Network Working Group
Request for Comments: 1582
Category: Standards Track

G. Meyer
Spider Systems

February 1994

Extensions to RIP to Support Demand Circuits

Status of this Memo

Abstract

Running routing protocols on connection oriented Public Data Networks, for example X.25 packet switched networks or ISDN, can be expensive if the standard form of periodic broadcasting of routing information is adhered to. The high cost arises because a connection has to all practical intents and purposes be kept open to every destination to which routing information is to be sent, whether or not it is being used to carry user data.

Routing information may also fail to be propagated if the number of destinations to which the routing information is to be sent exceeds the number of channels available to the router on the Wide Area Network (WAN).

This memo defines a generalized modification which can be applied to Bellman-Ford (or distance vector) algorithm information broadcasting protocols, for example IP RIP, Netware RIP or Netware SAP, which overcomes the limitations of the traditional methods described above.

The routing protocols support a purely triggered update mechanism on demand circuits on WANs. The protocols run UNMODIFIED on LANs or fixed point-to-point links.

Routing information is sent on the WAN when the routing database is modified by new routing information received from another interface. When this happens a (triggered) update is sent to a list of destinations on other WAN interfaces. Because routing protocols are datagram based they are not guaranteed to be received by the peer router on the WAN. An acknowledgement and retransmission mechanism is provided to ensure that routing updates are received.

The WAN circuit manager advises the routing applications on the reachability and non-reachability of destinations on the WAN.

Acknowledgements

I would like to thank colleagues at Spider, in particular Richard Edmonstone, Tom Daniel and Alam Turland, Yakov Rekhter (IBM), Martha Steenstrup (BBN), and members of the RIP-2 working group of the IETF for stimulating discussions and comments which helped to clarify this memo.

Conventions

The following language conventions are used in the items of specification in this document:

- MUST -- the item is an absolute requirement of the specification. MUST is only used where it is actually required for interoperation, not to try to impose a particular method on implementors where not required for interoperability.

- SHOULD -- the item should be followed for all but exceptional circumstances.

- MAY or optional -- the item is truly optional and may be followed or ignored according to the needs of the implementor.

The words "should" and "may" are also used, in lower case, in their more ordinary senses.

Table of Contents

1. Introduction

Routers are used on connection oriented networks, such as X.25 packet switched networks and ISDN networks, to allow potential connectivity to a large number of remote destinations. Circuits on the Wide Area Network (WAN) are established on demand and are relinquished when the traffic subsides. Depending on the application, the connection between any two sites for user data might actually be short and relatively infrequent.

Practical experience of routing shows that periodic 'broadcasting' of routing updates on the WAN is unsatisfactory on several counts:

- Running a routing protocol like RIP is expensive if the standard form of transmitting routing information to every next hop router every 30 seconds is adhered to. The more routers there are wishing to exchange information the worse the problem. If there are N routers on the WAN, N * (N - 1) routing updates are sent over N * (N - 1)/2 connections every broadcast period.

  The expense arises because a circuit has to be kept open to each destination to which routing information is to be sent. Routing updates are sufficiently frequent that little benefit is obtainable on most networks by attempting to set up a call purely for the duration of the routing update. (There are often minimum call charges, or there is a charge to set up a call irrespective of what data is sent.)

  The option of reducing the 'broadcast' frequency, while reducing the cost, would make the system less responsive.

- The number of networks to be connected (N) on the WAN can easily exceed the number of simultaneous calls (M) which the interface can support. If this happens the routing 'broadcast' will only reach M next hop routers, and (N - M) next hop routers will not receive the routing update.

  A basic rate ISDN interface can support 2 simultaneous calls, and even the number of logical channels most users subscribe to on an X.25 network is not large. There is no fundamental reason why

routing protocols should restrict the user to routing between so few sites.

- Since there is no broadcast facility on the WAN, 'broadcasting' of routing information actually consists of sending the updates separately to all known locations. This means that N routing updates are queued for transmission on the WAN link (in addition to any data which might be queued).

  Routers take a pragmatic view on queuing datagrams for the WAN. If the queue length gets too long, so that datagrams at the end of the queue would take too long be transmitted in a reasonable time (say 1 to 2 seconds) the router may start discarding them. On an X.25 network, with slow line speeds (perhaps 9600 baud), it may not take too many routing updates to fulfill this condition if the link is also actively being used to carry user data.

This memo addresses all the above problems.

The approach taken is to modify the routing protocols so as to send information on the WAN only when there has been an update to the routing database OR a change in the reachability of a next hop router is indicated by the task which manages connections on the WAN.

Because datagrams are not guaranteed to get through on all WAN media, an acknowledgement and retransmission system is required to provide reliability.

This memo describes the modifications required for Bellman-Ford (or distance vector) algorithm information broadcasting protocols, such as IP RIP [1,2] or Netware RIP and SAP [3] on the WAN. The protocols run unmodified on Local Area Networks (LANs) or fixed point-to-point links, and so interoperate transparently with implementations adhering to the original specifications.

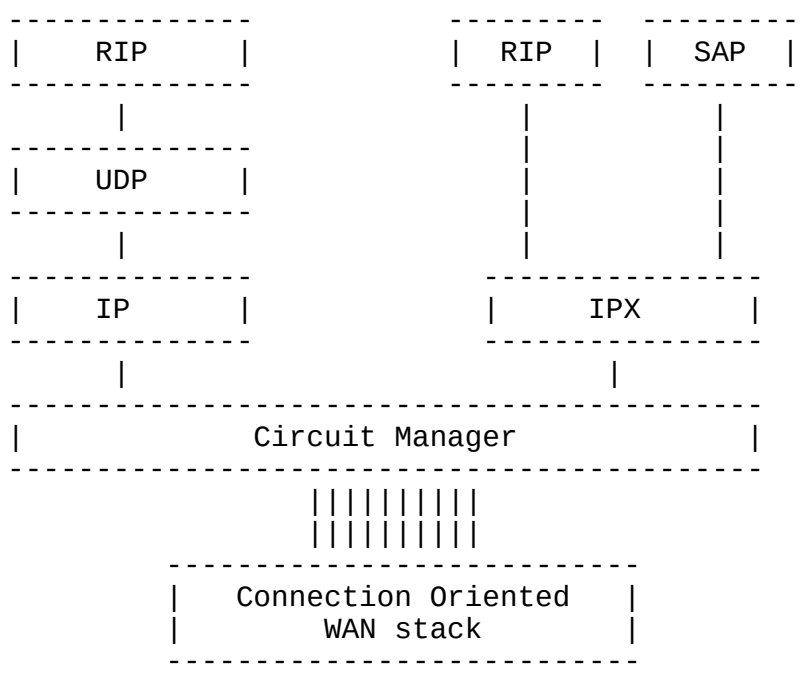
2. Running Routing Protocols on the WAN

2.1 Overview

Multiprotocol routers are used on connection oriented Wide Area Networks (WANs), such as X.25 packet switched networks and ISDN networks, to interconnect LANs. By using the multiplexing properties of the underlying WAN technology, several LANs can be interconnected simultaneously through a single physical interface on the router.

A circuit manager provides an interface between the connectionless network layers (IP, IPX, CLNP etc) and the connection oriented WAN (X.25 or ISDN). Figure 1 shows a schematic representative stack

showing the relationship between routing protocols, the network layers, the circuit manager and the connection oriented WAN.



A WAN circuit manager will support a variety of network layer protocols, on its upper interface. On its lower interface, it may support one or more subnetworks. A subnetwork may support a number of Virtual Circuits.

Figure 1. Representative Multiprotocol Router stack

The router has a translation table which relates the network layer address of the next hop router to the physical address used to establish a Virtual Circuit (VC) to it. Datagrams may be encapsulated in a header to distinguish the network layer protocol [5].

The circuit manager takes datagrams from the connectionless network layer protocols and (if one is not currently available) opens a VC to the next hop router. A VC can carry all traffic between two end-point routers for a given network layer protocol (or with appropriate encapsulation all network layer protocols). An idle timer is used to close the VC when the datagrams stop arriving at the circuit manager.

Running routing protocols on the WAN has traditionally consisted of making small modifications to the methods used on LANs. Where

routing information would be broadcast periodically on a LAN interface, it is converted to a series of periodic updates sent to a list of addresses on the WAN.

This memo targets two areas:

- Eliminating the overkill inherent in periodic transmission of routing updates.

- Overcoming the bandwidth limitations on the WAN: the number of simultaneous VCs to next hop routers and restricted data throughput which the WAN link can support.

The first of these is overcome by transmitting routing updates (called routing responses) only when required:

- Firstly, when a specific request for a routing update has been received.

- Secondly, when the routing database is modified by new information from another interface.

  Update information received in this way is not normally propagated on other interfaces immediately, but is delayed for a few seconds to allow information from several updates to be grouped.

- Thirdly, when the circuit manager indicates that a destination has changed from an unreachable (circuit down) to a reachable (circuit up) state.

Because of the inherent unreliability of a datagram based system, both routing requests and routing responses require acknowledgement, and retransmission in the event of NOT receiving an acknowledgement.

To overcome the bandwidth limitations the routing application can perform a form of self-imposed flow control, to spread routing updates out over a period of time.

2.2 Presumption of Reachability

If a routing update is received from a next hop router on the WAN, entries in the update are thereafter always considered to be reachable, unless proven otherwise:

- If in the normal course of routing datagrams, the circuit manager fails to establish a connection to the next hop router, it notifies the routing application that the next hop router is not

reachable through an internal circuit down message.

The routing application then goes through a process of timing out database entries to make them unreachable in the routing sense.

- o If the circuit manager is subsequently able to establish a connec tion to the next hop router, it will notify the routing applica tion that the next hop router is reachable through an internal circuit up message.

  The routing application will then exchange messages with the next hop router so as to re-prime their respective routing databases with up-to-date information.

Handling of circuit up and circuit down messages requires that the circuit manager takes responsibility for establishing (or reestablishing) the connection in the event of a next hop router becoming unreachable. A description of the processes the circuit manager adopts to perform this task is outside the scope of this memo.

2.3 WAN Router list

The routing task MAY be provided with a list of routers to send routing updates to on the WAN. It will comprise of the logical addresses of next hop routers for which the router has a logical to physical address mapping. Entries in the list SHOULD be categorized (on a per-peer basis) as follows:

- o Running the standard routing protocol, namely transmitting updates periodically with the packet formats used in LAN broadcasts.

  This option is supported to allow interoperability with existing routing implementations, and might also be appropriate if some of the destinations are using Permanent Virtual Circuits (PVCs) rather than SVCs.

- o Running the triggered update routing protocol proposed in this memo.

Omitting an address from both of these categories is equivalent to not running the routing protocols.

If routing packets arrive from a destination not supporting the appropriate variant they MUST be discarded.

2.4 Triggered Updates and Unreliable Delivery

If triggered update information is sent to next hop routers on the WAN only once it can fail to arrive for one of the following reasons:

- A free VC resource might not be available, because of a restricted number of X.25 logical channels or ISDN B-channels.
- The transmit queue might be full - requiring the datagram to be discarded.
- The VC might be pre-empted (in favour of establishing a VC to another next hop router) while the datagram is in a queue, resulting in the queue being flushed and the datagram discarded.
- In cases where the method of transport is not guaranteed, for example with PPP where there is no acknowledgement and retransmission of HDLC frames, a corrupted frame will result in the loss of the datagram.

2.5 Guaranteeing delivery of Routing Updates

To guarantee delivery of routing updates on the WAN an acknowledgement and retransmission scheme MUST be used:

- Send a routing update to a next hop router on the WAN.
- The other router responds with an acknowledgement packet.

  The original router receives the acknowledgement.
- Otherwise the original router retransmits the update until an acknowledgement is received.

  Retransmission timer values are covered in section 7.

  In cases where the routing database is modified before an acknowledgement is received a new routing update with an updated sequence number is sent out. If an acknowledgement for the old routing update is received it is ignored.
- A router only updates its routing database when it receives a complete update, which may consist of several fragments. Each fragment is individually acknowledged.

The above mechanism caters for cases where the datagram is lost because of a frame error or is discarded because of an over-full

queue. The routing update and acknowledgement will eventually both get through.

In cases where the circuit manager cannot establish a connection, a mechanism is provided to allow the circuit manager to inform the routing task of the failure to make a connection so that it can suppress retransmissions until a circuit becomes available.

2.6 The Routing Database

A requirement of using triggered updates for propagating routing information is that NO routing information ever gets LOST or DISCARDED.

The routing database MUST adopt one of the following strategies:

- It must keep ALL alternative routing information it learns from any routing updates from the LAN and the WAN, so that if the best route disappears an alternative route (if available) can replace it as the new best route.

- If the amount of memory this consumes is problematic the routing application must keep SOME alternative routing information - say a best route and two alternatives.

  If the router ever has to discard routing information about a route it should note the fact. If the routes that have been kept disappear because they have become unreachable, the router MUST issue a request on all interfaces to try and obtain discarded alternatives.

  It is recommended that the request is issued BEFORE all routes to a destination have been lost.

Entries in the routing database can either be permanent or temporary. Entries learned from broadcasts on LANs are temporary. They will expire if not periodically refreshed by further broadcasts.

Entries learned from a triggered response on the WAN are ’permanent’. They MUST not time out in the normal course of events. The entries state MUST be changed to ’temporary’ by the following events:

- The arrival of a routing update containing the entry set to unreachable.

  The normal hold down timer MUST be started, after which the entry disappears from the routing database.

- The arrival of a routing update with the entry absent.

  If the hold down timer is not already running, the entry MUST be set to unreachable and the hold down timer started.

- A message sent from the circuit manager, to indicate that it failed to make a connection in normal running.

  The routing table MUST be scanned for all routes via that next hop router. Aging of these routing entries MUST commence. If the aging timer expires the entry MUST be set to unreachable and the hold down timer started. If the hold down timer expires the entry disappears from the routing database.

- If the interface goes down, the circuit manager should indicate that all circuits on that interface have gone down.

Database timer values are covered in section 7.

2.7 New Packet Types

To support triggered updates, three new packet types MUST be supported:

TRIGGERED REQUEST

A request to the responding system to send all appropriate elements in its routing database.

A triggered request is retransmitted at periodic intervals until a triggered response is received.

Routing requests are transmitted in the following circumstances:

- Firstly when the router is powered on.
- Secondly when the circuit manager indicates a destination has been in an unreachable (circuit down) state for an extended period and changes to a reachable (circuit up) state.
- Thirdly in the event of all routing update fragments failing to arrive within a set period.
- It may also send triggered requests at other times to compensate for discarding non-optimal routing information.

TRIGGERED RESPONSE

A message containing all appropriate elements of the routing database. An appropriate element is an entry NOT learned from the interface to which the routing information is being sent out. This is known as "split horizon".

Stability is improved by adding "poisoned reverse" on routes learned from a destination. This consists of also including some routes learned from a destination in routing updates sent back to that destination, but setting the routes as unreachable. A route is only poisoned if it is the best route (rather than an inferior alternative route) in the database.

A triggered response message may be sent in response to a triggered request, or it may be an update message issued because of a change in the routing database.

A triggered response message MUST be sent in response to a triggered request message even if there are no routes to propagate. This would be the case for a host which had a WAN interface only, but which wished to run the triggered update protocol.

A triggered response is retransmitted at periodic intervals until a triggered acknowledgement is received.

TRIGGERED ACKNOWLEDGEMENT

A message sent in response to every triggered response packet received.

Triggered response and triggered acknowledgement packets MUST contain additional fields for a sequence number, fragment number and number of fragments.

If a triggered request or response is not acknowledged after 10 retransmissions, routes to the destination should be marked as unreachable for the duration of a hold down timer before being deleted.

The destination should then be polled at a lower frequency using triggered request packets. When a triggered response is received, the router should prime the next hop router my sending its routing database through triggered response packets.

Strictly speaking polling should occur indefinitely to guarantee database integrity. However the administrator MAY wish the router to cease polling after a few attempts believing that the lack of response is due to a mis-configuration of the next hop router. The destination should be marked as NOT supporting the mechanism and no further routing messages should be sent to that destination.

Before marking the destination as not supporting the mechanism, at least 5 triggered request polls (without acknowledgement) should be sent.

If a destination marked as not supporting the mechanism, subsequently sends a valid 'triggered' message, the destination should be marked as supporting the mechanism once more (to allow for the next hop router's configuration being changed). It should be sent a triggered request and a triggered response to obtain and propagate up-to-date routing information.

2.8 Fragmentation

If a routing update is sufficiently large, the information MUST be fragmented over several triggered response packets:

- Each fragment MUST be individually acknowledged with a triggered acknowledgement packet.

  The sender of the routing update MUST periodically retransmit fragments which have not been acknowledged (or until the destination is marked as not supporting the mechanism).

- A router receiving fragments MUST re-assemble them before updating its routing database.

- If all fragments are not received within four times the retransmit period, they MUST be discarded.

  A triggered request packet MUST then be sent to the originator of the routing update.

  On receiving the triggered request packet, the originator of the routing update MUST retransmit ALL fragments.

- If a fragment with an updated sequence number is received, ALL fragments with the earlier sequence number MUST be discarded.

  An updated sequence number is defined as any sequence number that is different. There is no concept of the value of the sequence number conveying its age.

Fragmentation timer values are covered in section 7.

2.9 Preventing Queue Overload

In order to prevent too many routing messages being queued at a WAN interface, the routing task MAY operate a scheme whereby 'broadcasting' of a triggered request or triggered response to a WAN interface is staggered. All routing requests or routing responses are not sent to ALL next hop routers on the interface in a single batch:

- The routing task should limit the number of outstanding triggered request messages for which a triggered response has not been received.

- The routing task should limit the number of outstanding triggered response messages for which a triggered acknowledgement has not been received.

As outstanding messages are appropriately acknowledged, further messages can be sent out to other next hop routers, until all next hop routers have been sent the message and have acknowledged it.

The maximum number of outstanding messages transmitted without acknowledgement is a function of the link speed and the number of other routing protocols operating the triggered update mechanism.

Messages should always be acknowledged immediately (even if it causes the limit to be exceeded), since a connection is almost certainly available. This has the potential benefit of allowing the VC to close sooner (on its idle timer).

Sending all triggered request fragments to a destination at once is also beneficial.

3. IP Routing Information Protocol Version 1

This section should be read in conjunction with reference [1].

IP RIP is a UDP-based protocol which generally sends and receives datagrams on UDP port number 520.

To support the mechanism outlined in this proposal the packet format for RIP version 1 [1] is modified as shown in Figure 2.

Every Routing Information Protocol datagram contains the following:

COMMAND Commands supported in RIP Version 1 are: request (1), response (2), traceon (3), traceoff (4), SUN reserved (5). The fields sequence number, fragment number and number of fragments MUST NOT be included in packets with these command values.

The following new commands (with values in brackets) are required:

TRIGGERED REQUEST (6)

A request for the responding system to send all of its routing database.

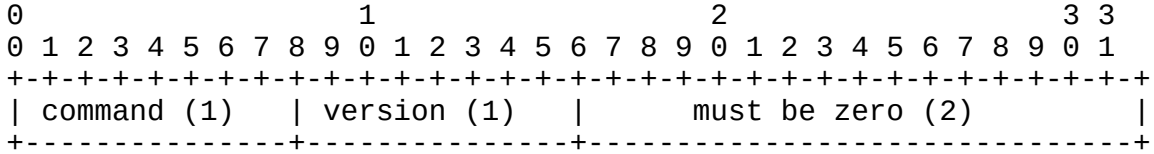
Only the first 4 octets of the packet format shown in figure 2 are sent, since all routing information is implied by this request type.

TRIGGERED RESPONSE (7)

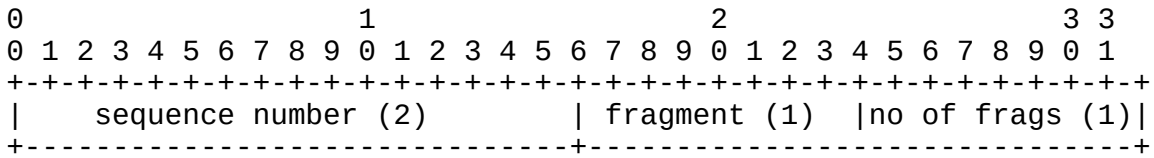
A message containing all of the sender's routing database, excluding those entries learned from the interface to which the routing information is being sent.

This message may be sent in response to a triggered request, or it may be an update message resulting from a change in the routing database.

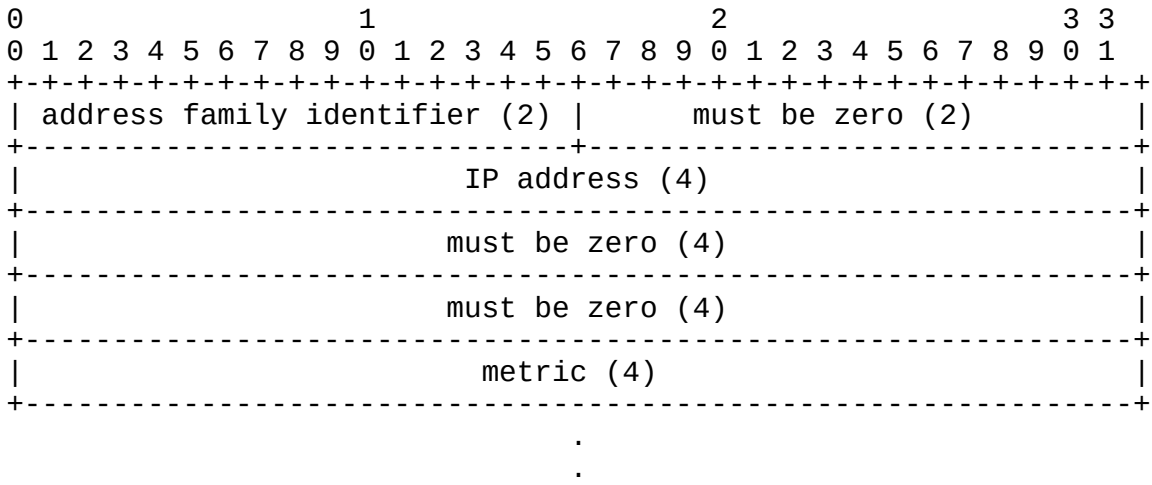
A triggered response message MUST be sent in response to a triggered request message even if there are no routes to propagate. This would be the case for a host which had a WAN interface only, but which wished to run the triggered update protocol.



The following new fields are inserted for some commands



Followed by up to 25 routing entries (each 20 octets)



The format of an IP RIP datagram in octets, with each tick mark representing one bit. All fields are in network order.

The four octets: sequence number (2), fragment number (1) and number of fragments (1) are not present in the original RIP specification. They are only present if command takes the values 7 or 8.

Figure 2. IP Routing Information Protocol packet format

TRIGGERED ACKNOWLEDGEMENT (8)

A message sent in response to every triggered response packet received.

Only the first 8 octets of the packet format shown in figure 2 are sent.

VERSION In this instance Version 1.

SEQUENCE NUMBER

This is a new field inserted if command takes the values 7 or 8.

The sequence number MUST be incremented every time updated information is sent out on a WAN. The sequence number wraps round at 65535.

When a triggered acknowledgement is sent the sequence number is set to the same value as the triggered response packet being acknowledged.

The sequence number MUST be identical over fragments. If a fragment is retransmitted the sequence number MUST not change.

FRAGMENT NUMBER

The fragment number is one for the first fragment of a routing update, and is incremented for each subsequent fragment. A fragment can contain up to 25 routing entries.

When a triggered acknowledgement is sent the fragment number is set to the same value as the triggered response packet being acknowledged.

NUMBER OF FRAGMENTS

In a triggered response packet this indicates the number of packets required to complete the routing update.

This field has no relevance for triggered acknowledgement packets so should be set to zero.

For triggered response packets the rest of the datagram contains a list of destinations, with information about each. Each entry in this list contains the address family identifier (2 for IP), a destination network or host, and the metric for it. The packet format is intended to allow RIP to carry routing information for several different protocols, identifiable by the family identifier.

The IP address is the usual Internet address, stored as 4 octets in network order. The metric field contains a value between 1 and 15 inclusive, specifying the current metric for the destination, or the value 16 (representing 'infinity'), which indicates that the destination is not reachable. Each route sent by a router supersedes any previous route to the same destination from the same router.

The maximum datagram size is 508 octets, excluding UDP and IP headers.

4. IP Routing Information Protocol Version 2

An enhancement to IP RIP to include subnetting has recently become available [2]. This section only describes differences from that RFC.

The triggered update mechanism can be supported by including the triggered request (6), triggered response (7) and triggered acknowledgement (8) commands described in the previous section.

The sequence number, fragment number and number of fragments fields are included in triggered response and triggered acknowledgement commands.

The triggered request packet should also contain the 4 extra octets corresponding to the sequence number, fragment number and number of fragments fields - but set to zero.

Because additional security information is included in RIP Version 2 packets, this MUST be appended to the triggered request and triggered acknowledgement packets, as well as being present in the triggered response packet.

The version number becomes 2. Other aspects of packet layout follow reference [2].

5. Netware Routing Information Protocol

This section should be read in conjunction with references [3], since it only describes differences from the specification.

Netware [3] is the trade name of Novell Research's protocols for computer communication which are derived and extended from Xerox Network System's (XNS) protocols [4].

Netware supports a mechanism that allows routers on an internetwork to exchange routing information using the Routing Information Protocol (RIP) which runs over the Internetwork Packet Exchange (IPX) protocol using socket number 453h.

Netware RIP and IP RIP share a common heritage, in that they are both based on XNS RIP, but there is some divergence, mostly at the packet format level to reflect the differing addressing schemes.

The triggered update mechanism can be applied to Netware RIP. To support the mechanism outlined in this proposal the packet format for Netware RIP is modified as shown in Figure 3.

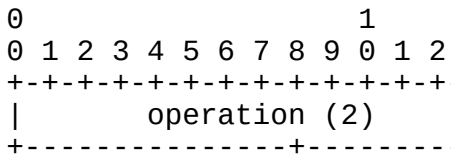
Every datagram contains the following:

RIP OPERATION

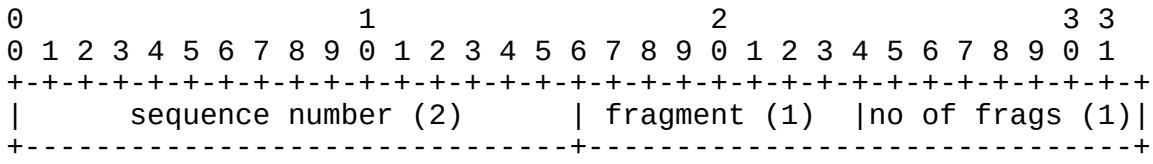
Operations supported in standard Netware RIP are: request (1) and response (2).

The fields sequence number, fragment number and number of fragments MUST NOT be included in packets with these operation values.

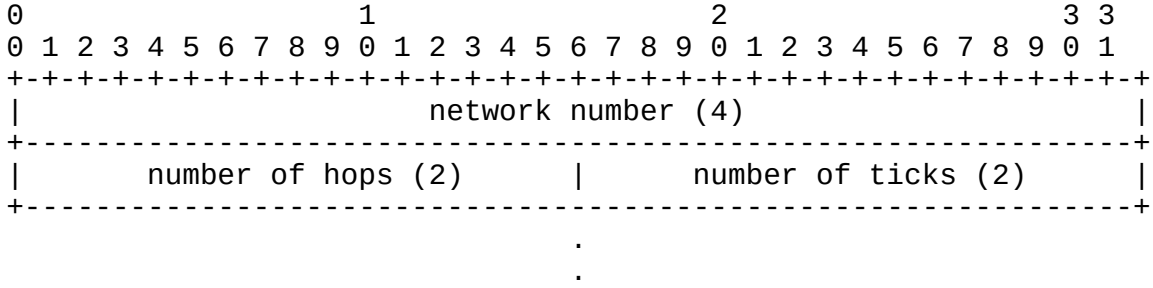
The following new operations are required (with values chosen to be the same as for IP RIP commands):



The following new fields are inserted for some operations



Followed by up to 50 routing entries (each 8 octets)



The format of a Netware RIP datagram in octets, with each tick mark representing one bit. All fields are in network order.

The four octets: sequence number (2), fragment number (1) and number of fragments (1) are not present in the original RIP specification. They are only present if operation takes the values 7 or 8.

Figure 3. Netware Routing Information Protocol packet format

TRIGGERED REQUEST (6)

A request for the responding system to send all of its routing database.

Only the first 2 octets of the packet format shown in figure 3 are sent, since all routing information is implied by this request type.

TRIGGERED RESPONSE (7)

A message containing all of the sender's routing database, excluding those entries learned from the interface to which the routing information is being sent.

This message may be sent in response to a triggered request, or it may be an update message resulting from a change in the routing database.

A triggered response message MUST be sent in response to a triggered request message even if there are no routes to propagate. This would be the case for a host which had a WAN interface only, but which wished to run the triggered update protocol.

TRIGGERED ACKNOWLEDGEMENT (8)

A message sent in response to every triggered response packet received.

Only the first 6 octets of the packet format shown in figure 3 are sent.

SEQUENCE NUMBER

This is a new field inserted if operation takes the values 7 or 8.

The sequence number MUST be incremented every time updated information is sent out on a WAN. The sequence number wraps round at 65535.

When a triggered acknowledgement is sent the sequence number is set to the same value as the triggered response packet being acknowledged.

The sequence number MUST be identical over fragments. If a fragment is retransmitted the sequence number MUST not change.

FRAGMENT NUMBER

The fragment number is one for the first fragment of a routing update, and is incremented for each subsequent fragment. A fragment can contain up to 50 routing entries.

When a triggered acknowledgement is sent the fragment number is set to the same value as the triggered response packet being acknowledged.

NUMBER OF FRAGMENTS

In a triggered response packet this indicates the number of packets required to complete the routing update.

This field has no relevance for triggered acknowledgement packets so should be set to zero.

For triggered response packets the rest of the datagram contains a list of networks, with information about each. Each entry in this list contains a destination network, and the number of hops and number of ticks for each.

The maximum datagram size is 406 octets, excluding the IPX header (a further 30 octets).

6. Netware Service Advertising Protocol

This section should be read in conjunction with references [3], since it only describes differences from the specification.

Netware [3] also supports a mechanism that allows servers on an internetwork to advertise their services by name and type using the Service Advertising Protocol (SAP) which runs over the Internetwork Packet Exchange (IPX) protocol using socket number 452h.

SAP operates on similar principals to running RIP. Routers act as SAP agents, collecting service information from different networks and relay it to interested parties.

To support the triggered update mechanism outlined in this proposal the packet format for Netware SAP is modified as shown in Figure 4.

Every Service Advertising Protocol datagram contains the following:

SAP OPERATION

Operations supported in standard Netware SAP are: general service query (1), general service response (2), nearest service query (3) and nearest service response (4).

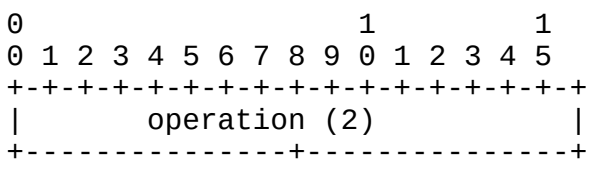
The fields sequence number, fragment number and number of fragments MUST NOT be included in packets with these operation values.

The following new operations are required:

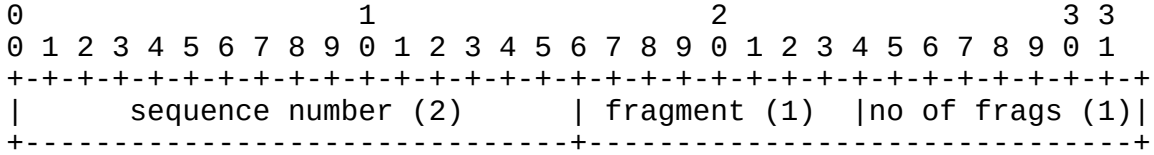
TRIGGERED GENERAL SERVICE QUERY (6)

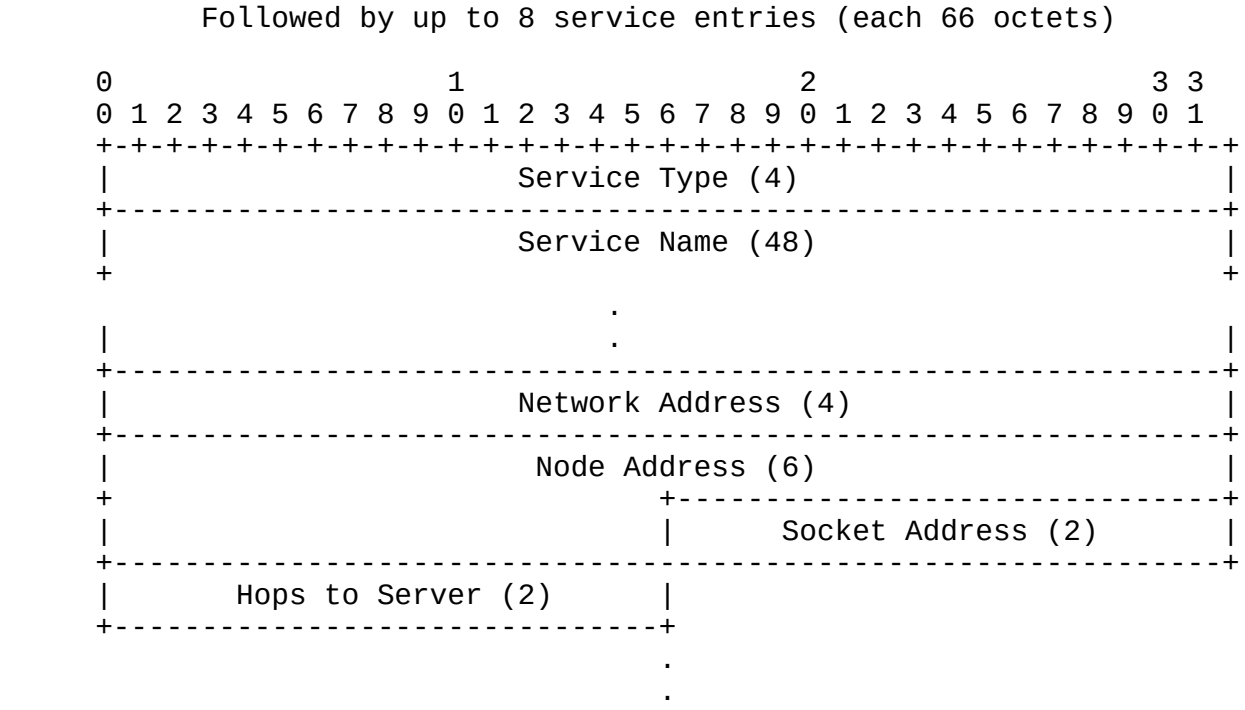
A request for the responding system to send the identities of all servers of all types.

Only the first 2 octets of the packet format shown in figure 4 are sent, since all service types are implied by this request type.



The following new fields are inserted for some operations





The format of a Netware SAP datagram in octets, with each tick mark representing one bit. All fields are in network order.

The four octets: sequence number (2), fragment number (1) and number of fragments (1) are not present in the original SAP specification. They are only present if operation takes the values 7 or 8.

Figure 4. Netware Service Advertising Protocol packet format

TRIGGERED GENERAL SERVICE RESPONSE (7)

A message containing all of the sender's services table, excluding those entries learned from the interface to which the service advertising information is being sent out.

This message may be sent in response to a triggered general service query, or it may be an update message resulting from a change in the service advertising database.

A triggered general service response message MUST be sent in response to a triggered general request message even if there are no services to advertise. This would be the case for a router with a LAN network which had work stations but no servers on it.

TRIGGERED GENERAL SERVICE ACKNOWLEDGEMENT (8)

A message sent in response to every triggered general service response packet received.

Only the first 6 octets of the packet format shown in figure 4 are sent.

SEQUENCE NUMBER

This is a new field inserted if operation takes the values 7 or 8.

The sequence number MUST be incremented every time updated information is sent out on a WAN. The sequence number wraps round at 65535.

When a triggered general service acknowledgement is sent the sequence number is set to the same value as the triggered general service response packet being acknowledged.

The sequence number MUST be identical over fragments. If a fragment is retransmitted the sequence number MUST not change.

FRAGMENT NUMBER

The fragment number is one for the first fragment of a triggered general service response update, and is incremented for each subsequent fragment. A fragment can contain up to 8 service entries.

When a triggered general service acknowledgement is sent, the fragment number is set to the same value as the triggered general service response packet being acknowledged.

NUMBER OF FRAGMENTS

In a triggered response packet this indicates the number of packets required to complete the service update.

This field has no relevance for triggered acknowledgement packets so should be set to zero.

For triggered general service response packets the rest of the datagram contains a list of services, with information about each. Each entry in this list contains the service type, service name, full address (network, node and socket), and the number of hops to the server.

The maximum datagram size is 534 octets, excluding the IPX header (a further 30 octets).

7. Timers

A number of timers are supported to handle the triggered update mechanism:

- o Database timers.
- o Retransmission timer.
- o Reassembly timer.

In this section appropriate timer values for IP RIP are suggested.

For other routing protocols, only the database timer should need to take different values. The database timer values are chosen to match equivalent timer operation for using the protocol on a LAN. The behaviour of a routing entry when a timer is running becomes indistinguishable from a routing entry learned from a broadcast update.

Implementations MAY make timer values configurable - and hence different from the values suggested here - but interoperability requires that all timers on a sub-network should be the same in all routers.

7.1 Database Timers

Routes learned by a triggered response command (7) are normally considered to be permanent - that is they do NOT time out unless activated by one of the following events:

- o If the circuit manager indicates that a next hop router cannot be contacted, all routes learned from that next hop router should start timing out as if they had (just) been learned from a conventional response command (2).

Namely each route exists while the database entry timer is running and is advertised on other interfaces as if still present. The route is then advertised as unreachable while a further hold down timer is allowed to expire, at which point the entry is deleted.

If the circuit manager indicates that the next hop router can be contacted while the database entry timer is running, the routes are reinstated as permanent entries.

If the database entry timer has expired and the circuit manager indicates that the next hop router is reachable, the routing application MUST issue a triggered request. The routes will be reinstated on the basis of any triggered response packet(s) received.

- If a triggered response packet is received in which a route is marked unreachable, the hold down timer MUST be started and the entry is advertised as unreachable on other interfaces. On expiry of the hold down timer the entry is deleted.

  If a triggered response packet is received in which an existing route is ABSENT, the hold down timer MUST also be started and the entry is advertised as unreachable on other interfaces. On expiry of the hold down timer the entry is deleted.

For IP RIP the hold down timer should always run for 120 seconds, to be consistent with RIP usage on broadcast networks. The database entry timer should by default run for 180 seconds. The network can be made more responsive by reducing the database entry timer value. However, making this timer too short can lead to network instabilities. The duration of the database entry timer allows a period of grace in which contention for network resources can be resolved by the circuit manager.

7.2 Retransmission Timer

The routing task runs a retransmission timer:

- When a triggered request is sent it will be retransmitted periodically while a triggered response packet is not received.

- When a triggered response is sent a note of the sequence number and fragment number(s) of the routing update is kept.

  Fragments will be retransmitted at periodic intervals while a triggered acknowledgement packet is not received for the appropriate fragment.

With call set up time on the WAN being of the order of a second, a value of 5 seconds for the retransmission timer is appropriate.

If no response is received after 10 retransmissions, routes via the next hop router are marked as unreachable, the hold down timer MUST be started and the entry is advertised as unreachable on other interfaces. On expiry of the hold down timer the entry is deleted.

The next hop router is then polled using a triggered request packet at 60 second intervals. If a response is received the routers should exchange routing information using triggered response packets.

It may not be desirable to poll indefinitely, since a lack of response (when a circuit is up) is most likely caused by incorrect configuration of the next hop router. An administrator definable number of polls (5 or greater) should be provided.

If the circuit manager indicates that the next hop router is unreachable, the retransmission is suppressed until the circuit manager indicates that the next hop router is reachable once more. Counting of the number of retransmissions continues from where it left off prior to the circuit down indication.

7.3 Reassembly Timer

When a router receives a triggered response update it MUST acknowledge each fragment. If the routing update is fragmented over more than one packet, the receiving router MUST store the fragments until ALL fragments are received.

On receiving the first fragment a timer should be started. If all fragments of the routing update are not received within that period they are discarded - and a triggered request is sent back to the originator (with retransmissions if necessary). The originator MUST then resend ALL triggered response fragments.

The reassembly timer should be set to four times the value of the retransmission timer. With a suggested retransmission timer value of 5 seconds, the suggested reassembly timer value SHOULD be 20 seconds.

Implementations MAY allow the reassembly timer and retransmission timer to be configurable (in the 1:4 ratio), but interoperability will be compromised on WANs where all participating routers DO NOT support the same values for these timers.

Fragments MUST also be discarded if a new fragment with a different sequence number is received. A triggered request MUST not be sent in this instance.

8. Implementation Considerations

In the implementation described in this memo, it is assumed that there is a close binding between the circuit manager and the routing applications - that they are in some way the same 'program'. This is not necessarily true of all products which are routers.

In particular there are UNIX host implementations in which the routing application is distinct from the kernel, where the circuit manager is likely to be installed. In such systems it is possible to stop (or crash) the routing applications independently of what is happening in the kernel.

Other implementations might have the circuit manager on a separate card which again may give the circuit manager a life of its own.

In implementations where the applications and circuit manager have independent lives, a keep-alive mechanism MUST be provided between the applications and the circuit manager, so that if the application or network layer dies and is subsequently re-started they can resynchronize their state tables.

Ideally, when an application dies, the circuit manager should close all existing VCs appropriate to the application and make no further outgoing calls and reject incoming calls until the application is running again.

If the circuit manager is using some form of encapsulation, several applications may be sharing the same VC. If this is the case the circuit manager may wish to filter out datagrams for the appropriate network layer if only one of the applications is affected. But this is not an ideal solution.

Conversely if the application believes the circuit manager has died, it should mark all routes via the circuit manager as unreachable and advertise them on other interfaces for the duration of the hold down timer before deleting them.

9. Security Considerations

Security is provided my a number of aspects:

- The circuit manager is required to be provided with a list of physical addresses to enable it to establish a call to the next hop router on an X.25 SVC or ISDN B-channel.

  The circuit manager SHOULD only allow incoming calls to be accepted from the same well defined list of routers.

Elsewhere in the system there will be a set of logical address and physical address tuples to enable the network protocols to run over the correct circuit. This may be a lookup table, or in some instances there may be an algorithmic conversion between the two addresses.

- o The routing (or service advertising) task MUST be provided with a list of logical addresses to which triggered updates are to be sent on the WAN. The list MAY be a subset of the list of next hop routers maintained by the circuit manager.

  There MAY also be a separate list of next hop routers to which traditional broadcasts of routing (or service advertising) updates should be sent. Next hop routers omitted from either list are assumed to be not participating in routing (or service advertising) updates.

  The list (or lists) doubles as a list of routers from which routing updates are allowed to be received from the WAN. Any routing information received from a router not in the appropriate list MUST be discarded.

10. References

[1] Hedrick. C., "Routing Information Protocol", STD 34, RFC 1058, Rutgers University, June 1988.

[2] Malkin. G., "RIP Version 2 - Carrying Additional Information", RFC 1388, Xylogics, January 1993.

[3] Novell Incorporated., "IPX Router Specification", Version 1.10, October 1992.

[4] Xerox Corporation., "Internet Transport Protocols", Xerox System Integration Standard XSIS 028112, December 1981.

[5] Malis. A., Robinson. D., and R. Ullmann, "Multiprotocol Interconnect on X.25 and ISDN in the Packet Mode", RFC 1356, BBN Communications, Computervision Systems Integration, Process Software Corporation, August 1992.

11. Author's Address

Gerry Meyer
Spider Systems
Stanwell Street
Edinburgh EH6 5NG
Scotland, UK

Phone: (UK) 31 554 9424
Fax: (UK) 31 554 0649
EMail: gerry@spider.co.uk