets, this information must be recorded inside the GRA headers of reverse GRA packets to provide unambiguous **transport session identification**. Since these IP address fields will only be present in reverse GRA packets, a **direct/reverse**

discriminator has been provided in the common portion of the GRA header. The same transport session identification scheme applies to RMTs with the difference that an NLA-scoped host session identifier is carried directly in the GRA header in place of "borrowing" port information from the transport.

Note also that in reverse packets, the destination address is further required in case the reverse GRA packet is to be subject to multicast forwarding as part of the corresponding GRA function definition such as in the case of a suppression service.

The operand field may be one of two types: it may consist of a fixed number of fixed-format operands, or a variable number of variable-format operands. While variable operands may provide a degree of feature flexibility, the overhead associated with parsing variable operand lists may tax the time constraints on forwarding-time processing of GRA, so a **fixed-operand/variable-operand discriminator** has been provided in the common portion of the GRA header so that network elements can efficiently detect more complex GRA headers and make an early determination as to how to handle them.

5.1.1. Header Format in the Context of UDP

A GRA header in the context of UDP must be formatted as follows:

```
IP HEADER including the IP GRA OPTION

UDP HEADER
.................................................................
.          Source Port           .      Destination Port         .
.................................................................
.           Check Sum            .         TPDU Length           .
.................................................................

The common portion of the GRA header:
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|G|R|V|-|-|-|V N| Header Length |  Function ID   | Instance #    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

... followed immediately, in reverse GRA packets only,
    by transport-session-identifying IP address information:
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-...-+
|            Network Layer Source (IP) Address              ... |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-...-+
|    Network Layer Destination (IP Multicast Group) Address ... |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-...-+

... followed immediately by the GRA-function-specific operand portion
    of the GRA header:
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-...-+
|                         Operands                          ... |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-...-+

... followed immediately by the APDU.
```

Figure 1 - GRA Header Format in the Context of UDP

5.1.2. Header Format in the Context of an RMT

A GRA header in the context of an RMT must be formatted as follows:

```
IP HEADER including the IP GRA OPTION

The common portion of the GRA header:
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            NLA-scoped host session identifier(s)              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|G|R|V|-|-|-|V N| Header Length |  Function ID  | Instance #    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

... followed immediately, in reverse GRA packets only,
    by transport-session-identifying IP address information:
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-...-+
|            Network Layer Source (IP) Address              ... |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-...-+
|    Network Layer Destination (IP Multicast Group) Address ... |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-...-+

... followed immediately by the GRA-function-specific operand portion
    of the GRA header:
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-...-+
|                         Operands                          ... |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-...-+

... followed immediately by the TPDU.
```

Figure 2 - GRA Header Format in the Context of an RMT

5.1.3. Header Contents

`NLA-Scoped Host Session Identifier (HSI)`
: For RMTs, the logical equivalent of the UDP source and destination ports.

`G-bit`
: MUST be set. Signals the presence of a prepended **G**RA header to UDP applications expecting APDUs after the UDP header or to RMTs expecting TPDUs after the IP header.

`R-bit`
: Clear on direct GRA packets, set on **R**everse GRA packets.

`V-bit`
: Clear on fixed-operand GRA packet, set on **V**ariable-operand GRA packets.

`VN-bits`
: GRA **V**ersion **N**umber, 0x0 in this instance.

`Header Length`
: Total length of the GRA header in longwords.

`Function ID (GFID)`

The GFID is the name of the transport-session-specific instance of the GRA function to be used to process the GRA header.

`GFID Instance Number (GFIN)`

The GFIN is the session-specific instance number of the GRA function to be used to process the GRA header.

`Source IP Address`

`Destination IP (Multicast Group) Address`

Reverse GRA packets must bear the network-layer addresses of the source and destination for the identification of the transport session within which the GRA packets are to be processed.

Note that while these addresses are rendered here as IPv4 addresses, their actual format in the context of UDP or a given RMT is determined by the network layer protocol in which UDP or the RMT is operating.

`Operands`

The operands (including a GSID; see below) as defined for one of the GRA services in the GRA function specification for the given GFID.

5.2. Procedures

5.2.1. Network Elements

Network elements that explicitly detect the IP GRA option (TBD) in a packet must then locate and parse the GRA header first to determine the packet's admissibility and second to determine its handling. The GRA header is always located directly following the network header, but network elements must make a transport-protocol-specific determination of its format which will be as specified in Figure 1 for UDP and as specified in Figure 2 for all other transport protocols.

> **NOTA BENE:** There's a strong assumption here that network elements do IP option detection BEFORE they make any determinations based on the destination network layer address in the general forwarding path.

If the G-bit is not set, the network element must discard the packet.

If the packet is a reverse GRA unicast packet, the network element must determine whether the packet bears a destination address addressing the network element itself. If a reverse GRA unicast packet is not addressed to the network element itself, it must continue to be processed just as if the IP GRA option had NOT been detected.

If a reverse GRA unicast packet is addressed to the network element itself, the network element must first correct for potential unicast routing asymmetries. In particular, since a reverse GRA unicast packet may traverse a non-contiguous GRA hop and therefore arrive at the upstream GRA neighbour on other than the interface to which it was addressed, the packet must, upon receipt, be reassociated by the network element with the interface to which it was addressed. Note that the destination address check is essential since reverse packets are seen not just by the GRA neighbour to which they are unicast but by any intervening GRA capable routers as well (due to the IP GRA option).

Network elements must then verify that they have corresponding definitions for the version number and the GFID. Packets that fail this verification must continue to be processed just as if the IP GRA option had NOT been detected.

Admissibility is further determined by matching the source and destination addresses (found in the IP header in the case of direct GRA packets and inside the GRA header in the case of reverse GRA packets), and the GFID (and possibly the incoming interface) against any local access lists protecting access to GRA functionality. Packets blocked by such lists must continue to be processed just as if the IP GRA option had NOT been detected.

For packets passed by such lists, the GRA function corresponding to the GFID must be invoked and passed the packet itself as well as the identity of the incoming interface.

The network element may make local determinations on the processing priority of such GRA packets based on any local conditions or any attributes of the GRA packet itself. More specifically, a network element is not constrained to order GRA packets relative to non-GRA packets in any way. Network elements must, however, process admissible GRA packets matching a particular (transport-session-scoped) GFID in relative order.

All further processing of a GRA packet is GRA-function-specific as defined by the GRA function specification for the GFID. In particular, it is incumbent upon the GRA function to determine the ultimate fate of the packet itself.

Network elements must provide a multicast forwarding service which permits selective forwarding on a specified subset of the interfaces on a multicast route.

5.2.2. End Systems

End systems are defined for the purposes of this section as either host-resident UDP applications or RMT stacks.

End systems originating GRA packets must format them as specified in Figure 1 for UDP applications and as specified in Figure 2 for all other transport protocols. The underlying network layers of these end systems must provide the ability to selectively add the IP GRA option to GRA packets submitted for transmission by these end systems.

Upon receipt, end systems must explicitly detect the presence of a GRA header in a packet based on the state of the G-bit and must then locate and parse the GRA header first to determine the packet's admissibility and second to determine its handling. Note that the location of the G-bit depends on whether the end system is a UDP application or an RMT.

End systems must then verify that they have corresponding definitions for the version number and the GFID. Packets that fail this verification must continue to be processed ignoring the GRA header altogether.

All further processing of a GRA packet is GRA-function-specific as defined by the GRA function specification for the GFID.

6. Constraints on GRA Function Specifications

General constraints on GRA function specifications are described here since those constraints stem primarily from consideration of the processing environment in the forwarding path of a network element. Individual GRA function specifications must adhere to the constraints stated here, and these constraints may be expanded in light of any irrational exuberance detected in GRA function specifications themselves.

Static and dynamic GRA functions
The lower half of the GFID space is reserved for static, well-known GRA function specifications. The upper half of the GFID space is reserved for dynamic, proprietary GRA function specifications when it becomes possible to define these through a GRA control protocol. Thus the upper half of the GFID space is scoped by the transport session.

Maximum number of GRA services
No more than 8 GRA services may be defined for a single GRA function. The definition of the GRA service identifier (GSID) is a matter purely local to the GRA function specification except to say that GSIDs must be encoded in the operand portion of the GRA header.

GRA neighbour discovery service
If a GRA function specifies a GRA service for reverse packets, it must also specify a GRA service for GRA neighbour discovery in the session. Such a GRA service may also be used as a general session information service to provide any other session-specific parameters required by the GRA function.

Packet Access
The GRA header MUST be self-contained. That is, all parameters required by a given GRA service MUST be provided in the GRA header itself other than those provided in the network header (the UDP header in the case of UDP applications). More specifically, there is no capability defined for GRA to specify operands by offset into the packet. This may necessitate duplicating information carried elsewhere in the packet. Note that when a GRA header is present, an RMT is free to refer to the GRA header for values that normally would be carried elsewhere in the transport header (**e.g.** the HSI).

Packet Modification
A GRA packet may not be modified in any way by a network element other than to write to the GRA header and then only in the same format in which the header was received. GRA specifies no packet encapsulation or packet decapsulation capabilities. In addition, this constraint precludes packet accumulation.

Packet Generation
GRA Function specifications may specify the generation and forwarding of a packet consisting only of a GRA header defined by the GRA function specification and derived from the copy of a received GRA header. These packets may be used for delayed or periodic GRA-related events in the session.

Fixed and Variable Operands
An implementation of GRA must support GRA headers with either a fixed number of operands in fixed formats, or a variable number of operands in variable formats. Fixed format GRA headers

should be used in GRA service specifications where forwarding-time processing performance is the priority. Variable format GRA headers should be used in GRA service specifications where syntactic flexibility is the priority. Variable format GRA headers must be self-describing to the extent that the type, length, and value of all operands can be determined dynamically. While variable-format operands may vary across GRA packets, they may not vary from the format established for a given packet by its originator.

Operands Semantics
While operands may represent a whole range of meaningful (to the transport) attributes, they are not interpreted by GRA in other than the context of the operators of each type supported by GRA. That is, the numerical and relational operators treat operands as unsigned integers, and the logical operators treat operands as bit fields (**e.g.** for the purposes of aggregation).

Designated KEY Operand
A GRA function specification may define a single distinguished operand, up to 32 bits in length, to be a key for the purposes of indexing sub-service-specific state.

GRA Function State
A GRA function specification may specify up to 8 GRA-function-local variables, 8 GRA-function-local timers, and 1 interface list (where an interface list may record no more interfaces than are recorded on the multicast route corresponding to the session).

GRA Service State
For each GRA service, a GRA function specification may specify up to 8 GRA-service-local variables, 8 GRA-service-local timers, storage for 1 full copy of the GRA header defined by the GRA service, and 1 GRA-service-local interface list. GRA header copies are intended for such functions as delayed forwarding and elimination.

Keyed GRA Service State
For each instance of a key, a GRA service specification may specify up to 8 sub-service-local variables, 8 sub-service-local timers, storage for 1 full copy of the GRA header defined by the GRA service, and 1 sub-service-local interface list.

Forwarding Functions
A GRA service specification may specify unicast and or multicast forwarding for a received GRA packet or for a GRA packet generated by the GRA service and consisting only of the GRA header defined by the GRA service. Unicast forwarding may be to an arbitrary unicast destination. Multicast forwarding may only be to the multicast group corresponding to the multicast distribution tree for the session possibly augmented by an interface list for the purposes of selective forwarding.

Bandwidth
GRA is not intended for use on data packets involved in a transport's basic data conveyance function. Rather, it is intended for control and exception processing within a session. Neither the frequency of GRA packets nor their total data rate within a session may be more than 5% in the steady state.

7. Security Considerations

Until a GRA control protocol can be defined to securely configure and manage access to GRA functionality in network elements, network elements must restrict access to GRA functionality through positive access lists explicitly configured to permit access by the source/destination/GFID triplet (and possibly by interface as well) where these addresses are found in the IP header in the case of direct GRA packets and inside the GRA header in the case of reverse GRA packets.

When and where available, GRA will rely on general mechanisms for receiver access control and for source and receiver authentication rather than mandating its own.

8. Requirements from other Building Blocks

The only aspect of GRA functionality not completely defined by GRA function specifications is the number of separate instances of the application of a given GRA function within a session as represented by the GFIN. The expectation is that only one instance is typically required. In any case, a UDP application or an RMT must not specify the use of more than 8 instances of a given GFID.

9. Codepoint Considerations

(This section is in compliance with RFC3269)

10. IANA Considerations

As part of defining the new IP GRA option, an new IP option type will be requested from IANA.

It is further proposed here that the GFID name space be under IANA administration.

11. Definitions

The definitions in this section apply throughout this document.

Distribution Tree
The multicast distribution tree of network elements defined by network-layer multicast routing information for a given multicast transport session rooted at a single root network element (at the host that is the source of the data in the transport session), and fanning out (possibly through transit network elements) to one or more leaf network elements (at the hosts that are the receivers of the data in the transport session).

Downstream
In the direction away from the source and toward receivers.

Upstream
In the direction away from receivers and toward the source.

GRA packet
A packet bearing a GRA header.

GRA hop
The single logical hop between any two GRA-capable network elements adjacent (in a strictly upstream/downstream sense) to each other in the distribution tree (*i.e.*, not separated in the distribution tree by any other GRA-capable network element).

GRA Neighbours
Two GRA-capable network elements separated by a single GRA hop.

Direct Path
The path taken by a packet from a source to a receiver as determined by IP routing

Reverse Path
The path taken by a packet from a receiver to a source as it is forwarded through the sequence of upstream GRA neighbours between the receiver and the source.

Direct GRA Packets
GRA packets being forwarded by IP routing on the direct path.

Reverse GRA Packets
GRA packets being forwarded GRA-hop-by-GRA-hop on the reverse path.

Suppression
Potential transmitters of some packet refrain from the transmission of that packet because they detect its duplicate.

Elimination
A potential forwarder of some packet refrains from forwarding that packet because it has a record of already having forwarded a duplicate of that packet.

Aggregation
Multiple identically sized operands are combined in a bit-wise fashion that results in a compositely encoded operand of the same size.

Accumulation
Multiple operands are concatenated together to form an operand list whose length is greater than length of any of the constituent operands.

12. Expired Drafts

Two expired Internet Drafts, one an architecture spec of sorts, the other a functional spec of sorts, may be of interest to the reader and can be found at:

http://www.ietf.org/proceedings/01aug/I-D/draft-ietf-rmt-gra-arch-02.txt
http://www.ietf.org/proceedings/01aug/I-D/draft-ietf-rmt-gra-fspec-00.txt

A set of slides elaborating on the functional spec can be found at:
http://www.ietf.org/proceedings/01aug/slides/rmt-2.pdf

Authors' Addresses

Tony Speakman
speakman@cisco.com

Lorenzo Vicisano
lorenzo@cisco.com