基於Linux的身份和位置分離映射機制設計與實現
Design and Implementation of Separation and Mapping Scheme of Identifier and Locator Based on Linux

孫照輝 Zhao-Hui Sun, 楊水根Shui-Gen Yang, 邱峰Feng Qiu, 張宏科 Hong-Ke Zhang, 秦雅娟 Ya-Juan Qin
北京交通大學電子資訊工程學院
03211079@bjtu.edu.cn, 04111026@bjtu.edu.cn, 03211142@bjtu.edu.cn, hkzhang@center.njtu.edu.cn

摘要

在傳統互聯網中TCP/IP協議體系中，IP位址既代表節點的身份標識又代表節點的位置標識。IP位址的這種雙重語義不利於支持移動性、安全性和路由可擴展性等功能。將IP位址的身份標識和位置標識分離已成為下一代互聯網的設計思路。一體化網路通過引入接入標識、交換路由標識和接入交換路由器，把網路劃分為使用接入標識的接入層和使用交換路由標識的核心層，從而達到身份與位置分離。本文主要基於Linux平臺設計和實現一體化網路中的身份與位置分離映射機制，具體包括協定棧資料包轉發流程設計、映射關係存儲結構設計等。

關鍵字：一體化網路、接入標識、交換路由標識、標識分離映射。

Abstract

In the traditional TCP/IP stack, the IP address represents both the identifier and locator roles of a node. This dual nature makes it difficult to support mobility, security, routing scalability and other functions well. The new idea of separating the identifier and locator roles of the IP address becomes a hot research top in next generation Internet. In order to achieve the separation of identifier and locator, the universal network introduces access identifier, switch route identifier, and access switch router to divide the network into the access layer using access identifier and the backbone layer using switch route identifier. This paper mainly designs and implements the separation and mapping scheme of identifier and locator in the universal network based on Linux platform, including the packet forwarding process and the data structure of mapping relationship tables.

Keywords: Universal Network, Access Identifier, Switch route Identifier, Separation and Mapping of Identifier and Locator.

1 緒論

隨著互聯網業務不斷擴展，通過互聯網承載多種業務成為今後互聯網的發展趨勢。這使得互聯網面向資料傳輸的初始設計面臨極大的挑戰。例如，IP位址面向固定終端的設計出發點限制了互聯網對移動性的支持；平等自治的初始設計思想使得互聯網網路層中缺少對終端身份進行驗證的基礎性安全支持，引發了冒充等安全問題；核心路由表加速膨脹也成為互聯網發展的隱患。

在傳統互聯網TCP/IP協定體系中，IP位址具有雙重功能：網路層使用IP位址作為節點的位置標識，用於路由；傳輸層使用IP位址作為節點的身份標識，用於建立傳輸層連接[1]。IP位址的這種雙重性不利於支持節點的移動性、安全性、多家鄉等。近來，國內外研究者紛紛提出了將節點的身份標識和位置標識分離的設計思想，如互聯網工程部（Internet Engineering Task Force, IETF）的主機身份標識協定（Host Identity Protocol, HIP）[2]、思科公司的位置／身份分離協議（Locator/ID Separation Protocol, LISP）[3][4]、一體化網路[5][6]等。

在一體化網路中，使用接入標識（Access Identifier, AID）代表節點的身份資訊，使用交換路由標識（Switch Route Identifier, SRID）代表節點的位置資訊。所有終端都經一體化網路的核心網路邊界路由器——接入交換路由器（Access Switch Router, ASR）接入網路。通信時，通信雙方使用AID來標識資料包的源和目的；當該資料包到達ASR時，ASR將資料包中的AID都替換為對應的SRID，然後由SRID在核心網路把資料包路由到通信對端的ASR；通信對端的ASR接收到該資料包後，進行一次SRID到原來AID的逆替換，並轉發給通信對端。通過上述機制，一體化網路克服了互聯網原始設計中由於IP位址二義性帶來的種種弊端，能夠較好的滿足移動性、安全性、路由可擴展等多方面的需求。

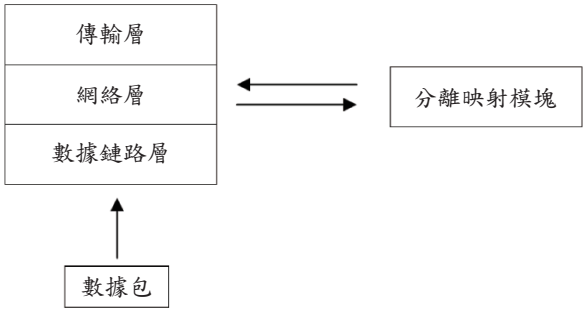
本論文主要設計並實現了在Linux平臺下身份和位置的分離映射機制。本論文的組織結構如下：第2節介紹分離映射的設計方案，第3節介紹在Linux平臺下分離映射機制的實現，第4節進行了實驗驗證，第5節對論文進行總結。

2 分離映射的設計方案

2.1 總體設計思路

為了保證和現有TCP/IP協議棧的相容性，一體化網路的接入標識和交換路由標識採用了32位元或128位元的標識長度。但標識不同於普通IP位址，每一位都有它特殊含義，某些位元定義所屬的標識類型，某些位元定義所歸屬的域，對於接入標識，有些位元還表示是否屬於移動終端等等。此外，接入標識也不像IP位址必須要有網路首碼，它僅表示用戶的身份。無論是對接入標識的分析還是對交換路由標識，都是在ASR上進行的。

ASR要對接收到的資料包的源和目的標識進行操作，提前必須截獲資料包。本方案通過在網路層增加一個鉤子函數，串列於原協定棧，將資料包引入到所設計的分離映射模組中。分離映射模組在協定棧中的位置如下圖一所示。



圖一 協定棧裏分離映射模組的位置

資料包進入到分離映射模組後，在進行接入標識和交換路由標識映射之前，必然涉及到接入標識和交換路由標識映射關係的雙向查找。因此，ASR需要建立接入標識和交換路由標識之間對應關係的存儲表。

考慮到映射表的存儲複雜性、查詢處理時間和網路的負擔，ASR只保存部分映射關係；其餘映射關係由一體化網路中另一實體映射伺服器負責管理。

ASR管理下的本地終端使用的交換路由標識交由ASR進行分配和管理，方便這些本地終端接入網路時ASR查詢它們的映射關係。當本地終端發起到其他終端通信時，為了實現標識替換，ASR同樣需要知道目的終端的映射關係。故設計兩張表：一張存儲本地接入終端的映射關係，可稱為本地映射關係表；一張存儲通信對端的映射關係，即對端映射關係表，這樣既方便查詢又方便管理。

2.2 分離映射模組資料包轉發流程設計

一體化網路以ASR為分界劃分為接入網和核心網，這種劃分是接入標識和交換路由標識分離映射的基礎。ASR上與接入網相連的介面定義為接入口，配置接入標識；與核心網相連的介面定義為核心口，配置交換路由標識。一體化網路對終端的身份和位置進行分離，既要避免接入標識流入核心網，又要避免交換路由標識流入接入網。在此，通過網卡介面就可從物理上的隔斷實現這種隔離。

資料包處理流程設計中首先要對資料包進行合法性檢查，之後進行不同處理。判斷的三個關鍵因素：資料包接收介面、源標識、目的標識。根據這三個屬性的組合判斷資料包合法性如表1所示。「＊」號表示這裏取任意值。接收介面的屬性有：接入口（A）或者核心口（R）。源標識和目的標識的屬性有：本地接入的接入標識（LC）、非本地接入的接入標識（GC）、本地接入的交換路由標識（LR）、非本地接入的交換路由標識（GR）。資料包合法性檢查遵循以下原則：接入網一側網路介面上收到資料包的源標識和目的標識都應該是接入標識；核心網一側網路介面上收到的資料包的源標識和目的標識都應該是交換路由標識。所以在下表中，3、4、6、7、8、9、11、12都是非法的。

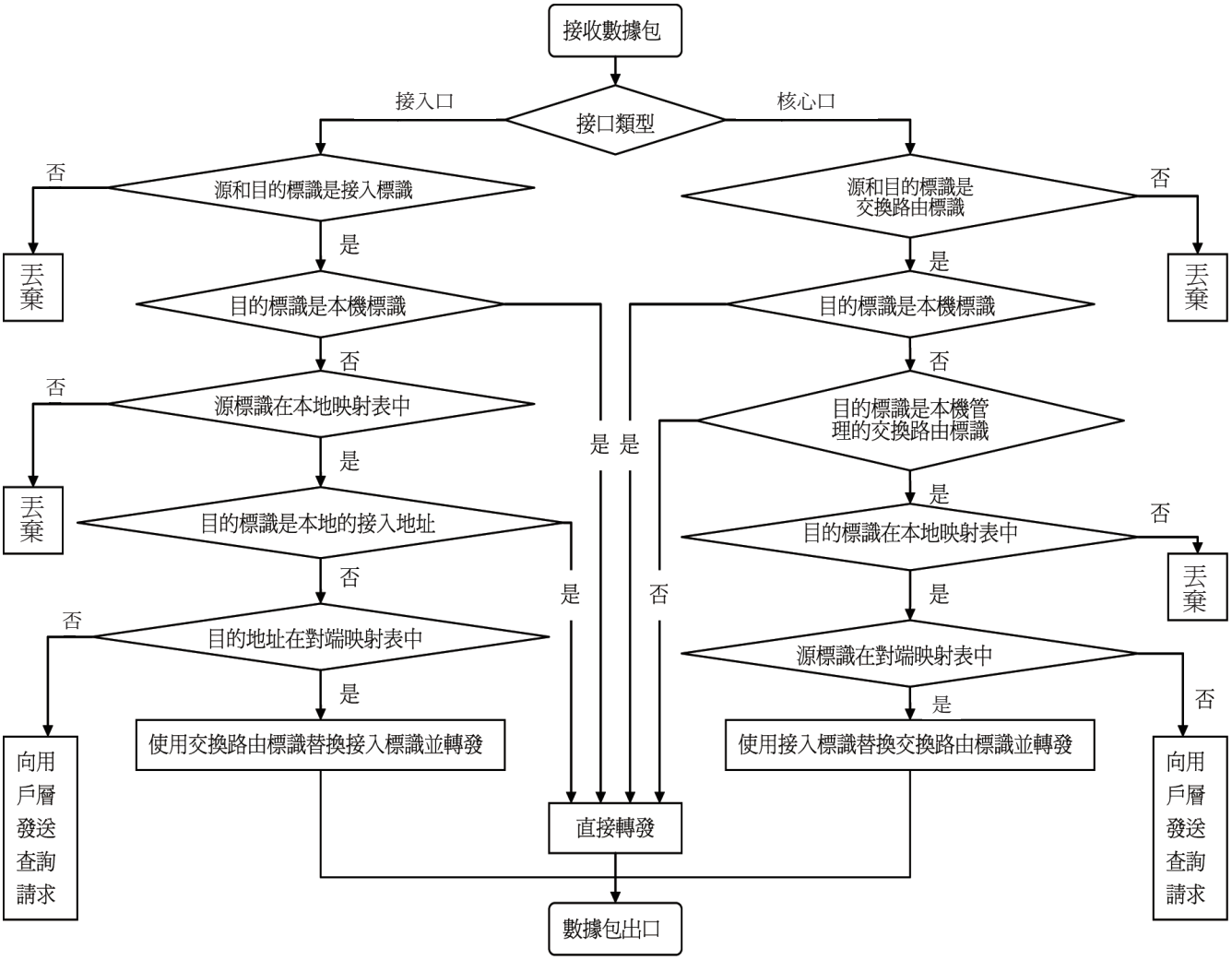
表1 資料包合法性判斷表

序號	接收介面	源標識	目的標識	合法性
1	A	LC	LC	合法
2	A	LC	GC	合法
3	A	LC	LR	非法
4	A	LC	GR	非法
5	A	GC	*	非法
6	A	LR	*	非法
7	A	GR	*	非法
8	R	LC	*	非法
9	R	GC	*	非法
10	R	LR	*	非法
11	R	GR	LC	非法
12	R	GR	GC	非法
13	R	GR	LR	合法
14	R	GR	GR	合法

這樣的設計不僅可以從網路拓撲層面確保終端身份和位置的分離，也方便了網路管理，同時還杜絕了一些安全攻擊。

3 標識分離映射在ASR上的實現

要實現網路層上的標識分離映射，必然需要修改原有的TCP/IP協定棧，考慮到目前路由器大多都採用Linux或Unix作業系統，而且Linux的內核源代碼是完全開源的，有利於編程開發，故針對以上的方案設計選擇在Linux環境下進行編程實現。本實現將分離映射的功能以Linux內核模組（Kernel Module）的方式加載入Linux作業系統內運行。Linux內核模組是內核的一部分，可以在Linux作業系統運行期間動態載入或者卸載。這樣，根據用戶需求，可以動態裝入對應功能的內



圖二　資料包處理流程

核模組，既保證了內核的最小化，又有高度的靈活性。內核模組部分主要涉及分離映射替換模組和映射表模組。此外，ASR上還需要運行用戶層程式用來提供對外的介面，例如與伺服器等其他網路實體的通信；還可用於配置內核模組中的資料，如更改標識映射表等。

3.1分離映射模組

分離映射模組主要功能是在系統內核中進行標識檢查及替換，即ASR中的接入標識和交換路由標識映射過程。首先，在Linux系統協定棧的IP層負責接收所有IP資料包的函數ip_rcv中增加一個鉤子函數，如果此內核模組已載入，系統對資料包的操作就會通過該函數指標指向分離映射功能模組的函數入口。在此模組內首先檢查資料包源標識和目的標識是否需要進行替換，若需要替換再進一步實現映射。資料包將會在這裏被直接丟棄、轉發或替換轉發。

在資料包進入到分離映射函數後，該函數對資料包進行如下步驟的操作：

步驟(1)：檢查映射功能是否打開，如果關閉則無需替換，按照正常系統處理流程進行操作。

步驟(2)：判斷資料包來自於哪類介面，如果是接入口，此時資料包中的源和目的標識都應該是接入標識，若不是，其可能是冒充的非法資料包，直接丟棄。

步驟(3)：根據身為接入標識的源標識檢查該資料包是否是發送給本ASR，若是，直接返回，資料包轉入原來的IP協議層，進行正常的資料操作流程。如果不是ASR自身的，則遍曆內核中本地映射表，將查詢到的源標識的映射關係保存下來，以供後續使用；若沒有查詢到，說明該資料包來自一個新用戶，需要該終端發送認證請求，通過認證後才為它分配映射關係。

步驟(4)：檢查資料包中的目的標識。分成三種情況：第一，如果目的標識也是本地映射表裏的接入標識，表明這是同一個ASR管理域內兩個終端的通信，資料包無需進入核心網，直接把資料包轉發即可；第二，如果目的標識不在本地映射表裏，而在對端映射表中，則用查詢得到的源和目的的交換路由標識替換接入標識，再轉發資料包。此時，網路對這個資料包完成了身份與位置的分離，代表身份資訊的接入標識被代表位置資訊的交換路由標識完全替代，此後資料包進入到核心網；第三，如果目的標識也不在對端映射表中，就需要向用戶層發送映射關係查詢。

被替換後的資料包依靠交換路由標識路由到通信對端所接入的ASR下，再進行由交換路由標識到接入標

識的逆替換。通信對端的ASR同樣進行與上述相類似的逆替換操作：

步驟(1)：對於來自路由口的資料包，只有源和目的標識都是交換路由標識才合法，否則丟棄該資料包。

步驟(2)：再查看資料包是否是發給本ASR，確認是就將資料包返回給系統繼續處理。

步驟(3)：以目的標識去查詢本地映射表。若沒有對應的映射關係則丟棄該資料包，說明通信對端已經不在此ASR下接入，已離開或者關機；若能查詢到目的標識的映射關係對，再進一步判斷源標識是否存儲在對端映射表中。

步驟(4)：如果能在對端映射表中查到源標識的映射關係對，則將查詢得到的源和目的的接入標識替換對應的交換路由標識，完成了逆替換，資料包攜帶著表示身份資訊的接入標識被轉發到通信對端；如果沒查詢不到映射關係，則需要向用戶層進一步查詢。

整個資料包處理流程如圖二所示。

在內核實際操作標識替換時，直接把sk_buff結構體中包含的資料包裏的源和目的標識用所需的標識替換掉，並且需要重新對資料包包頭進行校驗和計算。

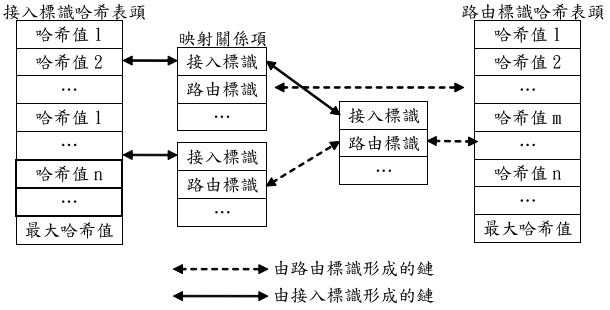
3.2 標識映射表模組

映射表存放的映射關係供替換資料包包頭的源和目的標識使用。查找映射表就是根據已知的一個標識，遍曆映射表找到與其對應的標識。只有當映射表中條目是最新的，才能確保標識被正確替換，資料包才能夠正確無誤的到達通信對端，所以需要通過用戶配置、計時器定時更新或來自其他網路實體的通告消息等多種方式來維護更新標識映射表。

映射表存儲是本分離映射系統的一個關鍵部分。從2.1可以看出，映射分離處理函數對每個收到的資料包都要查找映射關係，針對源和目的標識至少查兩次。故標識映射關係的存儲和查找效率對資料包處理品質和速度有決定性影響，也是分離映射協定棧性能的重要指標。

映射表的實現採用了哈希存儲的方法[7]。哈希表就是在存儲位置和它的關鍵字之間建立一個確定的對應關係f，使每個關鍵字和一個唯一的存儲位置相對應。哈希表的查找演算法時間複雜度是O（1），即查找時間只跟f（key）的計算時間有關，與存儲的表項的數量無關。在哈希表的構建中，f稱為哈希函數，它的選取是影響查找效率的關鍵，本方案選用了Linux 2.6.11內核中的哈希函數，其執行效率能夠滿足需要。另外，在哈希表的構建中，因為一般哈希函數都是一個壓縮函數，所以會出現對不同的關鍵字可能得到同一個哈希索引值的衝突情況。可通過在發生衝突的關鍵字索引後面建立一個單向鏈表，將得到相同哈希的結果的資料內容放入此鏈表。

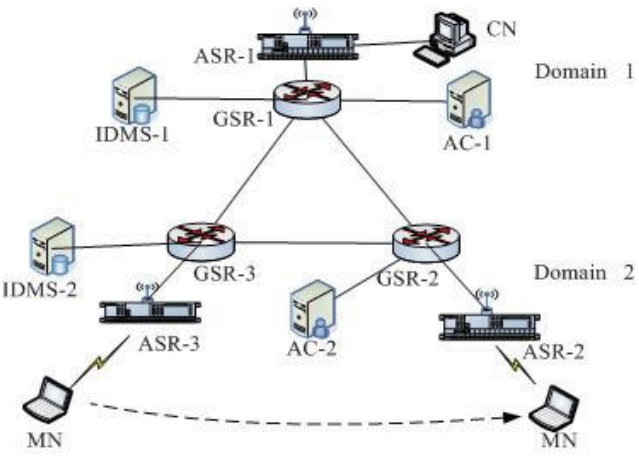
而對於分離映射的hash表在此處卻存在一個問題：一般的哈希表都是單關鍵字，即只能根據表項中的一項內容來查找。但是本方案中既需要以接入標識查找交換路由標識，又需要以交換路由標識查找接入標識，意味著雙關鍵字。為了解決該問題，改用雙哈希表頭：一個表頭hash1是以接入標識作為關鍵字，一個表頭hash2是以交換路由標識作為關鍵字。每一個表項，既要根據接入標識掛在hash1上，又要根據交換路由標識掛在hash2上，類似於十字鏈表的組織結構。雙哈希表的存儲結構如圖三所示。



圖三　雙哈希表存儲結構

4 實驗驗證

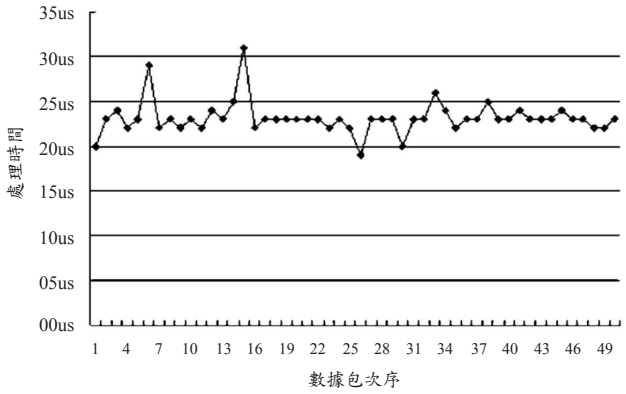
由於分離映射是在系統內核中增加的一個獨立模組，其要對每個資料包的源和目的標識進行分析，自然會增加整個處理資料流程的時間。本文在以下實驗環境中進行測試：



圖四　實驗環境平臺

移動終端MN發送到通信對端CN的資料包經過中間的ASR-3和ASR-1進行標識替換轉發給CN，其中在ASR-3上隨機抽取的50個替換轉發資料包，圖五就是資料包分離映射的處理時延。

可發現資料包的處理時間分佈在18us-32us之間，更多的集中在20us-25us之間。從測試效果上看，分離映射處理時延幾乎沒有給通信帶來影響，資料傳送的時間及流量都不受限。上面的實驗資料是基於兩張映射表只有200個條目的情況，將本地和對端映射表條目都增加到10000-20000條時，分離映射的處理時間沒有明顯增加，總體時間分佈在18us-40us之間，平均值在30us左右。由此可見雙哈希映射表的查詢效率很高。



圖五　處理分離映射時延

5 結論

基於Linux平臺，本文設計並實現了一種適用於一體化網路體系下的身份和位置分離映射機制，並對分離映射模組中資料包轉發流程和映射關係表進行詳細而全面的分析。該分離映射機制的設計和實現能較好實現終端身份資訊和位置資訊的有效分離。

基金專案

國家863計畫（2007AA01Z202）。

參考文獻

[1] H. Balakrishnan, K. Lakshminarayanan, and S. Ratnasamy, et al., *"A Layered Naming Architecture for the Internet [J],"* ACM SIGCOMM Computer Communication Review, Vol. 34, No. 4, 2004, pp.343-352.

[2] R. Moskowitz, *"Host Identity Protocol (HIP) Architecture,"* RFC 4423 [S], IETF, May 2006.

[3] D. Farinacci, V. Fuller, and D. Oran., *"Locator/ID Separation Protocol (LISP), Internet draft, draft-farinacci-lisp-03 [S],"* IETF, Aug. 2007.

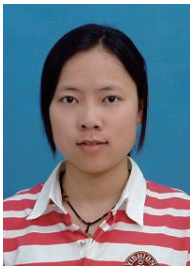
[4] D. Farinacci, *"Locator/ID Separation Protocol (LISP), Internet draft. draft-farinacci-lisp-06 [S],"* IETF, Feb. 2008.

[5] 張宏科、蘇偉，*新網路體系基礎研究——一體化網路與普適服務 [J]*，電子學報，Vol. 35，No. 4，2007，4月，pp.593-598。
Zhang Hongke, Su Wei, *"Fundamental research on the architecture of new network -- Universal network and pervasive services [J],"* Chinese Journal of Electronics, Vol. 35, No. 4, Apr. 2007, pp.593-598.

[6] 董平、秦雅娟、張宏科，*支援普適服務的一體化網路研究[J]*，電子學報，Vol.35，No.4，2007，4月，pp.599-606。
Dong Ping, Qin Yajian, and Zhang Hongke, *"Research on universal network supporting pervasive services [J],"* Chinese Journal of Electronics, Vol. 35, No. 4, Apr. 2007, pp.599-606.

[7] H. C. Thomas, E. Charles, *"Introduction to Algorithms, Second Edition [M],"* Cambridge, Massachusetts London, England, the MIT Press, 2001.

作者簡歷



孫照輝（**Zhao-Hui Sun**），目前是北京交通大學下一代互聯網研究中心在讀碩士生。主要研究方向為IP網路的路由及移動技術、下一代互聯網的移動切換及子網接入等。

楊水根（**Shui-Gen Yang**），目前是北京交通大學下一代互聯網研究中心在讀博士研究生。主要研究方向為移動IP、移動互聯網、下一代網路的移動性管理等。

邱峰（**Feng Qiu**），目前是北京交通大學下一代互聯網研究中心在讀博士研究生。主要研究方向為IP網路的路由及移動技術，移動互聯網、下一代網路的移動性管理等。

張宏科（**Hong-Ke Zhang**），北京交通大學教授，博士生導師。目前主要從事下一代資訊網路關鍵理論與技術的研究工作，並作為首席科學家主持國家973專案「一體化網路與普適服務體系基礎研究」的研究工作。

秦雅娟（**Ya-Juan Qin**）女，山西晉城人，博士，博士生導師。2003年獲北京郵電大學工學博士學位。近年來主要從事互聯網體系結構、移動互聯網路由與交換、寬頻無線通信等領域的技術研究，主持或主研完成多項國家863、國家自然科學基金及國家發改委專案。