MILSA: A New Evolutionary Architecture for Scalability, Mobility, and Multihoming in the Future Internet

Jianli Pan, *Member, IEEE,* Raj Jain, *Fellow, IEEE,* Subharthi Paul, *Member, IEEE,* and Chakchai So-in, *Member, IEEE*

Abstract—**Many challenges to the Internet including global routing scalability have drawn significant attention from both industry and academia, and have generated several new ideas for the next generation. MILSA (Mobility and Multihoming supporting Identifier Locator Split Architecture) [1] and related enhancements [2, 3] are designed to address the naming, addressing, and routing scalability challenges, provide mobility and multihoming support, and easy transition from the current Internet. In this paper, we synthesize our research into a multiple-tier realm-based framework and present the fundamental principles behind the architecture. Through detailed presentation of these principles and different aspects of our architecture, the underlying design rationale is justified. We also discuss how our proposal can meet the IRTF RRG design goals [4]. As an evolutionary architecture, MILSA balances the high-level long-run architecture design with ease of transition considerations. Additionally, detailed evaluation of the current inter-domain routing system and the achievable improvements deploying our architecture is presented that reveals the roots of the current difficulties and helps to shape our deployment strategy.**

Index Terms—**Routing Scalability, Naming, Addressing, Mobility, Multihoming, MILSA, Future Networks, Next Generation Internet, Clean Slate Architecture, Transition, ID Locator Split.**

I. Introduction

THE INITIAL Internet design is a successful example of the balance between effectiveness and complexity. Many existing protocols succeeded because of their simplicity. However, the original designers couldn't expect such broad expansion of Internet as today. Newer contexts have introduced newer challenges. One of the typical examples is that the initial Internet was designed for a trusted community of universities and research institutions. However, broad commercial applications have made this assumption invalid leading to a series of security issues. Apart from the security flaw, other typical disadvantages of the current Internet design include difficulty in supporting routing scalability, mobility and multihoming, renumbering, traffic engineering, and policy enforcements [5]. Some of these issues may be correlated, thus making it infeasible to try to solve them one by one by putting ad-hoc patches to the architecture.

A. Routing Scalability

Routing scalability issue due to the expansion of the global routing tables was initially alleviated by progress in hardware technology and Classless Inter-Domain Routing (CIDR) [59]. However, for multihoming, traffic engineering, and renumbering benefits, address aggregation rules for scalable routing are often disregarded. This pushed BGP routers in the Default Free Zone (DFZ) to their capacity limit. However, ISPs may face different technical or economic constraints in upgrading their hardware devices. Also notice that the size has other implications such as bigger update churns, longer convergence time, and routing instability. Two new trends of the Internet may make things worse: one is the new IPv6 address space and the other is that potentially billions of mobile small-sized handheld or even "smart dust" hosts are expected to connect through Internet (so-called "Internet of Things" [56]). Moreover, the scaling problem also leads to other challenges related to security, control, and management. From the architectural design perspective, it is also broadly believed that the overloaded IP address semantics of "identifier" (ID) and "locator" is one of the major reasons for the scaling problem [5].

B. Mobility and Multihoming

In the current Internet, a connection between two end-hosts is uniquely identified by the IP addresses and TCP ports 4-tuple. When the mobile host changes attachment to the Internet and gets a new IP address, previous sessions are broken, i.e., there is no consistent and portable identity attached to the end-host other than the IP address. This semantic overloading breaches the independence between the layers in the protocol stack and application may use the IP address directly. The second concern is that the caching mechanisms in DNS cannot provide fast address updates for mobile users.

Multihoming can be host or site based. Site multihoming has more impact on global routing scalability. Typically, IPv4 multihoming can be done by Provider Independent (PI) or Provider Aggregatable (PA) addresses. Both approaches depend on the global routing system to fulfill the functionality and both violate the basic CIDR aggregation rules and thus hurt the routing scalability [49]. Detailed multihoming evaluation will be presented in Section VII.D. Multihoming may also correlate with traffic engineering since some multihoming actions are always due to traffic engineering requests.

Manuscript received 30 October 2009; revised 1 May 2010. This work was supported in part by a grant from Intel Corporation and NSF CISE Grant #1019119.

The authors are with the Department of Computer Science and Engineering, Washington University in Saint Louis, MO 63130 (e-mail: {jp10, jain, pauls}@cse.wustl.edu and so-in@ieee.org).

Digital Object Identifier 10.1109/JSAC.2010.101012.

C. Renumbering

Renumbering is very costly in the current Internet. When users or sites change service providers, their IP address block is generally changed leading to painful, costly and error-prone re-configurations. Such renumbering can be avoided by using PI addresses but PI addresses cause scalability issue.

D. Traffic Engineering

As discussed above, traffic engineering (including load balancing) and multihoming are always correlated. Currently, traffic engineering is often achieved by injecting more-specific prefixes into the global routing table, which negatively impacts routing scalability. Moreover, this approach cannot achieve custom-built finer granular policies.

E. Policy Enforcements

Conceptually, policies are mostly expressed as high-level requirements that are applied to the routing and data plane to realize the user/application level requirements. However, due to the ID locator overloading and the AS overloading (will be addressed shortly in this paper), there is no clear and efficient way to do policy enforcements without introducing problems in routing scalability or configuration and management.

Given all these challenges in the current Internet, different solutions in the past have aimed at just a few of them ignoring the fact that they are all related. Therefore, we try to put all of them into a holistic evolutionary architecture and try to balance the long term requirements with short term transition needs. Routing scalability is addressed with the first and most urgent priority in our architecture. Meanwhile, we gain all the other benefits through this evolutionary architecture.

The rest of this paper is organized as follows. Section II describes some important related research work in this area. MILSA design principles and model are presented in Section III. Detailed design issues and underlying rationale discussion is in Section IV. Section V is the hybrid transition mechanism of the architecture. In Section VI, we show how MILSA can meet most of the design goals of RRG. A detailed evaluation on the current global routing system is presented in Section VII. Conclusions and future works follow in Section VIII.

II. Related Work

Regarding the above challenges, there are many research efforts from both academia and industry which lead to many new solutions with different features.

A. Proposals for Separation: Host or Network

There is an on-going debate or dilemma on two competing directions. One is called "core-edge separation" which is relatively an easy-to-deploy strategy for routing scalability requiring no changes to the end hosts. Typical solutions include: LISP, SIX/ONE, APT, IVIP, DYNA, and TRRP (all from RRG [10]). Critics believe that from architecture view the tunneling in the core network looks awkward. Also there is no natural way of handling host mobility and multihoming; and handling the path-MTU problem is difficult [11]. However, the core-edge separation doesn't need any upgrade or even awareness from user side which is a big advantage in deployment compared with the host-based solutions. Of the core-edge solutions, LISP [12] is being carried out by a working group in IETF and many people are contributing to it. SIX/ONE [13] is another good example which also provides insight in transition of IPv4/IPv6, host/network cooperative solution for edge network multihoming, and incremental deployment capability. APT [14] did a good job in trying to make the mapping between the delivery address space and transit address space efficient, and minimize the delay and cause minimum negative influence to the current Internet.

The other direction is called "ID locator split" in which the IDs are decoupled from locators in the hosts' network stacks. Sample solutions include HIP [15], Shim6 [16], I3 [17], and Hi3 [18]. This type of scheme is advantageous in host mobility, multihoming, renumbering, etc. However, it is criticized to require host changes and has compatibility issues with the current applications.

Actually both these two categories try to decouple the "ID" from "locator" in some sense though through two different ways, i.e., decoupling in the host or in the network edge.

B. Schemes for Aggregation or Different Conceptual Routing

There are also several related ideas from both academia and industry that are worth discussing. GSE [19] proposed the preliminary ID locator split idea by separating ID from locator in IPv6 address space. Although it was not adopted at that time, it provides useful ideas regarding the separation. Huston [20, 21] presented the original insight on the routing scalability challenges. Virtual Aggregation [22, 35] is a good idea for temporarily alleviating the routing table FIB size problem with small cost to borrow time for new solutions. Atom policy [23] introduces an intermediate level between the Autonomous System (AS) and prefixes to improve the aggregation. There are also endeavors to find alternative inter-domain routing protocols other than BGP. HLP [24] presents a hybrid link-state and path-vector protocol that can reduce the churn-rate of route updates and achieve better convergence and scalability. There is also research on compact routing [25] which allows developing routing algorithms to meet the limits on routing table size, stretch, overhead, etc. NIRA [26] aims to provide users the ability to choose the route by themselves which is radically different from the current routing protocols. Nimrod [27] also tries to present scalable routing framework by representing and manipulating routing related information at multiple levels of abstraction.

C. Solutions for Mobility and Multihoming

There are also several papers on mobility and multihoming architectures. Mobile IPv4 [6] and Mobile IPv6 [7] are simplified versions of the host based ID locator split solutions in which the home address is used as an ID and care-of-address is used as a locator. These solutions suffer from triangular routing. SIP [8] tries to put all the functions in the application layer, and therefore, does not apply to all applications. TTR mobility [28] presents a very good mobility framework with economical consideration. For multihoming, besides the SHIM6, which is basically an IPv6 host-based ID locator split multihoming solution, there are also papers [29, 30, 31] on site multihoming which have implications on the global routing scalability. NEMO [9] is an example of site mobility. In this paper, however, we will mainly focus on host mobility.

D. Key BGP Technologies Related to Our Research

BGP itself as a de facto inter-domain routing protocol is also a research topic. There are several proposals for improving BGP to accommodate the emerging challenges. For example, inference of AS relationships out of the global routing table by GAO [32] and [33, 38] is an important foundation step towards the AS-level evaluation we have taken in this paper. Wang's work [34] helps us understand the transient failure and its implication to the BGP. Griffin's significant work on BGP wedgies [36] and other dynamics help us understand some basic problems and limitations of BGP. Bonaventure [37] also presents insightful thinking on building the next generation routing system. RCP [53] tries to ease the configuration and management in the AS by centralized policy and path selection decision instead of by distributed and highly meshed BGP links in local AS. Some significant work on routing policies theory and languages [39] also incite our thinking on the BGP AS overloading problem and the potential way out.

E. New Contributions and Relation to Our Previous Work

This journal version paper is a summarization and refinement of our previous work in three conference papers [1, 2, 3]. In papers [1, 2], we proposed a primitive host-based ID locator split architecture targeting at host mobility and multihoming improvements. In paper [3], we considered more on the network side potential solution for routing scalability and how our host-based solution [1, 2] can be deployed in a compatible and evolvable way.

In this paper, however, we have refined the previous work and have added new contributions as follows:

1) We generalize previously separate ideas and designs into a multi-tier evolutionary framework which mimics the evolution of the biological world,
2) Our design goals and design principles are clearly specified to guide our further design,
3) We formalize the key new concepts such as tier, realm, identifier (ID), and locator to facilitate multiple-level policy enforcements and security mechanisms; we also clarify their difference and relationship with the current concepts such as IP address, layer, and domain name,
4) Incremental deployment models, routes and strategies are presented in this paper; different incentives such as scalability, mobility, and multihoming, and their impacts on different deployment routes are presented,
5) We present detailed evaluation on the inter-domain routing system based on the real routing table data, which reveals the most up-to-date status we are faced with; we also evaluate and analyze how our architecture can potentially reduce the routing table size gradually under different deployment models,

All in all, we have tried to synthesize the previous conference papers into a new framework which is guided by new design goals and principles, enhanced by new concepts and components, and powered by the evaluation and analysis based on the real routing table data.

III. MILSA Model, Principles and Related Terminology

A. Key Terminologies

Tier: Tier represents the basic dependence of communication entities. Depending on the functionality and resource dependency relationship in the architecture, entities are divided into different tiers such as: *application/user/data/service* (Tier 3), networking *end-hosts* (Tier 2), and *routing infrastructure* (Tier 1), as shown in Fig. 1. A simple illustration of tier is that "*a service (Tier-3 object) resides on a host (Tier-2 object) which is attached to the routing infrastructure network (Tier-1 object)*". Notice that the host/interface is the common entity that the higher-tier objects need to affiliate to. The tiers are not necessarily limited to 3 and they can be extended to accommodate future requirements. Every entity in the network belongs to a tier and carries out tier-specific functions.

Realm: Realm consists of entities of the same tier grouped together according to their common affiliation or policies. For example, all the hosts belonging to a single organization form a realm. Similar realms exist for the other tiers. Each realm is supposed to have a Realm Manager (RM) that controls the assignment and resolution of IDs. Objects in a realm wishing to communicate with other objects have to follow a set of policies set by the RM.

Identifier: Identifier is the identity assigned to an object by its realm authority (generally RMs). It is a general term to identify the entities in the realms. Its format can be flat, hierarchical, or descriptive. Depending on which tier the ID holders belong to, the IDs can be divided into different types such as User-IDs (Tier 3), Host-IDs (Tier2), Routing-infrastructure-IDs (Tier 1), etc. Note that the Routing-infrastructure-ID is also called "locator" which is the ID of the point of attachment to the routing infrastructure tier, and it is also explained as follows.

Locator: Locator assigned by the routing infrastructure authority uniquely identifies the current location of the object. Locators are used for routing only and all the high-level semantic initially put upon the IP address is separated into IDs. Note that we will no longer use the term "address" in our solution since it is generally believed to be overloaded; instead, we use ID and locator separately. More often, locator is associated to the network interface that uniquely identifies a network attachment point that can be located.

B. Design Principles and Arguments

We need to emphasize that the original Internet design principles match the original design goals and the changing of the design goals leads to changes of design principles. The book by John Day [60] is a valuable resource to refer to for the basic patterns of the network architecture. It also includes many discussions on the original principles, history lessons, and basic reasoning in the process of Internet development.

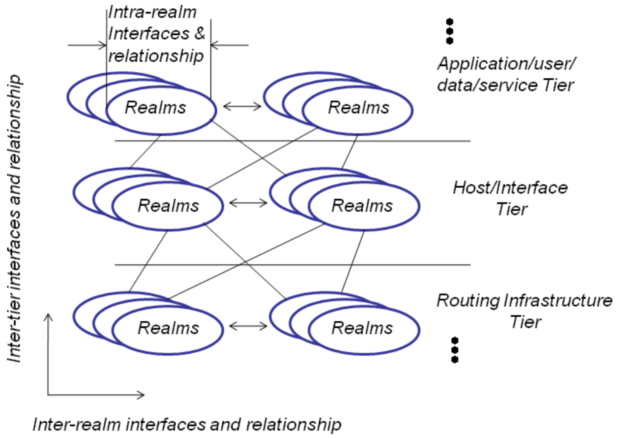


Fig. 1. Multi-level separation and the interfaces and relationships

1) Design Principles: Regarding the future Internet architecture to be an evolutionary design or a clean-slate design [40, 41], we have the following design principles:

Principle 1: "Evolutional Kernel": *We keep the "evolutional kernel" of the current Internet such as: layering, packet switching, and end-to-end argument [57].*

Principle 2: Variation and Diversity: *The variation and diversity in the architectural, protocol, technical or application level is allowed for a plethora of mutations and let the environment select the most competitive ones.* For example, we allow the ID locator split and core-edge separation to coexist in the architecture for transition, and let environmental contexts select.

Principle 3: Fitness and Synergy: *Given the variation and diversity mutation, survival or not of the solution depends on the fitness of the solution to environments.* In our solution, we try to make our design fit the basic call or the most urgent problems such as routing scalability, mobility and multihoming.

With these design principles in mind, we have the following multi-tier separation design decisions in MILSA.

2) Multi-Tier Separation: We observe that one of the key reasons leading to the ossification of the current Internet is the semantic overloading of multiple logical tiers. Moreover, given the perspective that the future Internet should interconnect many different technologies, the scalability requirement and convergence trends require the architecture to be open to accommodate significantly different networks, and to provide interfaces among different tiers. Our multi-tier separation is designed to match this call in the long run.

To be specific, typical separations are as follows (Fig.1):

1. Separation of application/user/data/service, host, and routing infrastructure tiers [58]

We picked these three tiers as the typical ones for the separation because that they represent the basic dependence and ownership of communication entities. Hosts are the common entities that the higher tier objects need to affiliate to. For example, the application/user/data/service usually should provide service or get access to data or service through an end-host. That is to say, physically they can coexist in one machine, but logically they should be separate to avoid trouble and to enable security or higher tier goals.

Due to the current intermixing of these tiers, difficulties arise in scalability, policy enforcements, etc. A typical good attempt of trying to address the separation of routing infrastructure provider and the service provider is the CABO [42]. After the separation of the tiers, the objects in each tier are grouped into realms. Thus, we have application/user/data/service realms, host realms, and routing infrastructure realms. A typical example is that Washington University provides email service to all the faculty and students; The University may use the routing infrastructure from AT&T or Verizon to provide Internet access. Thus, the bundle of service and data provided by the university belong to one or many tier-3 realms. The hosts used to access the service may belong to one or many host realms (Tier2). The network infrastructure may be provided by AT&T, Verizon, etc. and thus belongs to one or more routing infrastructure realms (Tier1). It is also possible that in the campus area, the routing infrastructure is owned by the university itself and in this case an organization may provide multiple realms in different tiers; however, they are logically separate. Note that the direct benefits of the multi-tier separation are individual tier's policy enforcements, commercial relationships, application, service architecture setup, etc.

2. Separation of Identifier Space from Routing Locator Space

Locators in the core MILSA networks obey the topological aggregation law to enable scalability. During the transition period, the conventional IP addresses will be treated as IDs by RMs and mapped into locators for global routing in the core routing system. Note that ID locator split or core-edge separation [3] is only the first step toward the multi-tier separation.

3. Separation of Control and Management from Data Plane

Though control and data can be in-band, they should logically be separate. The consequence of intermixing can be the inefficiency of the signaling and control of the network, difficulty in configuration and management, and insecurity.

4. Separation of AS Semantic Overloading

In our architecture, we decouple the "AS overloading". The basic idea is separating the host-realm's AS policy from the routing policy, so that any commercial policy of AS will not mess with routing, and the locator aggregation can be guaranteed. Easy configuration and managements can also be achieved. More details will be presented in Section VII.C.2.

Effective definition and implementation of inter-tier and inter-realm interfaces and protocols are important for multi-tier separation. We observe that basically there are three types of relationships and interfaces: inter-tier, inter-realm, and intra-realm, which are shown in Fig. 1. A simple example is that if we do ID locator split, we in fact separate the host realms from routing infrastructure realms, thus interaction functionality is required to bridge the two tiers, which is the global mapping system between IDs and locators. More inter-tier, inter-realm, and intra-realm interfaces may be defined in the future when the convergence of Internet among heterogeneous networks becomes a requirement or new services emerge.

In summary, multi-tier separation is an important feature of the MILSA architecture. Given the fact that there are many kinds of existing mutations of separation in different tiers, we generalize the idea and incorporate this idea into our

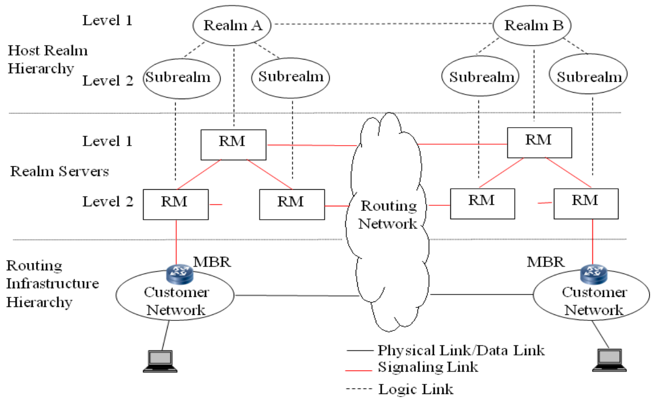


Fig. 2. A first-step MILSA example structure

architectural design. We argue that the multi-tier separation is the way to the ultimate scalability and many other benefits as discussed above.

C. MILSA Model

MILSA's design consists of three different functional planes. In the ***data plane***, the overloaded IP address is decoupled as ID and locator and upper layer protocols are bound to ID instead of locator. ***Control plane*** is in charge of the mapping from ID to locator and performs the locator-based routing, and some other function related to the interaction between end-host and the routing infrastructure such as three-tier mapping and object delegation. ***Management plane*** function is responsible for the management of objects and realms in various tiers.

MILSA follows a control and data plane separation principle to gain efficiency, controllability and manageability similar to that of the conventional telecommunication networks. Dedicated RMs form the control plane, while the data plane consists of the MILSA Border Router (MBR) hierarchy.

Fig. 2 shows a simplified two-tier MILSA architecture. This separates Host-ID space from routing locator space, which can be seen as the first step to the future MILSA multi-tier separation. The realms in each tier can be hierarchical. For example, a two level-hierarchy for host realms is shown in Fig. 2. The host RMs have a hierarchy similar to the host realms. Although not shown in the figure, routing infrastructure may also have a hierarchy. The host RMs map Host-ID to the locators. Signaling (control) links are set up between RMs. The hierarchical trust relationship between different groups of objects is depicted in the realm hierarchy. Realm hierarchy is mapped into the RM hierarchy by a one-to-one or one-to-many mapping (many RMs may serve the same realm for robust failure tolerance or load spreading). Fig. 2 only shows one-to-one mapping. Trust relationships are set up among RMs and they can authenticate and act as proxies for each other. MILSA objects can have multiple IDs belonging to different realms. Hosts can have multiple locators to support multihoming.

However, for future multi-tier separation, user/app/data may also have their own realms and RMs to negotiate trust or policy with other realms in different tiers and the mapping can be done between IDs of different tiers just like the mapping between Host-ID and locator. Note that realms become the basic operation unit of configuration, management, and policy enforcement. By physically and logically interacting with other realms from other tiers, the networks carry out multiple functions.

IV. MILSA: Key Design Details

Based on the design principles and reference model, we now present the details of the key design features of the architecture.

A. ID Locator Split Argument and Design

1) Argument: Actually, a successful ID locator split prototype exists in 2G/3G networks which have been proven to be scalable and good at handling layer-2 mobility. For example, a given mobile phone number of “123-456-7890” is actually an ID instead of a locator. When the mobile phone moves to the other states, the number remains unchanged but is assigned a temporary locator, which is hierarchical and transparent to the end-users. For IP networks, the static cache-based DNS structure cannot ensure fast update when users move and change their locators. By doing ID locator split, however, we can maintain the session portability and avoid these problems through an effective global mapping system. However, it seems that it requires a new host network stack to be installed and may affect the current applications [43]. The extra distributed global mapping system will also introduce costs. That's why some people argue against this separation on the host side. However, in the long run, we believe that an ID locator split is inevitable in order to support better host mobility and multihoming, renumbering, better policy enforcement, and more diverse upper-layer applications. What we can do is to design and plan a good transition strategy with evolution in mind that can provide the flexibility in accommodating different alternative solutions, and allow them to evolve to either direction when the environment makes the “natural selection”. That's why MILSA presents the hybrid design allowing the two strategies to coexist and evolve.

2) Design: To split IDs from locators, we introduce a new Identifier Sub-layer (IS) into the network layer. As shown in Fig. 3, the upper layers only use ID for session binding and the location information is transparent to upper layers. The lower layers don't know about the ID used in upper layers. IS also performs mapping from ID to locators by interacting with the RMs. If host multihoming is enabled, the IS maintains the mapping state, keeps monitoring the reachability of all the links, and interacts with RMs. Multiple ID-to-locator mappings are set up in the RM, each of which represents one active locator. We put IS below IPSec's AH and ESP headers so that the IPSec need not be aware of the locator changes due to mobility or multihoming. The fragmentation and reassembly header is also above the IS to make reassembly robust when using different locators for different fragments if there is a broken multi-path routing.

B. Different IDs and Realms

In MILSA, we have different IDs corresponding to different tiers as shown in the Fig. 4. User-IDs, Data-IDs, and Service-IDs are application-level IDs similar to the DNS names,

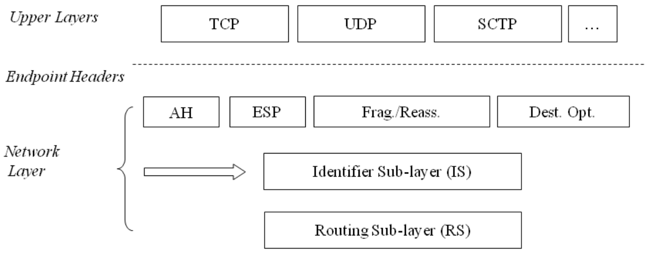


Fig. 3. ID locator split in the protocol stack

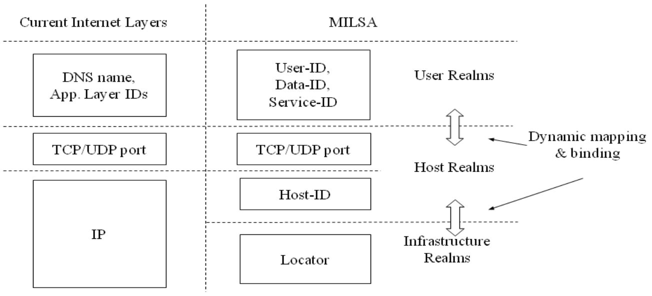


Fig. 4. IDs in Locator, Host, and User Realms

but have more meanings in helping set up user realms and enforcing policies among them. Host-ID, however, is the ID to represent the hosts on which different users run different applications. The current Internet uses the IP address as the session ID as well as routing locator which makes it difficult for host mobility and session portability. In MILSA, the host-ID is decoupled from locator to solely represent the hosts in host realms, and the locator is not used for the session identity. There is a dynamic IDs mapping and binding relationship between the IDs belonging to different tiers.

C. ID and Locator Structure

1) MILSA Identifier (MID): We have different kind of MIDs such as: User-ID, Host-ID, Routing-infrastructure-ID (locator) for different tiers. Objects in each tier have a set of IDs that are registered with the RMs. The bindings of the IDs from different tiers can be dynamic. For example, if we consider the following scenario: *"a user A roams and uses a host from B hotel which uses routing infrastructure provided by service provider C; A tries to access data D remotely"*. So in this scenario, user A has his/her User-ID which is bound to the Host-ID he is using (belongs to hotel B) and further is bound to the routing locator provided by the ISP of the hotel (provider C). The correspondents always send packets to one or more User-, Data-, or Service-IDs, and these IDs are further translated to multiple Host-IDs owned by or temporarily leased to the corresponding user/data/application. Each Host-ID may be translated into a set of locators due to the possible mobility of the hosts or multihomed hosts with more than one interfaces and hence locators. The IDs can be designed as locally valid and unique or globally valid and unique depending on the specific requirements. In current Internet, we have unicast and multicast addresses. Correspondingly, MILSA has unicast and multicast MIDs.

The MIDs for different tiers have different design requirements. For example, intuitively User-ID should be designed suitable for human memorization and usage; Host-ID should incorporate more hierarchical information denoting the position it resides in the logical realm, and possibly flat strings for security processing purpose; Routing-infrastructure-ID should be close to the current IP address framework with CIDR aggregation. Correspondingly, in our design, formats of MIDs are flexible. They can be ***descriptive, hierarchical, flat***, or the combination of these three. They have different features which may be desirable in different circumstances. For example, flat ID is fast for machine computation and easy to be applied to cryptographic usage, but it is not easy for human understanding and memorization, and not very suitable for naming of distributed applications with multiple levels. Hierarchical ID, however, is generally more understandable by human and suitable for multi-level distributed system. Descriptive ID is more useful in the high-level attributes-based circumstances where the desired objects' attributes may not be known completely, or it can be expanded dynamically. For example, consider the "printer" case in which we may need printing service in a specific location. Here, we do not need to specify the detailed Service-ID of the printer. Instead, we can specify our requirements by giving a series of attributes describing our requirements such as:

*"[university = wustl [building=bryan] [service = printer [type = color [resolution =1024*768]]]]"*

The MILSA network will select the most suitable printing service for the user according to the preference and policy. However, for host/interface, we may need the combination of hierarchical and flat IDs to gain the benefits in realm control, policy enforcements, and security. The hierarchical ID (generally good for Host-ID) used in MILSA enables security and AAA policy enforcements among different realms. HIP's [15] flat IDs are not suitable for this purpose. It also lacks a powerful control plane to carry out efficient ID to locator mappings. Thus, in MILSA, we introduce a host MID system which combines the features of hierarchical and flat ID. The host MID contains a flat encrypted part for security mechanisms similar to HIP. The mapping from ID to locator is done by a hierarchical RMs structure using a hybrid Push/Pull design to ensure mapping lookup and update performance, and the control plane is logically separated from the data forwarding plane. We argue that these new features are important for the long-term evolution.

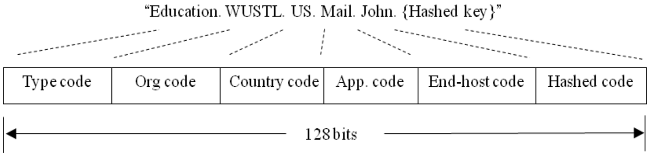


Fig. 5. A example of fitting the MID into 128 bits code

An example of host MID is shown in Fig. 5. However, note that it is not the actual proposed fields for a MID, instead, it is just a simple example illustrating how the host MID can be encoded into a structure compatible with the current IPv6 address paradigm.

2) Locator: In Rekhter's law [5] it is stated that "The addressing can follow the topology or the topology can follow addressing. Choose One." The current Internet violates this law

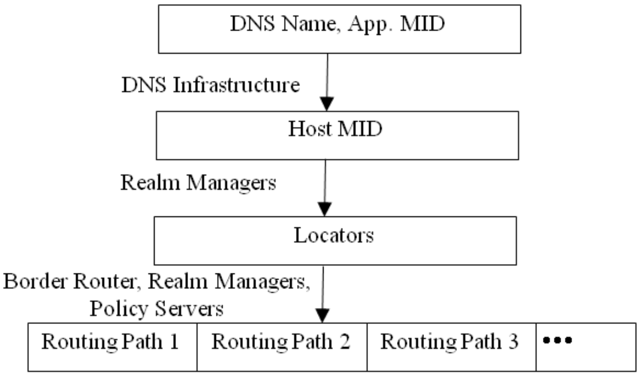


Fig. 6. Name resolutions and mappings

for scalable routing. Therefore, we require that the locators (addresses) in the new architecture obey the topological aggregation law. This requirement basically eliminates the necessity of using Provider Independent (PI) addresses for renumbering. Since locator is purely used for packet forwarding without any higher-layer meaning, the control and data plane split can also be achieved. However, to ensure the locator aggregation, efficient and automatic IP address allocation and configuration is needed. Also note that in MILSA, the locator in the core networks should also be 128 bit, and to distinguish locator from ID (both 128 bits), we can designate several special bits and encode them accordingly.

D. Three Level Mapping

We also need to clarify how MIDs are used in MILSA. We assume using the easy-to-understand DNS name or IDs in the application layer. So we allow the mapping from the general DNS name to the host MID. Note that this mapping is not very dynamic and can be implemented by adding a new Resource Record (RR) type into DNS. After getting the host MID for the given DNS name, it can be further resolved into the current locator of the object by the RM hierarchy. However, other protocols such as LISP-DHT [44] to achieve potentially greater efficiency in this overlay network is open for future design. Fig. 6 illustrates the three level mapping and the entities or systems involved in initiating or assisting the mapping. For the mapping from the locator to the routing paths, we allow the cooperation among the three planes to assist the decision.

E. Mobility and Multihoming

Mobility and Multihoming can be both host based and site based. Instead of using the triangular registration mechanism like in NEMO [9], site mobility in our solution makes use of the group ID and the RM-based global mapping system to keep the users and sessions of mobile network portable across the network. However, we will not address this case in this paper. Site multihoming will be addressed in Section VII.D. Here, we will focus on the two basic host mobility and host multihoming cases.

1) Mobility: We discuss the mobility issue in three cases:

1.1 Pre-Communication Mobility

If there are no on-going sessions with other correspondents, every time the end hosts change locator due to mobility, they should update their locators in their RM.

1.2 Mid-Communication Mobility

If two end hosts are talking and one end host moves and gets a new locator, it may want the correspondent to send subsequent packets to its new locator. In this case, the mobile host can directly notify the new locator to the correspondent's IS layer. The handover can be fulfilled with the assistance of lower layer (such as link layer) handover technologies. At the same time, the mobile host should update his locator with his RM just as in case 1.1. Note that the upper-layer sessions are bound to the MIDs instead of locators and thus won't break up when locators change. MILSA mobility model supports both peers moving and changing locators at the same time.

1.3 Roaming

Suppose a roaming user needs RM from another realm because there is no such service available close to him from its own realm. In this case, first of all, trust relationship is required between the two realms. Secondly, the roaming user should pass some AAA procedure. Then the foreign RM can act as the proxy for the home RM and control messages destined to the home RM to query for the current location of the roaming user will be directed to the foreign RM. Note that only control messages go through this triangular route (only for the first packet), but not the data path.

This mobility model has several advantages: First, control and data separation facilitates the update of the binding. Second, ID locator split makes the locator changes transparent to the upper layers. Third, there is no triangular routing problem. Fourth, roaming function is supported.

2) Host Multihoming: As discussed in Section IV.A.2, if multihoming is enabled, IS will maintain the mapping context, and more than one locator can be active for the end host and multiple MID to locator mapping entries should be registered with the RMs. IS will keep monitoring the state of these links, and update the status to the RM so that the overlay RM structure can find the current active locators of the end host. Based on the policy, the traffic may use one of the locators, or use them both for load spreading. When a link failure is detected by IS during the communication, IS will notify the correspondent to switch to another locator. It will also update the mapping entries in the RM. The second case is that if it is not in communication with other nodes, it will simply update the mapping entries at the RM.

With cooperation from RMs and IS, multihoming in MILSA is efficient and the policy can be configured flexibly by the users. Also note that assisted by the IS and RMs, simultaneous mobility and multihoming also becomes possible in MILSA.

F. Multicast and Manycast

The current Internet basically doesn't support multicast well. IP multicast is not widely deployed due to scalability and other problems. Multicast in MILSA is MID-based instead of address-based which makes MILSA multicast like "deliver this information to these end-hosts" instead of "deliver these

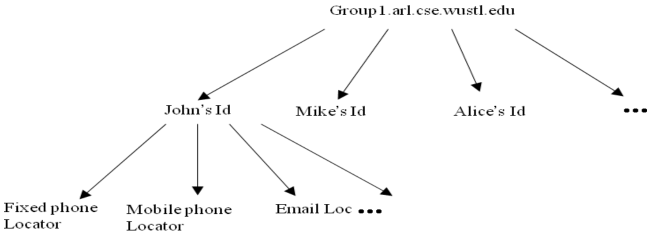


Fig. 7. Simple many-cast example

packets to these addresses". In the basic MILSA multicast design, we designate a specific ***multicast MID*** for a multicast group. The locator bound to this MID should be a locator of a MILSA Border Router (MBR) instead of an end-host. This MBR is in charge of maintaining a state list of the group. The end-hosts who want to join this multicast MID group register their MID and corresponding locator with the MBR which owns the group MID. After the multicast packets arrive at the MBR, it will look up the group state and replicate the packets to the members. In practice, to facilitate this procedure, we can use dedicated multicast routers.

In MILSA, multicast server doesn't replicate the packets directly to each locator since multiple copies of the packets in the same route is not optimal. The topologically aggregated locators in the state list form a tree comprised of multicast servers. We can also use the multicast server's locator in its upper level multicast server's state list to replicate packets to the whole sub-zone instead of every locator in the state list.

We also have ***manycast*** in MILSA to enable the packets to be delivered to a user with different locators for different devices or services. Fig. 7 gives a simple manycast example. Note that MILSA keeps the global routing system unaware of the multicast thus avoids the system scalability problem.

V. MILSA Transition Mechanism

Evolution of the Internet needs enough incentives or even the competence and compromise among different interest groups. So, to make the MILSA's future multi-tier separation possible, we will justify a first-step prototype of the separation between ID and locator in this section.

A. Non-technical Incentives

Regarding the pros and cons of the ID locator split and core-edge separation, there is an on-going debate on which way to go including strategies other than these two. Thus, to reduce the future potential risk, we propose a hybrid transition mechanism that can unify the "common essence" between the two strategies and make them coexist and complement each other. Moreover, the architecture can easily evolve into any of the two directions in the future when the environment makes the selection. Thus, in MILSA, the legacy hosts can coexist and talk to the new MILSA hosts regardless of whether they use PI or PA addresses.

History has shown that every change in the Internet needs good incentives and timing. It's reasonable to require only the entities actually feeling pain to change. For example, ISP (routers) changes for scalability, and end-user (hosts) changes for host mobility and multihoming. Those users who do not need host mobility and multihoming services may continue using the legacy host stack. They can be upgraded to MILSA stack when they actually need these services and are ready to pay for it. MILSA's hybrid transition design actually provides this option for users to choose and to bear the cost. As time goes by, it is possible that enough incentives are available to attract all the users to upgrade to the new networking stack. This idea also fit well into our design principles discussed in Section III.

B. Technical Discussion

To allow the two strategies to coexist, the "common essence" that we make use of is the global mapping system that is required in both strategies. We envisage an IPv6 world in the near future where the core routing system will also be IPv6 based. We also expect that by doing our hybrid design and AS decoupling, the core routing can be scalable and the current aggregation status of IPv6 network confirms this vision [45]. IPv4 address in the edge can still be used but treated by the architecture as ID, i.e., for transition purpose, the IPv4 as overloaded carrier can still be used as usual in the edge network, however, it will be treated differently and separately in the edge networks. Thus, no matter whether PI or PA addresses used, or in the future the ID used, they can all be compatible in the new architecture. Moreover, the legacy hosts and the new MILSA host with the new stack can all function in the new architecture.

For communications between two new MILSA hosts which implement the ID locator split in their networking stack, they talk to each other directly using the aggregatable locators after their IDs are mapped into locators. To allow legacy hosts, however, we divide the Internet into core and edge in order to separate the global routing from the edge routing. The edge network, generally a stub Autonomous System (AS), uses a series of aggregatable or un-aggregatable prefixes and is attached to one (for stub network with single service provider) or more (for multihomed stub network) transit ASs. Between the stub-ASs and transit-ASs is the MBR that performs the core-edge separation, responds to mapping queries and restructures the received packets using the global routable locators in the core networks. Notice that MBR is used only in legacy stub networks to act as an "proxy" between the legacy networks and the new networks. There is no need to deploy MBR in MILSA-aware stub networks since all hosts are aware of the MID and there is no need for proxy.

Different from HRA [46] that eliminates the global reachability of the local IP addresses, in order to ensure backward compatibility, MILSA ensures that irrespective of whether the stub network uses legacy PI or PA addresses, they will still be globally reachable. However, these global unique addresses or prefixes will no longer appear in the global routing tables. Instead, the prefix will be bound to a group MID (routing infrastructure realm MID) [2] and then to an entry point locator of the MBR, i.e., a triple binding of "legacy prefix-MID-MBR locator" will be set up and maintained by the RM structure. Through this triple binding, the legacy prefix acts similar to a globally reachable ID. Since there can be many

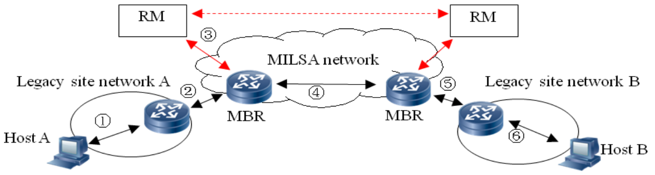


Fig. 8. A legacy host talks to a legacy host

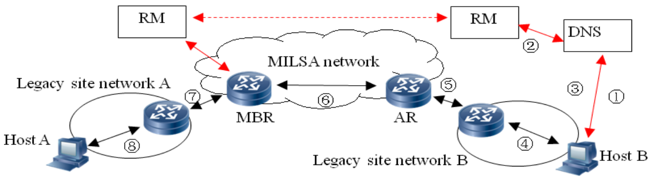


Fig. 9. A legacy host (without MBR) talks to a legacy host (with MBR)

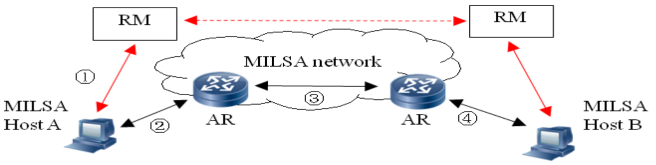


Fig. 10. A MILSA host talks to a MILSA host

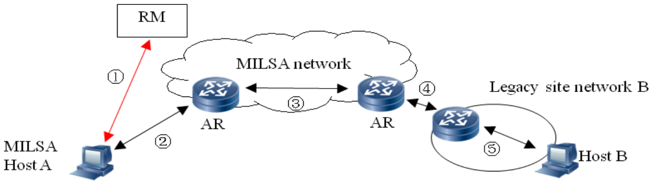


Fig. 11. A MILSA host talks to a legacy host

unaggregatable prefixes for an AS, many legacy prefixes can be bound to the same group MID and then to one or more entry point locators. The group MID actually represents the specific AS. Ideally, one group MID and one prefix block is enough for perfect aggregation in an AS. Since in a legacy AS, there is no split of IDs and locators, to let them exist and function in the new architecture, we need the routing infrastructure realm MID to represent the AS as an organization in the new network. Notice that this “organization” is different from the host realm since it is overloaded with two tier semantics. In summary, for legacy hosts not implementing the ID locator split, the provider network side but not the legacy host side bears the responsibility of deploying MBR and realizing the split. We need to emphasize that, to avoid confusion and possible misuse, the new MILSA aggregatable locator format should be distinguished from any of the legacy hosts’ prefixes by specifying certain special bits. Note that we use indirection instead of tunneling between core and edge.

After the above changes, the legacy hosts and the new MILSA hosts will coexist in the Internet. We now discuss how they can talk to each other.

1) Legacy hosts to legacy hosts: Regardless of whether the legacy hosts are IPv4 or IPv6 capable, they will all be globally reachable through the triple bindings registered in the global mapping system (as shown in Fig. 8), and the traffic will go through the entry point MBR through one of its MILSA locators for inter-domain routing. When a MBR is deployed for a legacy network using IPv4 addresses (PI or PA), they are mapped to the entry point MILSA aggregatable locator. Thus, the DFZ global routing table size will reduce by one (and by N if N prefixes were announced by this site). Then the size can be reduced step by step by deploying more and more MBR routers for the legacy networks. Note that the edge networks can still use legacy IPv4 addresses without harming routing scalability and theoretically all the IPv4 addresses are portable like MIDs.

The hosts in the legacy networks with MBR can talk to the MILSA hosts. However, since the MBRs are deployed incrementally, those site networks that have not deployed MBR yet need to talk to MILSA networks or to sites with MBR. As shown in Fig. 9 for the sites with MBRs, their PI prefixes are no longer used for global routing and are not in the global DFZ routing table any more, host A may not be easily reached by host B through the PI addresses and host B does not know anything about the MID. Note that A can talk to B since B’s address is still in the global routing table. For B to initiate a communication to A, we need some mechanism to route host B’s packets destined to host A’s legacy PI address to the closest MBR, which acts as a proxy between them and the MILSA networks. One possible solution is that we can get assistance from DNS. For example, suppose host A has a DNS name, then when host B queries DNS for host A, DNS server will retrieve the corresponding Host-ID (the group MID of the triple binding registered for the PI prefixes) and get the MBR locator of host A from RM, and return it to host B. Then host B can send out packets (by tunneling possibly). In Fig. 9, AR is general Access Router that is not MILSA-aware.

2) MILSA Hosts to MILSA Hosts: In this case, the MILSA host gets the receiver’s latest MILSA locator corresponding to the given Host-ID, puts them in the packets and sends out. Since the source Host-ID and destination Host-ID, and source locator and destination locator are all included in the packets, the traffic in the reverse direction will go through a similar procedure as shown in Fig. 10.

3) MILSA Hosts to Legacy Hosts: If MILSA host A wants to talk to legacy host B that has a legacy PI/PA address and its site has an MBR, host A can easily distinguish the legacy address from MILSA ID. Thus host A sends out a query to the RM server to get the MBR locator, and then encapsulates and sends out the packet to the MBR of host B. Host B’s MBR extracts the original address and does local routing to deliver the packets to host B. If host B’s site does not have an MBR (shown in Fig. 11), which means that the site prefix is still globally visible in the DFZ routing table, in this case, host A will not find any valid mapping from the RM. Host A uses its own MILSA locator as the source address and constructs the packets in the legacy format and sends to host B.

In the opposite direction, for legacy hosts talking to MILSA hosts, the packets will go directly to the MILSA locator of host A. However, since MILSA’s locator can be dynamic and host B may have no idea of MID, the communication can be assisted by the DNS. The procedure is similar to Fig. 9.

For MILSA hosts talking to legacy IPv4 hosts, the “dual stack lite” [48] or tunneling [47] mechanisms may apply.

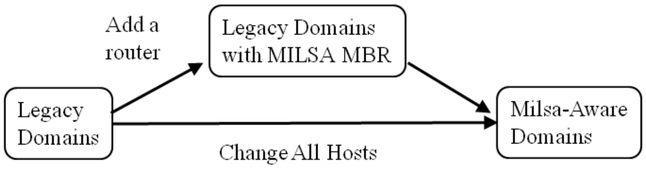


Fig. 12. MILSA transition map

However, the topic of IPv4/IPv6 coexistence is out of the scope of this paper.

C. Transition Map and Scenarios

In ideal case, after the transition process is finished, all the hosts are the MILSA-aware hosts in which ID locator split and new MIDs are implemented in the host stacks, and the whole network is MILSA network in which topological aggregation is achieved and the automatic locator re-allocation mechanism is successfully implemented. However, before migrating to this final status, our hybrid transition has a very clear transition route for legacy domain and all the legacy hosts who temporarily do not want or cannot afford to the new services, and as the mechanisms discussed above we allow all the different types of hosts to talk to each other during the transition. The basic idea is to deploy MBR which bridges between the legacy domain and the MILSA-aware domain. The transition map is shown in Fig. 12. By adding MBR to the legacy domains gradually, incremental deployability can be achieved. Firstly, the routing scalability issue can be resolved step by step and the total inter-domain routing table size can be reduced gradually. Secondly, new features such as mobility, multihoming, and traffic engineering will be widely supported as more and more MBRs are deployed and more domains evolve into MILSA-aware domains.

As shown in Fig. 13, six communication scenarios exist during the transition period: (1) between two MILSA-aware domains, (2) between MILSA-aware domains and legacy domains with MILSA MBR deployed, (3) between MILSA-aware domains and legacy domains, (4) between two legacy domains, (5) between two legacy domains with MILSA MBRs, and (6) between legacy domain without MBR and legacy domain with MBR.

VI. MILSA's Answers to the RRG Design Goals

In this section, we analyze and discuss how MILSA's first step can meet the design goals [4] set by IRTF RRG.

A. Routing Scalability

MILSA's hybrid design adopts short-term core-edge separation as well as ID locator split to tackle routing scalability challenges. Legacy IPv4 to IPv6 aggregatable address indirection in MBR makes it possible to continue using the PI addresses transparently without affecting the global routing system. The ID locator split mechanism further eliminates the necessity of using the PI addresses. Only topologically aggregated PA addresses are used in backbone routing and the size of DFZ global routing table is kept small. We can also deploy this mechanism incrementally by using the strategies we will present in Section VII.

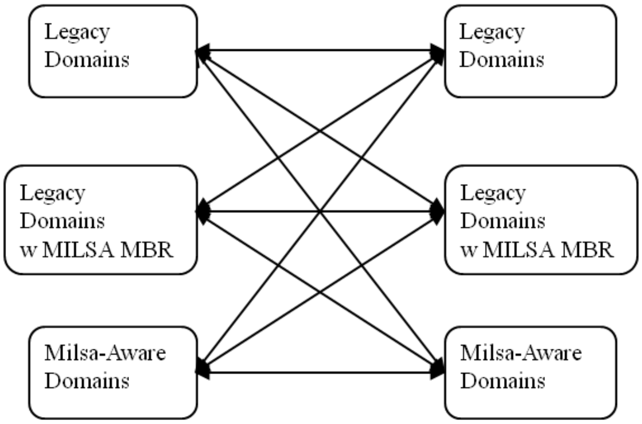


Fig. 13. MILSA transition scenarios

B. Traffic Engineering

Traffic engineering including load balancing in the current Internet is fulfilled by injecting more-specific prefixes into the global routing table. In MILSA, a given ID can be mapped to different locators to support multihoming. These locators may be preferred with different priority or sequence for load-balancing or load-spreading. Both end-host and the RM can participate in the selection of the locator based on user's policy, and the RM can be used for traffic engineering of incoming packet flows.

C. Mobility and Multihoming

For mobility, since upper layer protocols are bound to ID instead of IP addresses, sessions are portable for mobile users whose locators change due to mobility. MILSA supports two mobility models as discussed in Section IV.E. One is a simple model with end-to-end secure locator updates. However, to support initial communication with the ID and to allow both peers to be mobile, global RM mapping system is needed. MILSA mobility performance can be improved with the help of layer 2 handover mechanisms and potential cross-layer designs. Note that the global RM structure also helps in supporting global roaming and object delegation. Multihoming assisted by IS and the global RM structure is easier for both IPv4 and IPv6, and multihoming no longer harms the routing scalability. The ID locator split can work closely with RM to support scalable multihoming, load balancing or spreading.

D. Simplified Renumbering

Renumbering is no longer costly in MILSA. When users change service providers and get different locator blocks, their IDs remain unchanged. The renumbering will be taken care by the RMs to rebuild the ID to locator mappings. PI addresses are no longer in the global routing table. IP addresses used for packets filter, access control list, or management will be replaced by MID.

E. Decoupling Location and Identification

In MILSA, the two namesspaces of location and ID are completely decoupled both in host side and network side.

F. Routing Quality

Latency and reliability can be used to determine the routing quality. The topological locator based routing mechanism allows MBRs to select paths with shorter delay or better performance according to inter-realm trust or other policies. Furthermore, the hybrid design reduces the size of global routing table and decreases the packet forwarding delays. Since the edge address changes are transparent to the global routing system, the routing table updating frequency can also be reduced, which increases the routing stability. However, the first packet of a new session may suffer from the latency of the mapping system. The mapping from DNS name to MID is done by DNS which is a proactive pull system. Since this mapping is static to some extent, a caching mechanism can help reduce this latency. The mapping from MID to locator is fulfilled by the dedicated RM structure (also a proactive pull system) that has predetermined locations in the backbone network, which can help reduce the latency. Proactive push systems can avoid extra delays at the cost of higher state requirement by maintaining a complete mapping database at or close to the sender. In the future, mapping systems with features of hybrid push/pull design may be investigated.

G. Routing Security

Security is considered in several aspects of our design. MILSA uses DNS and the RM system for mapping, and MBR for packet routing. DNS is well proven to be secure in handling brutal attacks. RM is also transparent or invisible to the end-hosts. Inter-host trust and inter-realm trust are defined to provide end-to-end and inter-realm security to prevent potential DDoS attacks or limit them in a small scale. Participation of trust relationship and policies in deciding the optimal routing path can also reduce the potential indirection attacks. Moreover, since the edge network addresses are kept out of global routing system, it is hard for the attackers to inject bogus mappings into the RMs for eavesdropping, redirection, and flooding attacks.

H. Incremental Deployability

As discussed in the above sections, MILSA's evolutionary network architecture and the hybrid transition design are highly incrementally deployable. With the evaluation results on AS imbalance in Section VII, we can make our deployment strategies accordingly to gain the fastest reduction for the routing table size. However, it is also important to realize that each step of Internet evolution needs enough incentives or motivations, both from technical and non-technical aspects, and it is also important to make sure that each step will pay off, i.e., with reasonable and acceptable balance between the costs and benefits. Since in previous sections we mainly focused on technical incentives, here we will also consider and discuss several major non-technical incentives: the demands from ISPs and users about ***scalability, mobility, multihoming***, and possible combination of these demands. Correspondingly, we can consider several different deployment models based on different patterns driven by different incentives. They can generally be categorized into ***"top-down route"*** and ***"bottom-up route"***.

TABLE I
DEPLOYMENT INCENTIVES AND MAJOR MOTIVATORS

Incentives type	Top-down route		Bottom-up route	Mixed route
	Scalability driven	Mobility driven	Multihoming driven	Mixed factors
Major motivators	Medium and small size ISPs	Mobile ISPs especially big Mobile ISPs	Stub-ASs or networks having multihoming demands	All the ISPs

1) Scalability-Driven Deployment Model: The details about the evaluation of the current status of the routing scalability are in Section VII. Based on our observation, less than 20% of the ISP ASs announced more than 80% of the total prefixes in the DFZ routing table and most of the medium-sized and small-sized ISPs have the motivations to protect their current investments on the existing routers. It is understandable that they are reluctant or unable to keep upgrading their inter-domain routers with a speed or rate close to the exponential expansion of the global routing table size. However, they are too important to be given up due to the facts that the consistency of inter-domain routing system needs them to be able to handle global routing updates as good as the other bigger ISPs. Hence, they are more motivated to solve their problems with lowest costs possible.

In this scalability-driven deployment model (as shown in Table I), these medium and small-sized ISPs can simply deploy a MBR at the edge of its domain which acts as a data plane edge router working together with RMs which act as control plane carrying out legacy address to MID and locator mapping. Due to the control-data-separation design in MILSA architecture, MBR and RM distribute their loads separately based on functionalities which put no extra load upon the MBR, and the MBR doesn't have to be as powerful as the current inter-domain routers. Moreover, with the wide deployment of such MBRs in different ISPs, a beneficial cycle can be formed between the deployment and the reduction in the DFZ entries, which in reverse will expedite the deployment and the evolution.

In summary, if scalability is the major incentive driving the MILSA deployment:

(a) Costs: as a precondition, design to distinguish IDs and locators in IPv6 space is needed; legacy IPv4-to-MID bindings are needed to be set up in DNS by adding new RRs; a RM is needed for the legacy domain to cooperate with those of the MILSA-aware domains,

(b) Benefits: gradual reduction of the DFZ routing table size; help MILSA-aware domains expand, and a beneficial cycle can be formed to support sustainable evolution.

2) Mobility-driven Deployment Model: As we discussed in Section II, current solutions supporting layer-3 mobility are not as successful as expected. Also as discussed previously, MILSA's multi-tier separation design ultimately will support many high-level mobility services including the host layer-3 mobility. Many ISPs, especially bigger mobile telecommunication ISPs, have the motivation and incentives to provide better upper-layer mobility services to the customers, which also form a mobility-driven deployment model for MILSA architecture (as shown in Table I).

Compared with the scalability-driven deployment model, the mobility-driven model is a little different. To be more specific, the deployment and evolution motivations mainly come from those mobile ISPs especially big mobile ISPs instead of the medium and small sized ISPs. Accordingly, the mobile subscribers who need new mobility-supported service will install the new MILSA host stack as discussed in Section IV which splits the identity from location. From the network side, MILSA-aware domains will be equipped with RMs which carry out full MID-to-locator mapping, and RMs also interact with each other to set up inter-realm trust relationship, enforce the realm policy, and help build end-to-end security support. In addition, general bindings between DNS names and MIDs are registered in the DNS. One of the basic principles for MILSA mobility is that it goes through a "he who pays, gets" style which put no extra cost burden upon legacy hosts without mobility demand. Successful mobility stories of the MILSA domains will expedite the deployment of the RMs and host stacks which can also form a beneficial cycle to support better evolution, and in reverse it also will reduce the DFZ routing table size gradually.

In summary, if mobility is the major incentive driving the MILSA deployment:

(a) Costs: new MILSA hosts and stacks; RMs sub-system carrying out MID-to-locator mappings; general bindings between DNS names and MIDs in the DNS,

(b) Benefits: layer-3 or higher-level mobility support; no extra cost for legacy hosts without mobility demand; gradual reduction of the DFZ routing table size; beneficial cycle for sustainable evolution.

3) Multihoming-driven Deployment Model: IPv4 Multihoming using PI addresses violates the CIDR aggregation and also leads to the increase of DFZ routing table size. The details about the evaluation of the current status of the multihoming are in Section VII. Based on our observation, the number of multihomed ASs has exceeded the number of single-homed ASs in the current Internet and is still increasing at a very high rate. Accordingly, the stub-ASs and stub networks having multihoming demand are the major motivators in the multihoming-driven deployment model (as shown in Table I).

In the multihoming-driven deployment model, the stub-ASs using PI addresses will have these addresses mapped into aggregatable locators by the deployed RMs. The PI addresses will no longer be injected into the DFZ routing table, hence, the DFZ routing table size will be significantly reduced. Similarly, the MBRs are needed to be deployed in the edge of the ASs to map the legacy IPv4 PI or PA addresses into aggregatable MILSA locators which also can reduce the DFZ routing table size. Intuitively, the reduction benefits by this model are less than the benefits brought by the scalability-driven deployment model (more details in section VII). However, it can meet the increasing multihoming demands from more and more stub-ASs. Successful multihoming stories of stub-ASs will expedite the deployment of the RMs and also form a beneficial cycle for the evolution. It will also reduce the DFZ routing table size gradually, though this effect is somewhat smaller than the scalability-driven deployment model.

In summary, if multihoming is the major incentive driving the MILSA deployment:

(a) Costs: design to distinguish IDs and locators in IPv6 space is needed; legacy IPv4-to-MID bindings are needed to be set up in DNS by adding new RRs; RMs and MBRs are needed to be deployed for the legacy domain,

(b) Benefits: more multihoming support for stub-ASs without compromising the global routing scalability; also gradual reduction of the DFZ routing table size; help MILSA-aware domains expand, and a beneficial cycle can also be formed to support sustainable evolution.

4) Mixed Model: Obviously, the above three models are different in where to deploy the new MILSA designs with different benefits and costs. Different incentives can lead to different deployment priorities and decisions. However, if there are no strict priorities among these different incentives, we also present a coarse, general and mixed deployment example model which is separated into several gradual steps:

(a) Deployment of IPv4-to-MID binding in DNS, and MID-to- locator global mapping system. We need to add a new RR in DNS, and add a triple binding maintained in the RM infrastructure.

(b) Deployment of MBR and the interaction with RM structure.

By finishing these two steps, the DFZ routing table size can be reduced gradually and the backward compatibility can be guaranteed. The major incentives underlying is routing scalability and the major motivators are medium and small ISPs. If deploying the RMs and MBR in the stub-ASs, the multihoming can be achieved in which case the major incentives is multihoming and the major motivators are the stub-ASs and customer networks with multihoming demands.

(c) Deployment of the data plane ID locator split, end-to-end mobility and security support.

(d) Host realm assignment and management, inter-realm trust setup, and DNS name to MID mapping registration in DNS.

By finishing the above two steps, the new feature of host mobility can be expected. The major incentives underlying is mobility and the major motivators are mobile ISPs especially big mobile ISPs.

(e) The AS overloading decoupling, new mechanism for better BGP routing policy enforcement, efficient automatic address allocation mechanism and topological aggregation.

(f) Secure signaling of the three-plane cooperation, policies, and an integrated service model, etc.

The above two steps are more advanced and are in preparation for the evolution to future multi-tier separation.

Note again that in the very first transitional period, we allow end-hosts to choose to support MILSA or not. The deployment is also open to potential new technologies, and may co-work with other solutions such as Virtual Aggregation (VA) [22, 35].

In summary, the mobility-driven and scalability-driven models seems to follow a ***"top-down route"*** and the multihoming-driven model follows a ***"bottom-up route"*** which are the two different options we can have when considering the deployment of MILSA.

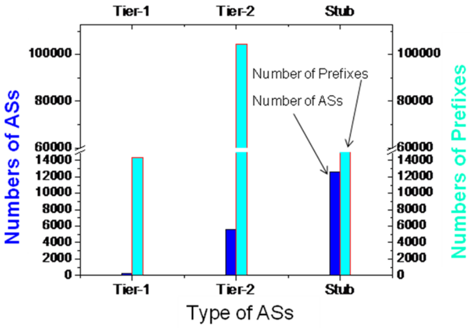


Fig. 14. AS and Prefix Distribution

VII. Evaluation and Analysis

The goal of this section is to present the real status of the global inter-domain routing system, and to discuss how and why our MILSA architecture can address these situations potentially in the short term as well as in the long term.

A. Methodology

The basic methodology and thinking we follow in this section is to "understand the reality better" before we "do something that makes sense" to justify a sound new architecture. Thus, we first evaluate and analyze the up-to-date status related to BGP routing table's FIB size, and the corresponding BGP trends. After that, we specifically focus on the AS level Aggregation Degree (AD) analysis which can reveal underlying causes of the routing table growth. Thirdly, the implications of multihoming underlying the scalability issue are discussed as multihoming is a significant factor in the routing table size expansion.

Then, based on the above work, we do further evaluations and analysis of the future potential changes and our deployment strategies. Though MILSA architecture is designed for multiple benefits, we here focus on the effect on reducing the inter-domain routing table size. We evaluate and analyze how the three deployment models discussed in Subsection VI.H can affect the routing table size and how it can be reduced in different deployment scenarios with different speeds.

Note that all of the raw data for evaluation and analysis are from three sources: Oregon RouteView Project [50], CIDR Report [51], and CAIDA [52]. During the evaluation, we also review the previous work that has been done addressing related issues. We partly use and compare some of the results of these papers, however, our work is different in the sense that we do the evaluation mostly through a different angle (i.e., AS level), and we are more concerned with the potential implications of applying our solution and evolution into the future Internet.

B. BGP FIB Table and the Prefixes

1) Routing Table Size and Its Contributors: Firstly, we simply calculate the rough routing table sizes based on the number of prefixes. We found that it is increasing very fast. Specifically, the total prefixes size has increased by 6 times during the last 12 years (from 52K in 1998 to 301K in 2009), and there is about 20% increase every year. The exponential increase brings a series of impacts to the inter-domain routing system. For example, it can lead to longer BGP convergence time, more signaling traffic, more memory to store the routing table, more CPU computation capacities, and even more power consumption and heat dissipation from the routers. It also makes it difficult to keep the whole system consistent. Conceptually, it would become even worse in the perspective of future IPv6 world in which millions of nodes are expected to connect to the Internet. The exponential increasing trend of the size of the inter-domain routing table has been observed [20, 21] for long time. The underlying reasons are complicated [38]. To illustrate the different factors and contributors leading to the expansion of the inter-domain routing table size growth, we summarize the problem space by the following formal equation:

$$S \propto f[B \times T \times \text{Deagg}(A,\ P,\ M,\ TE,\ AF)] \qquad (1)$$

Here,
S: Size of the global routing table
B: Base size of the routing table entry
T: Topology complexity of the network
A: AS domain complexity
P: AS domain policy factor
M: Multihoming factor
TE: Traffic engineering and load balancing factor
AF: Address fragmentation factor [38]

The function Deagg reveals the influence of the address disaggregation due to miscellaneous contributors in the current Internet which weakens the CIDR effects.

2) AS/Prefixes Distribution, and Topological Aggregation: To present an overview of the difference among ASs, we do a coarse analysis based on a sample by using the approach of Dimitropoulos [61]. The result is shown in Fig. 14. From the figure, we observe that the stub-ASs cover about 65% of the total ASs available in the Internet, and about 0.4% percent of the ASs are transit-only ASs (tier-1 ASs mostly). Another 30% of the ASs are tier-2 ASs which provide some level of transit. The remaining is indiscernible due to the algorithm limitation and the complexity of the routing system. However, for the prefixes announced by these ASs, 70% of the total prefixes are contributed by the tier-1 and tier-2 service providers, and 20% are contributed by the stub-ASs (again, the remaining 10% are indiscernible). Another important observation of the Fig. 14 is that the ratio of the number of prefixes to the number of ASs for ISPs is significantly bigger than that of the stub-ASs. Hence, we conclude that the aggregation of prefixes by tier-1 and tier-2 ASs should be the most important parts for the global routing scalability.

We further evaluate the prefix length distribution in the routing table, as well as the increasing trends during the past ten years. We observe that the number of prefixes with length between /17 and /24, especially /24 is the biggest portion and with the fastest growth speed among all the prefixes. This can be explained by the broad usage of the class C addresses in the global routing, and the small and medium sized commercial ASs connected to the Internet with increasing requirements on multihoming, load balancing and routing policy enforcement which tamper the prefixes aggregation.

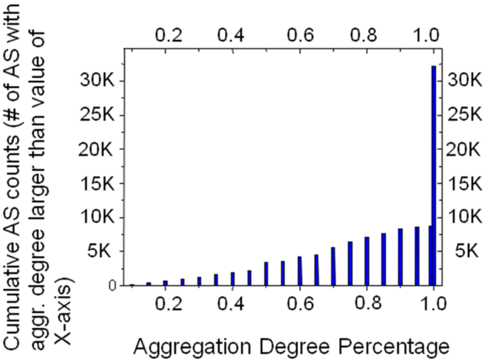


Fig. 15. Cumulative ASs count (# of ASs with aggregation degree larger than the value of the X-axis)

About improving aggregation, we also notice the evaluation results in [38] that the address fragmentation constitutes approximately 75% of the disaggregation of the prefixes. Our latest evaluation confirms the continuance of this trend. The topological aggregation rule so far is hard to be realized by the current addressing and routing system. However, the depletion of IPv4 address space also means that the total size can be bound. IPv6 can be a huge challenge for scalability without better aggregation. We argue that although the transition from IPv4 to IPv6 can be a pain, for MILSA, it is also a great opportunity to construct an IPv6 core with all the locator strictly topologically aggregated and with the help from efficient automatic address allocation technologies. As indicated in [45], IPv6 prefixes so far are well aggregated. Given the fact that the IPv6 locator space is big enough to accommodate future expansion, the 75% disaggregation due to fragmentation can be avoided.

C. Autonomous System (AS) and Aggregation Degree (AD)

1) AS Imbalance and Its Implication: We use the metric of Aggregation Degree (AD) to evaluate how well a specific AS performs aggregation inside itself. The AD is equal to the prefixes announced by the AS that were aggregated inside the AS divided by the total prefixes this AS announced in the global routing table. The AD ranges from 0 to 1 depending on how well the AS performs the aggregation. For example, suppose we have an AS which gets three prefixes from his customers and it aggregates two of them into one and announces totally two prefixes outside. According to our definition, the AD is 1/2.

To observe the distribution of ADs among different ASs, we first put the cumulative AD of the total AS space as X-axis (ranging from 0 to 1), and select out all the ASs that have higher ADs than the designated AD ratio. The result is shown in Fig. 15, and we can observe that about 1/4 of the total AS space (about 8,726 ASs) share a pretty even aggregation distribution between 0 and 1, the other 3/4 of the total AS space has the AD of 1 because most of them are small stub networks that announce only one prefix in the global routing table. Moreover, among the 1/4 of the ASs that announce more than 1 prefix, their cumulative percentage of AD and the AS count approximately match the linear trend which indicates that this portion is mostly the miscellaneous transit ASs (or ISPs) that do the aggregation of their customers and announce the prefixes to their upstream providers. In Fig. 17, however, we demonstrate the relationship between prefixes announced and the corresponding AS counts. We sum the prefixes announced by each AS and sort them by the number of prefixes they announced in a decreasing order. The distribution trends reveal a significant imbalance among these different ASs. Specifically, the top 5% (1,500 out of 301,659) ASs announced 60% (183,881 out of 301,659) of the total prefixes. The top 30% (9,000 out of 301,659) ASs announced almost 90% (263,267 out of 301,659) of the total prefixes. As a rough average, every AS contributes 9 prefixes to the global routing table. Moreover, we sort the ASs according to their AD in increasing order and calculate their cumulative AD. The results are shown in the Fig. 16. The cumulative AD ranges from 0.44 for the top 50 ASs to 0.61 for all the 32,141 ASs. Again, this result is consistent with what we observe in Fig. 17.

Since the top ASs mostly are transit-ASs or stub-ASs with more prefixes announcements and lower AD. Thus, we argue that for potential short-term solutions (such as Virtual Aggregation [22, 35]) aiming to resolve the aggregation and scalability issue, these transit AS with more prefixes announced and with lower AD should be considered with higher priority than the other ASs. This is one of the deployment strategies of our MILSA solution.

2) AS semantic overloading: The failure of achieving acceptable aggregation for these top transit-ASs mostly is due to reasons we observed in Section VII.B. However, from the architectural design perspective we also argue that one of the deeper underlying reasons is "AS overloading" as we name it. The AS concept is mainly used as a domain of connectivity in the current inter-domain routing system. However, we notice that AS is basically a group concept among organizations due to commercial connection or dependency. This overloading makes the configuration and management of the domain policies very awkward and inconvenient. The efficiency and consistency of the inter-domain routing is also impaired. Solutions, such as RCP [53], try to ease the configuration and management in the AS by centralizing policy and path selection decisions. However, RCP does not change the fact that the AS is still overloaded and the two different levels of domain policies are still mixed together. One of the most direct benefits of performing the separation is the easiness of applying routing policies and performing routing domain federation. Moreover, AS number and prefixes are the two different aggregation granularities that inter-domain routing system is based on. However, the fact is that the AS number in BGP is not a significant factor in improving the aggregation of the global routing table. Given all these temporary or short-term solutions attempting to improve the aggregation, we argue that without the presence of a complete separation of the overloaded semantics of AS, the scalability, configuration and management issues will go on in the future, and ID locator split solutions such as HIP [15] and LISP [10] themselves may not be enough for long-term evolution. In even longer term consideration, to prevent future potential ossification, two tiers may not be enough, thus we may need to divide them in more granularities. In the future, defining good inter-tier and

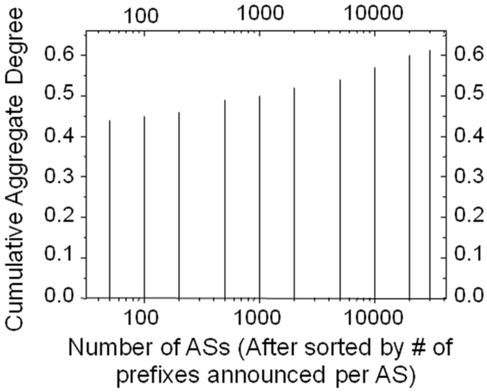


Fig. 16. Top-N ASs and their cumulative aggregation degree (ASs are sorted by their prefixes announced)

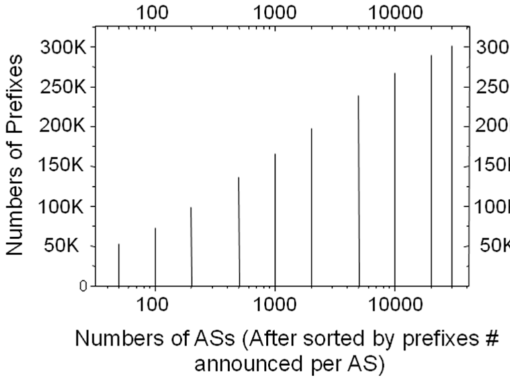


Fig. 17. Top-N ASs and their cumulative prefixes announced (ASs sorted by their prefixes announced)

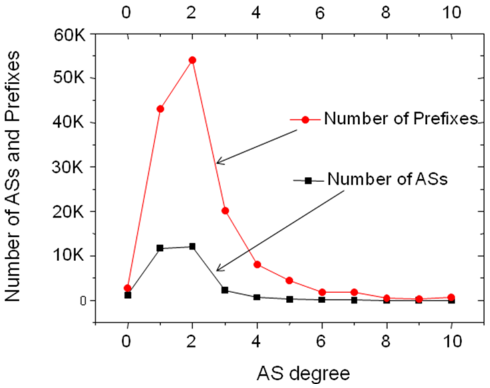


Fig. 18. Stub multihoming: Number of ASs and Prefixes

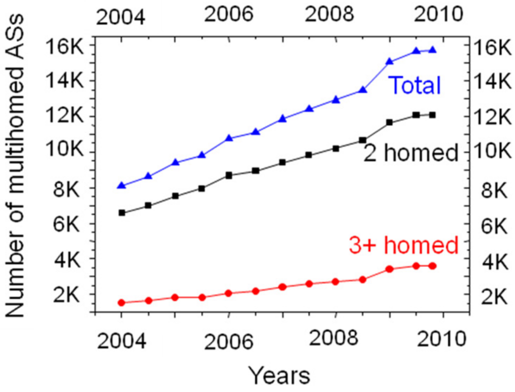


Fig. 19. Multihoming trends

inter-realm interaction protocols and interface are necessary to harvest the above benefits.

D. Site Multihoming Evaluation

We focus on AS level multihoming, i.e., site multihoming since it has a significant impact on the routing scalability. Since the transit ASs are generally ISPs that peer with more than one other ASs, they are “genuinely and naturally” multihomed. Thus, the multihoming under investigation in this section is that of the ASs around the edge of the networks, i.e., stub networks. To reveal the latest multihoming status, we extract the data from the latest routing table data. We first infer the AS relationships by combining several algorithms [32, 54]. After that, we define several rules to determine if a stub-AS is a multihomed one or not based on the inference results, which include: (a) The AS under investigation should have connection degree greater than or equal to one, (b) The AS under investigation should have at least two upstream providers offering connection, (c) The AS itself has no customer link, i.e., no down-link customer AS, (d) Prefixes announced by the AS should appear in one of its providers’ BGP routing table.

Based on the above rules, we have designed a rough algorithm (Algorithm M shown in Table II) to determine if an AS is multihomed and to calculate the number of multihomed stub networks and record their multihoming degree. We observe that among the total 33,485 ASs, there are totally 28,705 stubs-ASs. Among the stub-ASs, about 40.94% of them are single homed and 54.86% are multihomed (with multihomed degree larger or equal to 2). Notice that in our sample there are 1,204 (4.2%) “dead” ASs that connect to nowhere, which may be because of some on-going configuration or testing. We also computed statistics of the prefixes these multihomed ASs announced. We observe that the prefixes announced by multihomed stub sites constitute about 67% of the total prefixes announced by stub networks and 34% of the prefixes in the global routing table. It is easy to imagine if these portion cannot achieve acceptable AD, the total aggregation of the CIDR can suffer, hence lead to poor scalability of the routing system.

Further details on the comparison of the number of ASs and the prefixes they announced per multihomed degree are shown in Fig. 18. We can observe that 2-homed and 3-homed cases comprise the biggest parts regarding to the raw AS count as well as the prefixes announced in the global routing table. Statistical evaluation on the multihoming trends of the last 5 years is shown in Fig. 19 from which we see that the number of the multihomed stub-ASs has doubled over the last 5 years and corresponding prefixes announced by them has increased by 50%. The underlying implication is that research on how to facilitate the fast traffic switching and even load balancing can be of great interests to these types of stub-ASs. It can also be used for determining the deployment strategies for future potential solutions.

We also realize the limitations of the IPv4 multihoming [49] and the failure of the aggregation in IPv4. In MILSA,

TABLE II
ALGORITHM M TO DETERMINE MULTIHOMED SITES

```
Algorithm M: Determine multihomed sites
Objective BGP Routing table, denoted T;
Get the inference results S by algorithm [32, 55];
//Get stub-ASs set
Scan S iteratively to find the ASs with no
customer ASs;
Mark the ASs without customer ASs, get set Sstub;
//Filter out and mark the multihomed ASs set
Scan Sstub set iteratively to find ASs with
connection degree greater than 1;
Mark the ASs with the degree number, get set
Sstub(Degree);
The marked set Sstub(Degree) is the multihomed
stub-ASs set, and the degree number is the
multihoming degree;
```

we build multi-tier separated and extensible realms upon the fully aggregated IPv6 locator based routing system. Multihoming, load balancing and traffic engineering, and policy enforcements can be conceptually clear and right without any "overloading" semantic difficulties as in the current Internet.

E. MILSA Achievable Improvements

Though MILSA is designed to have multiple benefits as discussed in Section VI, here in this Subsection, we only focus on the effect on reducing the inter-domain routing table size by the three deployment models discussed in Subsection VI.H.

1) Total Routing Table Size Reduction: First we need to note that the current DFZ routing table size is about 301K and it is still increasing by 20% every year. So, in this part of the evaluation, to avoid confusion, we will not count this quasi-constant increase rate in. Instead, we evaluate how much reduction we can get for the global routing table size if we deploy MILSA from this static point of time.

Since the ISPs can be categorized into tier-1, tier-2, and other small ISPs, and if we go through the scalability-driven deployment model, the medium and small ISPs will be the major motivators of the new architecture. For evaluation purpose, based on the sample we got in Subsection VII.B.2, we assume that the major motivators come from the tier-2 service providers. Note that in the real case, there are no strict guidelines to decide whether an ISP is a small or big one. Instead, by analyzing the results of deploying MILSA first in tier-2 providers, we can know approximately how much reduction the scalability-driven deployment model can bring us. We also consider different deployment speeds and analyze how the trend will look like with each deployment speed. The results are shown in Fig. 20. In our sample, the tier-2 service providers totally announce 60% prefixes out of the whole 301K entries in the DFZ routing table. Note that we estimate the trends of three cases in which 10%, 20%, and 30% of the total tier-2 providers begin to deploy MILSA each year. We also curve the lines a little bit taking into consideration of facts that in the real case initially the reduction benefits can come a little bit slow than in the middle of the process, so is the case for the end of the lines which are close to the lower bound of the reduction. The lower bound is the ideal case that all the target ASs finish the deployment and their prefixes announced in the global routing table are significantly reduced.

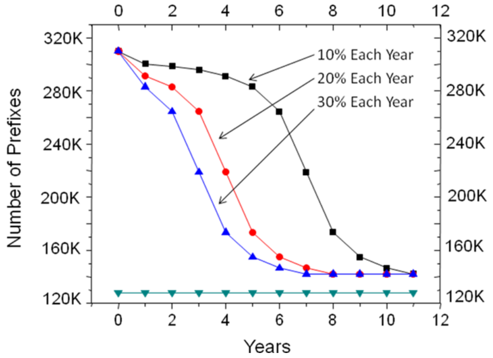


Fig. 20. DFZ routing table size reduction of scalability-driven deployment model, with different deployment speeds.

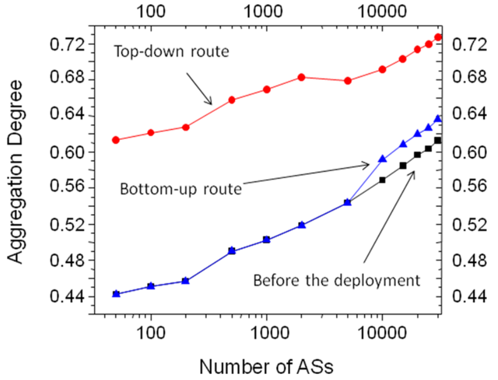


Fig. 21. Cumulative Aggregation Degree (AD) improvements of both "top-down route" and "bottom-up route" (Note: 1. ASs are sorted by their initial prefixes announced; 2. In each route, assume a first-step of 50% reduction of the unaggregated prefixes)

In a perfect case, each such AS need only one prefix in the global routing table.

Similarly, for mobility-driven deployment model, we can also evaluate the reduction trends in different deployment speeds. The shapes of the lines turn to be close to the above figure. The only difference here in this case is that we need to find which parts of the ASs to be considered as mobile ISPs that are likely to be potential motivators of this model. In our evaluation, we try several different cases in which 30% and 50%, of the tier-1 and tier-2 providers are related to the ISPs who provide mobile services to the subscribers and they are treated as potential motivators to the mobility-driven deployment model.

For the multihoming-driven deployment model, we did similar things. However, there are also some fundamental differences for this model. Firstly, multihoming's motivators are generally the stub-ASs that have multihoming demands. As we evaluated in Subsection VII.D, the multihomed ASs announce 34% prefixes out of the total DFZ routing table entries. So if we follow the multihoming-driven deployment model, intuitively this 34% reduction will be the upper-bound of the achievable benefit. However, note that there is also some "byproduct" of this model. As observed in [38], multihoming is one of the fastest increasing underlying drives for the 20% increase of DFZ total entries every year. So, if we follow the multihoming-driven deployment (***"bottom-up route"***) model, part of the total 20% increase every year is expected to be diminished. Hence, it is reasonable to believe that the total benefits will be larger than the 34% total routing table size reduction. Other benefits of multihoming that are not directly visible include better load-balancing and better local AS-level policy enforcement, which are also very important features provided by MILSA.

2) Aggregation Degree Improvement: We further evaluate how much the "top-down route" and "bottom-up route" can bring us in terms of the Aggregation Degree (AD) improvements. First, we draw the cumulative AD line as what we have done in Fig. 16. The cumulative AD ranges from 0.44 for the top 50 ASs to 0.61 for all the 32,141 ASs. Then we consider the "top-down route" in which we deploy MILSA firstly in the tier-1 and tier-2 ISPs for no matter scalability or mobility benefits. As evaluated before, the top-2000 ASs cover almost all the tier-1 and tier-2 ISPs, and the rest about 28,000 ASs are stub-ASs. So for the top-2000 ASs, we assume in the first step 50% of the unaggregated prefixes can be aggregated. This seems to be a reasonable assumption because we cannot be so confident that for every AS that decide to deploy MILSA and the MBR at the edge of their network all the unaggregated prefixes can be aggregated over a night. So it is fair to assume 50% of the unaggregated prefixes are aggregated at the first step. However, we can also control to aggregate the prefixes that previously failed to aggregate first. By doing so, gradual improvement of the AD can be achieved in a reasonable pace.

The evaluation results on AD improvements are shown in Fig. 21. From the figure, we can see that for "top-down route" the AD for the top-2000 ASs and the cumulative AD for all the ASs are significantly improved. We also notice the second small "hump" after the 2000 in the X-axis which is because we are following the "top-down route", and most of the stub-ASs cannot benefit from the deployment model. It is also interesting to note that the first small "hump" at about 200 in the X-axis is also the border between the tier-1 and tier-2 ISPs, and by drawing the lines for AD we can observe this fact clearly.

In comparison, for the "bottom-up route," we deploy MILSA in the stub-ASs especially multihomed stub-ASs first. As we discussed above, of the total about 28,000 stub-ASs, around 15,000 of them are multihomed which contribute about 100,000 entries in the DFZ routing table. So we estimate the AD improvement in these ASs and the result is shown in Fig. 21. Again, we also assume in the first step 50% of the unaggregated prefixes are aggregated. The result shows that there is a slight improvement in the cumulative AD of the whole ASs compared with the status before the deployment. Though as we discussed in Subsection VII.E.1 that the multihoming-driven deployment model ("bottom-up route") can bring some other beneficial byproducts, in terms of reducing the DFZ routing table size as fast as possible, it may not be as effective as the "top-down route". Thus, we can argue that for scalability benefit, deployment of MILSA firstly in the ISP side can be a wise choice.

One more observation based on Fig. 21 is that when time goes by, more and more unaggregated prefixes can be aggregated. So in this sense, finally as the MILSA deployment progresses, the lines shown in Fig. 21 will move upward and finally be close to 1 which is the upper bound of the AD in ideal case.

Finally, we have an observation on the whole evaluation process. We know that the prefix-based inter-domain routing system is a distributed cooperative system and there is some inter-relationship or inter-correlation between the aggregation of the stub-ASs and ISPs, and between the ISPs. So when considering the mixed factors, when do top-down or bottom-up routes, in the real deployment case, the resulting reduction effects may be a little different from what we observed, or may show a parallel or combinational shape of our results. Here in this paper, we consider the factors separately, to give a preliminary idea of the effects of the MILSA deployment.

F. Deployment Factor: Definition and Implications

Beside the two "macro-routes" we discussed above for the deployment, in a finer-grained scope, the imbalance across multiple ASs also has a significant impact on the deployment effectiveness. Thus, here, we give a coarse equation on all the factors that we should take into account when determining the priority of each AS deploying our transition mechanism, i.e., deploying the MBR at the edge of the stub-AS.

$$DF \propto f[T,\ CN,\ PX,\ BM,\ AD,\ AS,\ AF,\ P] \quad (2)$$

Each symbol denote a factor that needs to be taken into account when deciding the deployment priority,

DF: Deployment Factor

T: Type of the AS under investigation

CN: Cone number, the total number of ASs that act as customers of the AS, the customer of customer, and so on

PX: Prefixes announced by the AS

BM: BGP update message rate generated by the AS

AD: Aggregation Degree of the AS

AS: AS degree of the AS

AF: Address fragmentation degree of the given AS

P: Pain and incentives of the AS

Appropriate weights can be applied to different factors to calculate the final Deployment Factor. Through this equation (2), each AS has a DF which can be used to describe and evaluate the emergency degree of applying the new solution. As observed by [55], different ASs may have different capabilities and motivations to update their devices, and their consciousness and the time they feel pain will also vary. Thus, we also use the factor P to reflect this difference. By balancing all these factors, and applying different weights to these factors, we can design different deployment strategies that fit the practical requirements. For example, some rough rules can be listed to guide the deployment:

(a) Upgrade the ASs that have most urgent request first,

(b) Then upgrade those announce the most prefixes,

(c) Then upgrade those with lowest AD,

(d) Then upgrade those with biggest address fragmentation

The order of these rules may vary according to the real deployment strategies to reflect different considerations.

VIII. Conclusion

In this paper, we have tried to justify a holistic architectural thinking for all kinds of challenges we are facing, and tried to depict a rough view of where we can go by following different principles. One of our basic MILSA design goals is to address most of the challenges in one unified architectural view by appropriately balancing both short term and long term requirements, and integrating different design features for the transition. It is not a single attempt on a specific problem, but an attempt to address the basic roots of many problems. However, we must re-emphasize that we are not claiming that MILSA is omnipotent. On the contrary, we must admit that as a big whole architecture, behind many high level descriptions and discussions, especially about the concept of multi-tier separation, realms and routing domain separation, significant research and experiments are still needed to be done in the future.

ACKNOWLEDGMENT

Special thanks to all the people who contributed their discussions and those extraordinary candidate solutions to the RRG that helped us think and reshape our ideas to meet our design goals. Also many thanks to the reviewers who offered their valuable advices and helped us improve the paper.

REFERENCES

[1] J. Pan, S. Paul, R. Jain, et al., "MILSA: A Mobility and Multihoming Supporting Identifier Locator Split Architecture for Next Generation Internet," IEEE Globecom 2008, New Orleans, LA, December, 2008.

[2] J. Pan, S. Paul, R. Jain, et al., "Enhanced MILSA Architecture for Naming, Addressing, Routing and Security Issues in the Next Generation Internet," IEEE ICC 2009, Dresden, Germany, June, 2009.

[3] J. Pan, S. Paul, R. Jain, et al., "Hybrid Transition Mechanism for MILSA Architecture for the Next Generation Internet," Proceedings of the Second IEEE Workshop on the Network of the Future (FutureNet II), IEEE Globecom 2009, Honolulu, Hawaii, 30 November - 4 December, 2009.

[4] T. Li, "Design Goals for Scalable Internet Routing," draft-irtf-rrg-design-goals-01 (work in progress), July, 2007.

[5] D. Meyer, L. Zhang, K. Fall, "Report from IAB workshop on routing and addressing," RFC 4984, September, 2007.

[6] C. Perkins, Ed., "IP Mobility Support for IPv4", RFC 3344, August 2002.

[7] D. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6", RFC 3775, June, 2004.

[8] J. Rosenberg, H. Scheulzrinne, G. Camarillo, et al., "SIP: Session Initiation Protocol," RFC 3261, June, 2002.

[9] V. Devarapalli, R. Wakikawa, A. Petrescu, et al., "Network Mobility (NEMO) Basic Support Protocol," RFC 3963, January, 2005.

[10] Internet Research Task Force Routing Research Group Wiki page, 2008. http://trac.tools.ietf.org/group/irtf/trac/wiki/RoutingResearchGroup

[11] T. Li, "Internet Draft: Preliminary Recommendation for a Routing Architecture," draft-irtf-rrg-recommendation-00, February, 2009

[12] D. Farinacci, V. Fuller, et al, "Internet Draft: Locator/ID Separation Protocol (LISP)," draft-farinacci-LISP-05, September 29, 2009.

[13] C. Vogt, "Six/one router: a scalable and backwards compatible solution for provider-independent addressing," In ACM SIGCOMM MobiArch Workshop, Seattle, WA, 2008.

[14] D. Jen, M. Meisel, D. Massey, L. Wang, B. Zhang, and L. Zhang, "APT: A Practical Tunneling Architecture for Routing Scalability," Technical Report 080004, UCLA, 2008.

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

[16] E. Nordmark, M. Bagnulo, "Shim6: level 3 multihoming Shim protocol for IPv6," draft-ietf-shim6-proto-09, October, 2007.

[17] I. Stoica, D. Adkins, et al, "Internet Indirection Infrastructure," ACM SIGCOMM '02, Pittsburgh, Pennsylvania, USA, 2002.

[18] P. Nikander, et al, "Host Identity Indirection Infrastructure (Hi3)," in the Second Swedish National Computer Networking Workshop 2004 (SNCNW2004), Karlstad, Sweden, November, 2004.

[19] M. O'Dell, "GSE - An Alternate Addressing Architecture for IPv6," draft-ietf-ipngwg-gseaddr-00.txt, 1997.

[20] G. Huston, "Scaling Inter-domain Routing-A View Forward," Internet Protocol Journal, 4(4):2-16, December, 2001.

[21] G. Huston, "Analyzing the Internet BGP Routing Table," The Internet Protocol Journal, July, 2003.

[22] P. Francis, X. Xu, H. Ballani, "FIB Suppression with Virtual Aggregation and Default Routes," draft-francis-idr-intra-va-01.txt, September, 2008.

[23] Y. Afek, O. Ben-Shalom, and A. Bremler-Barr, "On the Structure and Application of BGP Policy Atoms," in ACM SIGCOMM Internet Measurement Workshop, November, 2002.

[24] L. Subramanian, M. Caesar, C. Ee, M. Handley, M. Mao, S. Shenker, I. Stoica, "HLP: a next-generation interdomain routing protocol," ACM SIGCOMM, August, 2005.

[25] L. J. Cowen, "Compact routing with minimum stretch," the 10th Annual ACM-SIAM Symposium on Discrete Algorithms (Jan. 1999), pp. 255-260.

[26] X. Yang, "NIRA: A new Internet routing architecture," Proc. ACM SIGCOMM FDNA 2003 Workshop, Karlsruhe, Germany, August, 2003.

[27] I. Castineyra, N. Chiappa, M. Steenstrup, "The Nimrod Routing Architecture," RFC 1992, August, 1996.

[28] R. Whittle, S. Russert, "TTR Mobility Extensions for Core-Edge Separation Solutions to the Internet's Routing Scaling Problem," August 25, 2008.

[29] C.de Launois, B. Quoitin, and O. Bonaventure, "Leveraging Network Performances with IPv6 Multihoming and Multiple Provider-Dependent Aggregatable Prefixes," Elsevier Computer Networks, volume 50, June, 2006.

[30] R. Atkinson, S. Bhatti, and S. Hailes, "A Proposal for Unifying Mobility with Multi-Homing, NAT, & Security," in ACM MobiWac, Greece, 2007.

[31] T. Bates, Y. Rekhter, "Scalable Support for Multihomed Multi-provider Connectivity," RFC 2260, January, 1998.

[32] L. Gao, "On Inferring Autonomous System Relationships in the Internet," IEEE Globe Internet, November, 2000.

[33] L. Subramanian, S. Agarwal, J. Rexford, and R. H. Katz, "Characterizing the Internet hierarchy from multiple vantage points," Proc. IEEE Infocom 2002, June, 2002.

[34] F. Wang, L. Gao, J. Wang, and J. Qiu, "On understanding of transient interdomain routing failures," Proc. ICNP 2005, Boston, MA, November, 2005.

[35] H. Ballani, P. Francis, T. Cao, and J. Wang, "ViAggre: Making Routers Last Longer!" Proc. Hotnets, 2008.

[36] T. G. Griffin and G. Huston, "BGP Wedgies," RFC 4264, November 2005.

[37] O. Bonaventure, "Reconsidering the Internet Routing Architecture," Internet Draft draft-bonaventure-irtf-rrg-rira-00.txt, March, 2007.

[38] T. Bu, L. Gao, and D. Towsley, "On characterizing BGP routing table growth," In IEEE Global Internet Symposium, November, 2002.

[39] T. G. Griffin, A. D. Jaggard, and V. Ramachandran, "Design principles of policy languages for path vector protocols," in Proceedings of ACM SIGCOMM'03. ACM Press, August 2003, pp. 61-72.

[40] C. Dovrolis, "What would Darwin think about clean-slate architectures?" ACM SIGCOMM Computer Communication Review. January 2008.

[41] A. Feldmann, "Internet clean-slate design: What and why?" SIGCOMM Computer Communication Review, vol. 37, no. 3, pp. 59-64, July 2007.

[42] N. Feamster, L. Gao, and J. Rexford, "How to lease the Internet in your spare time," ACM SIGCOMM Computer Communication Review, pp. 61-64, January, 2007.

[43] D. Thaler, "Why do we really want a ID/locator split anyway?" presented to MobiArch 2008, Seattle, WA, Auguest, 2008.

[44] L. Mathy, et al, "LISP-DHT: Towards a DHT to map identifiers onto locators," draft-mathy-lisp-dht-00, February, 2008.

[45] H. Zhang and M. Chen, "Forming an IPv6-only Core for Today's Internet," Proc. ACM SIGCOMM IPv6 Workshop, August, 2007.

[46] X. Xu, Dayong Guo, "Hierarchical Routing Architecture," Proc. 4th Euro-NGI Conference on Next Generation Internetworks, Krakow, Poland, 28-30 April, 2008.

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

[48] A. Durand, R. Droms, B. Haberman, J. Woodyatt, "Dual-stack lite broadband deployments post IPv4 exhaustion," draft-durand-softwire-dual- stack-lite-01, November 2008.

[49] J. Abley, K. Lindqvist, E. Davies, et al., "IPv4 Multihoming Practices and Limitations," RFC 4116, July, 2005.

[50] Oregon RouteViews, "University of Oregon RouteViews Project," Eugene, OR. [Online]. Available: http://www.routeviews.org

[51] T. Bates, et al., The CIDR Report. [Online]. http://www.cidr-report.org

[52] CAIDA: The Cooperative Association for Internet Data Analysis [Online]. Available: http://www.caida.org

[53] N. Feamster, H. Balakrishnan, J. Rexford, A. Shaikh, and J. van der Merwe, "The case for separating routing from routers," Proc. ACM SIGCOMM Workshop on Future Directions in Network Architecture, August, 2004.

[54] X. Dimitropoulos, D. Krioukov, M. Fomenkov, "AS Relationships: Inference and Validation," ACM SIGCOMM Computer Communication Review, January, 2007.

[55] B. Zhang, L. Zhang, "Evolution Towards Global Routing Scalability," draft-zhang-evolution-01.txt, March 3, 2009.

[56] Internet of Things [Online]. Available: http://en.wikipedia.org/wiki/Internet_of_Things

[57] J.H. Saltzer, D.P. Reed, D.D. Clark, "End-to-end arguments in system design," ACM Trans. Computer Systems (TOCS), Vol.2, Issue 4, pp 277-288, November, 1984.

[58] S. Paul, R. Jain, J. Pan, Mic Bowman, "A Vision of the Next Generation Internet: A Policy Oriented Perspective," British Computer Society (BCS) International Conference on Visions of Computer Science, Imperial College, London, September 22-24, 2008.

[59] V. Fuller, T. Li, "Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan," RFC 4632, August, 2006.

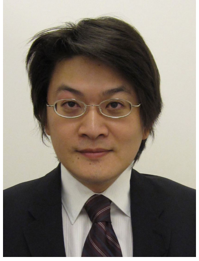
[60] J. Day, Patterns in Network Architecture: A Return to Fundamentals, Prentice Hall, Jan 2008.

[61] X. Dimitropoulos, D. Krioukov, G. Riley, K. Claffy, "Revealing the Autonomous System Taxonomy: The Machine Learning Approach," Passive and Active Measurements Workshop (PAM), March, 2006.

Jianli Pan received his B.E. in 2001 from Nanjing University in Posts and Telecommunications (NUPT), Nanjing, China, and M.S. in 2004 from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China. He is currently a Ph.D. student in the Department of Computer Science and Engineering in Washington University in Saint Louis, MO USA. His current research is on the next generation Internet architecture and related issues such as routing scalability, mobility, mulithoming, and Internet evolution. He is a student Member of IEEE .

Raj Jain is a Fellow of IEEE, a Fellow of ACM, a winner of ACM SIGCOMM Test of Time award, CDAC-ACCS Foundation Award 2009, and ranks among the top 50 in Citeseer's list of Most Cited Authors in Computer Science. Dr. Jain is currently a Professor of Computer Science and Engineering at Washington University in St. Louis. Previously, he was one of the Co-founders of Nayna Networks, Inc - a next generation telecommunications systems company in San Jose, CA. He was a Senior Consulting Engineer at Digital Equipment Corporation in Littleton, Mass and then a professor of Computer and Information Sciences at Ohio State University in Columbus, Ohio. He is the author of "Art of Computer Systems Performance Analysis," which won the 1991 "Best-Advanced How-to Book, Systems" award from Computer Press Association. His fourth book entitled " High-Performance TCP/IP: Concepts, Issues, and Solutions," was published by Prentice Hall in November 2003. Recently, he has co-edited "Quality of Service Architectures for Wireless Networks: Performance Metrics and Management," published in April 2010.

Subharthi Paul received his BS degree from University of Delhi, Delhi, India, and Masters degree in Software Engineering from Jadavpur University, Kolkata, India. He is presently a doctoral student in Computer Science and Engineering at Washington University in St. Louis, MO USA. His primary research interests are in the area of Future Internet Architectures.



Chakchai So-in received his B.Eng. and M.Eng. degrees from Kasetsart University, Thailand in 1999 and 2001. He also received M.S. and Ph.D. degrees from Washington University in St. Louis in 2006 and 2010. All are in computer engineering. In 2003, he was an intern in a CNAP at NTU and obtained CCNP and CCDP certifications. He was an intern at Cisco Systems, WiMAX Forum, and Bell Labs during summer 2006, 2008, and 2010, respectively. His research interests include architectures for future wireless networks, congestion control, protocols to support network and transport mobility, multi-homing, and privacy, and quality of service in broadband wireless access networks.