Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: December 10, 2010

F. Templin, Ed.
Boeing Research & Technology

June 8, 2010

Virtual Enterprise Traversal (VET)
draft-templin-intarea-vet-15.txt

Abstract

Enterprise networks connect hosts and routers over various link types, and often also connect to provider networks and/or the global Internet. Enterprise network nodes require a means to automatically provision addresses/prefixes and support internetworking operation in a wide variety of use cases including Small Office, Home Office (SOHO) networks, Mobile Ad hoc Networks (MANETs), ISP networks, multi-organizational corporate networks and the interdomain core of the global Internet itself. This document specifies a Virtual Enterprise Traversal (VET) abstraction for autoconfiguration and operation of nodes in enterprise networks.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on December 10, 2010.

Table of Contents

1. Introduction

Enterprise networks [RFC4852] connect hosts and routers over various link types (see [RFC4861], Section 2.2). The term "enterprise network" in this context extends to a wide variety of use cases and deployment scenarios. For example, an "enterprise" can be as small as a SOHO network, as complex as a multi-organizational corporation, or as large as the global Internet itself. ISP networks are another example use case that fits well with the VET enterprise network model. Mobile Ad hoc Networks (MANETs) [RFC2501] can also be considered as a challenging example of an enterprise network, in that their topologies may change dynamically over time and that they may employ little/no active management by a centralized network administrative authority. These specialized characteristics for MANETs require careful consideration, but the same principles apply equally to other enterprise network scenarios.

This document specifies a Virtual Enterprise Traversal (VET) abstraction for autoconfiguration and internetworking operation, where addresses of different scopes may be assigned on various types of interfaces with diverse properties. Both IPv4/ICMPv4 [RFC0791][RFC0792] and IPv6/ICMPv6 [RFC2460][RFC4443] are discussed within this context (other network layer protocols are also considered). The use of standard DHCP [RFC2131] [RFC3315] is assumed unless otherwise specified.

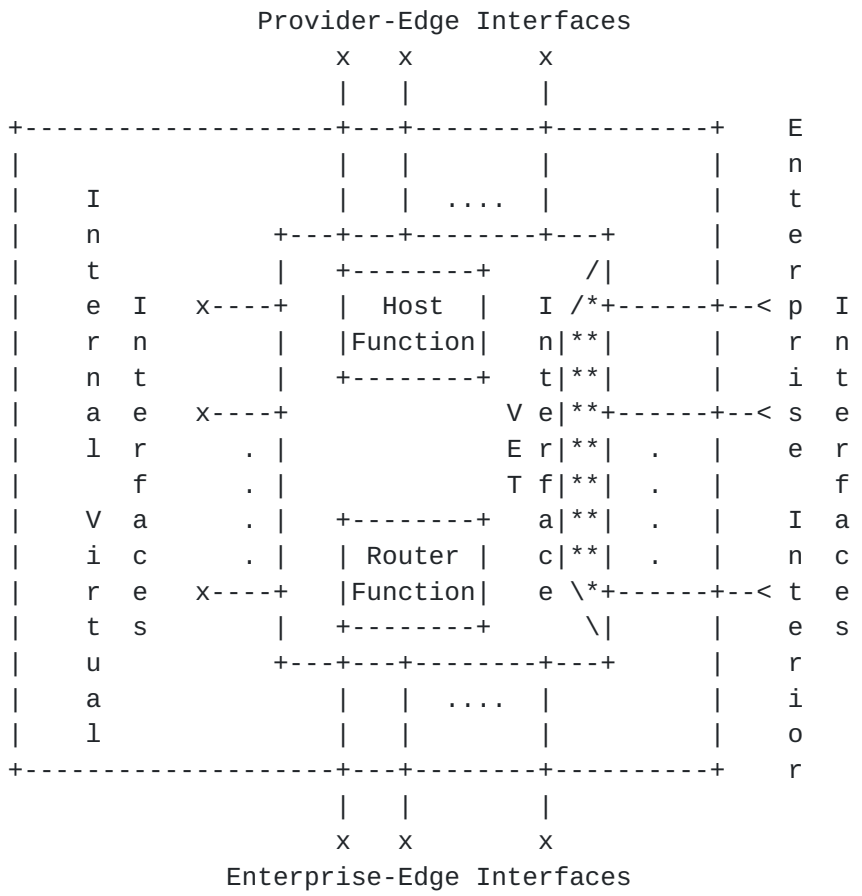

Figure 1: Enterprise Router (ER) Architecture

Figure 1 above depicts the architectural model for an Enterprise Router (ER). As shown in the figure, an ER may have a variety of interface types including enterprise-edge, enterprise-interior, provider-edge, internal-virtual, as well as VET interfaces used for encapsulating inner network layer protocol packets for transmission over outer IPv4 or IPv6 networks. The different types of interfaces are defined, and the autoconfiguration mechanisms used for each type are specified. This architecture applies equally for MANET routers, in which enterprise-interior interfaces correspond to the wireless multihop radio interfaces typically associated with MANETs. Out of scope for this document is the autoconfiguration of provider interfaces, which must be coordinated in a manner specific to the service provider's network.

Enterprise networks require a means for supporting both Provider-Independent (PI) and Provider-Aggregated (PA) addressing. This is especially true for enterprise network scenarios that involve mobility and multihoming. The VET specification provides adaptable mechanisms that address these and other issues in a wide variety of enterprise network use cases.

The VET framework builds on a Non-Broadcast Multiple Access (NBMA) [RFC2491] virtual interface model in a manner similar to other automatic tunneling technologies [RFC2529][RFC5214]. VET interfaces support the encapsulation of inner network layer protocol packets over IP networks (i.e., either IPv4 or IPv6). VET is also compatible with mid-layer encapsulation technologies including IPsec [RFC4301], and supports both stateful and stateless prefix delegation.

VET and its associated technologies (including the Subnetwork Encapsulation and Adaptation Layer (SEAL) [I-D.templin-intarea-seal]) are functional building blocks for a new Internetworking architecture based on the Internet Routing Overlay Network (IRON) [I-D.templin-iron] and Routing and Addressing in Networks with Global Enterprise Recursion (RANGER) [RFC5720] [I-D.russert-rangers]. Many of the VET principles can be traced to the deliberations of the ROAD group in January 1992, and also to still earlier initiatives including NIMROD [RFC1753] and the Catenet model for internetworking [CATENET] [IEN48] [RFC2775]. [RFC1955] captures the high-level architectural aspects of the ROAD group deliberations in a "New Scheme for Internet Routing and Addressing (ENCAPS) for IPNG".

VET is related to the present-day activities of the IETF INTAREA, AUTOCONF, DHC, IPv6, MANET, and V6OPS working groups, as well as the IRTF RRG working group.

2. Terminology

The mechanisms within this document build upon the fundamental principles of IP encapsulation. The term "inner" refers to the innermost {address, protocol, header, packet, etc.} *before* encapsulation, and the term "outer" refers to the outermost {address, protocol, header, packet, etc.} *after* encapsulation. VET also accommodates "mid-layer" encapsulations including the Subnetwork Encapsulation and Adaptation Layer (SEAL) [I-D.templin-intarea-seal], IPsec [RFC4301], etc.

The terminology in the normative references apply; the following terms are defined within the scope of this document:

Virtual Enterprise Traversal (VET)
: an abstraction that uses IP encapsulation to create overlays for traversing IPv4 and IPv6 enterprise networks.

enterprise network
: the same as defined in [RFC4852]. An enterprise network is further understood to refer to a cooperative networked collective of devices within a structured IP routing and addressing plan and

with a commonality of business, social, political, etc., interests. Minimally, the only commonality of interest in some enterprise network scenarios may be the cooperative provisioning of connectivity itself.

subnetwork
the same as defined in [RFC3819].

site
a logical and/or physical grouping of interfaces that connect a topological area less than or equal to an enterprise network in scope. From a network organizational standpoint, a site within an enterprise network can be considered as an enterprise unto itself.

Mobile Ad hoc Network (MANET)
a connected topology of mobile or fixed routers that maintain a routing structure among themselves over dynamic links. The characteristics of MANETs are defined in [RFC2501], Section 3, and a wide variety of MANETs share common properties with enterprise networks.

enterprise/site/MANET
throughout the remainder of this document, the term "enterprise network" is used to collectively refer to any of {enterprise, site, MANET}, i.e., the VET mechanisms and operational principles can be applied to enterprises, sites, and MANETs of any size or shape.

Enterprise Router (ER)
As depicted in Figure 1, an Enterprise Router (ER) is a fixed or mobile router that comprises a router function, a host function, one or more enterprise-interior interfaces, and zero or more internal virtual, enterprise-edge, provider-edge, and VET interfaces. At a minimum, an ER forwards outer IP packets over one or more sets of enterprise-interior interfaces, where each set connects to a distinct enterprise network.

Enterprise Border Router (EBR)
an ER that connects edge networks to the enterprise network and/or connects multiple enterprise networks together. An EBR is a tunnel endpoint router, and it configures a separate VET interface over each set of enterprise-interior interfaces that connect the EBR to each distinct enterprise network. In particular, an EBR may configure multiple VET interfaces - one for each distinct enterprise network. All EBRs are also ERs.

Enterprise Border Gateway (EBG)
an EBR that connects child enterprise networks to provider networks - either directly via a provider-edge interface or indirectly via another VET interface configured over a parent enterprise network. EBRs may act as EBGs on some VET interfaces and as ordinary EBRs on other VET interfaces. All EBGs are also EBRs.

VET host
any node (host or router) that configures a VET interface for host-operation only. Note that a node may configure some of its VET interfaces as host interfaces and others as router interfaces.

VET node
any node (host or router) that configures and uses a VET interface.

enterprise-interior interface
an ER's attachment to a link within an enterprise network. Packets sent over enterprise-interior interfaces may be forwarded over multiple additional enterprise-interior interfaces within the enterprise network before they are forwarded via an enterprise-edge interface, provider-edge interface, or a VET interface configured over a different enterprise network. Enterprise-interior interfaces connect laterally within the IP network hierarchy.

enterprise-edge interface
an EBR's attachment to a link (e.g., an Ethernet, a wireless personal area network, etc.) on an arbitrarily complex edge network that the EBR connects to an enterprise network and/or provider network. Enterprise-edge interfaces connect to lower levels within the IP network hierarchy.

provider-edge interface
an EBR's attachment to the Internet or to a provider network via which the Internet can be reached. Provider-edge interfaces connect to higher levels within the IP network hierarchy.

internal-virtual interface
an interface that is internal to an EBR and does not in itself directly attach to a tangible physical link (e.g., an Ethernet cable, a WiFi radio, etc.). Examples include a loopback interface, a virtual private network interface, or some form of tunnel interface.

VET link
: a virtual link that uses automatic tunneling to create an overlay network that spans an enterprise-interior routing region. VET links can be segmented (e.g., by filtering gateways) into multiple distinct segments that can be joined together by bridges or IP routers the same as for any link. Bridging would view the multiple (bridged) segments as a single VET link, whereas IP routing would view the multiple segments as multiple distinct VET links. VET link segments can further be partitioned into multiple logical areas, where each area is identified by a distinct set of EBGs.

  VET links in non-multicast enterprise networks are Non-Broadcast, Multiple Access (NBMA); VET links in enterprise networks that support multicast are multicast capable.

VET interface
: a VET node's attachment to a VET link. VET nodes configure each VET interface over a set of underlying enterprise-interior interfaces that connect to a routing region spanned by a single VET link. When there are multiple distinct VET links (each with their own distinct set of underlying interfaces), the VET node configures separate VET interfaces for each link.

  The VET interface encapsulates each inner packet in any mid-layer headers followed by an outer IP header, then forwards the packet on an underlying interface such that the Time to Live (TTL) - Hop Limit in the inner header is not decremented as the packet traverses the link. The VET interface therefore presents an automatic tunneling abstraction that represents the link as a single IP hop.

Provider Aggregated (PA) prefix
: a network layer protocol prefix that is delegated to an EBR by a provider network.

Provider-Independent (PI) address/prefix
: a network layer protocol prefix that is delegated to an EBR by an independent prefix registration authority.

Routing Locator (RLOC)
: a public-scope or enterprise-local-scope IP address that can appear in enterprise-interior and/or interdomain routing tables. Public-scope RLOCs are delegated to specific enterprise networks and routable within both the enterprise-interior and interdomain routing regions. Enterprise-local-scope RLOCs (e.g., IPv6 Unique Local Addresses [RFC4193], IPv4 privacy addresses [RFC1918], etc.) are self-generated by individual enterprise networks and routable

only within the enterprise-interior routing region.

ERs use RLOCs for operating the enterprise-interior routing protocol and for next-hop determination in forwarding packets addressed to other RLOCs. End systems can use RLOCs as addresses for end-to-end communications between peers within the same enterprise network. VET interfaces treat RLOCs as *outer* IP addresses during encapsulation.

Endpoint Interface iDentifier (EID)
a public-scope network layer address that is routable within an enterprise-edge or VET overlay network. EID prefixes are separate and distinct from any RLOC prefix space, but are mapped to RLOC addresses to support routing over VET interfaces.

EBRs use EIDs for operating the enterprise-edge or VET overlay network routing protocol and for next-hop determination in forwarding packets addressed to other EIDs. End systems can use EIDs as addresses for end-to-end communications between peers either within the same enterprise network or within different enterprise networks. VET interfaces treat EIDs as *inner* network layer addresses during encapsulation.

Note that an EID can also be used as an *outer* network layer address if there are nested encapsulations. In that case, the EID would appear as an RLOC to the innermost encapsulation.

The following additional acronyms are used throughout the document:

CGA - Cryptographically Generated Address
DHCP(v4, v6) - Dynamic Host Configuration Protocol
ECMP - Equal Cost Multi Path
FIB - Forwarding Information Base
ICMP - either ICMPv4 or ICMPv6
IP - either IPv4 or IPv6
ISATAP - Intra-Site Automatic Tunnel Addressing Protocol
NBMA - Non-Broadcast, Multiple Access
ND - Neighbor Discovery
NS/NA - Neighbor Solicitation/Advertisement
PIO - Prefix Information Option
PRL - Potential Router List
PRLNAME - Identifying name for the PRL
RIB - Routing Information Base
RIO - Route Information Option
RPF - Reverse Path Forwarding
RS/RA - Router Solicitation/Advertisement
SCMP - SEAL Control Message Protocol
SEAL - Subnetwork Encapsulation and Adaptation Layer

SLAAC - IPv6 StateLess Address AutoConfiguration

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119]. When used in lower case (e.g., must, must not, etc.), these words MUST NOT be interpreted as described in [RFC2119], but are rather interpreted as they would be in common English.

3. Enterprise Network Characteristics

Enterprise networks consist of links that are connected by Enterprise Routers (ERs) as depicted in Figure 1. ERs typically participate in a routing protocol over enterprise-interior interfaces to discover routes that may include multiple Layer 2 or Layer 3 forwarding hops. Enterprise Border Routers (EBRs) are ERs that connect edge networks to the enterprise network and/or join multiple enterprise networks together. Enterprise Border Gateways (EBGs) are EBRs that connect enterprise networks to provider networks.

Conceptually, an ER embodies both a host function and router function, and supports communications according to the weak end-system model [RFC1122]. The router function engages in the enterprise-interior routing protocol, connects any of the ER's edge networks to the enterprise networks, and may also connect the enterprise network to provider networks (see Figure 1). The host function typically supports network management applications, but may also support diverse applications typically associated with general-purpose computing platforms.

An enterprise network may be as simple as a small collection of ERs and their attached edge networks; an enterprise network may also contain other enterprise networks and/or be a subnetwork of a larger enterprise network. An enterprise network may further encompass a set of branch offices and/or nomadic hosts connected to a home office over one or several service providers, e.g., through Virtual Private Network (VPN) tunnels. Finally, an enterprise network may contain many internal partitions that are logical or physical groupings of nodes for the purpose of load balancing, organizational separation, etc. In that case, each internal partition resembles an individual segment of a bridged LAN.

Enterprise networks that comprise link types with sufficiently similar properties (e.g., Layer 2 (L2) address formats, maximum transmission units (MTUs), etc.) can configure a sub-IP layer routing service such that IP sees the network as an ordinary shared link the same as for a (bridged) campus LAN. In that case, a single IP hop is

sufficient to traverse the network without need for encapsulation. Enterprise networks that comprise link types with diverse properties and/or configure multiple IP subnets must also provide an enterprise-interior routing service that operates as an IP layer mechanism. In that case, multiple IP hops may be necessary to traverse the network such that care must be taken to avoid multi-link subnet issues [RFC4903].

In addition to other interface types, VET nodes configure VET interfaces that view all other nodes on the VET link as neighbors on a virtual NBMA link. VET nodes configure a separate VET interface for each distinct VET link to which they connect, and discover other EBRs on the link that can be used for forwarding packets to off-link destinations.

For each distinct enterprise network, a trust basis must be established and consistently applied. For example, in enterprise networks in which EBRs establish symmetric security associations, mechanisms such as IPsec [RFC4301] can be used to assure authentication and confidentiality. In other enterprise network scenarios, asymmetric securing mechanisms such as SEcure Neighbor Discovery (SEND) [RFC3971] may be necessary. Still other enterprise networks may find it sufficient to employ mechanisms (e.g., SEAL [I-D.templin-intarea-seal]) that can defeat off-path attacks.

Finally, in enterprise networks with a centralized management structure (e.g., a corporate campus network), an overlay routing/mapping service and a synchronized set of EBGs can provide sufficient infrastructure support for virtual enterprise traversal. In that case, the EBGs can provide a "default mapper" [I-D.jen-apt] service used for short-term packet forwarding until route-optimized paths between neighboring EBRs can be established. In enterprise networks with a distributed management structure (e.g., disconnected MANETs), peer-to-peer coordination between the EBRs themselves may be required. Recognizing that various use cases will entail a continuum between a fully distributed and fully centralized approach, the following sections present the mechanisms of Virtual Enterprise Traversal as they apply to a wide variety of scenarios.

4. Autoconfiguration

ERs, EBRs, EBGs, and VET hosts configure themselves for operation as specified in the following subsections.

4.1. Enterprise Router (ER) Autoconfiguration

ERs configure enterprise-interior interfaces and engage in any routing protocols over those interfaces.

When an ER joins an enterprise network, it first configures an IPv6 link-local address on each enterprise-interior interface and configures an IPv4 link-local address on each enterprise-interior interface that requires an IPv4 link-local capability. IPv6 link-local address generation mechanisms include Cryptographically Generated Addresses (CGAs) [RFC3972], IPv6 Privacy Addresses [RFC4941], StateLess Address AutoConfiguration (SLAAC) using EUI-64 interface identifiers [RFC4291] [RFC4862], etc. The mechanisms specified in [RFC3927] provide an IPv4 link-local address generation capability.

Next, the ER configures one or more RLOCs and engages in any routing protocols on its enterprise-interior interfaces. The ER can configure RLOCs via explicit management, DHCP autoconfiguration, pseudo-random self-generation from a suitably large address pool, or through an alternate autoconfiguration mechanism. The ER may optionally configure and assign a separate RLOC for each underlying interface, or it may configure only a single RLOC and assign it to a VET interface configured over the underlying interfaces (see Section 4.2.1). In the latter case, the ER can use the VET interface for link layer multiplexing and traffic engineering purposes as specified in Appendix B.

Alternatively (or in addition), the ER can request RLOC prefix delegations via an automated prefix delegation exchange over an enterprise-interior interface and can assign the prefix(es) on enterprise-edge interfaces. Note that in some cases, the same enterprise-edge interfaces may assign both RLOC and EID addresses if there is a means for source address selection. In other cases (e.g., for separation of security domains), RLOCs and EIDs are assigned on separate sets of enterprise-edge interfaces.

Pseudo-random self-generation of IPv6 RLOCs can be from a large public or local-use IPv6 address range (e.g., IPv6 Unique Local Addresses [RFC4193]). Pseudo-random self-generation of IPv4 RLOCs can be from a large public or local-use IPv4 address range (e.g., [RFC1918]). When self-generation is used alone, the ER continuously monitors the RLOCs for uniqueness, e.g., by monitoring the enterprise-interior routing protocol. (Note however that anycast RLOCs MAY be assigned to multiple enterprise interior interfaces; hence, monitoring for uniqueness applies only to RLOCs that are intended as unicast.)

DHCP generation of RLOCs uses standard DHCP procedures but may require support from relays within the enterprise network. For DHCPv6, relays that do not already know the RLOC of a server within the enterprise network forward requests to the 'All_DHCP_Servers' site-scoped IPv6 multicast group [RFC3315]. For DHCPv4, relays that do not already know the RLOC of a server within the enterprise network forward requests to the site-scoped IPv4 multicast group address 'All_DHCPv4_Servers', which SHOULD be set to 239.255.2.1 unless an alternate multicast group for the site is known. DHCPv4 servers that delegate RLOCs SHOULD therefore join the 'All_DHCPv4_Servers' multicast group and service any DHCPv4 messages received for that group.

A combined approach using both DHCP and self-generation is also possible when the ER configures both a DHCP client and relay that are connected, e.g., via a pair of back-to-back connected Ethernet interfaces, a tun/tap interface, a loopback interface, inter-process communication, etc. The ER first self-generates an RLOC from a temporary addressing range used only for the bootstrapping purpose of procuring an actual RLOC taken from a preferred addressing range. The ER then engages in the enterprise-interior routing protocol and performs a DHCP client/relay exchange using the temporary RLOC as the address of the relay. When the DHCP server delegates an actual RLOC address/prefix, the ER abandons the temporary RLOC and re-engages in the enterprise-interior routing protocol using an RLOC taken from the delegation.

In some enterprise network use cases (e.g., MANETs), assignment of RLOCs on enterprise-interior interfaces as singleton addresses (i.e., as addresses with /32 prefix lengths for IPv4, or as addresses with /128 prefix lengths for IPv6) MAY be necessary to avoid multi-link subnet issues. In other use cases, assignment of an RLOC on a VET interface as specified in Appendix B can provide link layer multiplexing and traffic engineering over multiple underlying interfaces using only a single IP address.

4.2. Enterprise Border Router (EBR) Autoconfiguration

EBRs are ERs that configure VET interfaces over distinct sets of underlying interfaces belonging to the same enterprise network; an EBR can connect to multiple enterprise networks, in which case it would configure multiple VET interfaces. In addition to the ER autoconfiguration procedures specified in Section 4.1, EBRs perform the following autoconfiguration operations.

4.2.1. VET Interface Initialization

EBRs configure a VET interface over a set of underlying interfaces belonging to the same enterprise network such that all other nodes on the VET link appear as single-hop neighbors from the standpoint of the inner network layer protocol through the use of encapsulation. If there are multiple inner network layer protocols (e.g., IPv4, IPv6, OSI, etc.) the EBR configures a separate VET interface for each protocol.

After the EBR configures a VET interface, it binds an RLOC to the interface to serve as the source address for outer IP packets then assigns link-local addresses to the interface if necessary. When IPv6 and IPv4 are used as the inner/outer protocols (respectively), the EBR autoconfigures an IPv6 link-local address on the VET interface using a modified EUI-64 interface identifier based on the IPv4 address (see Section 2.2.1 of [RFC5342]). Link-local address configuration for other inner/outer protocol combinations is through administrative configuration, random self-generation (e.g., [RFC4941], etc.) or through an unspecified alternate method. However, link-local address configuration for other inner/outer protocol combinations may not be necessary if a non-link-local address can be configured through other means (e.g., administrative configuration, DHCP, etc.).

The EBR next discovers a Potential Router List (PRL) that includes the RLOC addresses of EBGs. The PRL names the VET interface, and is specific to the address family of the inner network layer protocol (e.g., IPv4, IPv6, OSI, etc.). If there are multiple address families, then there will be multiple VET interfaces; each with its own PRL.

The PRL can be discovered through information conveyed in the enterprise-interior routing protocol, through a DHCP option [I-D.templin-isatap-dhcp], etc. In multicast-capable enterprise networks, EBRs can also listen for advertisements on the 'rasadv' [RASADV] multicast group address.

Whether or not other information is available, the EBR can resolve an identifying name for the PRL ('PRLNAME') formed as 'hostname.domainname', where 'hostname' is an enterprise-specific name string and 'domainname' is an enterprise-specific Domain Name System (DNS) suffix [RFC1035]. The EBR discovers 'PRLNAME' through manual configuration, the DHCP Domain Name option [RFC2132], 'rasadv' protocol advertisements, link-layer information (e.g., an IEEE 802.11 Service Set Identifier (SSID)), or through some other means specific to the enterprise network. The EBR can also obtain 'PRLNAME' as part of an arrangement with a private-sector PI prefix vendor (see:

Section 4.2.3).

In the absence of other information, the EBR sets the 'hostname' component of 'PRLNAME' to "isatapv2" and sets the 'domainname' component to an enterprise-specific DNS suffix (e.g., "example.com"). Isolated enterprise networks that do not connect to the outside world may have no enterprise-specific DNS suffix, in which case the 'PRLNAME' consists only of the 'hostname' component. (Note that the default hostname "isatapv2" is intentionally distinct from the convention specified in [RFC5214].)

After discovering 'PRLNAME', the EBR resolves the name into a list of RLOC addresses through a name service lookup. For centrally managed enterprise networks, the EBR resolves 'PRLNAME' using an enterprise-local name service (e.g., the DNS). For enterprises with no centralized management structure, the EBR resolves 'PRLNAME' using Link-Local Multicast Name Resolution (LLMNR) [RFC4795] over the VET interface. In that case, all EBGs in the PRL respond to the LLMNR query, and the EBR accepts the union of all responses.

4.2.2. Provider-Aggregated (PA) EID Prefix Autoconfiguration

EBRs that connect their enterprise networks to a provider network obtain Provider-Aggregated (PA) EID prefixes through stateful and/or stateless autoconfiguration mechanisms. The stateful and stateless approaches are discussed in the following subsections.

4.2.2.1. Stateful Prefix Delegation

For IPv4, EBRs acquire IPv4 PA EID prefixes via an automated IPv4 prefix delegation exchange, explicit management, etc.

For IPv6, EBRs acquire IPv6 PA EID prefixes via DHCPv6 Prefix Delegation exchanges with an EBG acting as a DHCP relay/server. In particular, the EBR (acting as a requesting router) can use DHCPv6 prefix delegation [RFC3633] over the VET interface to obtain prefixes from the server (acting as a delegating router). The EBR obtains prefixes using either a 2-message or 4-message DHCPv6 exchange [RFC3315]. For example, to perform the 2-message exchange, the EBR's DHCPv6 client forwards a Solicit message with an IA_PD option to its DHCPv6 relay, i.e., the EBR acts as a combined client/relay (see Section 4.1). The relay then forwards the message over the VET interface using VET encapsulation (see: Section 5.4) to an EBG which either services the request or relays it further. The forwarded Solicit message will elicit a reply from the server containing prefix delegations. The EBR can also propose a specific prefix to the DHCPv6 server per Section 7 of [RFC3633]. The server will check the proposed prefix for consistency and uniqueness, then return it in the

reply to the EBR if it was able to perform the delegation.

After the EBR receives IPv4 and/or IPv6 prefix delegations, it can provision the prefixes on enterprise-edge interfaces as well as on other VET interfaces configured over child enterprise networks for which it acts as an EBG. The EBR can also provision the prefixes on enterprise-interior interfaces to service any hosts attached to the link.

The prefix delegations remain active as long as the EBR continues to renew them before lease lifetimes expire. The lease lifetime also keeps the delegation state active even if communications between the EBR and delegation server are disrupted for a period of time (e.g., due to an enterprise network partition, power failure, etc.).

Stateful prefix delegation for non-IP protocols is out of scope.

4.2.2.2. Stateless Prefix Delegation

When IPv6 and IPv4 are used as the inner and outer protocols, respectively, a stateless IPv6 PA prefix delegation capability is available using the mechanisms specified in [RFC5569][I-D.ietf-softwire-ipv6-6rd]. EBRs can use these mechanisms to statelessly configure IPv6 PA prefixes that embed one of the EBR's IPv4 RLOCs.

Using this stateless prefix delegation, if the IPv4 RLOC changes the IPv6 prefix also changes and the EBR is obliged to renumber any interfaces on which sub-prefixes from the prefix are assigned. This method may therefore be most suitable for enterprise networks in which IPv4 RLOC assignments rarely change, or in enterprise networks in which only services that do not depend on a long-term stable IPv6 prefix (e.g., client-side web browsing) are used.

Stateless prefix delegation for other protocol combinations is out of scope.

4.2.3. Provider-Independent (PI) EID Prefix Autoconfiguration

EBRs can acquire Provider Independent (PI) prefixes to facilitate multihoming, mobility and traffic engineering without requiring site-wide renumbering events. These PI prefixes are made available to EBRs through provider-independent registries and without need to coordinate with Internet Service Provider networks.

EBRs that connect major enterprise networks (e.g., large corporations, academic campuses, ISP networks, etc.) to a parent enterprise network and/or the global Internet can acquire highly-

aggregated PI prefixes (e.g., an IPv6 ::/20, an IPv4 /16, etc.) through a registration authority such as the Internet Assigned Numbers Authority (IANA) or a major regional Internet registry. EBRs that connect small enterprise networks (e.g., SOHO networks, MANETs, etc.) to a parent enterprise network can acquire de-aggregated PI prefixes through arrangements with a PI prefix commercial vendor organization.

After an EBR receives PI prefixes, it can sub-delegate portions of the prefixes on enterprise-edge interfaces, on other VET interfaces for which it is configured as an EBG and on enterprise-interior interfaces to service any hosts attached to the link. The EBR can also sub-delegate portions of its PI prefixes to requesting routers within child enterprise networks. These requesting routers consider their sub-delegated portions of the PI prefix as PA, and consider the delegating routers as their points of connection to a provider network.

4.3. Enterprise Border Gateway (EBG) Autoconfiguration

EBGs are EBRs that connect child enterprise networks to provider networks via provider-edge interfaces and/or via VET interfaces configured over parent enterprise networks. EBGs autoconfigure their provider-edge interfaces in a manner that is specific to the provider connections, and they autoconfigure their VET interfaces that were configured over parent enterprise networks using the EBR autoconfiguration procedures specified in Section 4.2.

For each of its VET interfaces configured over a child enterprise network, the EBG initializes the interface the same as for an ordinary EBR (see Section 4.2.1). It then arranges to add one or more of its RLOCs associated with the child enterprise network to the PRL. In particular, for each VET interface configured over a child enterprise network the EBG adds the RLOCs to name service resource records for 'PRLNAME'.

EBGs respond to LLMNR queries for 'PRLNAME' on VET interfaces configured over child enterprise networks with a distributed management structure.

EBGs configure a DHCP relay/server on VET interfaces configured over child enterprise networks that require DHCP services.

To avoid looping, EBGs MUST NOT configure a default route on a VET interface configured over a child enterprise network interface.

4.4. VET Host Autoconfiguration

Nodes that cannot be attached via an EBR's enterprise-edge interface (e.g., nomadic laptops that connect to a home office via a Virtual Private Network (VPN)) can instead be configured for operation as a simple host connected to the VET interface. Such VET hosts perform the same VET interface initialization and border gateway discovery procedures as specified for EBRs in Section 4.2.1, but they configure their VET interfaces as host interfaces (and not router interfaces). Note also that a node may be configured as a host on some VET interfaces and as an EBR/EBG on other VET interfaces.

5. Internetworking Operation

Following the autoconfiguration procedures specified in Section 4, ERs, EBRs, EBGs, and VET hosts engage in normal internetworking operations as discussed in the following sections.

5.1. Routing Protocol Participation

ERs engage in any intra-enterprise routing protocols over enterprise-interior interfaces to exchange routing information for forwarding IP packets with RLOC addresses. EBRs and EBGs can additionally engage in any inter-enterprise routing protocols over VET, enterprise-edge and provider-edge interfaces to exchange routing information for forwarding IP packets with EID addresses. Note that the EID-based inter-enterprise IP routing domains are separate and distinct from any RLOC-based enterprise interior IP routing domains.

Routing protocol participation on non-multicast VET interfaces uses the NBMA interface model, e.g., in the same manner as for OSPF over NBMA interfaces [RFC5340], while routing protocol participation on multicast-capable VET interfaces uses the standard multicast interface model. EBRs use the list of EBGs in the PRL (see: Section 4.2.1) as an initial list of neighbors for inter-enterprise routing protocol participation.

5.1.1. PI Prefix Routing Considerations

EBRs that connect large enterprise networks to the global Internet advertise their EID PI prefixes directly into the Internet default-free RIB via the Border Gateway Protocol (BGP) [RFC4271] the same as for a major service provider network. EBRs that connect large enterprise networks to provider networks can instead advertise their EID PI prefixes into the providers' routing system(s) if the provider networks are configured to accept them.

EBRs that connect small enterprise networks to provider networks obtain one or more public IP address(es) (i.e., either EID or RLOC IP address) then dynamically register the mapping of their PI prefixes to the public IP address. The registrations are through secured end-to-end exchanges between the EBR and a server in the PI prefix vendor's network (e.g., through a vendor-specific short http(s) transaction). The PI prefix vendor network then acts as a virtual "home" enterprise network that connects its customer small enterprise networks to the Internet routing system with no arrangements needed with the customers' Internet Service Providers. The customer small enterprise networks in turn appear as mobile components of the PI prefix vendor's network, i.e., the customer networks are always "away from home".

Further details on routing for PI prefixes is discussed in "The Internet Routing Overlay Network (IRON)" [I-D.templin-iron] and "Fib Suppression with Virtual Aggregation" [I-D.ietf-grow-va].

5.2. Default Route Configuration and Selection

Configuration of default routes in the presence of VET interfaces must be carefully coordinated according to the inner and outer network protocols. If the inner and outer protocols are different (e.g., IPv6 within IPv4) then default routes of the inner protocol version can be configured with next-hops corresponding to default routers on a VET interface while default routes of the outer protocol version can be configured with next-hops corresponding to default routers on an underlying interface.

If the inner and outer protocols are the same (e.g., IPv4 within IPv4), care must be taken in setting the default route to avoid ambiguity. For example, if default routes are configured on the VET interface then more-specific routes could be configured on underlying interfaces to avoid looping. In a preferred method, however, multiple default routes can be configured with some having next-hops corresponding to (EID-based) default routers on VET interfaces and others having next-hops corresponding to (RLOC-based) default routers on underlying interfaces. In that case, special next-hop determination rules must be used (see: Section 5.4).

5.3. Address Selection

When permitted by policy and supported by enterprise interior routing, VET nodes can avoid encapsulation through communications that directly invoke the outer IP protocol using RLOC addresses instead of EID addresses for end-to-end communications. For example, an enterprise network that provides native IPv4 intra-enterprise services can provide continued support for native IPv4 communications

even when encapsulated IPv6 services are available for inter-enterprise communications. In other enterprise network scenarios, the use of EID-based communications (i.e., instead of RLOC-based communications) may be necessary and/or beneficial to support address scaling, NAT traversal avoidance, security domain separation, site multihoming, traffic engineering, etc. .

VET nodes can use source address selection rules (e.g., based on name service information) to determine whether to use EID-based or RLOC-based addressing. The remainder of this section discusses internetworking operation for EID-based communications using the VET interface abstraction.

5.4. Next Hop Determination

VET nodes perform normal next-hop determination via longest prefix match, and send packets according to the most-specific matching entry in the FIB. If the FIB entry has multiple next-hop addresses, the EBR selects the next-hop with the best metric value. If multiple next hops have the same metric value, the VET node can use Equal Cost Multi Path (ECMP) to forward different flows via different next-hop addresses, where flows are determined, e.g., by computing a hash of the inner packet's source address, destination address and flow label fields.

If the VET node has multiple default routes of the same inner and outer protocol versions, with some corresponding to EID-based default routers and others corresponding to RLOC-based default routers, it must perform source address based selection of a default route. In particular, if the packet's source address is taken from an EID prefix the VET node selects a default route configured over the VET interface; otherwise, it selects a default route configured over an underlying interface.

As a last resort when there is no matching entry in the FIB (i.e., not even default), VET nodes can discover next-hop addresses within the enterprise network through on-demand name service queries for the EID prefix taken from a packet's destination address (or, by some other inner address to outer address mapping mechanism). For example, for the IPv6 destination address '2001:DB8:1:2::1' and 'PRLNAME' "isatapv2.example.com" the VET node can perform a name service lookup for the domain name: '0.0.1.0.0.0.8.b.d.0.1.0.0.2.ip6.isatapv2.example.com'.

Name-service lookups in enterprise networks with a centralized management structure use an infrastructure-based service, e.g., an enterprise-local DNS. Name-service lookups in enterprise networks with a distributed management structure and/or that lack an

infrastructure-based name service instead use LLMNR over the VET interface. When LLMNR is used, the EBR that performs the lookup sends an LLMNR query (with the prefix taken from the IP destination address encoded in dotted-nibble format as shown above) and accepts the union of all replies it receives from other EBRs on the VET interface. When an EBR receives an LLMNR query, it responds to the query IFF it aggregates an IP prefix that covers the prefix in the query. If the name-service lookup succeeds, it will return RLOC addresses (e.g., in DNS A records) that correspond to next-hop EBRs to which the VET node can forward packets.

5.5. VET Interface Encapsulation/Decapsulation

VET interfaces encapsulate inner network layer packets in any necessary mid-layer headers and trailers (e.g., IPsec [RFC4301], etc.) followed by a SEAL header (if necessary) followed by an outer UDP header (if necessary) followed by an outer IP header. Following all encapsulations, the VET interface submits the encapsulated packet to the outer IP forwarding engine for transmission on an underlying interface. The following sections provide further details on encapsulation:

5.5.1. Inner Network Layer Protocol

The inner network layer protocol sees the VET interface as an ordinary network interface, and views the outer network layer protocol as an L2 transport. The inner- and outer network layer protocol types are mutually independent and can be used in any combination. Inner network layer protocol types include IPv6 [RFC2460] and IPv4 [RFC0791], but they may also include non-IP protocols such as OSI/CLNP [RFC0994][RFC1070][RFC4548].

5.5.2. Mid-Layer Encapsulation

VET interfaces that use mid-layer encapsulations encapsulate each inner network layer packet in any mid-layer headers and trailers as the first step in a potentially multi-layer encapsulation.

5.5.3. SEAL Encapsulation

Following any mid-layer encapsulations, VET interfaces that use SEAL add a SEAL header as specified in [I-D.templin-intarea-seal]. Inclusion of a SEAL header MUST be applied uniformly between all nodes on the VET link. Note that when a VET interface sends a SEAL-encapsulated packet to a VET node that does not use SEAL encapsulation, it may receive an ICMP "port unreachable" or "protocol unreachable" depending on whether/not an outer UDP header is included.

SEAL encapsulation is used on VET links that require path MTU mitigations due to encapsulation overhead and/or mechanisms for VET interface neighbor coordination. When SEAL encapsulation is used, the VET interface sets the 'Next Header' value in the SEAL header to the IP protocol number associated with either the mid-layer encapsulation or the IP protocol number of the inner network layer (if no mid-layer encapsulation is used).

The VET interface sets the other fields in the SEAL header as specified in [I-D.templin-intarea-seal]. For SEAL first-segments, the VET interface also sets the R and D flags in the SEAL header in order to control the operation of the SCMP Redirect function (see: Section 5.7.3). The VET interface sets R=1 in the SEAL header of each packet for which it is willing to receive a Redirect message from the neighbor, and sets D=1 in the SEAL header of each packet that it wishes the neighbor to discard the packet after sending a redirect (if necessary). The VET interface otherwise sets R=0 and D=0.

5.5.4. Outer UDP Header Encapsulation

Following any mid-layer and/or SEAL encapsulations, VET interfaces that use UDP encapsulation add an outer UDP header. Inclusion of an outer UDP header must be applied uniformly between all nodes on the VET link. Note that when a VET interface sends a UDP-encapsulated packet to a node that does not recognize the UDP port number, it may receive an ICMP "port unreachable" message.

VET interfaces use UDP encapsulation on VET links that may traverse Network Address Translators (NATs) and/or legacy networking gear that only recognizes certain network layer protocols, e.g., Equal Cost MultiPath (ECMP) routers, Link Aggregation Gateways (LAGs), etc. When UDP encapsulation is used, the VET interface encapsulates the mid-layer packet in an outer UDP header then sets the UDP port numbers as specified for the outermost mid-layer protocol (e.g., IPsec [RFC3947][RFC3948], etc.) When SEAL [I-D.templin-intarea-seal] is used as the outermost mid-layer protocol, the VET interface sets the UDP source port number to a hash calculated over the inner network layer {destination, source} values or (optionally) over the inner network layer {dest addr, source addr, protocol, dest port, source port} values. The VET interface uses a hash function of its own choosing, but it MUST be consistent in the manner in which the hash is applied..

For VET links configured over IPv4 enterprise networks, the VET interface sets the UDP checksum field to zero. For VET links configured over IPv6 enterprise networks, the VET interface instead calculates the UDP checksum and set the calculated value in the

checksum field as required for UDP operation over IPv6.

5.5.5. Outer IP Header Encapsulation

Following any mid-layer, SEAL and/or UDP encapsulations, the VET interface adds an outer IP header. Outer IP header construction is the same as specified for ordinary IP encapsulation (e.g., [RFC2003], [RFC2473], [RFC4213], etc.) except that the "TTL/Hop Limit", "Type of Service/Traffic Class" and "Congestion Experienced" values in the inner network layer header are copied into the corresponding fields in the outer IP header. The VET interface also sets the IP protocol number to the appropriate value for the first protocol layer within the encapsulation (e.g., UDP, SEAL, IPsec, etc.). When IPv6 is used as the outer IP protocol, the VET interface sets the flow label value in the outer IPv6 header the same as described in [I-D.carpenter-flow-ecmp].

5.5.6. Decapsulation

When a VET interface receives an encapsulated packet, it retains the outer headers and processes the SEAL header as specified in [I-D.templin-intarea-seal]. Following SEAL-layer reassembly (if necessary), the VET interface further examines the R and D bits in the SEAL header to determine whether Redirects are permitted and whether the packet is to be discarded following redirect determination (see: Section 5.7.3).

Next, if the packet will be forwarded from the receiving VET interface into a forwarding VET interface, the VET node copies the "TTL/Hop Limit", "Type of Service/Traffic Class" and "Congestion Experienced" values in the outer IP header received on the receiving VET interface into the corresponding fields in the outer IP header to be sent over the forwarding VET interface (i.e., the values are transferred between outer headers and *not* copied from the inner network layer header). This is true even if the packet is forwarded out the same VET interface that it arrived on, and necessary to support diagnostic functions (e.g., traceroute) and avoid looping.

During decapsulation, when the next-hop is via a non-VET interface, the "Congestion Experienced" value in the outer IP header is copied into the corresponding field in the inner network layer header.

5.6. Mobility and Multihoming Considerations

EBRs that travel between distinct enterprise networks must either abandon their PA prefixes that are relative to the "old" enterprise and obtain PA prefixes relative to the "new" enterprise, or somehow coordinate with a "home" enterprise to retain ownership of the

prefixes. In the first instance, the EBR would be required to coordinate a network renumbering event using the new PA prefixes [RFC4192][RFC5887]. In the second instance, an ancillary mobility management mechanism is required.

EBRs can retain their PI prefixes as they travel between distinct enterprise networks as long as they update their PI prefix to public IP address mappings with their PI prefix vendors. This is accomplished by performing the same PI prefix vendor-specific short transactions as specified in Section 5.1.1. In this way, EBRs can update their PI prefix to RLOC mappings in real time as their RLOCs change.

The EBGs of a multihomed enterprise network participate in a private inner network layer routing protocol instance between themselves (possibly over an alternate topology) to accommodate network partitions/merges as well as intra-enterprise mobility events.

5.7. Neighbor Coordination on VET Interfaces using SEAL

VET interfaces that use SEAL use the SEAL Control Message Protocol (SCMP) as specified in Section 4.5 of [I-D.templin-intarea-seal] to coordinate reachability, routing information, and mappings between the inner and outer network layer protocols. SCMP directly parallels the IPv6 Neighbor Discovery (ND) [RFC4191][RFC4861] and ICMPv6 [RFC4443] protocols, but operates from within the tunnel and supports operation for any combinations of inner and outer network layer protocols.

The following subsections discuss VET interface neighbor coordination using SCMP:

5.7.1. Router Discovery

VET hosts and EBRs can send SCMP Router Solicitation (RS) messages to one or more EBGs in the PRL to receive solicited SCMP Router Advertisements (RAs). They then process the RAs the same as for IPv6 ND RA messages, except that they ignore the 'M' and 'O' bits.

When an EBG receives an SCMP RS message on a VET interface, it prepares a solicited SCMP RA message. The RA includes Router Lifetimes, Default Router Preferences, PIOs and any other options/ parameters that the EBG is configured to include. If necessary, the EBG also includes Route Information Options (RIOs) formatted as specified in Section 5.7.3, i.e., the RIO may contain both IPv6 and non-IPv6 prefixes in RIOs as identified by an Address Family designator.

5.7.2. Neighbor Unreachability Detection

VET nodes perform Neighbor Unreachability Detection (NUD) on VET interface neighbors by monitoring hints of forward progress as evidence that a neighbor is reachable. SEAL includes an explicit acknowledgement mechanism that can provide hints of forward progress. When data packets are flowing, the VET node can periodically set the A bit in data packets to elicit Neighbor Advertisement (NA) messages from the neighbor. When no data packets are flowing, the VET node can send periodic Neighbor Solicitation (NS) messages for the same purpose.

Responsiveness to routing changes is directly related to the delay in detecting that a neighbor has gone unreachable. In order to provide responsiveness comparable to dynamic routing protocols, a reasonably short neighbor reachable time (e.g., 5sec) SHOULD be used.

Additionally, a VET node may receive outer IP ICMP "Destination Unreachable; net / host unreachable" messages from an ER on the path indicating that the path to a VET neighbor may be failing. The node SHOULD first check the packet-in-error to obtain reasonable assurance that the ICMP message is authentic. If the node receives excessive ICMP unreachable errors through multiple RLOCs associated with the same FIB entry, it SHOULD delete the FIB entry and allow subsequent packets to flow through a different route.

5.7.3. Redirect Function

A VET node (i.e., the redirectee) may receive a redirect message when it forwards packets over a VET interface to a neighboring VET node (i.e., the redirector). The redirector will forward the packet and return an SCMP Redirect message if necessary to inform the redirectee of a better next hop. Unlike ordinary ICMP redirects, the redirector sends an SCMP Redirect message (subject to rate limiting) whenever it receives a packet with R=1 in the SEAL header for which there is a better next hop on the same VET interface that it arrived on regardless of whether the inner source address of the packet was on-link. The redirector also discards packets with D=1 in the SEAL header after sending a redirect (if necessary) and before forwarding the packet to the next hop.

The SCMP Redirect message is formatted the same as for ordinary ICMPv6 redirect messages (see Section 4.5 of [RFC4861]), except that the Destination and Target Address fields are unnecessary and therefore omitted. The format of the SCMP Redirect message is shown in Figure 2

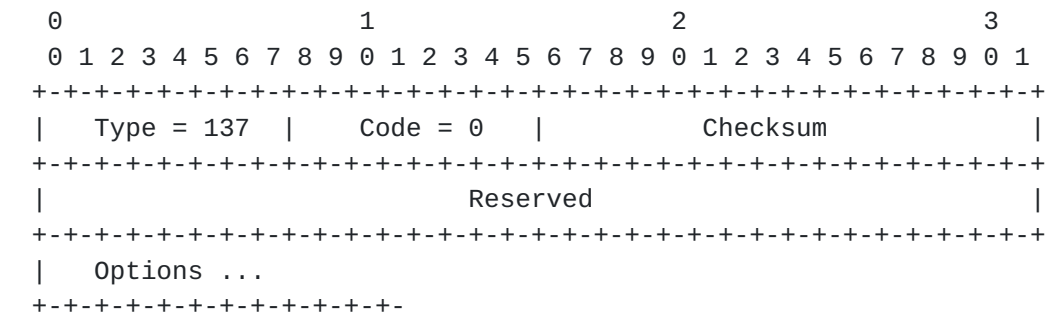

Figure 2: SCMP Redirect Message Format

The redirector then adds any necessary Options to the Redirect message. It first includes one or more Target Link-Layer Address Options (TLLAOs) (see: Section 4.6.1 of [RFC4861]) that include RLOCs corresponding to better next hops. The TLLAO formats for IPv4 and IPv6 RLOCs are shown in Figure 3 and Figure 4:

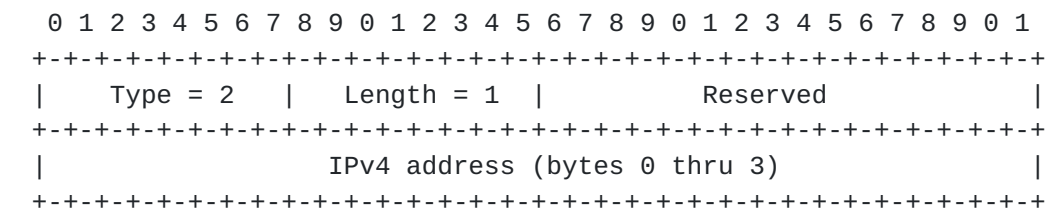

Figure 3: SCMP TLLAO Option for IPv4 RLOCs

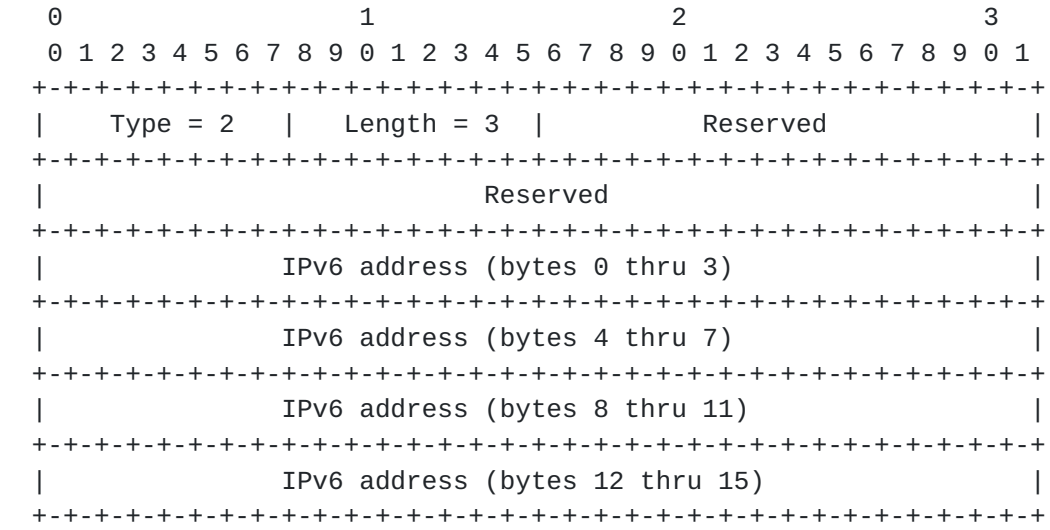

Figure 4: SCMP TLLAO Option for IPv6 RLOCs

The redirector next includes a Route Information Option (RIO) (see: [RFC4191]) that contains a prefix from its FIB that covers the

destination address of the original packet. SCMP uses a modified version of the RIO option formatted as shown in Figure 5:

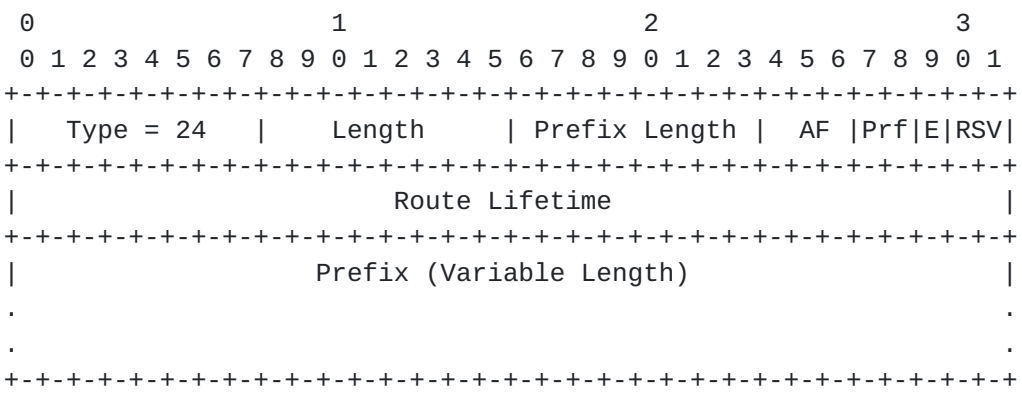

Figure 5: SCMP Route Information Option Format

In this modified format, the redirector prepares the Route Lifetime and Prefix fields in the RIO option the same as specified in [RFC4191]. It then sets the fields in the header as follows:

- the 'Type', 'Length' and 'Prf' fields are set the same as specified in [RFC4191].
- the 'RSV' field is set to 0.
- he 'Length' field is set to 1, 2, or 3 as specified in [RFC4191], or set to 4 if the 'Prefix Length' is greater than 128 in order to accommodate prefixes of non-IP protocols of up to 192 bits in length.
- the 'Prefix Length' field ranges from 0 to 192. The 'Prefix' field is 0, 8, 16 or 24 octets depending on the length, and the embedded prefix MAY be up to 192 bits in length.
- bits 24 - 26 are used to contain an 'Address Family (AF)' value that indicates the embedded prefix protocol type. This document defines the following values for AF:
  - 000 - IPv4
  - 001 - IPv6
  - 010 - OSI/CLNP NSAP
- the 'E' bit is set to 1 if this prefix is assigned to an End User Network, and set to 0 otherwise.

Following the RIO option, the redirector includes any other necessary options (e.g., SEND options) followed by a Redirected Header containing the leading portion of the packet that triggered the redirect as the final option in the message. The redirector then encapsulates the Redirect message the same as for any other SCMP message and sends it to the redirectee.

When the redirectee receives the Redirect, it first authenticates the message (i.e., by checking the SEAL_ID in the Redirected Header, by examining SEND options, etc.) then uses the EID prefix in the RIO with its respective lifetime to update its FIB. The redirectee also caches the IPv4 or IPv6 addresses in TLLAOs as the layer 2 addresses of potential next-hops.

The redirectee retains the FIB entry created as a result of receipt of an SCMP Redirect until the route lifetime expires, or until the redirected target neighbor becomes unreachable. In this way, RLOC liveness detection parallels IPv6 Neighbor Unreachability Detection as discussed in the next section.

5.7.4. Mobility

When a VET node moves to a new network point of attachment resulting in the change of an old RLOC to a new RLOC, it informs any correspondents of the change by sending specially-crafted SCMP Neighbor Advertisement (NA) messages. The VET node can ensure reliable delivery of the NA messages by setting the 'A' bit in the SEAL header in order to receive an explicit acknowledgement. The VET node SHOULD retry up to three times to get an explicit acknowledgement before abandoning the attempt.

The NA messages use the new RLOC as the outer IP source address and include the old RLOC in a Source Link Layer Address Option (SLLAO) formatted exactly as specified for TLLAOs in Section 5.7.3. When the neighbor receives the NA, it authenticates the message then replaces the old RLOC address with the new RLOC address. Methods for authenticating the NA are out of scope for this document.

5.8. Neighbor Coordination on VET Interfaces using IPsec

VET interfaces that use IPsec encapsulation use the Internet Key Exchange protocol, version 2 (IKEv2) [RFC4306] to manage security association setup and maintenance. The IKEv2 can be seen as a logical equivalent of the SEAL SCMP in terms of VET interface neighbor coordinations. In particular, IKEv2 also provides mechanisms for redirection [RFC5685] and mobility [RFC4555].

IPsec additionally provides an extended Identification field and

integrity check vector; these features allow IPsec to utilize outer IP fragmentation and reassembly with less risk of exposure to data corruption due to reassembly misassociations. On the other hand, IPsec entails the use of symmetric security associations and hence may not be appropriate to all enterprise network use cases.

5.9. Multicast

In multicast-capable deployments, ERs provide an enterprise-wide multicasting service (e.g., Simplified Multicast Forwarding (SMF) [I-D.ietf-manet-smf], Protocol Independent Multicast (PIM) routing, Distance Vector Multicast Routing Protocol (DVMRP) routing, etc.) over their enterprise-interior interfaces such that outer IP multicast messages of site-scope or greater scope will be propagated across the enterprise network. For such deployments, VET nodes can also provide an inner multicast/broadcast capability over their VET interfaces through mapping of the inner multicast address space to the outer multicast address space. In that case, operation of link-scoped (or greater scoped) inner multicasting services (e.g., a link-scoped neighbor discovery protocol) over the VET interface is available, but should be used sparingly to minimize enterprise-wide flooding.

VET nodes encapsulate inner multicast messages sent over the VET interface in any mid-layer headers (e.g., UDP, SEAL, IPsec, etc.) followed by an outer IP header with a site-scoped outer IP multicast address as the destination. For the case of IPv6 and IPv4 as the inner/outer protocols (respectively), [RFC2529] provides mappings from the IPv6 multicast address space to a site-scoped IPv4 multicast address space (for other encapsulations, mappings are established through administrative configuration or through an unspecified alternate static mapping).

Multicast mapping for inner multicast groups over outer IP multicast groups can be accommodated, e.g., through VET interface snooping of inner multicast group membership and routing protocol control messages. To support inner-to-outer multicast address mapping, the VET interface acts as a virtual outer IP multicast host connected to its underlying interfaces. When the VET interface detects that an inner multicast group joins or leaves, it forwards corresponding outer IP multicast group membership reports on an underlying interface over which the VET interface is configured. If the VET node is configured as an outer IP multicast router on the underlying interfaces, the VET interface forwards locally looped-back group membership reports to the outer IP multicast routing process. If the VET node is configured as a simple outer IP multicast host, the VET interface instead forwards actual group membership reports (e.g., IGMP messages) directly over an underlying interface.

Since inner multicast groups are mapped to site-scoped outer IP multicast groups, the VET node MUST ensure that the site-scope outer IP multicast messages received on the underlying interfaces for one VET interface do not "leak out" to the underlying interfaces of another VET interface. This is accommodated through normal site-scoped outer IP multicast group filtering at enterprise network boundaries.

5.10. Service Discovery

VET nodes can perform enterprise-wide service discovery using a suitable name-to-address resolution service. Examples of flooding-based services include the use of LLMNR [RFC4795] over the VET interface or multicast DNS (mDNS) [I-D.cheshire-dnsext-multicastdns] over an underlying interface. More scalable and efficient service discovery mechanisms are for further study.

5.11. Enterprise Network Partitioning

An enterprise network can be partitioned into multiple distinct logical groupings. In that case, each partition configures its own distinct 'PRLNAME' (e.g., 'isatapv2.zone1.example.com', 'isatapv2.zone2.example.com', etc.).

EBGs can further create multiple IP subnets within a partition by sending RAs with PIOs containing different IPv6 prefixes to different groups of nodes. EBGs can identify subnets, e.g., by examining RLOC prefixes, observing the enterprise interior interfaces over which RSs are received, etc.

5.12. EBG Prefix State Recovery

EBGs retain explicit state that tracks the inner PA prefixes delegated to EBRs within the enterprise network, e.g., so that packets are delivered to the correct EBRs. When an EBG loses some or all of its state (e.g., due to a power failure), it must recover the state so that packets can be forwarded over correct routes.

5.13. Support for Legacy ISATAP Services

EBGs support legacy ISATAP services according to the specifications in [RFC5214]. In particular, EBGs can configure legacy ISATAP interfaces and VET interfaces over the same sets of underlying interfaces as long as the PRLs and IPv6 prefixes associated with the ISATAP/VET interfaces are distinct.

6. IANA Considerations

There are no IANA considerations for this document.

7. Security Considerations

Security considerations for MANETs are found in [RFC2501].

The security considerations found in [RFC2529][RFC5214][I-D.nakibly-v6ops-tunnel-loops] also apply to VET.

SEND [RFC3971] and/or IPsec [RFC4301] can be used in environments where attacks on the neighbor coordination protocol are possible. SEAL [I-D.templin-intarea-seal] provides a per-packet identification that can be used to detect source address spoofing.

Rogue neighbor coordination messages with spoofed RLOC source addresses can consume network resources and cause VET nodes to perform extra work. Nonetheless, VET nodes SHOULD NOT "blacklist" such RLOCs, as that may result in a denial of service to the RLOCs' legitimate owners.

VET EBRs and EBGs observe the recommendations for network ingress filtering [RFC2827].

8. Related Work

Brian Carpenter and Cyndi Jung introduced the concept of intra-site automatic tunneling in [RFC2529]; this concept was later called: "Virtual Ethernet" and investigated by Quang Nguyen under the guidance of Dr. Lixia Zhang. Subsequent works by these authors and their colleagues have motivated a number of foundational concepts on which this work is based.

Telcordia has proposed DHCP-related solutions for MANETs through the CECOM MOSAIC program.

The Naval Research Lab (NRL) Information Technology Division uses DHCP in their MANET research testbeds.

Security concerns pertaining to tunneling mechanisms are discussed in [I-D.ietf-v6ops-tunnel-security-concerns].

Default router and prefix information options for DHCPv6 are discussed in [I-D.droms-dhc-dhcpv6-default-router].

An automated IPv4 prefix delegation mechanism is proposed in [I-D.ietf-dhc-subnet-alloc].

RLOC prefix delegation for enterprise-edge interfaces is discussed in [I-D.clausen-manet-autoconf-recommendations].

MANET link types are discussed in [I-D.clausen-manet-linktype].

The LISP proposal [I-D.ietf-lisp] examines encapsulation/ decapsulation issues and other aspects of tunneling.

Various proposals within the IETF have suggested similar mechanisms.

9. Acknowledgements

The following individuals gave direct and/or indirect input that was essential to the work: Jari Arkko, Teco Boot, Emmanuel Bacelli, Fred Baker, James Bound, Scott Brim, Brian Carpenter, Thomas Clausen, Claudiu Danilov, Chris Dearlove, Remi Despres, Gert Doering, Ralph Droms, Washam Fan, Dino Farinacci, Vince Fuller, Thomas Goff, David Green, Joel Halpern, Bob Hinden, Sascha Hlusiak, Sapumal Jayatissa, Dan Jen, Darrel Lewis, Tony Li, Joe Macker, David Meyer, Gabi Nakibly, Thomas Narten, Pekka Nikander, Dave Oran, Alexandru Petrescu, Mark Smith, John Spence, Jinmei Tatuya, Dave Thaler, Mark Townsley, Ole Troan, Michaela Vanderveen, Robin Whittle, James Woodyatt, Lixia Zhang, and others in the IETF AUTOCONF and MANET working groups. Many others have provided guidance over the course of many years.

10. Contributors

The following individuals have contributed to this document:

Eric Fleischman (eric.fleischman@boeing.com)
Thomas Henderson (thomas.r.henderson@boeing.com)
Steven Russert (steven.w.russert@boeing.com)
Seung Yi (seung.yi@boeing.com)

Ian Chakeres (ian.chakeres@gmail.com) contributed to earlier versions of the document.

Jim Bound's foundational work on enterprise networks provided significant guidance for this effort. We mourn his loss and honor his contributions.

11. References

11.1. Normative References

[I-D.templin-intarea-seal]
Templin, F., "The Subnetwork Encapsulation and Adaptation Layer (SEAL)", draft-templin-intarea-seal-15 (work in progress), June 2010.

[RFC0791] Postel, J., "Internet Protocol", STD 5, RFC 791, September 1981.

[RFC0792] Postel, J., "Internet Control Message Protocol", STD 5, RFC 792, September 1981.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2131] Droms, R., "Dynamic Host Configuration Protocol", RFC 2131, March 1997.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

[RFC2827] Ferguson, P. and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", BCP 38, RFC 2827, May 2000.

[RFC3315] Droms, R., Bound, J., Volz, B., Lemon, T., Perkins, C., and M. Carney, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 3315, July 2003.

[RFC3633] Troan, O. and R. Droms, "IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6", RFC 3633, December 2003.

[RFC3971] Arkko, J., Kempf, J., Zill, B., and P. Nikander, "SEcure Neighbor Discovery (SEND)", RFC 3971, March 2005.

[RFC3972] Aura, T., "Cryptographically Generated Addresses (CGA)", RFC 3972, March 2005.

[RFC4191] Draves, R. and D. Thaler, "Default Router Preferences and More-Specific Routes", RFC 4191, November 2005.

[RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, February 2006.

[RFC4443] Conta, A., Deering, S., and M. Gupta, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", RFC 4443, March 2006.

[RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, September 2007.

[RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless Address Autoconfiguration", RFC 4862, September 2007.

[RFC5342] Eastlake. , D., "IANA Considerations and IETF Protocol Usage for IEEE 802 Parameters", BCP 141, RFC 5342, September 2008.

11.2. Informative References

[CATENET] Pouzin, L., "A Proposal for Interconnecting Packet Switching Networks", May 1974.

[I-D.carpenter-flow-ecmp] Carpenter, B. and S. Amante, "Using the IPv6 flow label for equal cost multipath routing and link aggregation in tunnels", draft-carpenter-flow-ecmp-02 (work in progress), April 2010.

[I-D.cheshire-dnsext-multicastdns] Cheshire, S. and M. Krochmal, "Multicast DNS", draft-cheshire-dnsext-multicastdns-11 (work in progress), March 2010.

[I-D.clausen-manet-autoconf-recommendations] Clausen, T. and U. Herberg, "MANET Router Configuration Recommendations", draft-clausen-manet-autoconf-recommendations-00 (work in progress), February 2009.

[I-D.clausen-manet-linktype] Clausen, T., "The MANET Link Type", draft-clausen-manet-linktype-00 (work in progress), October 2008.

[I-D.droms-dhc-dhcpv6-default-router] Droms, R. and T. Narten, "Default Router and Prefix Advertisement Options for DHCPv6", draft-droms-dhc-dhcpv6-default-router-00 (work in progress), March 2009.

[I-D.ietf-dhc-subnet-alloc]
Johnson, R., Kumarasamy, J., Kinnear, K., and M. Stapp, "Subnet Allocation Option", draft-ietf-dhc-subnet-alloc-11 (work in progress), May 2010.

[I-D.ietf-grow-va]
Francis, P., Xu, X., Ballani, H., Jen, D., Raszuk, R., and L. Zhang, "FIB Suppression with Virtual Aggregation", draft-ietf-grow-va-02 (work in progress), March 2010.

[I-D.ietf-lisp]
Farinacci, D., Fuller, V., Meyer, D., and D. Lewis, "Locator/ID Separation Protocol (LISP)", draft-ietf-lisp-07 (work in progress), April 2010.

[I-D.ietf-manet-smf]
Macker, J. and S. Team, "Simplified Multicast Forwarding", draft-ietf-manet-smf-10 (work in progress), March 2010.

[I-D.ietf-softwire-ipv6-6rd]
Townsley, M. and O. Troan, "IPv6 Rapid Deployment on IPv4 Infrastructures (6rd)", draft-ietf-softwire-ipv6-6rd-10 (work in progress), May 2010.

[I-D.ietf-v6ops-tunnel-security-concerns]
Hoagland, J., Krishnan, S., and D. Thaler, "Security Concerns With IP Tunneling", draft-ietf-v6ops-tunnel-security-concerns-02 (work in progress), March 2010.

[I-D.jen-apt]
Jen, D., Meisel, M., Massey, D., Wang, L., Zhang, B., and L. Zhang, "APT: A Practical Transit Mapping Service", draft-jen-apt-01 (work in progress), November 2007.

[I-D.nakibly-v6ops-tunnel-loops]
Nakibly, G. and F. Templin, "Routing Loop Attack using IPv6 Automatic Tunnels: Problem Statement and Proposed Mitigations", draft-nakibly-v6ops-tunnel-loops-02 (work in progress), May 2010.

[I-D.russert-rangers]
Russert, S., Fleischman, E., and F. Templin, "Operational Scenarios for IRON and RANGER", draft-russert-rangers-03 (work in progress), June 2010.

[I-D.templin-iron]
Templin, F., "The Internet Routing Overlay Network

(IRON)", draft-templin-iron-03 (work in progress), June 2010.

[I-D.templin-isatap-dhcp] Templin, F., "Dynamic Host Configuration Protocol (DHCPv4) Option for the Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)", draft-templin-isatap-dhcp-06 (work in progress), December 2009.

[IEN48] Cerf, V., "The Catenet Model for Internetworking", July 1978.

[RASADV] Microsoft, "Remote Access Server Advertisement (RASADV) Protocol Specification", October 2008.

[RFC0994] International Organization for Standardization (ISO) and American National Standards Institute (ANSI), "Final text of DIS 8473, Protocol for Providing the Connectionless-mode Network Service", RFC 994, March 1986.

[RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, RFC 1035, November 1987.

[RFC1070] Hagens, R., Hall, N., and M. Rose, "Use of the Internet as a subnetwork for experimentation with the OSI network layer", RFC 1070, February 1989.

[RFC1122] Braden, R., "Requirements for Internet Hosts - Communication Layers", STD 3, RFC 1122, October 1989.

[RFC1753] Chiappa, J., "IPng Technical Requirements Of the Nimrod Routing and Addressing Architecture", RFC 1753, December 1994.

[RFC1918] Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, February 1996.

[RFC1955] Hinden, R., "New Scheme for Internet Routing and Addressing (ENCAPS) for IPNG", RFC 1955, June 1996.

[RFC2003] Perkins, C., "IP Encapsulation within IP", RFC 2003, October 1996.

[RFC2132] Alexander, S. and R. Droms, "DHCP Options and BOOTP Vendor Extensions", RFC 2132, March 1997.

[RFC2473] Conta, A. and S. Deering, "Generic Packet Tunneling in

IPv6 Specification", RFC 2473, December 1998.

[RFC2491] Armitage, G., Schulter, P., Jork, M., and G. Harter, "IPv6 over Non-Broadcast Multiple Access (NBMA) networks", RFC 2491, January 1999.

[RFC2501] Corson, M. and J. Macker, "Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations", RFC 2501, January 1999.

[RFC2529] Carpenter, B. and C. Jung, "Transmission of IPv6 over IPv4 Domains without Explicit Tunnels", RFC 2529, March 1999.

[RFC2775] Carpenter, B., "Internet Transparency", RFC 2775, February 2000.

[RFC3819] Karn, P., Bormann, C., Fairhurst, G., Grossman, D., Ludwig, R., Mahdavi, J., Montenegro, G., Touch, J., and L. Wood, "Advice for Internet Subnetwork Designers", BCP 89, RFC 3819, July 2004.

[RFC3927] Cheshire, S., Aboba, B., and E. Guttman, "Dynamic Configuration of IPv4 Link-Local Addresses", RFC 3927, May 2005.

[RFC3947] Kivinen, T., Swander, B., Huttunen, A., and V. Volpe, "Negotiation of NAT-Traversal in the IKE", RFC 3947, January 2005.

[RFC3948] Huttunen, A., Swander, B., Volpe, V., DiBurro, L., and M. Stenberg, "UDP Encapsulation of IPsec ESP Packets", RFC 3948, January 2005.

[RFC4192] Baker, F., Lear, E., and R. Droms, "Procedures for Renumbering an IPv6 Network without a Flag Day", RFC 4192, September 2005.

[RFC4193] Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses", RFC 4193, October 2005.

[RFC4213] Nordmark, E. and R. Gilligan, "Basic Transition Mechanisms for IPv6 Hosts and Routers", RFC 4213, October 2005.

[RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, December 2005.

[RFC4306] Kaufman, C., "Internet Key Exchange (IKEv2) Protocol", RFC 4306, December 2005.

[RFC4548] Gray, E., Rutemiller, J., and G. Swallow, "Internet Code Point (ICP) Assignments for NSAP Addresses", RFC 4548, May 2006.

[RFC4555] Eronen, P., "IKEv2 Mobility and Multihoming Protocol (MOBIKE)", RFC 4555, June 2006.

[RFC4795] Aboba, B., Thaler, D., and L. Esibov, "Link-local Multicast Name Resolution (LLMNR)", RFC 4795, January 2007.

[RFC4852] Bound, J., Pouffary, Y., Klynsma, S., Chown, T., and D. Green, "IPv6 Enterprise Network Analysis - IP Layer 3 Focus", RFC 4852, April 2007.

[RFC4903] Thaler, D., "Multi-Link Subnet Issues", RFC 4903, June 2007.

[RFC4941] Narten, T., Draves, R., and S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6", RFC 4941, September 2007.

[RFC5214] Templin, F., Gleeson, T., and D. Thaler, "Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)", RFC 5214, March 2008.

[RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, July 2008.

[RFC5569] Despres, R., "IPv6 Rapid Deployment on IPv4 Infrastructures (6rd)", RFC 5569, January 2010.

[RFC5685] Devarapalli, V. and K. Weniger, "Redirect Mechanism for the Internet Key Exchange Protocol Version 2 (IKEv2)", RFC 5685, November 2009.

[RFC5720] Templin, F., "Routing and Addressing in Networks with Global Enterprise Recursion (RANGER)", RFC 5720, February 2010.

[RFC5887] Carpenter, B., Atkinson, R., and H. Flinck, "Renumbering Still Needs Work", RFC 5887, May 2010.

Appendix A. Duplicate Address Detection (DAD) Considerations

A priori uniqueness determination (also known as "pre-service DAD") for an RLOC assigned on an enterprise-interior interface would require either flooding the entire enterprise network or somehow discovering a link in the network on which a node that configures a duplicate address is attached and performing a localized DAD exchange on that link. But, the control message overhead for such an enterprise-wide DAD would be substantial and prone to false-negatives due to packet loss and intermittent connectivity. An alternative to pre-service DAD is to autoconfigure pseudo-random RLOCs on enterprise-interior interfaces and employ a passive in-service DAD (e.g., one that monitors routing protocol messages for duplicate assignments).

Pseudo-random IPv6 RLOCs can be generated with mechanisms such as CGAs, IPv6 privacy addresses, etc. with very small probability of collision. Pseudo-random IPv4 RLOCs can be generated through random assignment from a suitably large IPv4 prefix space.

Consistent operational practices can assure uniqueness for EBG-aggregated addresses/prefixes, while statistical properties for pseudo-random address self-generation can assure uniqueness for the RLOCs assigned on an ER's enterprise-interior interfaces. Still, an RLOC delegation authority should be used when available, while a passive in-service DAD mechanism should be used to detect RLOC duplications when there is no RLOC delegation authority.

Appendix B. Link-Layer Multiplexing and Traffic Engineering

For each distinct enterprise network that it connects to, an EBR configures a VET interface over possibly multiple underlying interfaces that all connect to the same network. The VET interface therefore represents the EBR's logical point of attachment to the enterprise network, and provides a logical interface for link-layer multiplexing over its underlying interfaces as described in Section 3.3.4 of [RFC1122]:

> "Finally, we note another possibility that is NOT multihoming: one logical interface may be bound to multiple physical interfaces, in order to increase the reliability or throughput between directly connected machines by providing alternative physical paths between them. For instance, two systems might be connected by multiple point-to-point links. We call this "link-layer multiplexing". With link-layer multiplexing, the protocols above the link layer are unaware that multiple physical interfaces are present; the link-layer device driver is responsible for multiplexing and

routing packets across the physical interfaces."

EBRs can support such a link-layer multiplexing capability across the enterprise network in accordance with the Weak End System Model (see Section 3.3.4 of [RFC1122]). In particular, when an EBR autoconfigures an RLOC address, it can associate it with the VET interface only instead of assigning it to an underlying interface. The EBR therefore only needs to obtain a single RLOC address even if there are multiple underlying interfaces, i.e., it does not need to obtain one for each underlying interface. The EBR can then leave the underlying interfaces unnumbered, or it can configure a randomly chosen IP link-local address (e.g., from the prefix 169.254/16 [RFC3927] for IPv4) on underlying interfaces that require a configuration. The EBR need not check these link-local addresses for uniqueness within the enterprise network, as they will not normally be used as the source address for packets.

When the EBR engages in the enterprise-interior routing protocol, it uses the RLOC address assigned to the VET interface as the source address for all routing protocol control messages, however it must also supply an interface identifier (e.g., a small integer) that uniquely identifies the underlying interface that the control message is sent over. For example, if the underlying interfaces are known as "eth0", "eth1" and "eth7" the EBR can supply the token "7" when it sends a routing protocol control message over the "eth7" interface. This is necessary to ensure that other routers can determine the specific interface over which the EBR's routing protocol control message was sent, but the token need only be unique within the EBR itself and need not be unique throughout the enterprise network.

When the EBR discovers an RLOC route via the enterprise interior routing protocol, it configures a preferred route in the IP FIB that points to the VET interface instead of the underlying interface. At the same time, the EBR also configures an ancillary route that points to the underlying interface. If the EBR discovers that the same RLOC route is reachable via multiple underlying interfaces, it configures multiple ancillary routes (i.e., one for each interface). If the EBR discovers that the RLOC route is no longer reachable via any underlying interface, it removes the route in the IP FIB that points to the VET interface.

With these arrangements, all locally-generated packets with RLOC destinations will flow through the VET interface (and thereby use the VET interface's RLOC address as the source address) instead of through the underlying interfaces. In the same fashion, all forwarded packets with RLOC destinations will flow through the VET interface instead of through the underlying interfaces.

This arrangement has several operational advantages that enable a number of traffic engineering capabilities. First, the VET interface can insert the SEAL header so that ID-based duplicate packet detection is enabled within the enterprise network. Secondly, SEAL can dynamically adjust its packet sizing parameters so that an optimum Maximum Transmission Unit (MTU) can be determined. This is true even if the VET interface reroutes traffic between underlying interfaces with different MTUs.

Most importantly, the EBR can configure default and more-specific routes on the VET interface to direct traffic through a specific egress EBR (eEBR) that may be many outer IP hops away. Encapsulation will ensure that a specific eEBR is chosen, and the best eEBR can be chosen when multiple are available. Also, local applications see a stable IP source address even if there are multiple underlying interfaces. This link-layer multiplexing can therefore provide continuous operation across failovers between multiple links attached to the same enterprise network without any need for readdressing. Finally, the VET interface can forward packets with RLOC-based destinations over an underlying interface without any encapsulation if encapsulation avoidance is desired.

It must be specifically noted that the above arrangement constitutes a case in which the same RLOC may be used as both the inner and outer IP source address. This will not present a problem as long as both ends configure a VET interface in the same fashion.

It must also be noted that EID-based communications can use the same VET interface arrangement, except that the EID-based next hop must be mapped to an RLOC-based next-hop within the VET interface. For IPvX within IPvX encapsulation, as well as for IPv4 within IPv6 encapsulation, this requires a VET interface specific address mapping database. For IPv6 within IPv4 encapsulation, the mapping is accomplished through simple static extraction of an IPv4 address embedded within the IPv6 address.

Appendix C. Anycast Services

Some of the IPv4 addresses that appear in the Potential Router List may be anycast addresses, i.e., they may be configured on the VET interfaces of multiple EBRs/EBGs. In that case, each VET router interface that configures the same anycast address must provide equivalent packet forwarding and neighbor discovery services.

Use of an anycast address as the IP destination address of tunneled packets can have subtle interactions with tunnel path MTU and neighbor discovery. For example, if the initial fragments of a

fragmented tunneled packet with an anycast IP destination address are routed to different egress tunnel endpoints than the remaining fragments, the multiple endpoints will be left with incomplete reassembly buffers. This issue can be mitigated by ensuring that each egress tunnel endpoint implements a proactive reassembly buffer garbage collection strategy. Additionally, ingress tunnel endpoints that send packets with an anycast IP destination address must use the minimum path MTU for all egress tunnel endpoints that configure the same anycast address as the tunnel MTU. Finally, ingress tunnel endpoints should treat ICMP unreachable messages from a router within the tunnel as at most a weak indication of neighbor unreachability, since the failures may only be transient and a different path to an alternate anycast router quickly selected through reconvergence of the underlying routing protocol.

Use of an anycast address as the IP source address of tunneled packets can lead to more serious issues. For example, when the IP source address of a tunneled packet is anycast, ICMP messages produced by routers within the tunnel might be delivered to different ingress tunnel endpoints than the ones that produced the packets. In that case, functions such as path MTU discovery and neighbor unreachability detection may experience non-deterministic behavior that can lead to communications failures. Additionally, the fragments of multiple tunneled packets produced by multiple ingress tunnel endpoints may be delivered to the same reassembly buffer at a single egress tunnel endpoint. In that case, data corruption may result due to fragment misassociation during reassembly.

In view of these considerations, EBRs/EBGs that configure an anycast address should also configure one or more unicast addresses from the Potential Router List; they should further accept tunneled packets destined to any of their anycast or unicast addresses, but should send tunneled packets using a unicast address as the source address. In order to influence traffic to use an anycast route (and thereby leverage the natural fault tolerance afforded by anycast), ISATAP routers should set higher preferences on the default routes they advertise using an anycast address as the source and set lower preferences on the default routes they advertise using a unicast address as the source (see: [RFC4191]).

Appendix D. Change Log

(Note to RFC editor - this section to be removed before publication as an RFC.)

Changes from -14 to -15:

- new insights into default route configuration and next-hop determination

Changes from -13 to -14:

- fixed Idnits

Changes from -12 to -13:

- Changed "VGL" *back* to "PRL"
- More changes for multi-protocol support
- Changes to Redirect function

Changes from -11 to -12:

- Major section rearrangement
- Changed "PRL" to "VGL"
- Brought back text that was lost in the -10 to -11 transition

Changes from -10 to -11:

- Major changes with significant simplifications
- Now support stateless PD using 6rd mechanisms
- SEAL Control Message Protocol (SCMP) used instead of ICMPv6
- Multi-protocol support including IPv6, IPv4, OSI/CLNP, etc.

Changes from -09 to -10:

- Changed "enterprise" to "enterprise network" throughout
- dropped "inner IP", since inner layer may be non-IP
- TODO - convert "IPv6 ND" to SEAL SCMP messages so that control messages remain *within* the tunnel interface instead of being exposed to the inner network layer protocol engine.

Changes from -08 to -09:

- Expanded discussion of encapsulation/decapsulation procedures

- cited IRON

Changes from -07 to -08:

- Specified the approach to global mapping using virtual aggregation and BGP

Changes from -06 to -07:

- reworked redirect function
- created new section on VET interface encapsulation
- clarifications on nexthop selection
- fixed several bugs

Changed from -05 to -06:

- reworked VET interface ND
- anycast clarifications

Changes from -03 to -04:

- security consideration clarifications

Changes from -02 to -03:

- security consideration clarifications
- new PRLNAME for VET is "isatav2.example.com"
- VET now uses SEAL natively
- EBGs can support both legacy ISATAP and VET over the same underlying interfaces.

Changes from -01 to -02:

- Defined CGA and privacy address configuration on VET interfaces
- Interface identifiers added to routing protocol control messages for link-layer multiplexing

Changes from -00 to -01:

- Section 4.1 clarifications on link-local assignment and RLOC autoconfiguration.

- Appendix B clarifications on Weak End System Model

Changes from RFC5558 to -00:

- New appendix on RLOC configuration on VET interfaces.

Author's Address

Fred L. Templin (editor)
Boeing Research & Technology
P.O. Box 3707 MC 7L-49
Seattle, WA 98124
USA

Email: fltemplin@acm.org