Network Working Group X. Xu
Internet Draft Huawei
Intended status: Informational
Expires: February 2011 August 10, 2010

# Routing Architecture for the Next Generation Internet (RANGI)
# draft-xu-rangi-04.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 February 10, 2011.

## Copyright Notice

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 issue. 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.

Conventions used in this document

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 RFC-2119 [RFC2119].

Table of Contents

## 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 is expected to support the multi-homing, traffic-engineering, mobility and simplified renumbering features in a more scalable way.

This document describes a new ID/locator split architecture, 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 (ID) layer between the IPv6 network layer and the transport layer. As a result, the transport-layer associations (e.g., TCP connections) are no longer bound to IP addresses, but to the host IDs. Unlike HIP, RANGI adopts hierarchical and cryptographic host IDs which have delegation-oriented structure. As a result, the corresponding ID->locator mapping system for such identifiers has a reasonable business model and clear trust boundaries. In addition, RANGI uses special IPv4-embeded IPv6 addresses as locators. With such locators, site-controlled traffic-engineering and simplified renumbering can be easily achieved, meanwhile, the deployment cost of this new architecture is reduced greatly.

## 2. Architecture Description

### 2.1. Host Identifiers

In RANGI, host IDs are hierarchical and 128-bit long. As depicted in Figure 1, a host ID consists of two parts: the leftmost n bits (Note that the suitable value of "n" has not been determined yet, while the value of " n" is set to 64 in our current prototype) part is the Administrative Domain (AD) ID which has embedded organizational

affiliation and global uniqueness, and the remaining part (i.e., the rightmost 128-n bits)is the Local Host ID which is generated by computing a cryptographic one-way hash function from a public key of the ID owner and auxiliary parameters, e.g., the ID owner's AD ID. The binding between the public key and the host ID can be verified by re-computing the hash value and by comparing the hash with the host ID. As these identifiers are expected to be used along with IPv6 addresses at both applications and APIs, especially in the RANGI transition mechanisms defined in [RANGI-PROXY], it is desired to explicitly distinguish host IDs from IPv6 addresses (i.e., locators) and vice versa. Hence, a separate prefix for identifiers SHOULD be allocated by the IANA. As a result, several leftmost bits in the AD ID field SHOULD be reserved to fill this dedicated prefix.

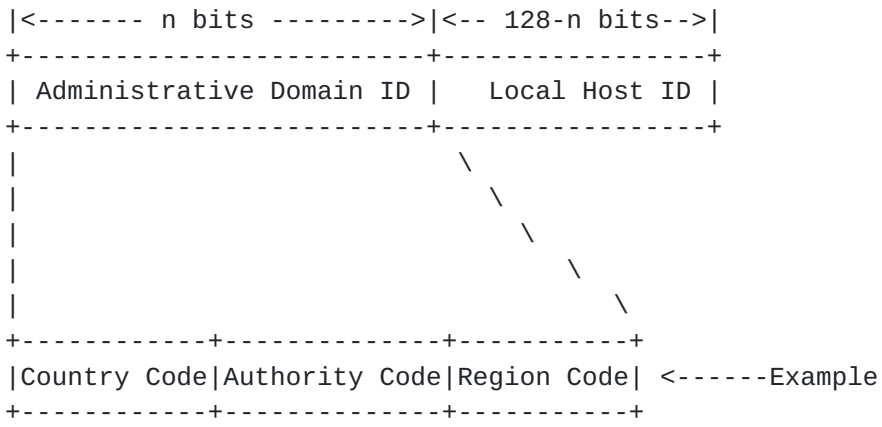


Figure 1. Host Identifier Structure

The approach of generating hierarchical RANGI host IDs is similar to that for Cryptographically Generated Addresses (CGA) [RFC3972]. The major difference is that the prefix of the RANGI host ID is AD ID, rather than ordinary IPv6 address prefix. In CGA, the process of generating a new address 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 contrast, the process of generating a hierarchical host ID in RANGI also takes three input values: the n-bit AD ID, the public key of the host ID owner and the security parameter Sec. Therefore, if we set the value of n to 64, the process of generating RANGI host IDs can be compatible with that for CGA.

The benefits of using hierarchical host IDs in RANGI include but not limited to: 1) manage the global identifier namespace in a scalable way; 2) hold a reasonable economic model and clear trust boundaries

in the corresponding ID->Locator mapping system; 3) ease the transition from the current Internet to RANGI.

In RANGI, the global uniqueness of host IDs is guaranteed through some registration mechanism. Since the AD IDs are globally unique and owned by the corresponding host ID registration and administrative authorities of different countries respectively, the Local Host IDs are only REQUIRED to be unique within the corresponding AD scope.

The resolution infrastructure for flat labels 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 hierarchical host IDs in RANGI has reasonable business model and clear trust boundaries since host IDs can be stored in the corresponding authoritative servers according to their organizational structures. To some extent, the business model of the ID->Locator mapping system in RANGI is similar to that for the Domain Name Service (DNS).

In the RANGI transition mechanisms described in [RANGI-PROXY], the identifiers of RANGI hosts are treated as ordinary IPv6 addresses by legacy IPv6 hosts. Upon receives a packet with the destination address being a host ID, the router SHOULD forward the packet according to the destination IPv6 address as normal. In the end, the packet will be forwarded to a dedicated proxy that is responsible for translating the packets between RANGI and IPv6. Since the identifiers are hierarchical and delegation-oriented aggregatable, such identifier-based routing during transition period will not cause any routing scalability issue. For more details, please refer to [RANGI-PROXY].

## 2.2. Host Locators

The host locators in RANGI are ordinary IPv6 addresses. Since the IPv4/IPv6 coexistence and transition will last for a long period, in order to reduce the deployment cost of this new routing and addressing architecture, RANGI uses specific IPv4-embeded IPv6 addresses as locators. As shown in Figure 2, the leftmost 96-bit part of a locator is called Locator Domain Identifier (LD ID), while the rightmost 32-bit part is filled with an IPv4 address which is REQUIRED to be unique within the scope of corresponding LD. LD IDs are used to globally identify each site network which is allowed to adopt independent IPv4 address space (either public or private IPv4 addresses). Actually, LD IDs are Provider-Assigned (PA) /96 IPv6 prefixes which are topologically aggregatable in provider networks.

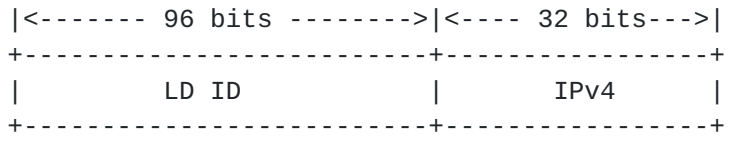


Figure 2. Host Locator Structure

Similar with the Intra-Site Automatic Tunnel Addressing Protocol (ISATAP) [RFC5214], this specific locator can be used for automatically tunneling IPv6 packets over IPv4 site networks.

### 2.3. Packet Formats

RANGI reuse the IPv6 packet format to maximum extent. The host IDs are filled as options in the Destination Option Header, whereas the locators are filled as IPv6 addresses in the IPv6 header. Packets sent from a RANGI host can be protected by attaching the public key and auxiliary parameters and by signing the packets with the corresponding private key. The protection works without a certification authority or any security infrastructure.

The details about the packet format and 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 host IDs to locators.

In RANGI, the mappings from Fully Qualified Domain Names (FQDNs) to host IDs are stored in the DNS system, while the mappings from host IDs to locators are stored in a distributed ID->Locator mapping system which can be built on the current DNS infrastrature. In a DNS based ID->Locator mapping system, if there are too many entries to be maintained by the authoritative servers of a given Administrative Domain (AD), Distribute Hash Table (DHT) technology can be used further to make these authoritative servers scale better. That is to say, the mappings maintained by a given AD will be distributed among a group of authoritative servers in a DHT fashion. As a result, the robustness feature of DHT is inherited naturally into the ID->Locator mapping system. Meanwhile, there is no trust issue since each AD authority runs its own DHT ring which maintains only mappings for the identifiers belonging to this AD.

A detailed mapping lookup example is given as follows:

1. A host ID will be transformed to a FQDN format string. Firstly, a host ID is expressed as "country-code.authority-code.region-code.local-host-ID" by inserting dots between adjacent fields, then by reversing the fields and attaching with the suffix "rangiid.arpa." it is transformed into a FQDN-format string as "local-host-ID.region-code.authority-code.country-code.rangiid.arpa."

2. The FQDN-format string is used as a key to locate the authoritative DNS server which maintains the desired resource records.

In order to facilitate such a lookup process, a new sub-domain " rangiid.arpa." needs to be inserted into the current domain name hierarchy. This sub-domain can delegate its own sub-domains according to the hierarchy of the FQDN-format string of the host ID. A new Resource Record (RR) named RANGI is also defined for the ID->Locator mappings, in which the NAME field is filled with the FQDN-format string of a host ID, while the RDATA field is filled with the corresponding locator information, including but not limited to an IPv6 address (i.e., locator) and its preference, and so on.

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 hierarchical host IDs, as used in RANGI, has reasonable business and trust models because hierarchical host IDs have clear organization affiliation.

To prevent the Man-in-the-Middle attacks during mapping lookups, the DNS Security Extensions (DNSSEC) [RFC535] is strongly recommended for the origin authentication and integrity assurance of the DNS data.

To prevent DNS recursive servers caching antique ID->Locator mapping information, the TTL of a RANGI RR for a mobile host SHOULD be set to 0 or a very small value. However, if a host (i.e., Correspondence Node) wants to cache the RR of the communicating host (i.e., Mobile Node), it can reset the TTL of that RR to a reasonable value internally.

The Secure DNS Dynamic Update mechanism defined in [RFC3007] is directly used for dynamically updating the ID->Locator mapping entries in the ID->Locator mapping system in a secure way.

## 2.5. Routing and Forwarding System

In RANGI, site networks (i.e., LDs) are connected to the IPv6 Internet via site border routers called Locator Domain Border Routers (LDBRs). LDBRd play the similar role as ISATAP [RFC5214] routers.

A simple RANGI routing procedure is illustrated in Figure 3. Host A (as source host) looks up the locator of host B (as destination host) through the ID->Locator mapping system before communicating with host B. Since these two hosts are located in different LDs, A will tunnel the packets destined for B to one of its local LDBRs, e.g., BR1. Otherwise, A will tunnel the packets destined for B directly towards B's IPv4 address. Once the packets arrive at the LDBR of the destination site, e.g., BR4, it will tunnel the IPv6 packets towards B's IPv4 address which is the last four octets of the destination locator.

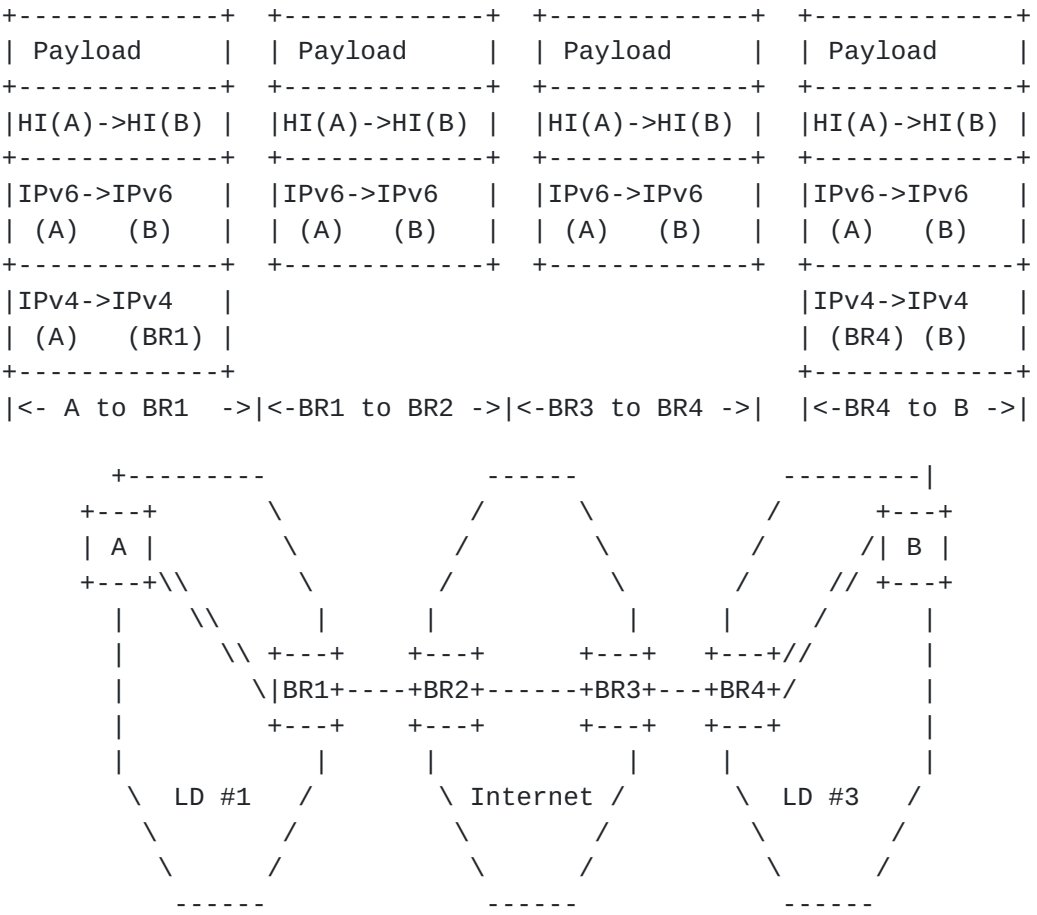


Figure 3. Routing Procedure

LDBRs are dual-stack routers which could be able to perform source-based policy routing and source address rewriting according to traffic-engineering policies on the outgoing packets.

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 site-scope well-known anycast address dedicated for LDBRs.

In RANGI, IPv6-over-IPv4 tunnels are deployed in the site networks. Hence, RANGI can achieve a smooth IPv4/IPv6 transition in the scope of site networks.

### 2.6. Site Multi-homing and Traffic-Engineering

In RANGI, each multi-homed site shall be assigned a /96 IPv6 prefix from each upstream ISP. Each host inside the multi-homed site, in turn, has multiple locators by concatenating the provider-assigned /96 IPv6 prefix with its locally unique IPv4 address. Hosts register the mappings from their identifiers to locators on the ID->Locator mapping system. As shown in Figure 4, host A is a RANGI host inside a multi-homed site, and it has two locators which are respectively synthesized from the LD IDs delegated from ISP1 and ISP2 and its IPv4 address. Host A chooses either one as the source locator of the outgoing packets. Upon receiving the packets, the site border router, BR1, performs source-based policy routing. For example, if the source locator is from ISP1, the packets will be forwarded to ISP1, otherwise, they will be forwarded to ISP2. In addition, BR1 could also rewrite the LD ID of the source locator to the one assigned from another ISP according to the configured traffic-engineering policy, and then forward the packets to the corresponding ISP according to source-based policy routing. Similar to the GSE [GSE], the site-controlled traffic-engineering by rewriting the source LD ID will impact the path (upstream ISP) selection for both outgoing packets and returned packets.

In addition, since each ID->locator mapping in the ID->Locator mapping system is associated with a preference. By setting different preference values for different locators of a given host which is located inside a multi-homed site network, the upstream ISP selection for the incoming traffic can also influenced.

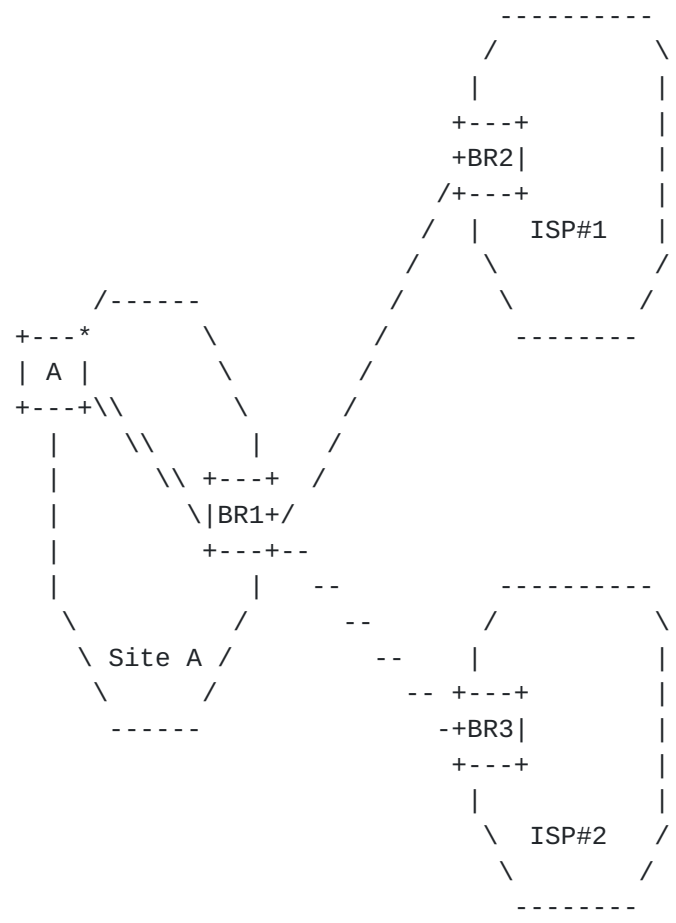


Figure 4. Site Multi-homing and Traffic-engineering

## 2.7. Host Mobility and Multi-homing

To some extent, host multi-homing is similar to host mobility since their effects on the network and on correspondents are identical.

In RANGI, when a host physically moves from one attachment point of network to another in the event of mobility or re-homing, it SHOULD inform its current correspondents of its new locator as soon as possible. Furthermore, it needs to update its locator information on the ID->Locator mapping authoritative server timely. In the case of simultaneous mobility, at least one of the communicating entities SHOULD resolve the correspondence node's new locator from the ID->Locator mapping system so as to continue their communication.

In order to allow legacy IPv6 hosts to initiate communicates with RANGI mobile hosts, many RANGI transit proxies SHOULD be deployed in the transit networks and each of them is dedicated to a bunch of

identifiers in a given AD scope and is responsible for translating packets from IPv6 and RANGI, and vice versa. For more details, please refer to the transit proxy mechanism defined in [RANGI-PROXY].

### 2.8. Network Mobility

To mitigate the registration burden on the ID->Locator mapping system triggered by network mobility, NEMO mechanism [RFC3963] is reused in RANGI to support network mobility. That is to say, the mobile router is responsible for updating its current locator on its home agent. As a result, network mobility event is transparent to the hosts inside that mobile network. Details about network mobility will be explored in the latter version of this draft.

## 3. Summary

RANGI achieves almost all of goals set by RRG, which are listed as follows:

1) Routing Scalability: Scalability is achieved by separating identifiers from locators.

2) Traffic Engineering: Hosts inside a multi-homed site can suggest the upstream ISP for outgoing and returned packets by using the appropriate source locator, while the local LDBRs have the final decision on the upstream ISP selection since they can perform site-controlled traffic-engineering through source locator rewritting.

3) Mobility and Multi-homing: Sessions will not be interrupted due to locator change in the case of mobility or re-homing.

4) Simplified Renumbering: When changing providers, the local IPv4 addresses of the site do not need to change. Hence the internal routers within the site don't need renumbering.

5) Decoupling Location and Identifier: Obvious.

6) Routing Stability: Since the locators are topologically aggregatable and the internal topology within the LD will not be disclosed outside, routing stability could be improved greatly.

7) Routing Security: RANGI reuses existing routing system and does not introduce any new security risk into the routing system.

8) Incremental Deployability: RANGI allows an easy transition from IPv4 networks to IPv6 networks. In addition, RANGI proxy allows RANGI-aware hosts to communicate to legacy IPv4 or IPv6 hosts, and vice-versa.

## 4. Security Considerations

TBD.

## 5. IANA Considerations

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

Two new options in the Destination Option Header need to be assigned for the host ID and its corresponding parameter date structure respectively.

## 6. Acknowledgments

The author would like to thank Raj Jain, Xuewei Wang and Dacheng Zhang for their valuable contributions. Thanks SHOULD also be given to Paul Francis, Lixia Zhang, Brain Carpenter, Dave Oran, Joel Halpern, and Tony Li for their insightful comments.

This research project is partially funded by the National"863" Hi-Tech Program of China.

## 7. References

[RAWS] D. Meyer, L. Zhang, and K. Fall. "Report from the IAB Workshop on Routing and Addressing", Internet draft, draft-iab-raws-report-01.txt, work in progress, February 2007.

[GOALS] T. Li, "Design Goals for Scalable Internet Routing", draft-irtf-rrg-design-goals-01, July 2007.

[RFC4423] R. Moskowitz and P. Nikander, "Host Identity Protocol (HIP) Architecture", RFC 4423, May 2006.

[RFC3972] T. Aura, "Cryptographically Generated Addresses (CGA)", RFC3972, Mar 2005.

[RFC3963] V. Devarapalli, R. Wakikawa, A. Petrescu and P. Thubert "Network Mobility (NEMO) Basic Support Protocol", RFC 3963, January 2005.

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

[RFC2136] P. Vixie, S. Thomson, Y. Rekhter, J. Bound, "Dynamic Updates in the Domain Name System (DNS UPDATE)", RFC 2136, April 1997.

[RFC2535] Eastlake, D., "Domain Name System Security Extensions", RFC 2535, March 1999.

[RFC3007] B. Wellington, "Secure Domain Name System Dynamic Update", RFC 3007, November 2000.

[H-DHT] L. Garces-Erice, E. Biersack, P. Felber, K. Ross, and G. Urvoy-Keller, "Hierarchical Peer-to-peer Systems", In Proc. Euro-Par 2003, Klagenfurt, Austria, 2003.

[GSE] M. O'Dell, "GSE-An Alternative Addressing Architecture for IPv6", Internet-Draft, Feb 1997.

[LNA] Hari Balakrishnan, Karthik Lakshminarayanan, Sylvia Ratnasamy,Scott Shenker, Ion Stoica and Michael Walfish, "A Layered Naming Architecture for the Internet", Proc. ACM SIGCOMM, Portland, Oregon, USA, August 30 - September 3, 2004.

[RANGI-PROXY] X. Xu, "Transition Mechanisms for Routing Architecture for the Next Generation Internet (RANGI)", draft-xu-rangi-proxy-01.txt, July 2009.

## 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 82882573
Email: xuxh@huawei.com