Network Working Group M. Crawford
Request for Comments: 2894 Fermilab
Category: Standards Track August 2000

Router Renumbering for IPv6

Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

IESG Note:

This document defines mechanisms for informing a set of routers of renumbering operations they are to perform, including a mode of operation in environments in which the exact number of routers is unknown. Reliably informing all routers when the actual number of routers is unknown is a difficult problem. Implementation and operational experience will be needed to fully understand the applicabilty and scalability aspects of the mechanisms defined in this document when the number of routers is unknown.

Abstract

IPv6 Neighbor Discovery and Address Autoconfiguration conveniently make initial assignments of address prefixes to hosts. Aside from the problem of connection survival across a renumbering event, these two mechanisms also simplify the reconfiguration of hosts when the set of valid prefixes changes.

This document defines a mechanism called Router Renumbering ("RR") which allows address prefixes on routers to be configured and reconfigured almost as easily as the combination of Neighbor Discovery and Address Autoconfiguration works for hosts. It provides a means for a network manager to make updates to the prefixes used by and advertised by IPv6 routers throughout a site.

Table of Contents

1. Functional Overview

Router Renumbering Command packets contain a sequence of Prefix Control Operations (PCOs). Each PCO specifies an operation, a Match-Prefix, and zero or more Use-Prefixes. A router processes each PCO in sequence, checking each of its interfaces for an address or prefix which matches the Match-Prefix. For every interface on which a match is found, the operation is applied. The operation is one of ADD, CHANGE, or SET-GLOBAL to instruct the router to respectively add the Use-Prefixes to the set of configured prefixes, remove the prefix which matched the Match-Prefix and replace it with the Use-Prefixes, or replace all global-scope prefixes with the Use-Prefixes. If the set of Use-Prefixes in the PCO is empty, the ADD operation does nothing and the other two reduce to deletions.

Additional information for each Use-Prefix is included in the Prefix Control Operation: the valid and preferred lifetimes to be included in Router Advertisement Prefix Information Options [ND], and either the L and A flags for the same option, or an indication that they are to be copied from the prefix that matched the Match-Prefix.

It is possible to instruct routers to create new prefixes by combining the Use-Prefixes in a PCO with some portion of the existing prefix which matched the Match-Prefix. This simplifies certain operations which are expected to be among the most common. For every Use-Prefix, the PCO specifies a number of bits which should be copied from the existing address or prefix which matched the Match-Prefix and appended to the use-prefix prior to configuring the new prefix on the interface. The copied bits are zero or more bits from the positions immediately after the length of the Use- Prefix. If subnetting information is in the same portion of the old and new prefixes, this synthesis allows a single Prefix Control Operation to define a new global prefix on every router in a site, while preserving the subnetting structure.

Because of the power of the Router Renumbering mechanism, each RR message includes a sequence number to guard against replays, and is required to be authenticated and integrity-checked. Each single Prefix Control Operation is idempotent and so could be retransmitted for improved reliability, as long as the sequence number is current, without concern about multiple processing. However, non-idempotent combinations of PCOs can easily be constructed and messages containing such combinations could not be safely reprocessed. Therefore, all routers are required to guard against processing an RR message more than once. To allow reliable verification that Commands have been received and processed by routers, a mechanism for duplicate-command notification to the management station is included.

Possibly a network manager will want to perform more renumbering, or exercise more detailed control, than can be expressed in a single Router Renumbering packet on the available media. The RR mechanism is most powerful when RR packets are multicast, so IP fragmentation is undesirable. For these reasons, each RR packet contains a "Segment Number". All RR packets which have a Sequence Number greater than or equal to the highest value seen are valid and must be processed. However, a router must keep track of the Segment Numbers of RR messages already processed and avoid reprocessing a message whose Sequence Number and Segment Number match a previously processed message. (This list of processed segment numbers is reset when a new highest Sequence Number is seen.)

The Segment Number does not impose an ordering on packet processing. If a specific sequence of operations is desired, it may be achieved by ordering the PCOs in a single RR Command message or through the Sequence Number field.

There is a "Test" flag which indicates that all routers should simulate processing of the RR message and not perform any actual reconfiguration. A separate "Report" flag instructs routers to send a Router Renumbering Result message back to the source of the RR Command message indicating the actual or simulated result of the operations in the RR Command message.

The effect or simulated effect of an RR Command message may also be reported to network management by means outside the scope of this document, regardless of the value of the "Report" flag.

2. Definitions

2.1. Terminology

Address
This term always refers to a 128-bit IPv6 address [AARCH]. When referring to bits within an address, they are numbered from 0 to 127, with bit 0 being the first bit of the Format Prefix.

Prefix
A prefix can be understood as an address plus a length, the latter being an integer in the range 0 to 128 indicating how many leading bits are significant. When referring to bits within a prefix, they are numbered in the same way as the bits of an address. For example, the significant bits of a prefix whose length is L are the bits numbered 0 through L-1, inclusive.

Match
An address A "matches" a prefix P whose length is L if the first L bits of A are identical with the first L bits of P. (Every address matches a prefix of length 0.) A prefix P1 with length L1 matches a prefix P2 of length L2 if L1 >= L2 and the first L2 bits of P1 and P2 are identical.

Prefix Control Operation
This is the smallest individual unit of Router Renumbering operation. A Router Renumbering Command packet includes zero or more of these, each comprising one matching condition, called a Match-Prefix Part, and zero or more substitution specifications, called Use-Prefix Parts.

Match-Prefix
This is a Prefix against which a router compares the addresses and prefixes configured on its interfaces.

Use-Prefix
The prefix and associated information which is to be configured on a router interface when certain conditions are met.

Matched Prefix
The existing prefix or address which matched a Match-Prefix.

New Prefix
A prefix constructed from a Use-Prefix, possibly including some of the Matched Prefix.

Recorded Sequence Number
The highest sequence number found in a valid message MUST be recorded in non-volatile storage.

Note that "matches" is a transitive relation but not symmetric. If two prefixes match each other, they are identical.

2.2. Requirements

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 [KWORD].

3. Message Format

There are two types of Router Renumbering messages: Commands, which are sent to routers, and Results, which are sent by routers. A third message type is used to synchronize a reset of the Recorded Sequence Number with the cancellation of cryptographic keys. The three types of messages are distinguished the ICMPv6 "Code" field and differ in the contents of the "Message Body" field.

```
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   /                IPv6 header, extension headers                 /
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   /                 ICMPv6 & RR Header (16 octets)                /
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   /                       RR Message Body                         /
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Router Renumbering Message Format

Router Renumbering messages are carried in ICMPv6 packets with Type = 138. The RR message comprises an RR Header, containing the ICMPv6 header, the sequence and segment numbers and other information, and the RR Message Body, of variable length.

All fields marked "reserved" or "res" MUST be set to zero on generation of an RR message, and ignored on receipt.

All implementations which generate Router Renumbering Command messages MUST support sending them to the All Routers multicast address with link and site scopes, and to unicast addresses of link-local and site-local formats. All routers MUST be capable of receiving RR Commands sent to those multicast addresses and to any of their link local and site local unicast addresses. Implementations SHOULD support sending and receiving RR messages addressed to other unicast addresses. An implementation which is both a sender and receiver of RR commands SHOULD support use of the All Routers multicast address with node scope.

Data authentication and message integrity MUST be provided for all Router Renumbering Command messages by appropriate IP Security [IPSEC] means. The integrity assurance must include the IPv6 destination address and the RR Header and Message Body. See section 7, "Security Considerations".

The use of authentication for Router Renumbering Result messages is RECOMMENDED.

3.1. Router Renumbering 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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |     Type      |     Code      |            Checksum           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                        SequenceNumber                         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | SegmentNumber |     Flags     |            MaxDelay           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           reserved                            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Fields:

Type 138 (decimal), the ICMPv6 type value assigned to Router Renumbering

Code
  0 for a Router Renumbering Command
  1 for a Router Renumbering Result
255 for a Sequence Number Reset.
The Sequence Number Reset is described in section 5.

Checksum The ICMPv6 checksum, as specified in [ICMPV6]. The checksum covers the IPv6 pseudo-header and all fields of the RR message from the Type field onward.

SequenceNumber
An unsigned 32-bit sequence number. The sequence number MUST be non-decreasing between Sequence Number Resets.

SegmentNumber
An unsigned 8-bit field which enumerates different valid RR messages having the same SequenceNumber. No ordering among RR messages is imposed by the SegmentNumber.

Flags A combination of one-bit flags. Five are defined and three bits are reserved.

```
                                        +-+-+-+-+-+-+-+-+
                                        |T|R|A|S|P| res |
                                        +-+-+-+-+-+-+-+-+
```

The flags T, R, A and S have defined meanings in an RR Command message. In a Result message they MUST be copied from the corresponding Command. The P flag is meaningful only in a Result message and MUST be zero in a transmitted Command and ignored in a received Command.

```
T   Test command --
    0 indicates that the router configuration is to be
      modified;
    1 indicates a "Test" message: processing is to be
      simulated and no configuration changes are to be
      made.

R   Result requested --
    0 indicates that a Result message MUST NOT be sent
      (but other forms of logging are not precluded);
    1 indicates that the router MUST send a Result
      message upon completion of processing the Command
      message;

A   All interfaces --
    0 indicates that the Command MUST NOT be applied to
      interfaces which are administratively shut down;
    1 indicates that the Command MUST be applied to all
      interfaces regardless of administrative shutdown
      status.

S   Site-specific -- This flag MUST be ignored unless
    the router treats interfaces as belonging to
    different "sites".
    0 indicates that the Command MUST be applied to
      interfaces regardless of which site they belong
      to;
    1 indicates that the Command MUST be applied only to
      interfaces which belong to the same site as the
      interface to which the Command is addressed.  If
      the destination address is appropriate for
      interfaces belonging to more than one site, then
      the Command MUST be applied only to interfaces
      belonging to the same site as the interface on
      which the Command was received.

P   Processed previously --
    0 indicates that the Result message contains the
      complete report of processing the Command;
```

1 indicates that the Command message was previously processed (and is not a Test) and the responding router is not processing it again. This Result message MAY have an empty body.

MaxDelay An unsigned 16-bit field specifying the maximum time, in milliseconds, by which a router MUST delay sending any reply to this Command. Implementations MAY generate the random delay between 0 and MaxDelay milliseconds with a finer granularity than 1ms.

3.2. Message Body -- Command Message

The body of an RR Command message is a sequence of zero or more Prefix Control Operations, each of variable length. The end of the sequence MAY be inferred from the IPv6 length and the lengths of extension headers which precede the ICMPv6 header.

3.2.1. Prefix Control Operation

A Prefix Control Operation has one Match-Prefix Part of 24 octets, followed by zero or more Use-Prefix Parts of 32 octets each.

3.2.1.1. Match-Prefix Part

```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |    OpCode     |   OpLength    |    Ordinal    |   MatchLen    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |    MinLen     |    MaxLen     |           reserved            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   +-                                                             -+
   |                                                               |
   +-                         MatchPrefix                         -+
   |                                                               |
   +-                                                             -+
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Fields:

OpCode An unsigned 8-bit field specifying the operation to be performed when the associated MatchPrefix matches an interface’s prefix or address. Values are:

1 the ADD operation

2 the CHANGE operation

3 the SET-GLOBAL operation

OpLength The total length of this Prefix Control Operation, in units of 8 octets. A valid OpLength will always be of the form 4N+3, with N equal to the number of UsePrefix parts (possibly zero).

Ordinal An 8-bit field which MUST have a different value in each Prefix Control Operation contained in a given RR Command message. The value is otherwise unconstrained.

MatchLen An 8-bit unsigned integer between 0 and 128 inclusive specifying the number of initial bits of MatchPrefix which are significant in matching.

MinLen An 8-bit unsigned integer specifying the minimum length which any configured prefix must have in order to be eligible for testing against the MatchPrefix.

MaxLen An 8-bit unsigned integer specifying the maximum length which any configured prefix may have in order to be eligible for testing against the MatchPrefix.

MatchPrefix The 128-bit prefix to be compared with each interface’s prefix or address.

3.2.1.2. Use-Prefix Part

```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |    UseLen     |    KeepLen    |   FlagMask    |    RAFlags    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                        Valid Lifetime                         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                      Preferred Lifetime                       |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |V|P|                         reserved                          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   +-                                                             -+
   |                                                               |
   +-                          UsePrefix                          -+
   |                                                               |
   +-                                                             -+
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Fields:

UseLen — An 8-bit unsigned integer less than or equal to 128 specifying the number of initial bits of UsePrefix to use in creating a new prefix for an interface.

KeepLen — An 8-bit unsigned integer less than or equal to (128-UseLen) specifying the number of bits of the prefix or address which matched the associated Match-Prefix which should be retained in the new prefix. The retained bits are those at positions UseLen through (UseLen+KeepLen-1) in the matched address or prefix, and they are copied to the same positions in the New Prefix.

FlagMask — An 8-bit mask. A 1 bit in any position means that the corresponding flag bit in a Router Advertisement (RA) Prefix Information Option for the New Prefix should be set from the RAFlags field in this Use-Prefix Part. A 0 bit in the FlagMask means that the RA flag bit for the New Prefix should be copied from the corresponding RA flag bit of the Matched Prefix.

RAFlags — An 8 bit field which, under control of the FlagMask field, may be used to initialize the flags in Router Advertisement Prefix Information Options [ND] which advertise the New Prefix. Note that only two flags have defined meanings to date: the L (on-link) and A (autonomous configuration) flags. These flags occupy the two leftmost bit positions in the RAFlags field, corresponding to their position in the Prefix Information Option.

Valid Lifetime
A 32-bit unsigned integer which is the number of seconds for which the New Prefix will be valid [ND, SAA].

Preferred Lifetime
A 32-bit unsigned integer which is the number of seconds for which the New Prefix will be preferred [ND, SAA].

V A 1-bit flag indicating that the valid lifetime of the New Prefix MUST be effectively decremented in real time.

P A 1-bit flag indicating that the preferred lifetime of the New Prefix MUST be effectively decremented in real time.

UsePrefix The 128-bit Use-prefix which either becomes or is used in forming (if KeepLen is nonzero) the New Prefix. It MUST NOT have the form of a multicast or link-local address [AARCH].

3.3. Message Body -- Result Message

The body of an RR Result message is a sequence of zero or more Match Reports of 24 octets. An RR Command message with the "R" flag set will elicit an RR Result message containing one Match Report for each Prefix Control Operation, for each different prefix it matches on each interface. The Match Report has the following format.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         reserved          |B|F|    Ordinal    |  MatchedLen   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         InterfaceIndex                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-                                                             -+
|                                                               |
+-                        MatchedPrefix                        -+
|                                                               |
+-                                                             -+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Fields:

B   A one-bit flag which, when set, indicates that one or more fields in the associated PCO were out of bounds. The bounds check is described in section 4.2.

F   A one-bit flag which, when set, indicates that one or more Use-Prefix parts from the associated PCO were not honored by the router because of attempted formation of a forbidden prefix format, such as a multicast or loopback address.

Ordinal   Copied from the Prefix Control Operation whose MatchPrefix matched the MatchedPrefix on the interface indicated by InterfaceIndex.

MatchedLen   The length of the Matched Prefix.

InterfaceIndex
             The router’s numeric designation of the interface on which the MatchedPrefix was configured. This MUST be the same as the value of ipv6IfIndex which designates that index in the SNMP IPv6 MIB General Group [IPV6MIB].

It is possible for a Result message to be larger than the Command message which elicited it. Such a Result message may have to be fragmented for transmission. If so, it SHOULD be fragmented to the IPv6 minimum required MTU [IPV6].

4. Message Processing

Processing of received Router Renumbering Result messages is entirely implementation-defined. Implementation of Command message processing may vary in detail from the procedure set forth below, so long as the result is not affected.

Processing of received Router Renumbering Command messages consists of three conceptual parts: header check, bounds check, and execution.

4.1. Header Check

The ICMPv6 checksum and type are presumed to have been checked before a Router Renumbering module receives a Command to process. In an implementation environment where this may not be the case, those checks MUST be made at this point in the processing.

If the ICMPv6 length derived from the IPv6 length is less than 16 octets, the message MUST be discarded and SHOULD be logged to network management.

If the ICMPv6 Code field indicates a Result message, a router which is not a source of RR Command messages MUST discard the message and SHOULD NOT log it to network management.

If the IPv6 destination address is neither an All Routers multicast address [AARCH] nor one of the receiving router’s unicast addresses, the message MUST be discarded and SHOULD be logged to network management.

Next, the SequenceNumber is compared to the Recorded Sequence Number. (If no RR messages have been received and accepted since system initialization, the Recorded Sequence Number is zero.) This comparison is done with the two numbers considered as unsigned integers, not as DNS-style serial numbers. If the SequenceNumber is less than the Recorded Sequence Number, the message MUST be discarded and SHOULD be logged to network management.

Finally, if the SequenceNumber in the message is greater than the Recorded Sequence Number or the T flag is set, skip to the bounds check. Otherwise the SegmentNumber MUST now be checked. If a correctly authenticated message with the same SequenceNumber and SegmentNumber has not already been processed, skip to the bounds check. Otherwise, this Command is a duplicate and not a Test Command. If the R flag is not set, the duplicate message MUST be discarded and SHOULD NOT be logged to network management. If R is set, an RR Result message with the P flag set MUST be scheduled for transmission to the source address of the Command after a random time uniformly distributed between 0 and MaxDelay milliseconds. The body of that Result message MUST either be empty or be a saved copy of the Result message body generated by processing of the previous message with the same SequenceNumber and SegmentNumber. After scheduling the Result message, the Command MUST be discarded without further processing.

4.2. Bounds Check

If the SequenceNumber is greater than the Recorded Sequence Number, then the list of processed SegmentNumbers and the set of saved Result messages, if any, MUST be cleared and the Recorded Sequence Number MUST be updated to the value used in the current message, regardless of subsequent processing errors.

Next, if the ICMPv6 Code field indicates a Sequence Number Reset, skip to section 5.

At this point, if T is set in the RR header and R is not set, the message MAY be discarded without further processing.

If the R flag is set, begin constructing an RR Result message. The RR header of the Result message is completely determined at this time except for the Checksum.

The values of the following fields of a PCO MUST be checked to ensure that they are within the appropriate bounds.

OpCode must be a defined value.

OpLength must be of the form 4N+3 and consistent the the length of the Command packet and the PCO’s offset within the packet.

MatchLen must be between 0 and 128 inclusive

UseLen, KeepLen
in each Use-Prefix Part must be between 0 and 128 inclusive, as must the sum of the two.

If any of these fields are out of range in a PCO, the entire PCO MUST NOT be performed on any interface. If the R flag is set in the RR header then add to the RR Result message a Match Report with the B flag set, the F flag clear, the Ordinal copied from the PCO, and all other fields zero. This Match Report MUST be included only once, not once per interface.

Note that MinLen and MaxLen need not be explicitly bounds checked, even though certain combinations of values will make any matches impossible.

4.3. Execution

For each applicable router interface, as determined by the A and S flags, the Prefix Control Operations in an RR Command message must be carried out in order of appearance. The relative order of PCO processing among different interfaces is not specified.

If the T flag is set, create a copy of each interface's configuration on which to operate, because the results of processing a PCO may affect the processing of subsequent PCOs. Note that if all operations are performed on one interface before proceeding to another interface, only one interface-configuration copy will be required at a time.

For each interface and for each Prefix Control Operation, each prefix configured on that interface with a length between the MinLen and MaxLen values in the PCO is tested to determine whether it matches (as defined in section 2.1) the MatchPrefix of the PCO. The configured prefixes are tested in an arbitrary order. Any new prefix configured on an interface by the effect of a given PCO MUST NOT be tested against that PCO, but MUST be tested against all subsequent PCOs in the same RR Command message.

Under a certain condition the addresses on an interface are also tested to see whether any of them matches the MatchPrefix. If and only if a configured prefix "P" does have a length between MinLen and MaxLen inclusive, does not match the MatchPrefix "M", but M does match P (this can happen only if M is longer than P), then those addresses on that interface which match P MUST be tested to determine whether any of them matches M. If any such address does match M, process the PCO as if P matched M, but when forming New Prefixes, if KeepLen is non-zero, bits are copied from the address. This special case allows a PCO to be easily targeted to a single specific interface in a network.

If P does not match M, processing is finished for this combination of PCO, interface and prefix. Continue with another prefix on the same interface if there are any more prefixes which have not been tested against this PCO and were not created by the action of this PCO. If no such prefixes remain on the current interface, continue processing with the next PCO on the same interface, or with another interface.

If P does match M, either directly or because a configured address which matches P also matches M, then P is the Matched Prefix. Perform the following steps.

If the Command has the R flag set, add a Match Report to the Result message being constructed.

If the OpCode is CHANGE, mark P for deletion from the current interface.

If the OpCode is SET-GLOBAL, mark all global-scope prefixes on the current interface for deletion.

If there are any Use-Prefix parts in the current PCO, form the New Prefixes. Discard any New Prefix which has a forbidden format, and if the R flag is set in the command, set the F flag in the Match Report for this PCO and interface. Forbidden prefix formats include, at a minimum, multicast, unspecified and loopback addresses. [AARCH] Any implementation MAY forbid, or allow the network manager to forbid other formats as well.

For each New Prefix which is already configured on the current interface, unmark that prefix for deletion and update the lifetimes and RA flags. For each New Prefix which is not already configured, add the prefix and, if appropriate, configure an address with that prefix.

Delete any prefixes which are still marked for deletion, together with any addresses which match those prefixes but do not match any prefix which is not marked for deletion.

After processing all the Prefix Control Operations on all the interfaces, an implementation MUST record the SegmentNumber of the packet in a list associated with the SequenceNumber.

If the Command has the R flag set, compute the Checksum and schedule the Result message for transmission after a random time interval uniformly distributed between 0 and MaxDelay milliseconds. This interval SHOULD begin at the conclusion of processing, not the beginning. A copy of the Result message MAY be saved to be retransmitted in response to a duplicate Command.

4.4. Summary of Effects

The only Neighbor Discovery [ND] parameters which can be affected by Router Renumbering are the following.

A router’s addresses and advertised prefixes, including the prefix lengths.

The flag bits (L and A, and any which may be defined in the future) and the valid and preferred lifetimes which appear in a Router Advertisement Prefix Information Option.

That unnamed property of the lifetimes which specifies whether they are fixed values or decrementing in real time.

Other internal router information, such as the time until the next unsolicited Router Advertisement or MIB variables MAY be affected as needed.

All configuration changes resulting from Router Renumbering SHOULD be saved to non-volatile storage where this facility exists. The problem of properly restoring prefix lifetimes from non-volatile storage exists independently of Router Renumbering and deserves careful attention, but is outside the scope of this document.

5. Sequence Number Reset

It may prove necessary in practice to reset a router’s Recorded Sequence Number. This is a safe operation only when all cryptographic keys previously used to authenticate RR Commands have expired or been revoked. For this reason, the Sequence Number Reset message is defined to accomplish both functions.

When a Sequence Number Reset (SNR) has been authenticated and has passed the header check, the router MUST invalidate all keys which have been used to authenticate previous RR Commands, including the key which authenticated the SNR itself. Then it MUST discard any saved RR Result messages, clear the list of recorded SegmentNumbers and reset the Recorded Sequence Number to zero.

If the router has no other, unused authentication keys already available for Router Renumbering use it SHOULD establish one or more new valid keys. The details of this process will depend on whether manual keying or a key management protocol is used. In either case, if no keys are available, no new Commands can be processed.

A SNR message SHOULD contain no PCOs, since they will be ignored. If and only if the R flag is set in the SNR message, a router MUST respond with a Result Message containing no Match Reports. The header and transmission of the Result are as described in section 3.

The invalidation of authentication keys caused by a valid SNR message will cause retransmitted copies of that message to be ignored.

6. IANA Considerations

Following the policies outlined in [IANACON], new values of the Code field in the Router Renumbering Header (section 3.1) and the OpCode field of the Match-Prefix Part (section 3.2.1.1) are to be allocated by IETF consensus only.

7. Security Considerations

The Router Renumbering mechanism proposed here is very powerful and prevention of spoofing it is important. Replay of old messages must, in general, be prevented (even though a narrow class of messages exists for which replay would be harmless). What constitutes a sufficiently strong authentication algorithm may change from time to time, but algorithms should be chosen which are strong against current key-recovery and forgery attacks.

Authentication keys must be as well protected as any other access method that allows reconfiguration of a site’s routers. Distribution of keys must not expose them or permit alteration, and key validity must be limited in terms of time and number of messages authenticated.

Note that although a reset of the Recorded Sequence Number requires the cancellation of previously-used authentication keys, introduction of new keys and expiration of old keys does not require resetting the Recorded Sequence Number.

7.1. Security Policy and Association Database Entries

The Security Policy Database (SPD) [IPSEC] of a router implementing this specification MUST cause incoming Router Renumbering Command packets to either be discarded or have IPsec applied. (The determination of "discard" or "apply" MAY be based on the source address.) The resulting Security Association Database (SAD) entries MUST ensure authentication and integrity of the destination address and the RR Header and Message Body, and the body length implied by the IPv6 length and intervening extension headers. These requirements are met by the use of the Authentication Header [AH] in transport or tunnel mode, or the Encapsulating Security Payload [ESP] in tunnel mode with non-NULL authentication. The mandatory-to-implement IPsec authentication algorithms (other than NULL) seem strong enough for Router Renumbering at the time of this writing.

Note that for the SPD to distinguish Router Renumbering from other ICMP packets requires the use of the ICMP Type field as a selector. This is consistent with, although not mentioned by, the Security Architecture specification [IPSEC].

At the time of this writing, there exists no multicast key management protocol for IPsec and none is on the horizon. Manually configured Security Associations will therefore be common. The occurrence of "from traffic" in the table below would therefore more realistically be a wildcard or a fixed range. Use of a small set of shared keys per management station suffices, so long as key distribution and storage are sufficiently safeguarded.

A sufficient set of SPD entries for incoming traffic could select

Field	SPD Entry	SAD Entry
Source	wildcard	from traffic
Destination	wildcard	from SPD
Transport	ICMPv6	from SPD
ICMP Type	Rtr. Renum.	from SPD
Action	Apply IPsec	
SA Spec	AH/Transport Mode	

or there might be an entry for each management station and/or for each of the router's unicast addresses and for each of the defined All-Routers multicast addresses, and a final wildcard entry to discard all other incoming RR messages.

The SPD and SAD are conceptually per-interface databases. This fact may be exploited to permit shared management of a border router, for example, or to discard all Router Renumbering traffic arriving over tunnels.

8. Implementation and Usage Advice for Reliability

Users of Router Renumbering will want to be sure that every non-trivial message reaches every intended router. Well-considered exploitation of Router Renumbering's retransmission and response-directing features should make that goal achievable with high confidence even in a minimally reliable network.

In one set of cases, probably the majority, the network management station will know the complete set of routers under its control. Commands can be retransmitted, with the "R" (Reply-requested) flag set in the RR header, until Results have been collected from all routers. If unicast Security Associations (or the means for creating them) are available, the management station may switch from multicast to unicast transmission when the number of routers still unheard-from is suitably small.

To maintain a list of managed routers, the management station can employ any of several automatic methods which may be more convenient than manual entry in a large network. Multicast RR "Test" commands can be sent periodically and the results archived, or the management station can use SNMP to "peek" into a link-state routing protocol such as OSPF [OSPFMIB]. (In the case of OSPF, roughly one router per area would need to be examined to build a complete list of routers.)

In a large dynamic network where the set of managed routers is not known but reliable execution is desired, a scalable method for achieving confidence in delivery is described here. Nothing in this section affects the format or content of Router Renumbering messages, nor their processing by routers.

A management station implementing these reliability mechanisms MUST alert an operator who attempts to commence a set of Router Renumbering Commands when retransmission of a previous set is not yet completed, but SHOULD allow the operator to override the warning.

8.1. Outline and Definitions

The set of routers being managed with Router Renumbering is considered as a set of populations, each population having a characteristic probability of successful round-trip delivery of a Command/Result pair. The goal is to estimate a lower bound, P, on the round-trip probability for the whole set. With this estimate and other data about the responses to retransmissions of the Command, a confidence level can be computed for hypothesis that all routers have been heard from.

If the true probability of successful round-trip communication with a managed router were a constant, p, for all managed routers then an estimate P of p could be derived from either of these statistics:

> The expected ratio of the number of routers first heard from after transmission (N + 1) to the number first heard from after N is (1 - p).
>
> When N different routers have been heard from after M transmissions of a Command, the expected total number of Result messages received is pNM. If R is the number of Results actually received, then P = R/MN.

The two methods are not equivalent. The first suffers numerical problems when the number of routers still to be heard from gets small, so the P = R/MN estimate should be used.

Since the round-trip probability is not expected to be uniform in the real world, and the less-reliable units are more important to a lower-bound estimate but more likely to be missed in sampling, the sample from which P is computed is biased toward the less-reliable routers. After the Nth transmission interval, N > 2, neglect all routers heard from in intervals 1 through F from the reliability estimate, where F is the greatest integer less than one-half of N. For example, after five intervals, only routers first heard from in the third through fifth intervals will be counted.

A management station implementing the methods of this section should allow the user to specify the following parameters, and default them to the indicated values.

Ct The target delivery confidence, default 0.999.

Pp A presumptive, pessimistic initial estimate of the lower bound of the round-trip probability, P, to prevent early termination. (See below.) Default 0.75.

Ti The initial time between Command retransmissions. Default 4 seconds. MaxDelay milliseconds (see section 3.1) must be added to the retransmission timer. Knowledge of the routers’ processing time for RR Commands may influence the setting of Ti. Ti+MaxDelay is also the minimum time the management station must wait for Results after each transmission before computing a new confidence level. The phrase "end of the Nth interval" means a time Ti+MaxDelay after the Nth transmission of a Command.

Tu The upper bound on the period between Command retransmissions. Default 512 seconds.

The following variables, some a function of the retransmission counter N, are used in the next section.

T(N) The time between Command transmissions N and N+1 is V*T(N) + MaxDelay, where V is random and roughly uniform in the range [0.75, 1.0]. T(1) = Ti and for N > 1, T(N) = min(2*T(N-1), Tu).

M(N) The cumulative number of distinct routers from which replies have been received to any of the first N transmissions of the Command.

F=F(N) FLOOR((N-1)/2). All routers from which responses were received in the first F intervals will be effectively omitted from the estimate of the round-trip probability computed at the Nth interval.

R(N,F) The total number of RR Result messages, including duplicates, received by the end of the Nth interval from those routers which were NOT heard from in any of the first F intervals.

p(N) The estimate of the worst-case round-trip delivery probability.

c(N) The computed confidence level.

An asterisk (*) is used to denote multiplication and a caret (^) denotes exponentiation.

If the difference in reliability between the "good" and "bad" parts of a managed network is very great, early c(N) values will be too high. Retransmissions should continue for at least Nmin = log(1-Ct)/log(1-Pp) intervals, regardless of the current confidence estimate. (In fact, there’s no need to compute p(N) and c(N) until after Nmin intervals.)

8.2. Computations

Letting A = N*(M(N)-M(F))/R(N,F) for brevity, the estimate of the round-trip delivery probability is p(N) = 1-Q, where Q is that root of the equation

$$Q^N - A*Q + (A-1) = 0$$

which lies between 0 and 1. (Q = 1 is always a root. If N is odd there is also a negative root.) This may be solved numerically, for example with Newton’s method (see any standard text, for example [ANM]). The first-order approximation

$$Q1 = 1 - 1/A$$

may be used as a starting point for iteration. But Q1 should NOT be used as an approximate solution as it always underestimates Q, and hence overestimates p(N), which would cause an overestimate of the confidence level.

If necessary, the spurious root Q = 1 can be divided out, leaving

$$Q^{(N-1)} + Q^{(N-2)} + ... + Q - (A-1) = 0$$

as the equation to solve. Depending on the numerical method used, this could be desirable as it’s just possible (but very unlikely) that A=N and so Q=1 was a double root of the earlier equation.

After N > 2 (or N >= Nmin) intervals have been completed, Compute the lower-bound reliability estimate

$$p(N) = R(N,F)/((N-F)*(M(N) - M(F))).$$

Compute the confidence estimate

$$c(N) = (1 - (1-p(N))^{N})^{(M(N) - M(F) + 1)}.$$

which is the Bayesian probability that M(N) is the number of routers present given the number of responses which were collected, as opposed to M(N)+1 or any greater number. It is assumed that the a priori probability of there being K routers was no greater than that of K-1 routers, for all K > M(N).

When c(N) >= Ct and N >= Nmin, retransmissions of the Command may cease. Otherwise another transmission should be scheduled at a time V*T(N) + MaxDelay after the previous (Nth) transmission, or V*T(N) after the conclusion of processing responses to the Nth transmission, whichever is later.

One corner case needs consideration. Divide-by-zero may occur when computing p. This can happen only when no new routers have been heard from in the last N-F intervals. Generally, the confidence estimate c(N) will be close to unity by then, but in a pathological case such as a large number of routers with reliable communication and a much smaller number with very poor communication, the confidence estimate may still be less than Ct when p’s denominator vanishes. The implementation may continue, and should continue if the minimum number of transmissions given in the previous paragraph have not yet been made. If new routers are heard from, p(N) will again be non-singular.

Of course no limited retransmission scheme can fully address the possibility of long-term problems, such as a partitioned network. The network manager is expected to be aware of such conditions when they exist.

8.3. Additional Assurance Methods

As a final means to detect routers which become reachable after missing renumbering commands during an extended network split, a management station MAY adopt the following strategy. When performing each new operation, increment the SequenceNumber by more than one.

After the operation is believed complete, periodically send some "no-op" RR Command with the R (Result Requested) flag set and a SequenceNumber one less than the highest used. Any responses to such a command can only come from router that missed the last operation. An example of a suitable "no-op" command would be an ADD operation with MatchLen = 0, MinLen = 0, MaxLen = 128, and no Use-Prefix Parts.

If old authentication keys are saved by the management station, even the reappearance of routers which missed a Sequence Number Reset can be detected by the transmission of no-op commands with the invalid key and a SequenceNumber higher than any used before the key was invalidated. Since there is no other way for a management station to distinguish a router's failure to receive an entire sequence of repeated SNR messages from the loss of that router's single SNR Result Message, this is the RECOMMENDED way to test for universal reception of a SNR Command.

9. Usage Examples

This section sketches some sample applications of Router Renumbering. Extension headers, including required IPsec headers, between the IPv6 header and the ICMPv6 header are not shown in the examples.

9.1. Maintaining Global-Scope Prefixes

A simple use of the Router Renumbering mechanism, and one which is expected to to be common, is the maintenance of a set of global prefixes with a subnet structure that matches that of the site's site-local address assignments. In the steady state this would serve to keep the Preferred and Valid lifetimes set to their desired values. During a renumbering transition, similar Command messages can add new prefixes and/or delete old ones. An outline of a suitable Command message follows. Fields not listed are presumed set to suitable values. This Command assumes all router interfaces to be maintained already have site-local [AARCH] addresses.

```
IPv6 Header
   Next Header = 58 (ICMPv6)
   Source Address = (Management Station)
   Destination Address = FF05::2 (All Routers, site-local scope)

ICMPv6/RR Header
   Type = 138 (Router Renumbering), Code = 0 (Command)
   Flags = 60 hex (R, A)
First (and only) PCO:

   Match-Prefix Part
       OpCode = 3 (SET-GLOBAL)
       OpLength = 4 N + 3 (assuming N global prefixes)
       Ordinal = 0 (arbitrary)
       MatchLen = 10
       MatchPrefix = FEC0::0

   First Use-Prefix Part
       UseLen = 48 (Length of TLA ID + RES + NLA ID [AARCH])
       KeepLen = 16 (Length of SLA (subnet) ID [AARCH])
       FlagMask, RAFlags, Lifetimes, V & P flags -- as desired
       UsePrefix = First global /48 prefix

   . . .

   Nth Use-Prefix Part
       UseLen = 48
       KeepLen = 16
       FlagMask, RAFlags, Lifetimes, V & P flags -- as desired
       UsePrefix = Last global /48 prefix
```

This will cause N global prefixes to be set (or updated) on each applicable interface. On each interface, the SLA ID (subnet) field of each global prefix will be copied from the existing site-local prefix.

9.2. Renumbering a Subnet

A subnet can be gracefully renumbered by setting the valid and preferred timers on the old prefix to a short value and having them run down, while concurrently adding adding the new prefix. Later, the expired prefix is deleted. The first step is described by the following RR Command.

```
   IPv6 Header
      Next Header = 58 (ICMPv6)
      Source Address = (Management Station)
      Destination Address = FF05::2 (All Routers, site-local scope)

   ICMPv6/RR Header
      Type = 138 (Router Renumbering), Code = 0 (Command)
      Flags = 60 hex (R, A)
```

First (and only) PCO:

```
   Match-Prefix Part
       OpCode = 2 (CHANGE)
       OpLength = 11 (reflects 2 Use-Prefix Parts)
       Ordinal = 0 (arbitrary)
       MatchLen = 64
       MatchPrefix = Old /64 prefix

   First Use-Prefix Part
       UseLen = 0
       KeepLen = 64 (this retains the old prefix value intact)
       FlagMask = 0, RAFlags = 0
       Valid Lifetime = 28800 seconds (8 hours)
       Preferred Lifetime = 7200 seconds (2 hours)
       V flag = 1, P flag = 1
       UsePrefix = 0::0

   Second Use-Prefix Part
       UseLen = 64
       KeepLen = 0
       FlagMask = 0, RAFlags = 0
       Lifetimes, V & P flags -- as desired
       UsePrefix = New /64 prefix
```

The second step, deletion of the old prefix, can be done by an RR Command with the same Match-Prefix Part (except for an OpLength reduced from 11 to 3) and no Use-Prefix Parts. Any temptation to set KeepLen = 64 in the second Use-Prefix Part above should be resisted, as it would instruct the router to sidestep address configuration.

10. Acknowledgments

This protocol was designed by Matt Crawford based on an idea of Robert Hinden and Geert Jan de Groot. Many members of the IPNG Working Group contributed useful comments, in particular members of the DIGITAL UNIX IPv6 team. Bill Sommerfeld provided helpful IPsec expertise. Relentless browbeating by various IESG members may have improved the final quality of this specification.

11. References

[AARCH] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 2373, July 1998.

[AH] Kent, S. and R. Atkinson, "IP Authentication Header", RFC 2402, November 1998.

[ANM] Isaacson, E. and H. B. Keller, "Analysis of Numerical Methods", John Wiley & Sons, New York, 1966.

[ESP] Kent, S. and R. Atkinson, "IP Encapsulating Security Payload (ESP)", RFC 2406, November 1998.

[IANACON] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 2434, October 1998.

[ICMPV6] Conta, A. and S. Deering, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6)", RFC 2463, December 1998.

[IPSEC] Kent, S. and R. Atkinson, "Security Architecture for the Internet Protocol", RFC 2401, November 1998.

[IPV6] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

[IPV6MIB] Haskin, D. and S. Onishi, "Management Information Base for IP Version 6: Textual Conventions and General Group", RFC 2466, December 1998.

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

[ND] Narten, T., Nordmark, E. and W. Simpson, "Neighbor Discovery for IP Version 6 (IPv6)", RFC 2461, December 1998.

[OSPFMIB] Baker, F. and R. Coltun, "OSPF Version 2 Management Information Base", RFC 1850, November 1995.

12. Author’s Address

Matt Crawford
Fermilab MS 368
PO Box 500
Batavia, IL 60510
USA

Phone: +1 630 840 3461
EMail: crawdad@fnal.gov

Appendix -- Derivation of Reliability Estimates

If a population S of size k is repeatedly sampled with an efficiency p, the expected number of members of S first discovered on the nth sampling is

$$m = [1 - (1-p)^n] * k$$

The expected total number of members of S found in samples, including duplicates, is

$$r = n * p * k$$

Taking the ratio of m to r cancels the unknown factor k and yields an equation

$$[1 - (1-p)^n] / p = nm/r$$

which may be solved for p, which is then an estimator of the sampling efficiency. (The statistical properties of the estimator will not be examined here.) Under the substitution $p = 1-q$, this becomes the first equation of Section 8.2.

With the estimator p in hand, and a count m of members of S discovered after n samplings, we can compute the a posteriori probability that the true size of S is $m+j$, for $j >= 0$. Let H_j denote the hypothesis that the true size of S is $m+j$, and let R denote the result that m members have been found in n samplings. Then

$$P\{R \mid H_j\} = [(m+j)!/m!j!] * [1-(1-p)^n]^m * [(1-p)^n]^j$$

We are interested in $P\{H_0 \mid R\}$, but to find it we need to assign a priori values to $P\{H_j\}$. Let the size of S be exponentially distributed

$$P\{H_j\} / P\{H_0\} = h^{(-j)}$$

for arbitrary h in $(0, 1)$. The value of h will be eliminated from the result.

The Bayesian method yields

$$P\{H_j \mid R\} / P\{H_0 \mid R\} = [(m+j)!/m!j!] * [h*(1-p)^n]^j$$

The reciprocal of the sum over $j >= 0$ of these ratios is

$$P\{H_0 \mid R\} = [1-h*(1-p)^n] ^ {(m+1)}$$

and the confidence estimate of Section 8.2 is the h -> 1 limit of this expression.

Full Copyright Statement

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.