Routing Architecture for the Next Generation Internet (RANGI)
draft-xu-rangi-01.txt

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on January 13, 2010.

Copyright Notice

Copyright (c) 2009 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents in effect on the date of publication of this document (http://trustee.ietf.org/license-info). Please review these documents carefully, as they describe your rights and restrictions with respect to this document.
Abstract

IRTF Routing Research Group (RRG) is exploring a new routing and addressing architecture to address the issues with the current Internet, e.g., mobility, multi-homing, traffic engineering, and especially the routing scalability. This document describes a new identifier (ID)/locator split based routing and addressing architecture, called Routing Architecture for the Next Generation Internet (RANGI), in an attempt to deal with the above problems.

Table of Contents

1. Introduction..................................................3
2. Architecture Description.....................................3
  2.1. Host Identifiers.........................................3
  2.2. Host Locators...........................................5
  2.3. Packet Formats..........................................6
  2.4. ID/Locator Mapping Resolution..........................7
  2.5. Routing and Forwarding System..........................8
    2.5.1. Host Behavior......................................8
    2.5.2. LDBR Behavior......................................8
    2.5.3. Non-LDBR Behavior.................................9
    2.5.4. Forwarding Procedures.............................9
  2.6. Multi-homing and Traffic-Engineering....................11
  2.7. Mobility...............................................13
3. Summary......................................................13
4. Security Considerations......................................14
5. IANA Considerations..........................................14
6. Acknowledgments..............................................14
7. References...................................................14
1. Introduction

The Default Free Zone (DFZ) routing table size has been growing at an increasing and potentially alarming rate for several years, which has detrimental impact on the routing system scalability and the routing convergence performance. This so-called routing scalability issue has drawn significant attention from both industry and academia. After much discussion following the IAB Routing and Addressing workshop [RAWS] in Amsterdam, a common conclusion was reached that the explosive growth in the DFZ routing table is mainly caused by the wide adoption of multi-homing, traffic engineering and provider-independent address. However, the underlying reason for this issue is the overloading of IP address semantics of both identifiers and locators. This overloading makes it impossible to renumber IP addresses in a topologically aggregatable way.

At present, the IRTF Routing Research Group (RRG) is chartered to explore a new routing and addressing architecture which could support the multi-homing, traffic-engineering, mobility and simplified renumbering features in a more scalable way.

This document describes an ID/locator split approach, called Routing Architecture for the Next Generation Internet (RANGI), which aims to deal with the above issues. Similar with Host Identity Protocol (HIP) [RFC4423], RANGI also introduces a host identifier layer between the IP network layer and the transport layer. As a result, the transport-layer associations (i.e., TCP connections and UDP associations) are no longer bound to IP addresses, but to the host identifiers. The major difference from the HIP is that the host identifier in RANGI is hierarchical and cryptographic host identifiers which have organizational structure. As a result, the ID/locator mapping system for such identifiers has reasonable business model and trust boundaries. In addition, RANGI uses special IPv6 addresses with IPv4 address embedded in the last four octets as locators. With this special locator, site-controlled traffic-engineering can be easily supported and the renumbering works during ISP change are also simplified greatly. Besides, the deployment cost of the new architecture is reduced greatly and the new architecture could be deployed in a more incremental way.

2. Architecture Description

2.1. Host Identifiers

Similar to the HIP, RANGI is also a kind of ID/locator split architecture. It introduces a 128-bit hierarchical host identifier.
As depicted in Figure 1, this identifier is composed of two parts, the leftmost n-bit part is the Administrative Domain (AD) ID which has embedded organizational affiliation and global uniqueness, and the remaining part is the local host ID which is a hash over the AD ID and the public key of the ID owner. In order to distinguish identifiers from locators (IPv6 addresses) in the transition mechanisms [RANGI PROXY], identifiers use a specific prefix, which is to be allocated from the IPv6 address space by IANA. Hence several leftmost bits in the AD ID field should be reserved for this purpose.

```
|<-------- n bits -------->|-- 128-n bits-->|
+--------------------------+-----------------+
| Administrative Domain ID | Local Host ID |
+--------------------------+-----------------+
    |                     |       |
    |                     |       |
    |---------------------|-------|
| Country ID | Authority ID | Region ID |<------Example
+--------------------------+
```

Figure 1. Host Identifier Structure

The purposes of the hierarchical host ID in RANGI are 1) to ease the management of the global identifier namespace resource; 2) to hold the economic and trust model in the ID/locator mapping system; 3) to ease the transition from the current Internet to the RANGI.

In RANGI, the global uniqueness of host IDs should be guaranteed through some registration mechanism. Since the AD IDs are globally unique and owned by the host ID registration and administrative authorities from different countries, the second part of the host ID (i.e., the local host ID) just needs to be unique within the corresponding AD scope.

The resolution infrastructure for flat label has no "pay-for-your-own" model, as names are stored at essentially random nodes (See Layered Naming Architecture (LNA) [LNA]). In contrast, the resolution infrastructure for the hierarchical host IDs in RANGI has reasonable business model and clear trust boundaries since they are stored in the corresponding nodes according to the organizational structures of the host IDs. To some extent, the business model of the ID/locator mapping system is similar with that for the Domain Name Service (DNS)
In the transition mechanisms for RANGI described in [RANGI-PROXY], identifiers are treated as IPv6 addresses by legacy hosts, and routers will forward the IPv6 packets with destination addresses being identifiers. Since the identifiers are hierarchical and aggregatable, the identifier based routing will not cause any routing scalability issue. For more information, please refer to [RANGI-PROXY].

The approach of generating hierarchical host identifiers in RANGI is much similar to the Cryptographically Generated Addresses (CGA) [RFC3972]. The major difference is that the prefix of the RANGI host identifier is AD ID, rather than ordinary IPv6 address prefix. In the CGA, the process of generating a new CGA takes three input values: a 64-bit subnet prefix, the public key of the address owner as a DER-encoded ASN.1 structure of the type SubjectPublicKeyInfo and the security parameter Sec, which is an unsigned three-bit integer. In RANGI, the process of generating a new host identifier also takes three input values: the n-bit AD ID (the suitable value of "n" has not been determined), the public key of the host ID owner and the security parameter Sec.

2.2. Host Locators

Host locators in RANGI are just IPv6 addresses. Since the IPv4/IPv6 coexistence and transition will last for a long period of time, in order to reduce the deployment cost of the new routing and addressing architecture, RANGI uses specific IPv6 addresses with IPv4 address embedded in the last 4 octets as locators. As shown in Figure 2, the leftmost 96-bit part of the locator is Locator Domain Identifier (LD ID), and it is used to globally identify each autonomous site network which adopts independent IPv4 address space. Actually, the LD IDs are provider-assigned (PA) /96 IPv6 prefixes which are topologically aggregatable. The rightmost 32-bit part is the host’s IPv4 address which is locally unique in the corresponding LD scope.
Similar with the Intra-Site Automatic Tunnel Addressing Protocol (ISATAP) [RFC5214], this specific locator can be used to automatically tunnel IPv6 packets over the destination IPv4 site networks since the IPv4 address part of the destination locator is used as the destination address of the tunnel header.

2.3. Packet Formats

RANGI-aware hosts reuse IPv6 protocol stack and packet format to maximum extent. The host identifier simply appears as a Next Header in the IPv6 header, as depicted in Figure 3, whereas the locator is filled as the IPv6 address in the IPv6 header, as shown in Figure 4.

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             Reserved                   | Protocol Type |     Checksum    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      +-----------------------------------|
|                        |                             |
| Source Host ID         +|
|                        +|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     +-----------------------------|
|                        |                             |
| Destination Host ID    +|
|                        +|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3. Host Identifier Header Format
Figure 4. Host Locator Header Format

The details regarding how to use IPsec to carry the data traffic will be described in the latter version of this draft or in a separate draft.

2.4. ID/Locator Mapping Resolution

ID/locator split implies a need for storing and distributing the mappings from identifiers to locators.

Within RANGI, the mappings from Fully Qualified Domain Names (FQDN) to host identifiers are stored in the DNS system, while the mappings from host identifiers to locators are stored in a distributed id/locator mapping system (e.g., a hierarchical Distributed Hash Table (DHT) system, or a reverse DNS system).

When using the hierarchical DHT system as the ID/locator mapping system, the AD ID of the host ID is regarded as a key in the top-level DHT ring to locate the corresponding bottom-level DHT ring, whereas the local host ID part is used as a key in the corresponding bottom-level DHT ring to determine the corresponding peer which
stores the desired ID/locator mappings. In practice, the top-level DHT ring can split further into multiple levels according to the hierarchical structure of the AD ID. To some extent, such a mapping system is a combination of the DNS and the DHT, and the top-level is more like the DNS since the routing is based on the former part of the host identifier (i.e., hierarchical AD ID), whereas the bottom-level is a DHT because the routing is based on the latter part of the host identifier (i.e., the flat local host ID).

The resolution infrastructure for flat names has no "pay-for-your-own" model, as the flat names are stored at essentially random nodes. In contrast, the resolution infrastructure for the hierarchical host identifiers, as used in RANGI, has reasonable business and trust model, since the hierarchical host identifiers have clear organization affiliation and the identifier resources can be managed with clear administrative boundaries.

2.5. Routing and Forwarding System

Within the RANGI, LDs are connected via Locator Domain Border Routers (LDBRs). A LDBR has at least one locally unique IPv4 address in each LD to which it is connected. The adjacent LDBRs exchange LD ID or LD prefix (Specific LD IDs can be aggregated into one LD prefix) reachability information with an inter-LD routing protocol. In fact, the Border Gateway Protocol (BGP) for IPv6 address family can be used directly as the inter-LD routing protocol.

2.5.1. Host Behavior

Generally, a source host will obtain the locator of the destination host from the ID/locator mapping system before initializing a communication. If the LD of the source host is identical with that of the destination host, the source host will send the packets directly to the destination host over an IPv4 tunnel. Otherwise, the source host will forward the packets towards the local LDBR over IPv4 tunnels.

Hosts can get the IPv4 addresses of their local LDBRs in several ways, e.g., a new Dynamic Host Configuration Protocol (DHCP) option, or a well-known LD-scope anycast address specific for LDBRs.

2.5.2. LDBR Behavior

LDBRs establish BGP sessions among them so as to exchange LD reachability information (i.e., IPv6 routing information with the mask length less than /96). LDBR can forward IPv6 packets according
2.5.3. Non-LDBR Behavior

The non-LDBRs just need to support IPv4 forwarding capability. So there is no need to upgrade them.

2.5.4. Forwarding Procedures

RANGI introduces a two-level routing mechanism which is composed of LD ID based routing and IPv4 address based routing. The former is used for inter-LD routing while the latter is used for intra-LD routing.

A simple RANGI routing procedure is illustrated in Figure 5. Source host A looks up the current locator of the destination host B through the ID/Locator mapping system. After obtaining that information, host A will tunnel the packets with destination address being host B to one of its local LDBRs. The LDBR shall find the next-hop LDBR based on the IPv6 globally routable locator, and forward the packets to it. For the intermediate transit networks, if the Non-LDBR routers which the packets have to traverse are legacy IPv4 routers, the ingress LDBR (for that locator domain) forwards the packet to the egress LDBR of the same locator domain over IPv4 tunnels.
For the intra-LD routing, RANGI uses the IPv6-over-IPv4 tunnels between LDBRs (or between LDBRs and hosts). Hence, RANGI can achieve a smooth IPv4/IPv6 transition. Once the internal routers within LDs are upgraded to IPv6, the requirement of IPv6 over IPv4 tunnel between the LDBRs or between LDBRs and hosts will be eliminated and the packets will be delivered by the LDBRs and the internal IPv6 routers hop by hop without tunneling as shown in Figure 6.
2.6. Multi-homing and Traffic-Engineering

In RANGI, each multi-homed stub LD shall be assigned a LD ID by each upstream ISP. In fact, these LD IDs are /96 IPv6 prefixes which are topologically aggregatable in provider networks. Each host within the multi-homed site, in turn, has multiple locators by concatenating the provider-assigned LD IDs with its locally unique IPv4 address. These hosts register the mappings from their identifiers to locators with the ID/Locator mapping system. As shown in Figure 7, host A which is located in a multi-homed site, has two LD IDs, LD ID_1 and LD ID_2, assigned separately from ISP1 and ISP2. Host A chooses either one as the source LD ID of the outgoing packets. Upon receiving the packets, the site exit LDBR, BR1, implements source-based policy routing. For example, if the source LD is LD ID_1, the packets will be forwarded into the ISP1’s network, otherwise, they will be forwarded into ISP2’s network.
The site-controlled traffic-engineering works as follows:

I. The source host sends out packets with the source LD ID being one of its LD IDs (assuming LD ID 1 being used).

II. The packets are intercepted by the LDBR BR1, and according to the traffic-engineering policy, the source LD IDs of the packets may be re-written from LD ID_1 to LD ID_2. Then BR1 forwards the packets into ISP2’s network due to source-based policy routing.

III. Once the packets arrive at the destination host, that host will use the source LD IDs in the received packets as the destination
LD IDs in the response packets. So the response packets will also enter the site A through the ISP2’s network.

IV. The source host could accept this change and use LD ID 2 as the source LD ID in the subsequent packets.

Similar to the GSE [GSE], the site-controlled traffic-engineering by rewriting the source LD ID will have effect on the path (upstream ISP) selection for both the outgoing packets and the incoming packets. In addition, the multi-homing and traffic-engineering usages in RANGI will not cause any routing scalability issue.

2.7. Mobility

In RANGI, when a host physically moves from one attachment to another, its host ID remains unchanged. The host needs to register the new locator with the ID/locator mapping system. Meanwhile, it should notify the corresponding entity of its new locator as soon as possible.

3. Summary

RANGI achieves almost all of goals set by RRG as follows:

1. Routing Scalability: Scalability is achieved by separating the identifiers from locators. Global routing is done based on provider assigned locators.

2. Traffic Engineering: LDBRs can overwrite the LD ID part of the source locator and implement source-based policy routing according to the source LD ID.

3. Mobility and Multi-homing: Applications and transport layers are bound to host IDs and so the sessions will not be interrupted due to locator change in cases of mobility or multi-homing.

4. Simplified Renumbering: When changing providers, the local locators (IPv4 part of the locators) do not change. The non-LDBR within the LD don’t need renumbering.

5. Decoupling Location and Identifier: Obvious.

6. Routing Quality: Since LDBRs only exchange LD reachability and the topology within LD will not be disclosed outside, the routing stability could be improved greatly.
7. Routing Security: RANGI reuses current routing system and does not introduce any new security risk into the routing system.

8. Incremental Deployability: non-LDBR within LD can still be IPv4-only. RANGI allows easy transition from IPv4 networks to IPv6 networks. In addition, the transition mechanisms for RANGI defined in [RANGI-PROXY] allows RANGI hosts to communicate with the legacy IPv4 or IPv6 hosts, and vice versa.

4. Security Considerations

TBD.

5. IANA Considerations

A specific prefix for host identifiers needs to be assigned from the IPv6 address space.

6. Acknowledgments

The author would like to thank Raj Jain for his valuable comments and reviews. Thanks should also be given to several experts who commented on the earlier versions of this proposal including Dave Oran, Joel Halpern, and Tony Li.

7. References


Authors’ Addresses

Xiaohu Xu
Huawei Technologies,
No.3 Xinxi Rd., Shang-Di Information Industry Base,
Hai-Dian District, Beijing 100085, P.R. China
Phone: +86 10 82836073
Email: xuxh@huawei.com