清 华 大 学

综 合 论 文 训 练

题目：基于AS编址的互联网可扩展路由机制的仿真评价与测试

系　　别：计算机科学与技术系

专　　业：计算机科学与技术

姓　　名：王庆

指导教师：王之梁副研究员

2015年 6月23日

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：学校有权保留学位论文的复印件，允许该论文被查阅和借阅；学校可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存该论文。

(涉密的学位论文在解密后应遵守此规定)

签　名：王庆　导师签名：王立梅　日　期：2015.6.23

中文摘要

互联网的不断完善和扩展，网络设备和终端设备数量的不断增加，IPv4地址处于枯竭的状态，从IPv4过渡到IPv6的技术逐渐成熟，IPv6 地址本身的诸多优势，都预示未来IPv6网络将会迅速发展，由此带来的IPv6网络路由的可扩展性问题值得研究。

为了解决IPv6网络带来的可扩展路由问题，有学者提出一种层次化路由机制，通过基于无类AS编址的CABA(Classless AS Based Addressing)方案以及基于自治系统号的域间路由机制A-BGP(ASN based BGP) 实现。本文对该CABA下的A-BGP协议进行仿真评价与测试，验证其对路由可扩展性的作用。该CABA编址方案将自治系统号嵌入到IPv6 的地址中。该A-BGP协议仅需向外宣告一条嵌入自治系统号的前缀，这将极大程度的减少全局路由表的大小，同时也会减少UPDATE包的数目以及路由震荡的频率。

本文主要进行了以下工作：

1. 通过评估基于CABA编址的A-BGP下FIB表压缩情况，验证其对路由可扩展性的作用。结果显示基于CABA编址的A-BGP下的FIB表项是现网络中FIB表的10%。
2. 通过SIMBGP仿真平台对比自治系统在基于CABA编址的A-BGP协议和现网络环境中宣告UPDATE的数目，验证A-BGP对路由可扩展性的作用。结果显示当前网络自治系统向外宣告的前缀越多，采用新型A-BGP机制发出的UPFATE数目越少。
3. 在软件路由器Quagga上实现了A-BGP机制并采用Docker进行了试验床测试。结果显示在Tier1层级自治系统构成的拓扑下，全局BGP路由表项从现网络环境中的1万多条减少到17条。

关键词：自治系统；域间路由；可扩展路由

ABSTRACT

With Internet improving and expanding and the number of network terminal equipments increasing, IPv4 address is exhausted. The technical from IPv4 to IPv6 is mature and there are many advantages in IPv6, so IPv6 network will develop rapidly in the future, which may result in IPv6 network routing scalability issue.

To solve this issue, someone provides a hierarchical routing mechanism, which includes the new IPv6 addressing scheme based AS, called CABA (Classless AS Based Addressing) and the new inter-domain routing mechanism called A-BGP (ASN based BGP). There are some simulations and tests to verify the IPv6 network routing scalability under this new mechanism in this thesis. The new IPv6 addressing scheme embedded the autonomous system number. The autonomous system only needs to broadcast one prefix in A-BGP. This could not only greatly reduce the size of the global routing table, but also reduce the number of BGP update packets and routing flapping frequency and convergence time.

This thesis mainly for the following work:

1. Compare the number of current FIB table and the number of new FIB table based CABA and A-BGP to verify the IPv6 network routing scalability. The result shows the new FIB table is 10% of the current.
2. Conduct simulations by SIMBGP on the CABA and A-BGP to verify IPv6 network routing scalability. The result shows that autonomous system of announcing more prefixs can reduce more UPDATE packets under new mechanism.
3. Realize A-BGP by Quagga and test in a virtual network environment made by Docker to verify the IPv6 network routing scalability. The result shows BGP routing table reduces from about 10000 to 17 under the new mechanism A-BGP in the topology of Tier 1.

Keywords: Autonomous System; Inter-domain; Scalable Routing

目 录

主要符号表

CABA	基于无类AS的编址方案(Classless AS Based Addressing)
A-BGP	基于自治系统号的域间路由机制(ASN based BGP)
RIB	路由信息表(Routing Information Base)
FIB	转发信息库(Forwarding Information Base)
NLRI	网络可达信息(Network Layer Reachability Information)
PI	独立供应商(Provider Independent)
PA	聚合供应商(Provider Aggregatable)
HLP	混合链路状态和路径向量协议(Hybrid Link-state And Path-Vector Procotol)
ISP	互联网服务提供商(Internet Service Provider)
GIRO	基于地理信息的域间路由(Geographically Informed Inter-Domain Routing)
GSE	全球区域的终端系统标识(Global, Site, and End-system address elements)
IANA	互联网数字分配机构(Internet Assigned Number Authority)
APNIC	亚太互联网信息中心(Asia-Pacific network information centra)
CNNIC	中国互联网信息中心(China Internet Network Information Center)
RIR	区域互联网注册管理机构(Regional Internet Registry)
ASN	自治系统号(Autonomous System Number)
LIMA	Less-Is-More Architecture
CAIDA	Center For Applied Internet Data Analysis
DFZ	Default-Free Zone

第 1 章　引言

1.1　研究背景

随着网络的不断扩大，越来越多的可路由地址块加入了全球BGP的路由表，2015年6月IPv4路由表的规模达到了58 万条[1]。目前IPv4地址分配基本结束，IPv6 地址逐渐在网络中使用，同时现今网络IPv6向IPv4过渡技术逐渐成熟，都表明IPv6网络将是未来网络的核心，由此产生大规模的网络编址带来的路由可扩展性问题。

当前网络路由的可扩展性较差，主要体现两个方面：路由表规模非线性增长、路由结构扁平化造成路由层次间隔离性较差。针对这两个问题，有学者提出一种层次化路由机制，通过基于无类AS编址的CABA(Classless AS Based Addressing)方案以及基于自治系统号的域间路由机制A-BGP(ASN based BGP) 实现。本文对该CABA下的A-BGP协议进行仿真评价与测试，验证其对路由可扩展性的作用。

1.2　主要工作

本文对基于AS编址的互联网可扩展路由机制进行了仿真评价与测试，主要进行了以下工作：

1. 对域间路由协议BGP、BGP-4、BGP4+以及可扩展路由机制的研究现状、现有的IPv4和IPv6网络编址方案进行了综述；
2. 解释基于AS的无类IPv6编址方案和基于自治系统号的BGP协议的相关细节；
3. 为了验证论文中CABA和A-BGP对路由可扩展性的影响，将当前网络环境下的FIB表与A-BGP机制下FIB表进行对比。首先把从Route Views[2]获取全局RIB表转换成FIB表，然后选取现网络环境下FIB表中的自治系统进行增量部署，将先网络中的FIB表转换成A-BGP下的FIB表，通过对比显示A-BGP下的FIB表项数目是现网络中FIB 表项数目的10%；
4. 通过SIMBGP仿真平台，统计全网拓扑结构下自治系统向外宣布前缀

的update数目。在A-BGP机制下自治系统只需要向外宣布一条嵌有自治系统号的前缀，而不是现今网络中一个自治系统可能向外宣告多条前缀，最多甚至有4000多条。通过分析显示将CABA编址方案和A-BGP路由策略完全部署到现互联网环境中，向外宣布前缀越多的自治系统，向外发布UPDATE 包数减少越多；

5. 在Docker软件上部署多台软件路由器，在由Tire1自治系统构成的简单拓扑上实现A-BGP机制，每台软件路由器为一个自治系统，向外宣布一条路由，查看生成的BGP路由表表项数目。结果显示在CABA编址下A-BGP路由策略中的全局路由表项有17项，远少于现网络环境中约1万条的全局路由表项。

1.3 论文结构

本文共包含七章：

- 第一章：引言，介绍研究背景和主要工作；
- 第二章：相关研究综述，对域间路由协议、可扩展路由机制的研究现状、现有的IP网络编址方案进行了综述；
- 第三章：基于AS编址的互联网可扩展路由机制框架，描述了基于AS的无类IPv6编址方案和基于自治系统号的BGP协议；
- 第四章：基于CABA编址的FIB表压缩评估，将当前网络环境下的FIB表与A-BGP机制下FIB表进行对比，对路由的可扩展性进行评估；
- 第五章：基于AS编址的A-BGP域间路由的仿真评价，通过SIMBGP仿真平台，统计全网拓扑结构下自治系统向外宣布前缀的update数目；
- 第六章：基于AS编址的A-BGP域间路由的实现与测试，在Docker软件上部署多台软件路由器，在Tier1自治系统构成的拓扑上实现A-BGP机制，查看生成的BGP 路由表表项数目；
- 第七章：总结和进一步研究工作。

第 2 章　相关研究综述

2.1　引言

本章对域间路由协议、可扩展路由机制的研究现状、现有的IP网络编址方案进行了综述，首先介绍域间路由的基础知识，然后介绍目前解决路由可扩展性的方法和缺陷，以及现有的IPv4标准编址方案、IPv6 标准编址方案、IPv6新型编址方案。

2.2　域间路由协议

2.2.1　BGP协议

BGP[3](Border Gateway Protocol)是在自治系统之间使用的外部网关路由协议。其在路由的过程中，需要考虑更多的政治、安全、经济因素，不同于内部网关协议，只需要将分组从源地址发送到目的地址。

我们可以将网络结构简化为自治系统间的拓扑图。根据BGP协议对中转流量的作用，我们可以将网络大致分为三类：

- 与BGP图只有一个连接的网络，中转流量不会经过这些网络，因为没有中转接收方，称之为末端网络（stub network）。
- 与BGP图有超过一个连接的网络，除非该网络限制中转，否则流量会经过这些网络进行中转，称之为多连接网络（multiconnected network）。
- 满足部分限制条件(比如交钱)之后，愿意处理第三方分组，称之为穿越网络（transit network）。

BGP路由器之间是通过建立TCP连接，使用TCP端口179进行通信的，可靠又能隐藏数据包在中转过程中其他自治系统的信息。BGP是距离矢量协议，每一台路由器需要维护它到目标的开销以及路径，能够轻易检测是否发生路由环路。每台路由器向外宣布自己的BGP最优路由信息，当一台路由器收到邻居的路径信息时，会与路由表中的路由信息进行对比，如果收到的路由信息比路由表中对应的路由信息更优，则更新路由表，否则丢弃路径信息包。所以，每台

路由器都有一个路由评价最优函数，根据路径的长度、是否该路径满足该BGP特有的限制等多方面的条件进行择优。

BGP邻居协商的过程中，使用四种类型的报文：

- OPEN:建立连接、协商参数。
- KEEPALIVE：周期发送，维护检查和Peer之间的连接。
- UPDATE：交换网络可达路由信息。
- NOTIFICATION：报告网络中发生的各类错误和特殊指令，发生错误TCP连接断开。

BGP邻居协商过程中有5种状态：

- Idle：BGP接收到启动事件的指令，初始化资源，转到Active状态。
- Active：和邻居建立TCP连接，本地路由器发OPEN包给邻居。如果接收到结束事件的指令，释放资源，转到Idle 状态。
- OpenSent： BGP收到OPEN包，检查数据是否正确。如果监测BGP帧头不正确返回到Active 状态。如果OPEN帧的内容错误，发送NOTIFICATION包，关闭TCP连接，返回到Idle状态。如果没有错误，发送OPEN CONFIRM消息，进入OpenConfirm状态。
- OpenConfirm：收到OPEN CONFIRM消息，进入Established状态，如果长时间没有收到OPEN CONFIRM消息，返回到Idle状态。
- Established：在该状态，路由器可以和自己的邻居之间发送UPDATE，KEEPALIVE，NOTIFICATION信息。如果收到断开连接的消息或者超过保持时间，返回到Idle状态。

2.2.2 BGP-4协议

BGP协议主要是为了在网络中交换网络可达信息（AS路径）。因为AS路径信息的存在，使得在路由的过程中，可以比较简单的检测出路由环路，同时使得一些路由策略能够被强制执行。

BGP-4[4]是BGP的扩展，BGP-4提出一个新的支持无类域间路由的机制。在该机制中支持宣告IP前缀，去除了地址类的概念。同时，该机制允许路由聚合和AS路径的聚合。BGP-4采用路径向量算法，综合了距离向量算法和链路

状态算法。BGP-4邻居协商时，使用四种类型的报文，报文的格式和BGP略有不同。在UPDATE 包中，宣布的前缀填写在NLRI(Network Layer Reachability Information) 部分，AS路径填写在Path Attribute部分。

BGP-4中有3个路由信息库：

- Adj-RIBs-In: 存储从收到UPDATE包学到的路由信息。
- Loc-RIB: 应用本地路由策略从Adj-RIBs-In中的路由信息中选出本地路由信息。
- Adj-RIBs-Out: 存储从peer学习到的最优路径信息，用于宣告。

BGP-4邻居协商过程中有6种状态，比BGP多一种Connect状态。

2.2.3 BGP4+协议

BGP4+[5]定义了两种BGP属性格式(MP_REACH_NLRI和MP_UNREACH_NLRI)，用于宣告或者撤销IPv6 路由可达信息的广播。

IPv6和IPv4之间的区别在于IPv6存在范围内的单播地址，特定环境中需要使用特定的地址范围。

IPv6定义了3种单播地址范围：

- global：下一跳属性也需要包含global范围地址。
- site-local：属于non-link-local，仅仅在一个区域范围内有效
- link-local：生成ICMP重定向包或者作为下一跳地址的时候，使用link-local。对于所有的直连路由，IPv6 路由器必须有一个link-local下一跳地址，而且该路由器和下一跳路由有相同的子网前缀。

综上，BGP4+在传播IPv6路由可达信息的过程中，如果BGP宣告路由器和下一跳IPv6地址在一个子网，那么UPDATE包里的下一跳属性应包括global地址和link-local 地址，因为同一个子网内的通信需要依赖link-local地址，即在MP_REACH_NLRI属性中的Next Hop Field的网络地址先写16bytes的link-local地址，再写16bytes的global地址。

2.3 可扩展路由机制研究现状

全球路由表的快速膨胀，使得互联网域间路由的可扩展性成为急需解决的问题，很多学者进行深入研究，提出了不改变现有路由方法的机制和新型可扩展路由机制。

2.3.1 不改变现有路由方法的机制

不改变现有路由方法的机制目前有两大类，ID/Locator分离机制和局部FIB压缩机制。

2.3.1.1 ID/Locator分离机制

根据分离机制的不同，可以将其分为三种[6]：

- 基于网络的ID/Locator分离机制，核心网络和边缘网络的分离方案。将网络划分两个部分：边缘网络采用PI地址，用其作为主机标识符和内部的路由寻址；核心网络采用PA地址空间。核心网络和边缘网络之间的转换通过映射来达到。
- 基于主机的ID/Locator分离机制,在主机协议栈上面加上表示层，完成ID/Locator之间的映射，主机标识应用的使用者，网络地址标识网络位置，应用连接不用和IP 地址绑定，只需要绑定主机标识。
- 基于主机和网络的ID/Locator分离机制，边缘网络可以使用写入报文头的目的地址，也可以根据服务提供商的需求重写目的地址，更改后续报文。

2.3.1.2 FIB聚合压缩技术

FIB聚合压缩技术通过在少数核心路由器中构建多个虚拟前缀，将这些虚拟前缀只宣布给没有构建虚拟前缀的非核心路由器，在FIB表中有相同下一跳的路由前缀会被聚合，从而减少非核心路由器的规模。

2.3.2 新型可扩展路由机制

目前技术比较成熟有层次化路由、地理路由、紧凑路由等。

- 层次化路由指的是基于网络层次结构的路由，不同的层次使用不同的路由算法。传统的层次化路由在不同的层级采用相同的IP前缀，限制了层次结构的可扩展性。鉴于此，有学者提出了基于AS的解决方案HLP(Hybrid

Link-state And Path-Vector Procotol)方案[7],在自治系统之间采用基于自治系统的路径向量算法，在自治系统内采用链路状态协议，通过划分层次结构，自治系统间的路径向量算法可以隐藏自治系统内的路由更新信息，增加了路由可扩展性。

- 地理路由是指利用地理信息来帮助编址和路由，目前有三大类：纯地理信息路由方案、基于地理信息的覆盖网络和ISP信息辅助的基于地理信息的路由机制。纯地理路由的典型代表是根据经纬度对IP地址进行编址。基于地理信息的覆盖网络核心是在网络上在部署一个有地理位置信息的网络，可以增量部署，但该方案不仅没有解决路由的可扩展性，而且使得全局路由表更大。ISP信息辅助的基于地理信息的路由机制典型代表是GIRO[8](Geographically Informed Inter-Domain Routing)，在IP地址中用AS号表示ISP 信息，可以通过IP 地址，了解AS号和ISP。
- 紧凑路由[9]是一些学者从图论算法的角度提出来解决路由可扩展的问题。

2.4 现有IP网络编址方案

目前IP网络的编址方案主要分为三大类：IPv4标准编址方案，IPv6标准编址方案，IPv6新型编址方案。

2.4.1 IPv4标准编址方案

IPv4地址从无类编址转化为有类编址。因为B类消耗快，C类增长快，导致路由表的规模增长。有学者提出了通过分配连续的C类地址块，在路由表中对其进行聚合的方案，缓解路由表增长的速度。但是该方法存在一定的局限性。当组织结构是多宿主时，前缀会通过所有运营商向外宣告，导致前缀不能被聚合，降低了寻址效率。此外，如果一个组织结构改变运营商又没有进行重新编址，原来的前缀将不能被新的运营商聚合，也会降低寻址效率。

2.4.2 IPv6标准编址方案

目前IPv6标准编址方案主要有三大类：基于地域[10]，基于运营商的[11]、基于GSE[12](Global, Site, and End-system address elements) 的编址方案。

- 基于地域：将IPv6地址分为国家标识、城市标识、站点标识、站点内部标识，它的优点是可扩展性好，基于地域编址的路由信息可以被聚合。如果

本地改变运营商，不需要重编号。缺点是对拓扑要求比较高，同一地址运营商之间需要相互连接；而且不支持灵活的路由策略。

- 基于运营商的编址结构：将IPv6地址分为注册机构ID，运营商ID，订阅者ID，订阅者内部ID。目前的注册机构有IANA、APNIC、CNNIC。注册机构ID是分配地址的注册机构的标识，运行商和订阅者ID根据策略决定。
- GSE层次化编址：层次化的编址结构，由可路由前缀（位置标识）、站点内子网标识、端系统标识（全球唯一的身份标识）三大部分组成。该编址方案提出了ID/Locator分离的思想，端系统是身份标识，可路由前缀是位置标识，解决了边缘网络的多宿主问题，因为有身份标识，可以定位到多宿主。具体的编址方案是把IPv6 分成6个区域。FP：001 现互联网IPv6的地址前缀；TLA ID：长度为13，核心路由器对活跃的TLA ID有一条路由；RES： 8位保留位， TLA、NLA值变大后使用；NLA ID：24位，每个TLA有24 位NLA空间；SLA ID：长度为16，用来标识一个组织内的子网（层次/平面）；Interface ID：64位（TLA是平面结构，则TLA最多8192个，为了限制DFZ(Default-Free Zone)路由表长度）。

2.4.3 IPv6新型编址方案

目前除了IPv6标准编址方案，还有很多IPv6新型编址方案，主要有三大类：

- 本地地址编址方案[13]，仅用于区域内部，将地址分为Unique Prefix ，Global Id， Local Id， Interface Id四个部分，Global和Local 随机分配Id，不支持全球路由。
- 嵌入地理位置的编址方案，比如可以嵌入经纬度，一个前缀对应的是一个位置，那么/64是比/48更大的地址块。
- 嵌入ASN的编制方案，目前比较经典的方案有两种：
  - M.Levy提出[14]一种根据自治系统号分配/48的IPv6地址，具体对IPv6的128 位地址划分如下表2.1 所示： IANA分配/16地址段，

表 2.1 基于ASN的地址分配方案

IANA/16	32 bit ASN	80 bits for IPv6/48 space

比如现在使用的从IPv6转换为IPv4的地址段2001::/16。因为ASN是

由RIR分配的32bits数字，所以分配32位。剩余的80bits 分配给用户定义的ID。该方案支持多宿主，但IANA分配的16位前缀所占位数过多，而且因为目前IPv6并没有广泛使用，所以这部分完全可以不要，浪费了16位的地址。剩余的80 位用作身份标识较多，因为mac地址目前有48位已足够使用，身份标识建议64位，那么总共有32位的地址仍可继续利用。这样的地址分配没有考虑到层次化的地址分配，平面式的结构，不容易聚合。

– J.Li等学者为了解决域间路由提出[15]一种新型的寻址和路由结构LIMA(Less-Is-More Architecture)，其不同于身份位置分离的策略，而是通过使用与位置无关的名称和与位置有关的地址进行路由。该方案中要求边缘网络必须是PA地址，而且边缘网络的路由信息不能扩散到全局路由表中，通过结合网络层次化的结构，减少全局路由表的大小和变动。LIMA需要依赖于一种新型的IPv6编址方案，具体如下2.2： 该地址分配方案充分利用层次化来寻址，前64位属于位置

表 2.2　LIMA：基于ASN的地址分配方案

Provider ASN-32	Stub ASN-32	Stub-local IDA-64

标识，后64位属于身份标识地址。在一级路由表中只需要维护服务提供商自治系统的路由信息，没有任何关于边界自治系统的路由信息，在服务提供商的边界路由器中的分离数据表中维护服务商自治系统号和对应的一些边界自治系统号信息，这样可以极大程度地减少全局路由表的大小，相当于把全局路由表的主干留下，枝干放到主干网络的路由器中，但当边界自治系统改变服务商时需要重新划分地址，还有多宿主，移动性和流量工程等方面的问题。总之，网络层级划分太多且不合理。

目前IPv6的地址分配和IPv4的相同，都是RIR(区域互联网注册管理机构)根据个人请求的空间需求分配IPv6地址块。IPv6网络目前处于萌芽的状态，可以使用未分配的IPv6地址块测试编址方案。

2.5 小结

本章对域间路由协议BGP、BGP-4、BGP+做了简单的介绍，了解其数据包格式以及协议的状态机，为之后的实验打下基础。其次，介绍了可扩展路由机制的研究现状，了解到目前解决可扩展路由问题的大体思路主要有两种身份，位置分离和层次路由，给本文的研究提供了思路。最后，对现有的IP网络编址方案进行了归纳总结，了解现有编址方案的优缺点，从而对本文设计的新型IPv6编址方案进行更全面的评价和思考。

第 3 章　基于AS编址的互联网可扩展路由机制框架

3.1　引言

本章将详细描述基于AS的无类编址方案CABA，以及在此编址方案下采用的基于自治系统号的域间路由机制A-BGP。

3.2　基于AS的无类编址方案

表 3.1　CABA：基于ASN的地址分配方案

8	32	24	64
Res	ASN	subnet ID	interface ID
保留位	域间可路由前缀	域内可路由前缀	接口ID

基于AS的无类编址方案如图3.1所示，该编址方案前8位为保留位，该方案目前只使用了整个IPv6地址的1/256,为以后IPv6地址可能出现的其他情况做好准备。从网络层次结构分析，该编址方案为层次化的编址结构，将IPv6 地址空间分为域间的AS 域，域内的子网域和接口ID，以此来分离域内和域间路由，可有效减少域内路由振荡对域间路由的影响。从地址聚合的角度分析，域间使用AS进行聚合，域内使用IP 前缀进行聚合，可能目前域间AS 聚合难度较大，但是域内IP 前缀聚合发展已经较为成熟。

关于域间AS聚合的问题，我们了解到目前对自治系统号的定义为32位，目前全球有5万多个AS，而32bits可以分配4294967296个自治系统号，未来将分配更多的自治系统号，在分配的过程中可以考虑分配给互连的自治系统可以聚合的自治系统号。

域间路由采用基于自治系统号的BGP外部网关协议；域内路由24位仍可参考IPv4的内部网关协议，兼容现今网络。

IPv4中0.0.0.0为默认网络，在该IPv6中0::0可以作为默认网络。

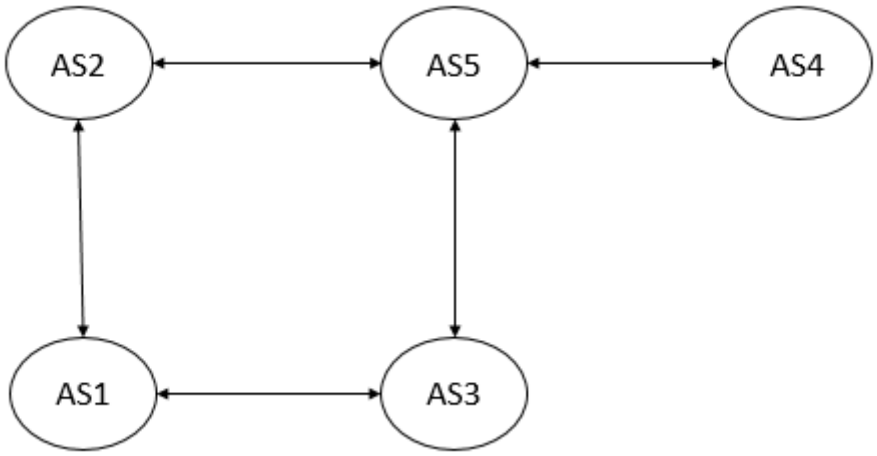

图 3.1　A-BGP路由示例-AS关系图

3.3　基于自治系统号的BGP协议

基于自治系统号的域间路由协议，将自治系统号作为域间路由的前缀。因为自治系统的变化是小概率事件，且域间路由与域内路由使用不同的地址段，相互独立，域内路由振荡不会影响域间路由，所以极大减少了路由通告的数量，压缩了域间全局路由表，提高了路由的可扩展性。原本一个自治系统可能宣告多个前缀，但在基于自治系统号的BGP协议中一个自治系统只需向外宣告一个前缀。同时，自治系统的多个前缀发生变动的概率远大于自治系统号改变的概率，所以在该基于自治系统号的BGP 协议中，路由振荡的频率会减少。

基于自治系统号的BGP协议域间路由的示例如下，自治系统之间的关系如图3.1：

- 自治系统4发布AS4/32的路由前缀，自治系统5收到自治系统4的通告，自治系统5的自治系统号可以和自治系统4的自治系统号相聚合，所以自治系统5将收到的路由信息进行聚合。
- 自治系统5发布聚合后的AS4/31和详细路由AS4/32，自治系统2和自治系统3收到来自自治系统5的通告。
- 自治系统2和3分别发布聚合后的AS4/31和详细路由AS4/32，自治系统1分别收到来自自治系统2和3的通告，共4条。

  如上图3.2，基于AS的路由聚合算法，选择前缀更短的路由，则路由信息3和4可作为可选路由。如果是单路径机制，选择下一跳ASN 较小的路

表 3.2　自治系统1收到的路由信息

路由信息编号	AS前缀	下一跳	AS路径
路由信息1	AS4/32	自治系统2	2，5，4
路由信息2	AS4/32	自治系统3	3，5，4
路由信息3	AS4/31	自治系统2	2，5
路由信息4	AS4/31	自治系统3	3，5

径，则路由信息3作为自治系统1的最佳路径。如果是多路径机制，最多选择两个非相交的路径，则路由信息3和4作为自治系统1的最佳路径。

基于自治系统号的BGP协议支持现有的通告IP前缀的路由协议，因此CABA编址方式可以在网络中进行增量部署。在自治系统部署了CABA编址的情况下，如果其邻居也部署了CABA编址，则基于自治系统号的BGP协议向邻居宣告基于ASN的前缀路由信息，如果其邻居没有部署CABA编址，则基于自治系统号的BGP协议向邻居宣告原有的IP前缀路由信息。

3.4　小结

基于AS编址的CABA将网络明确地划分成域内和域间两个层级，域内和域间路由的振荡相互独立，全局路由表中只有域间自治系统路由的信息，且每个自治系统仅向外公布一条嵌入自治系统号的前缀，可以合理推测基于CABA的A-BGP路由协议能有效减小全局路由表的大小和路由表的更新频率，提升路由的可扩展性。

第 4 章　基于CABA编址的FIB表压缩评估

4.1　引言

为了验证基于AS编址的互联网可扩展路由机制，需要通过仿真实验得到基于CABA编址的A-BGP路由下的BGP全局路由表，将其与现有的BGP全局路由表进行对比，验证基于CABA编址的A-BGP路由对网络路由可扩展性的影响。本章将详细介绍当前网络下FIB 表的生成过程以及在新型编址方案CABA下采用基于ASN的BGP协议时FIB表的生成情况，通过对比来评估新型编址环境下路由的可扩展性。

本文最初希望通过第5章SIMBGP仿真得到文中第3章新型域间路由方案下的BGP全局路由表。但是因为SIMBGP是单线程的事件驱动型仿真软件，在全网的拓扑结构下宣布一条前缀需要运行10min，假设内存满足运行需求，50000万个自治系统需要347天，所以本文决定直接对全网的FIB表进行处理，获取CABA编址下A-BGP路由后的全局FIB表。

表 4.1　随机部署：FIB压缩原始数据

路由器名称	当前网络FIB数目	部署20%	部署40%	部署60%	部署80%	部署100%
perth	8476	7551	4189	3360	3045	913
isc	573459	464036	353547	257451	170197	50034
linx	562937	455543	347739	253285	164897	49934
kixp	2329	2059	1255	1016	298	146
sydney	572895	462372	351338	255924	165075	50059
wide	558015	452474	345574	251150	165999	49882
eqix	561370	454054	347624	251743	164012	50002
saopaulo	575586	466723	357205	258361	167807	49953
nwax	558075	451568	346512	250927	162610	49950
telxatl	558182	451432	346063	250583	164080	49968
jinx	536466	432900	334158	241083	157437	49685
soxrs	38037	32766	26930	22203	17930	11880
sg	557282	450476	345647	249168	162759	49939

4.2 数据集介绍

实验的数据集来自Oregon大学的Route Views[2]工程搜集的BGP数据。进入FTP下载页面，选择13台路由器(路由器的名称为: perth, isc, linx, sydney, wide, eqix, saopaulo, nwax, telxatl, jinx, soxrs, sg, kixp)，获取其2015年4月3日0点的全局路由表。

4.3 当前网络下的FIB表

1. 从Oregon大学的Route Views[2]下载13台路由器的RIB数据，下载的原始数据是二进制格式，需要工具来解析。
2. 使用libbgpdump[16]解析二进制文件获得可读文本文件，通过$./bgpdump -M -Ooutputfileinputfile$命令获取解析后的文件，使用M参数简化路由表项(表项举例：TABLE_DUMP_V2—03/31/15 02:00:00—A—206.126.236.120—41095—0.0.0.0/0—41095 3356—IGP)。

表 4.2 根据宣告前缀数目部署：FIB压缩原始数据

路由器名称	当前网络FIB数目	部署20%	部署40%	部署60%	部署80%	部署100%
perth	8476	1977	1143	935	916	913
isc	573459	107546	65913	51985	50157	50034
linx	562937	106499	65513	5177	49984	49934
kixp	2329	306	208	181	160	146
sydney	572895	107023	65748	51974	50126	50059
wide	558015	105825	65335	51724	49955	49882
eqix	561370	106529	65531	51828	50049	50002
saopaulo	575586	575586	65382	51689	49953	49953
nwax	558075	106092	65372	51745	49979	49950
telxatl	558182	106262	65465	51788	50013	49968
jinx	536466	99756	62580	51125	49739	49685
soxrs	38037	15384	12858	12003	11881	11880
sg	557282	106131	65447	51749	49971	49939

3. 相同前缀在RIB表中可能有多条路由信息，但不考虑多路由机制，FIB表中存储的是该前缀的最优路由信息。在实际的网络环境中，BGP路由优先原则[17]有13条，因为在该实验只涉及域间路由，而且从RIB表项中得到的

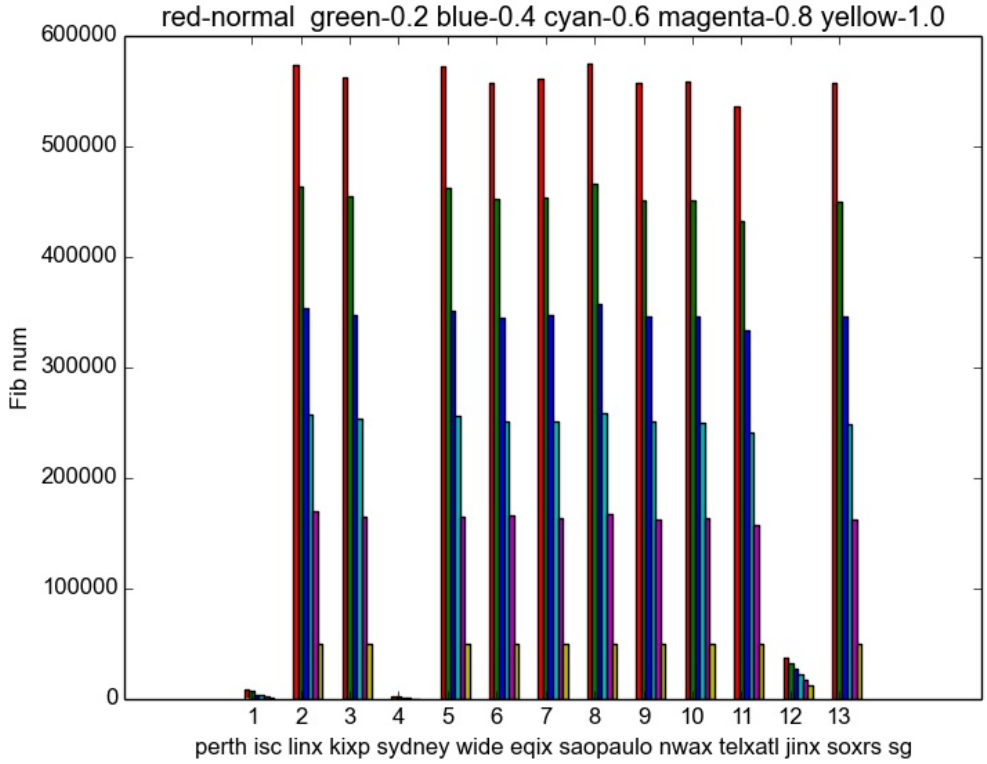

图 4.1　随机部署：FIB表项数目

路由信息有限，所以在RIB表转换FIB表时，遵循两个原则：优先路径的长度最短的路由；当路径长度相同的时候，优先下一跳自治系统号最小的路由信息。提取RIB表中所有前缀的最优路径得到FIB表。

4.4　基于CABA编址的FIB表的设计与生成

本章节将详细介绍基于AS新型编址的A-BGP协议下FIB表设计与生成的关键技术和增量部署的过程。

4.4.1　关键技术

将现有网络中的FIB表转换成基于CABA编址A-BGP协议下FIB表的主要过程如下所示：

1. 如果公布前缀的源自治系统部署了CABA编址，则将前缀替换成嵌入源自治系统的CABA 格式的IPv6/40前缀，前8位为保留位，选择IANA 未使用

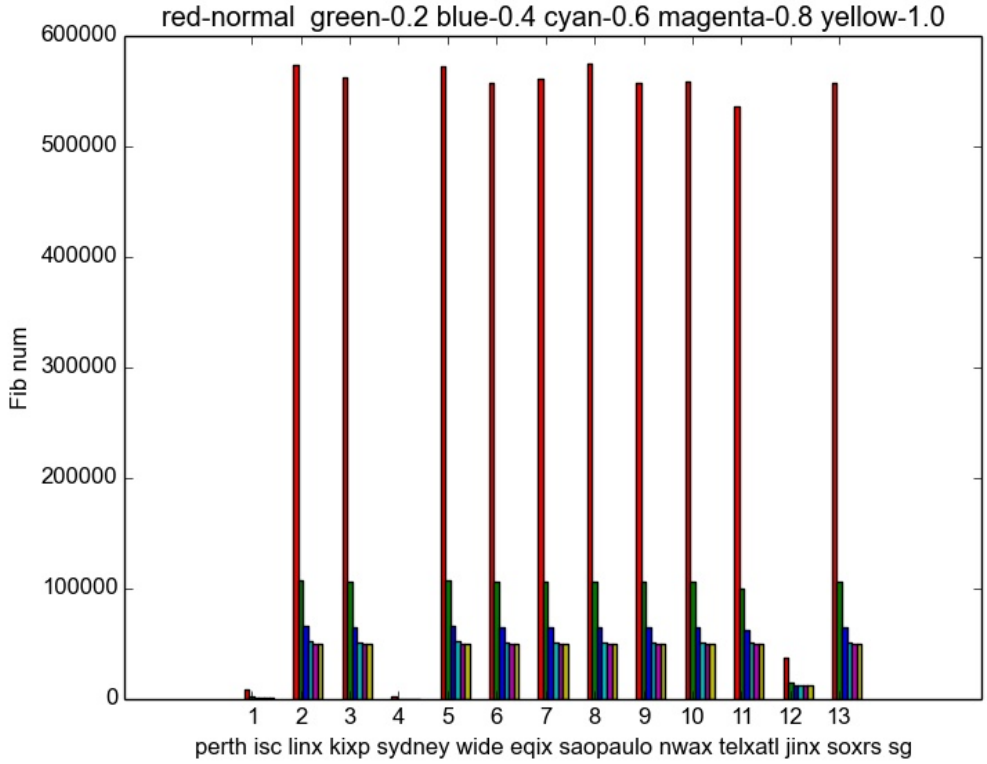

图 4.2　根据宣告前缀部署：FIB表项数目

的地址段10；之后32 位为源自治系统的ASN。

2. 因为在现有网络中的FIB表中，有些前缀是由聚合的AS集合宣告的，所以不能确定宣告这些前缀的自治系统号。经过对这13台路由器的FIB表中聚合AS 集合数目大于1 的表项进行统计，每个路由器约有50万条数据的FIB表中约有50条这样的表项，不足万分之一的数据。为了保证用CABA编址格式下的前缀替换原有前缀的准确性，我们舍弃这不足万分之一的聚合AS集合数目大于1的表项。
3. 替换前缀结束后，提取表中所有前缀的最优路径得到FIB表。评价最优路径依旧遵循两个原则：优先路径的长度最短的路由；当路径长度相同的时候，优先下一跳自治系统号最小的路由信息。

4.4.2　增量部署

现今互联网架构和规模已经很大，协议和策略已经很完备，并且渗透到经济社会生活的方方面面，将基于AS的新型编址CABA完全部署到互联网上是不

现实的，所以需要进行增量部署。增量部署的最小单元是自治系统，所以首先我们获取现今互联网所有自治系统的编号，其次根据不同增量部署的方案选择需要部署的自治系统编号。

获取现今互联网所有自治系统的编号：

- 我们得到13台路由器CABA编址下的FIB表，统计每张表的表项，显示该全局路由表中自治系统的个数。

表 4.3　不同路由器存储全局路由表中向外宣布前缀的源AS的数目

路由器名称	AS数目
perth	913
isc	50034
linx	49934
kixp	146
sydney	50059
wide	49882
eqix	50002
saopaulo	49953
nwax	49950
telxatl	49968
jinx	49685
soxrs	11880
sg	49939

从上表4.3可以看出perth、soxrs、kixp这三台路由器可能是边界路由器，所以我在剩余的10台路由器中随机一台路由器的FIB 表，获取增量部署时所有的ASN，随机路由器的结果为saopaulo路由器。

- 根据saopaulo路由器当前网络下的FIB表的数据，提取所有的ASN以及其自治系统宣布的前缀数目。

三种增量部署方案：

- 随机部署：随机选取ASN的20%,40%,60%,80%,100%作为部署CABA编址的自治系统。
- 根据宣告前缀数目部署：将ASN 根据宣告前缀的数目由大到小排序，选取排序后的20%,40%,60%,80%,100%。理论上当增量部署的程度相同时，部署宣告前缀越多的自治系统，路由表的压缩比例越多。

- 部署Tier1层级的自治系统，Tier1[18]为互联网层级结构的最顶层。

4.5 FIB表压缩情况的评估

本章节显示FIB表的压缩结果以及对结果进行分析。

表 4.4 部署Tier1自治系统：FIB压缩原始数据

路由器名称	当前网络FIB表项数目	部署Tier1后表项数目	部署Tier1后减少表项数目
perth	8476	8476	0
isc	573459	562564	10895
linx	562937	551601	11336
kixp	2329	2329	0
sydney	572895	562573	10322
wide	558015	547127	10888
eqix	561370	551601	9769
saopaulo	575586	564435	11151
nwax	558075	548597	9478
telxatl	558182	548524	9658
jinx	536466	527490	8976
soxrs	38037	37643	394
sg	557282	547893	9389

4.5.1 压缩结果

在增量部署的情况下，得到FIB表的原始数据：

本节我们首先看到表4.1中随机部署下路由表的压缩情况，和表4.2中根据宣告前缀数目部署下路由表的压缩情况。从数据能很明显看到在随机部署中，FIB表项随着部署比例的线性增加而线性减少，如图4.1，而根据宣告前缀数目部署中FIB表项在部署比例为20%时，大幅度减少，之后缓慢减少，如图4.2。说明一个自治系统向外宣告前缀越多，该自治系统采用CABA编址后，对FIB表的压缩贡献越大。由此我们可以设想，如果将Tier1的自治系统全部部署CABA编址，对FIB表的压缩程度也应该会很大，之后通过实验获得了表4.4的数据。

Tier1的自治系统号来源于CAIDA官网上的数据[19]，下载20150201时间的AS关系数据，在其关系数据里面有Tier1 的ASNs见表4.5，部署Tier1的17个自

表 4.5　20150201时间数据 Tier1自治系统号

174	209	286	701	1239	1299
2828	2914	3257	3320	3356	5511
6453	6461	6762	7018	12956	-

表 4.6　随机部署：FIB压缩掉数据占源数据的比例

路由器名称	当前网络FIB数目	部署20%	部署40%	部署60%	部署80%	部署100%
perth	8476	0.10913	0.50578	0.60359	0.64075	0.89228
isc	573459	0.19081	0.38348	0.55106	0.70321	0.91275
linx	562937	0.19077	0.38228	0.55007	0.70708	0.91130
kixp	2329	0.11593	0.46114	0.56376	0.87205	0.93731
sydney	572895	0.19292	0.38673	0.55328	0.71186	0.91262
wide	558015	0.18914	0.38071	0.54992	0.70252	0.91061
eqix	561370	0.19117	0.38076	0.55156	0.70784	0.91093
saopaulo	575586	0.18913	0.37941	0.55113	0.70846	0.91321
nwax	558075	0.19085	0.37909	0.55037	0.70862	0.91050
telxatl	558182	0.19125	0.38002	0.55107	0.70605	0.91048
jinx	536466	0.19305	0.37711	0.55061	0.70653	0.90738
soxrs	38037	0.13858	0.29201	0.41628	0.52862	0.68767
sg	557282	0.19166	0.37976	0.55289	0.70794	0.91039

治系统，核心网络FIB 表的表项可以减少约1万条，约占当前网络FIB表50万表项的2%。

4.5.2　结果分析

本节对三中部署方案的FIB表的压缩率进行计算绘图说明其变化趋势。

由图4.6可以看出，随机部署环境下，FIB压缩掉数据占源数据的比例(FIB表减少表项FIB表原有表项)随着部署比例的线性增加而线性增加，明显的趋势图如4.3。

由图4.7可以看出，根据宣告前缀数目部署环境下，FIB压缩掉数据占源数据的比例(FIB表减少表项FIB表原有表项)在部署比例为20%的情况下上升到很大的值，之后随着部署比例的线性增加而缓慢增加，明显的趋势图如4.4。

属于Tier1的自治系统有17个，部署之后，对于现在的约50万的FIB表数据，可以压缩1万的表项，占到了压缩比例的2%，如图4.8。17 个自治系统相

表 4.7　根据宣告前缀数目部署：FIB压缩掉数据占源数据的比例

路由器名称	当前网络FIB数目	部署20%	部署40%	部署60%	部署80%	部署100%
perth	8476	0.76675	0.86515	0.88969	0.89193	0.89228
isc	573459	0.81246	0.88506	0.90935	0.91254	0.91275
linx	562937	0.81082	0.88362	0.90803	0.91121	0.91130
kixp	2329	0.86861	0.91069	0.92228	0.93130	0.93731
sydney	572895	0.81319	0.88523	0.90928	0.91250	0.91262
wide	558015	0.81035	0.88292	0.90731	0.91048	0.91061
eqix	561370	0.81023	0.88327	0.90768	0.91084	0.91093
saopaulo	575586	0.81466	0.88641	0.91020	0.91321	0.91321
nwax	558075	0.80990	0.88286	0.90728	0.91044	0.91050
telxatl	558182	0.80963	0.88272	0.90722	0.91040	0.91048
jinx	536466	0.81405	0.88335	0.90470	0.90728	0.90738
soxrs	38037	0.59555	0.66196	0.68444	0.68765	0.68767
sg	557282	0.80956	0.88256	0.90714	0.91033	0.91039

对于目前网络50000个自治系统只占0.034%,是2%的158，也就是说平均部署一个Tier1的自治系统相当于部署58个普通层级的自治系统。根据理论推测，部署Tier1层级的自治系统可以大幅度的压缩FIB的原因，主要是Tier1层级的自治系统向外公布的前缀很多，见表4.9。

4.6　小结

总体来看，在全部部署的情况下，FIB表的表项数目是现网络环境下FIB表表项的1/10,增量部署的情况下，FIB表表项的压缩情况与部署自治系统向外宣布的前缀数目相关，部署自治系统向外宣告的前缀数目越多，FIB表表项的压缩情况越好。所以，我们在增量部署CABA编址时，可以考虑向外公布前缀较多的自治系统，这样对全局FIB 的压缩作用会比较明显。

表 4.8　部署Tier1自治系统：FIB压缩掉数据占源数据的比例

路由器名称	当前网络FIB表项数目	FIB压缩掉数据占源数据的比例	压缩掉数据条数
perth	8476	0.0	0
isc	573459	0.01900	10895
linx	562937	0.01978	11336
kixp	2329	0.0	0
sydney	572895	0.01802	10322
wide	558015	0.01951	10888
eqix	561370	0.01740	9769
saopaulo	575586	0.01937	11151
nwax	558075	0.01698	9478
telxatl	558182	0.01730	9658
jinx	536466	0.01673	8976
soxrs	38037	0.01036	394
sg	557282	0.01685	9389

表 4.9　20150201 saopaulo路由表中Tier1自治系统号公布的前缀数目

自治系统号	公布的前缀	自治系统号	公布的前缀
174	2755	209	1029
286	61	701	1402
1239	443	1299	209
2828	211	2914	602
3257	421	3320	559
3356	1291	5511	135
6453	168	6461	590
6762	78	7018	1161
12956	53	-	-

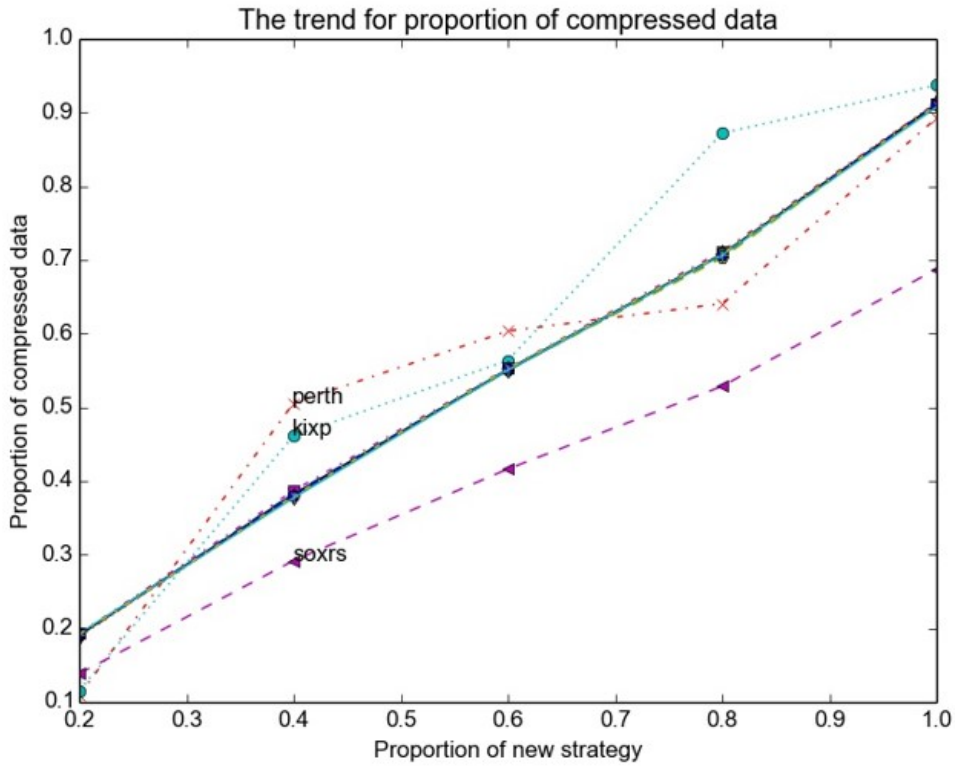

图 4.3　随机部署：FIB压缩掉数据占源数据的比例变化折线图

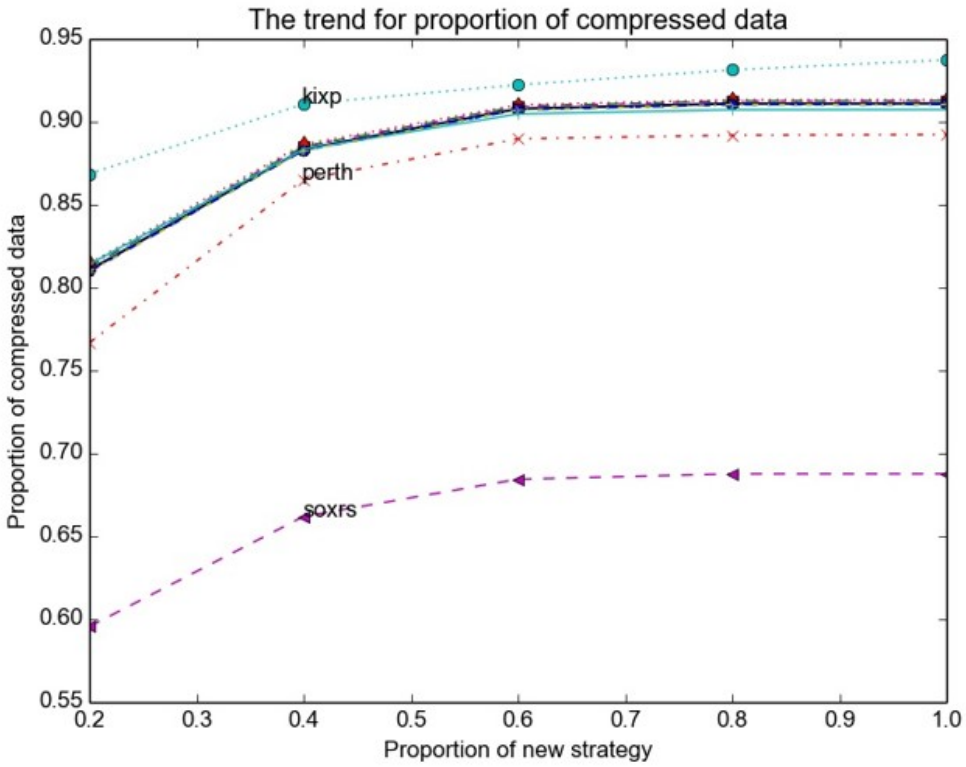

图 4.4　根据宣告前缀部署：FIB压缩掉数据占源数据的比例变化折线图

第 5 章　基于AS编址的A-BGP域间路由的仿真评价

5.1　引言

本章对基于AS编址的A-BGP路由机制进行仿真，验证其对路由可扩展性的作用。主要包括以下内容：仿真实验平台SIMBGP的介绍，仿真实验的过程以及结果分析。选取SIMBGP作为我们仿真实验的工具，分别统计和分析自治系统宣布CABA编址下的一条前缀和宣布多条现今网络宣布出去的前缀到互联网中，在网络路由信息稳定之后，互联网宣布的UPDATE 条数和收敛时间。

5.2　仿真实验平台介绍

5.2.1　平台特点及运行方法

- 平台特点：SIMBGP是一个超轻量级的BGP协议仿真器，忽略了所有底层的协议；该平台使用Python 语言实现，只有一个文件，运行简单；该平台可以在互联网的全网拓扑结构上进行仿真，满足实验要求；SIMBGP是一个事件驱动型的模拟器，使用者可以很容易的了解代码的结构以及在运行的过程中容易监控程序运行的程度，但是这也就意味着该仿真平台是单进程，向外宣布前缀只能一条一条向外宣布，不能同时执行多个事件，影响了批量模拟ANNOUNCE事件的效率。
- 运行方法：在linux系统的terminal窗口运行python simbgp.py config.txt，运行需要输入BGP配置文件，表示网络的拓扑结构等信息。

5.2.2　配置文件

仿真实验需要在全网的拓扑结构下进行，所以要生成表示全网拓扑结构关系的配置文件，大致可以分为两步：第一步：获取自治系统间的拓扑结构，自治系统的关系来自于CAIDA官网上的数据[19]。下载获得的数据部分如下：

1267|60280|-1

1267|20756|0

解读自治系统关系的文件需要参考文件[20],该格式的含义如下：

<provider-as> | <customer-as> |-1

<peer-as> | <peer-as> |0

第二步：为了简化拓扑结构，以自治系统为最小单位，每台路由器为一个自治系统。编写程序遵循SIMBGP的配置文件格式生成配置文件。

5.3 仿真实验流程设计

SIMBGP仿真实验的流程如下：

- 从CAIDA官网获得全网的拓扑结构，编写程序生成SIMBGP的输入配置文件，参考5.2.2。
- 在本文4.3章节获得了saopaulo路由器的当前网络下FIB表数据，统计每个自治系统公布出去的前缀。
- 在全网约50000自治系统中随机选取20个自治系统号。
- 自治系统A宣布嵌入自治系统号A的IPv6前缀，获取UPDATE数目和收敛时间。
- 自治系统A宣布saopaulo路由器的当前网络下FIB表数据中自治系统A公布出去的所有前缀，统计UPDATE数目和收敛时间。

5.4 仿真实验结果分析

SIMBGP仿真实验的结果如表5.1所示,对其结果进行分析：在仿真实验中，通过一台路由器代表一个自治系统，在这个自治系统中前缀只是一个标签，所以路由器向外宣告一条CABA编址下的前缀和向外宣告一条现今网络宣告出去的IPv4IPv6路由所宣布的UPDATE数目相同，所需要的收敛时间也是相同的。平均减少的UPDATE条数约有45万$*300= 1.35 * 10^8$，约减少300倍,收敛时间大部分约为100s。

可以观测到，CABA编址下宣告前缀与现今网络下宣告前缀相比，UPDATE数目差异主要和自治系统在现今网络下向外宣布的前缀数目相关。由图5.1可得，AS对应的前缀数目聚集在1-500之间，在0-200最为密集。

表 5.1　SIMBGP仿真实验结果

随机ASN	现今网络中宣告Prefix数目	UPDATE条数	收敛时间	UPDATE减少倍数
393393	17	445622	134.17	16
21021	32	436523	90.02	32
45773	53	514378	118.14	52
29684	73	490768	117.37	72
9535	114	506102	114.34	113
11003	121	475602	114.21	120
18978	144	434848	109.3	143
38370	174	460682	100.59	173
12418	189	449055	109.16	188
20299	216	512391	134.02	215
23674	260	496074	118.15	259
702	284	410536	106.48	283
3491	337	437333	90.87	336
812	403	422310	88.1	402
17762	453	460750	107.77	452
18207	595	468700	115.25	594
11139	619	419390	89.65	618
15003	839	440663	112.17	838
13188	1102	448425	111.64	1101
22394	1486	425741	110.53	1485
平均值	375.55	457795	109.60	375

由图5.2可以看出有1万多个自治系统只向外宣告了一个前缀，有4万多个前缀系统向外宣告了小于10个前缀，也就是说对于大部分的自治系统而言，部署CABA地址分配方案，向外宣布嵌有ASN的直到网络路由收敛的UPDATE的数目和没有部署向外宣布现今网络中的前缀直到网络路由收敛的UPDATE数目相比，并没有很大程度的减少。

由图5.3可以看出一个自治系统平均向外宣布约11条前缀，因为自治系统对应的前缀数目分布差异较大，所以平均值几乎没有意义。

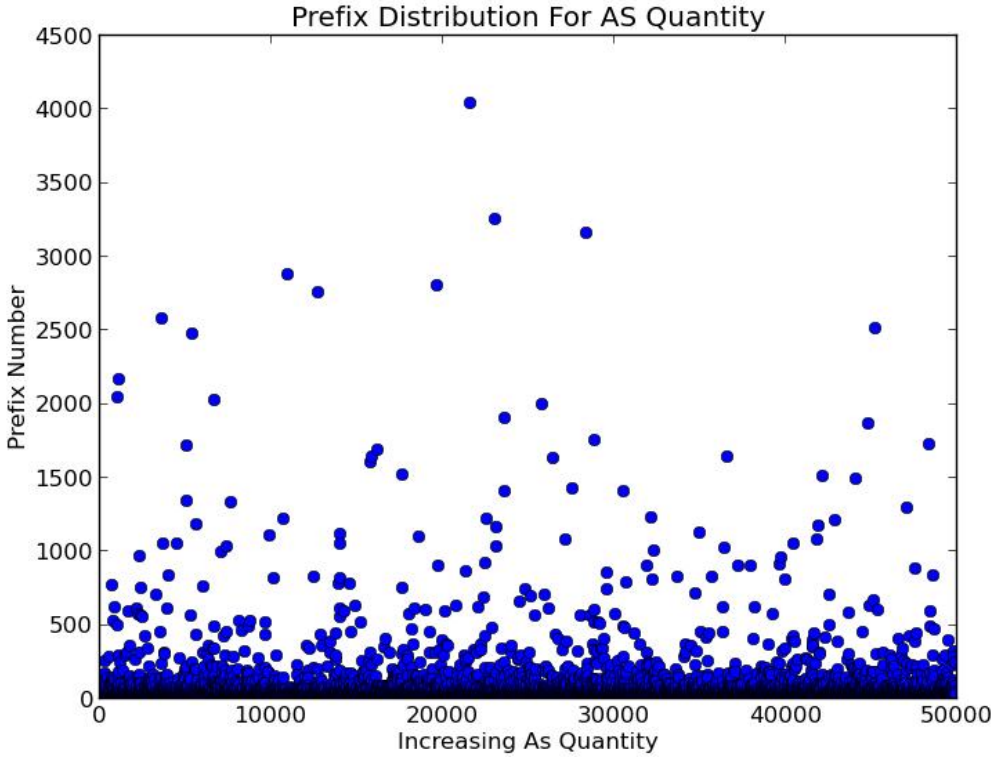

图 5.1　ASN对应宣告前缀数目的散点图

5.5　小结

综上所述，部署CABA方案的自治系统只需要向外宣布一条前缀，与现今网络中一个自治系统需要向外宣告n条前缀相比，UPDATE减少$(n-1)/n$倍。因为自治系统对应前缀数目的分布差异较大，所以增量部署CABA 方案在向外宣布较多前缀的自治系统上，更有利于减少UPDATE包数。

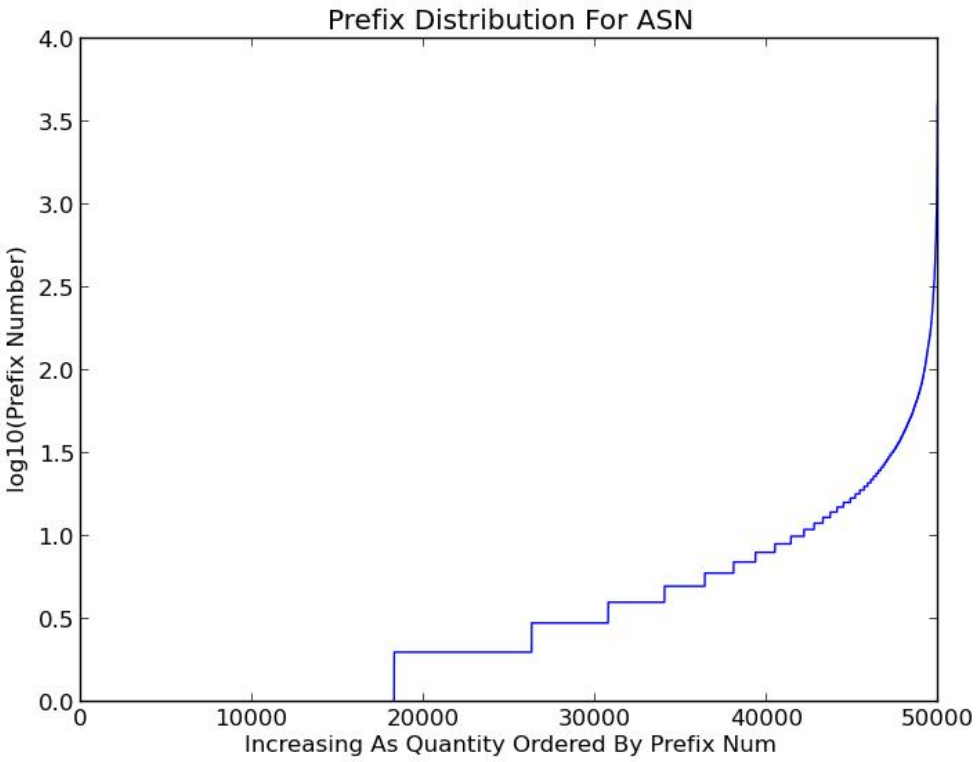

图 5.2　宣告前缀数目递增的ASN对应宣告前缀数目的LOG值

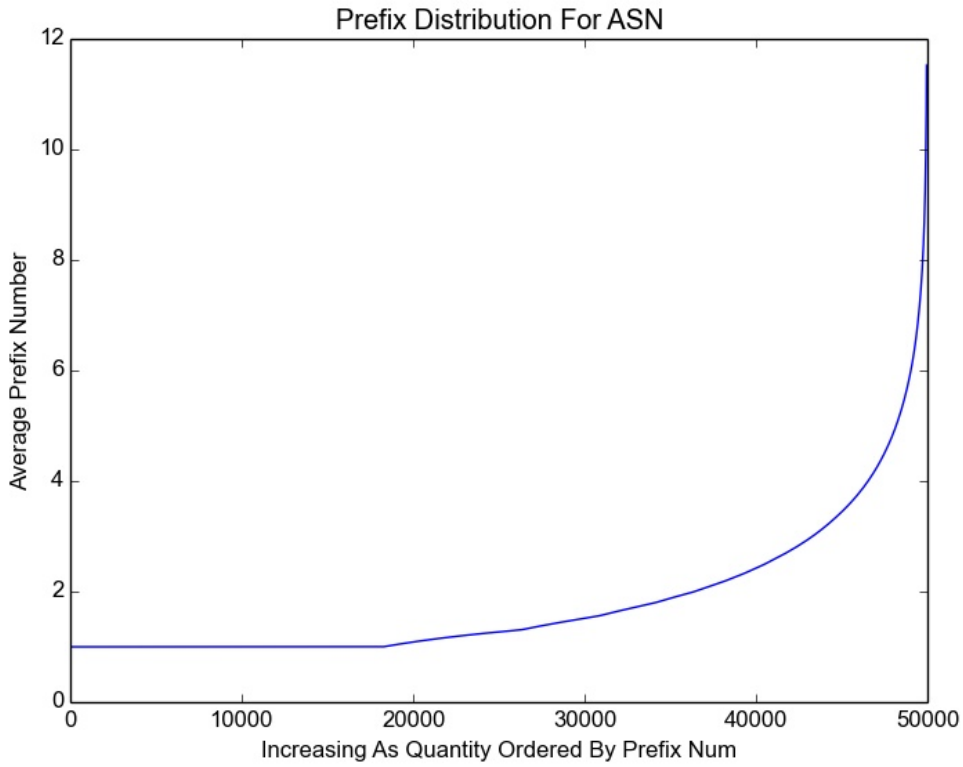

图 5.3　宣告前缀数目递增的ASN对应的平均宣告前缀数目的CDF图

第 6 章　基于AS编址的A-BGP域间路由的实现与测试

6.1　引言

本章节描述在软件路由器Quagga上实现A-BGP机制并采取Docker进行试验床测试的细节，通过比较现网络环境下和A-BGP协议下BGP 路由表项的大小，分析基于AS 编址的CABA 编址方案对路由可扩展性的作用。

6.2　试验环境

本章在Quagga软件路由器实现A-BGP路由机制，使用Docker平台部署多台软件路由器实现对A-BGP域间路由机制的测试。

6.2.1　Quagga软件路由器的核心思想和工作原理

Quagga[21]软件路由器的前身是Zebra，支持IPv6。不同网络协议运行不同的进程，相互独立；Zebra为核心进程，管理全局路由表；网络协议进程

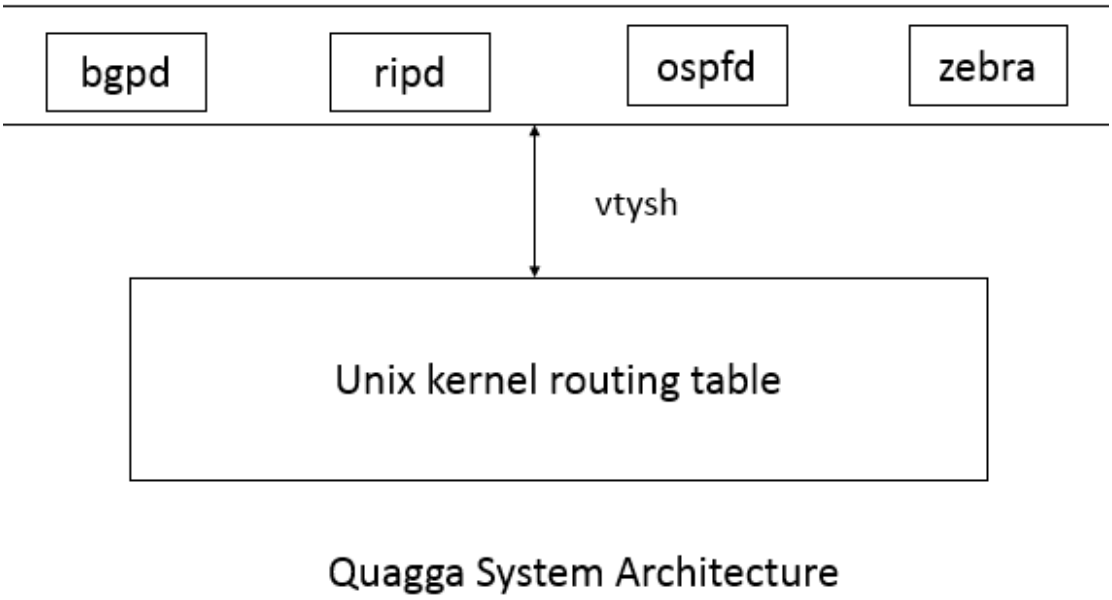

图 6.1　Quagga系统结构

通过vtysh进程和内核路由信息表进行交互；Bgpd运行BGP域间路由协议，见表6.1。

6.2.2 Docker系统的核心思想和工作原理

Docker[22]是一个实现了超轻量级的操作系统虚拟化平台。在Docker中，开发人员使用镜像来构建开发容器，通过将一台配置好环境的Docker容器保存成镜像进行复用。Docker中容器除了运行其中应用外，不消耗额外资源，大量节约了开发环境的部署、实际环境的测试等时间。Docker中有三个基本概念：镜像、容器和仓库，镜像是操作系统的安装包，仓库来管理存储所有的镜像，在容器中运行镜像就实现了操作系统的功能。

Docker 有很多优点，秒级启动容器，在一台实体机可运行几千台容器，每台容器装上镜像就如同一个虚拟机；将配好环境的容器保存成镜像进行复用，给一台容器配置软件路由器的环境，其余容器复用该环境，简单方便；运行中容器为一个进程，相互独立，可在实体机终端操控配置；Docker数据卷可以共享主机、镜像的文件夹，为实验数据的输入和提取提供方便。总之，Docker实验平台可以通过在主机上运行脚本部署实验环境，进行实验，实现在网络中部署成百上千个节点，满足实验的需求，也极大地提高了实验的可扩展性。

6.3 基于AS编址的A-BGP域间路由机制的实现方案

首先，熟悉Quagga软件，运行zebra和bgpd进程，打开vtysh进程。其次，通过修改Bgpd进程的配置文件，在Quagga软件路由器实现基于AS编址的A-BGP域间路由机制。每一个自治系统内仅有一台路由器，需要定义该路由器的自治系统号，路由器Id，邻居的IPv6地址和邻居的自治系统号，向外宣告嵌有自治系统号的基于AS的CABA 编址格式的IPv6 前缀。以自治系统174 为例，该自治系统有一台路由器，该路由器的配置文件如下：

```
hostname bgpd
password zebra
log stdout
line vty
router bgp 174
```

```
bgp router-id 0.0.1.74
neighbor 11:0::d1 remote-as 209
neighbor 11:1::11e remote-as 286
neighbor 11:2::2bd remote-as 701
neighbor 11:3::4d7 remote-as 1239
neighbor 11:4::513 remote-as 1299
neighbor 11:5::b0c remote-as 2828
neighbor 11:6::b62 remote-as 2914
neighbor 11:7::cb9 remote-as 3257
neighbor 11:8::cf8 remote-as 3320
neighbor 11:9::d1c remote-as 3356
neighbor 11:a::1154 remote-as 4436
neighbor 11:b::1587 remote-as 5511
neighbor 11:c::1935 remote-as 6453
neighbor 11:d::193d remote-as 6461
neighbor 11:e::1a6a remote-as 6762
neighbor 11:f::1b6a remote-as 7018
neighbor 11:10::329c remote-as 12956
address-family ipv6
network 1000:0000:ae00::64
neighbor 11:0::d1 activate
neighbor 11:1::11e activate
neighbor 11:2::2bd activate
neighbor 11:3::4d7 activate
neighbor 11:4::513 activate
neighbor 11:5::b0c activate
neighbor 11:6::b62 activate
neighbor 11:7::cb9 activate
neighbor 11:8::cf8 activate
neighbor 11:9::d1c activate
neighbor 11:a::1154 activate
```

neighbor 11:b::1587 activate
neighbor 11:c::1935 activate
neighbor 11:d::193d activate
neighbor 11:e::1a6a activate
neighbor 11:f::1b6a activate
neighbor 11:10::329c activate
exit-address-family

6.4 实验拓扑

实验的拓扑结构为现今网络Tier1层级[18]的拓扑结构，Tier1自治系统关系来源于CAIDA官网上20150501数据[19]，比表4.5显示的20150201时刻的tier1层级的自治系统增加了ASN为4436的自治系统。Tier1层级共有18个自治系统，为了简化实验，每个自治系统内仅有一个路由器，则共有18台路由器。18台路由器之间的peer-link 连接共有149 条，如果18 台路由器全连接有153 条，该拓扑结构中除自治系统(701, 12956), (2914, 12956), (3257, 12956), (4436, 12956)这四对自治系统内的路由器未连接，其余自治系统内的路由器两两连接构成Tier1的拓扑图，如图6.2。

6.5 实验网络环境配置

在Docker平台部署多个网络节点后，需要进行实验网络环境配置，步骤如下：

- 在本机上建立网桥
- Peer容器间的网络接口均连接到网桥
- 通过配置网桥的IPv6接口，建立Peer间连接

以图6.3拓扑，搭建如图6.4的网桥。以路由器1和路由器3之间的连接为例，路由器1的接口连接到网桥link-1-1，路由器3的接口连接到网桥link-1-3，通过将link-1-1和link1-3配置成一个网段的接口，连接路由器1和路由器3.

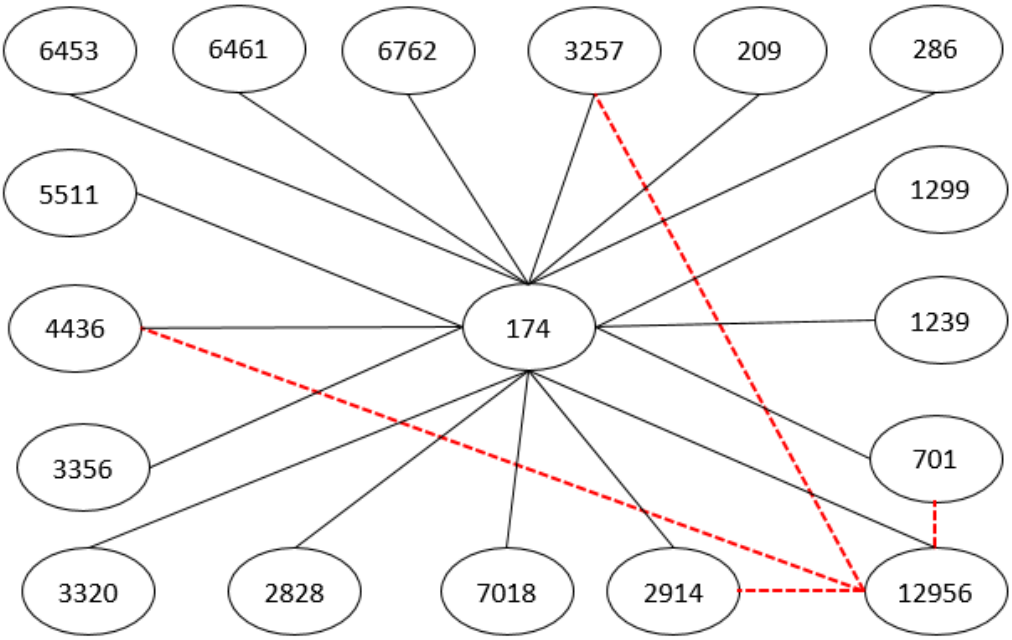

图 6.2　实验拓扑图：除虚线外其余全连接

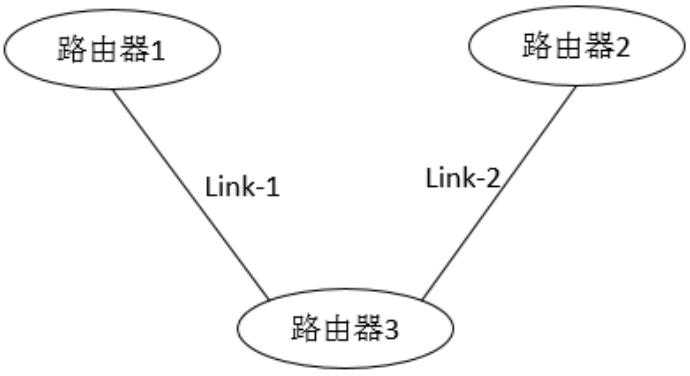

图 6.3　网络环境配置举例：拓扑图

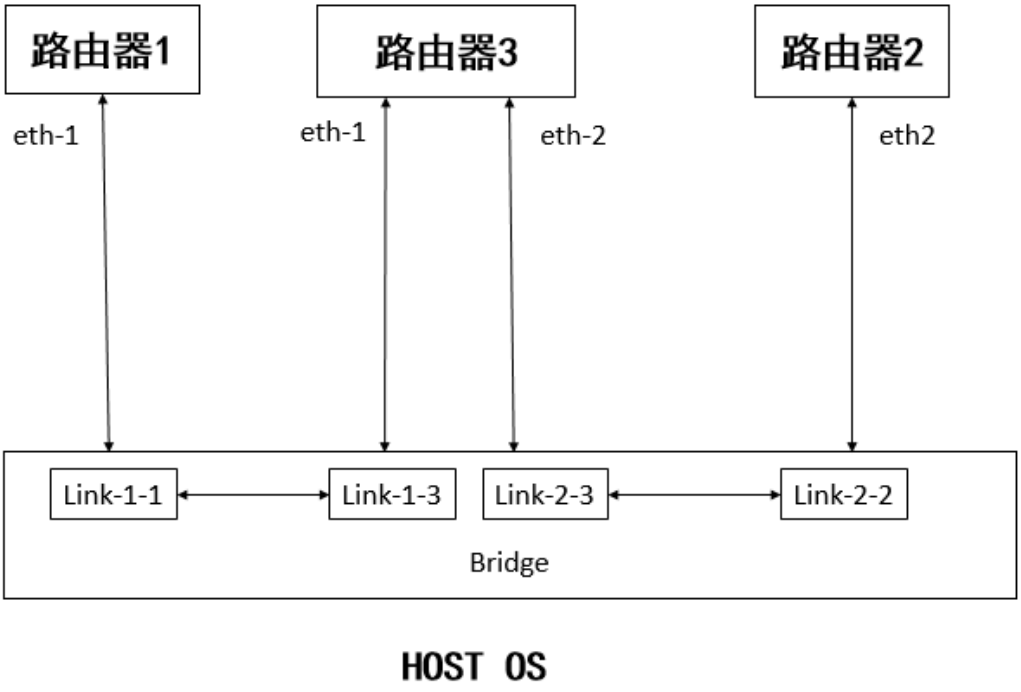

图 6.4　网络环境配置举例：网桥结构图

6.6　实验流程

基于AS编址的A-BGP域间路由机制实现与测试的实验流程如下：

1. 确定拓扑结构：Tier1层级自治系统构成的拓扑结构，Tier1层级有18个自治系统，每个自治系统内有一台路由器，所以共有18台路由器。
2. 配置运行环境：安装Docker软件，首先运行Ubuntu14.04系统的镜像于容器中，在容器中配置Quagga环境，保存该容器，并复制18台该容器，用于网络拓扑的18个节点。
3. 配置网络环境：在本机上搭建网桥，将容器的接口与网桥的接口相连接，不同的连接配置不同局域网，每一个Peer连接关系都需要在网桥上配置两个接口。
4. 生成Quagga配置文件：确定每台路由器对应的ASN，Router-Id，Neighbor-Info，Network-Prefix。
5. 在Docker试验床上运行所有部署Quagga软件路由器的zebra和bgpd进程，统计IPv6 Bgp Route的数目。

通过编写程序生成脚本来创建容器，启动容器，创建网桥，配置接口，生成Quagga路由器BGP配置文件，启动Quagga 中的进程。

本章实验首先尝试在两台物理主机上安装软件路由器，通过配置网络接口，向外宣告路由，之后查看其BGP路由表项。因为互联网的规模很大，所以小型拓扑不能说明问题，需要试验床来进行复杂拓扑的实验。多台物理主机不仅浪费资源，而且配置麻烦，可扩展性极差，所以选择可以实现超轻量级的虚拟化操作系统的Docker作为试验床。在Docker中运行多台Quagga软件路由器，搭建实验环境。

6.7 实验对比设计

单纯观察部署CABA地址方案的互联网拓扑结构生成的BGP路由表项数目，并不能说明问题。所以本章节设计了对比实验，来观察BGP路由项数目的变化趋势和变化程度。对比实验设计如下：

1. 给路由器配置基于AS的新型IPv6地址，向外宣布嵌入ASN的IPv6前缀。
2. 给路由器配置IPv4地址，向外宣布CAIDA官网统计出来ASN对应的所有IPv4前缀。

6.8 实验分析

在Docker中部署本章第6.4小节拓扑的网络环境，不同路由器属于不同自治系统，每个自治系统向外公布嵌有自治系统号的路由，最后18台路由器得到BGP路由表。自治系统174 内的路由器的BGP路由表见图6.5,因为该拓扑结构中有18台路由器，每台路由器向外公布一条前缀，自治系统174内的路由器的BGP表项有17条，我查看其他路由器的BGP路由表均为17条路由。

由于在自治系统向外宣告IPv4前缀的对比实验中，因为不能确定自治系统内的路由器向外宣告IPv4路由的IPv4网络接口，所以不能进行此对比实验。但是根据第 4章的方法，对Tier1层级的自治系统进行增量部署，得到其可以减少路由表项约1万条6.1。即Tier1自治系统构成的拓扑结构在A-BGP路由机制下的BGP 路由表项为17，远小于现网络中FIB表中与Tier1自治系统相关的约1万条表项。

表 6.1　部署Tier1-18个自治系统：FIB表项数目

路由器名称	FIB表项数目	部署后FIB表项数目	减少表项
perth	8476	8476	0
sc	573459	562072	11387
linx	561370	551028	10342
kixp	562937	551229	11708
sydney	2329	2329	0
wide	572895	562081	10814
eqix	558015	546635	11380
saopaulo	575586	563863	11723
nwax	558075	547953	10122
telxatl	558182	547548	10634
jinx	536466	527016	9450
soxrs	38037	37626	411
sg	557282	547401	9881

```
1  B>  1000:0:d100::/64 [20/0] via fe80::d424:76ff:fe77:374, eth0-0, 00:16:26
2  B>  1000:1:1e00::/64 [20/0] via fe80::68c0:ccff:fe34:bbec, eth0-1, 00:16:24
3  B>  1000:2:bd00::/64 [20/0] via fe80::f8ca:fff:fe49:5e4c, eth0-2, 00:16:27
4  B>  1000:4:d700::/64 [20/0] via fe80::bc94:86ff:fe75:bf82, eth0-3, 00:16:24
5  B>  1000:5:1300::/64 [20/0] via fe80::f44a:aaff:feee:7d1, eth0-4, 00:16:23
6  B>  1000:b:c00::/64 [20/0] via fe80::b2cf:ffff:feff:ffff, eth0-5, 00:16:26
7  B>  1000:b:6200::/64 [20/0] via fe80::b42f:ffff:feff:ffff, eth0-6, 00:16:27
8  B>  1000:c:b900::/64 [20/0] via fe80::807a:13ff:feaa:399b, eth0-7, 00:16:23
9  B>  1000:c:f800::/64 [20/0] via fe80::54e5:f7ff:fe69:18f0, eth0-8, 00:16:23
10 B>  1000:d:1c00::/64 [20/0] via fe80::e4d2:97ff:fe06:c0cb, eth0-9, 00:16:25
11 B>  1000:11:5400::/64 [20/0] via fe80::a4da:4dff:fe49:eebf, eth0-10, 00:16:25
12 B>  1000:15:8700::/64 [20/0] via fe80::b8a7:5fff:febf:cbfe, eth0-11, 00:16:23
13 B>  1000:19:3500::/64 [20/0] via fe80::54a2:d8ff:fee9:f206, eth0-12, 00:16:26
14 B>  1000:19:3d00::/64 [20/0] via fe80::d0d1:69ff:fe9e:381b, eth0-13, 00:16:26
15 B>  1000:1a:6a00::/64 [20/0] via fe80::186a:ffff:feff:ffff, eth0-14, 00:16:23
16 B>  1000:1b:6a00::/64 [20/0] via fe80::c852:3dff:fe1e:a3f0, eth0-15, 00:16:23
17 B>  1000:32:9c00::/64 [20/0] via fe80::309c:ffff:feff:ffff, eth0-16, 00:16:25
```

图 6.5　自治系统174内路由器的BGP路由表表项数据

6.9　小结

本章在软件路由器Quagga上实现了基于AS的A-BGP路由机制，在Docker试验床上进行A-BGP路由机制的测试。Tier1自治系统构成的拓扑结构在A-BGP路由机制下的BGP路由表项为17，远小于现网络中FIB 相关Tier1自治系统的约1万个表项。基于CABA 编址A-BGP路由机制下的路由表项的数目将会远小于现

有对应的BGP路由表项，因为在基于CABA编址的A-BGP路由机制中自治系统只需要向外宣布一条前缀，而现有网络对于一个自治系统，可能向外宣布多条前缀，最多可达4000多条，参考图5.2。

第 7 章　主要结论和进一步研究工作

7.1　主要结论

- 在全部部署的情况下，FIB表的表项数目是现网络环境下FIB表表项的1/10。增量部署的情况下，FIB表表项的压缩情况与部署自治系统向外宣布的前缀数目相关，部署自治系统向外宣告的前缀数目越多，FIB 表表项的压缩情况越好。
- 部署CABA方案的自治系统只需要向外宣布一条前缀，与现今网络中一个自治系统需要向外宣告n条前缀相比，update减少$(n-1)/n$倍。因为自治系统对应前缀数目的分布差异较大，所以增量部署CABA 方案在向外宣布较多前缀的自治系统上，更有利于减少UPDATE包数。
- Tier1自治系统构成的拓扑结构在A-BGP路由机制下的BGP路由表项为17，远小于现网络中FIB相关自治系统的约1万个表项。

7.2　进一步研究工作

1. 根据网络层级设计增量部署计算FIB压缩率:在第4章的实验中，可以根据网络层级进行部署，计算FIB的压缩率。
2. 根据层级结构设计增量部署进行SIMBGP仿真:仿真选取自治系统时，可以根据层级结构进行随机选取，而不是完全随机。
3. Docker部署更大的网络规模:目前Docker上只部署了18台路由器，为了接近网络的真实环境，可以部署几百台甚至上千台进行实验。
4. 部署实际网络环境，分析实验:本文所有的实验均是虚拟的网络环境，可能和真实的网络环境仍有出入，所以应该部署实际的网络环境，进行实验的分析。

插图索引

表格索引

参考文献

[1] Bgp table data. http://bgp.potaroo.net/index-bgp.html

[2] Route views. http://www.routeviews.org/

[3] Lougheed K, Rekhter Y. Border gateway protocol (bgp). request for comment rfc-1105. Network Information Center, 1989.

[4] Rekhter Y. T. li. a border gateway protocol 4 (bgp-4). request for comment rfc-1771. Network Information Center, 1995.

[5] Marques P R, Dupont F. Use of bgp-4 multiprotocol extensions for ipv6 inter-domain routing. 1999.

[6] 侯婕. 标识路由关键技术. 软件学报, 2010, 21(6):1326–1340

[7] Subramanian L, Caesar M, Ee C T, et al. HLP: A Next Generation Inter-Domain Routing Protocol. SIGCOMM, 2005

[8] Oliveira R, Lad M, Zhang B, et al. Geographically informed inter-domain routing. Network Protocols, 2007. ICNP 2007. IEEE International Conference on. IEEE, 2007. 103–112

[9] Krioukov D, Fall K, Brady A, et al. On compact routing for the internet. ACM SIGCOMM Computer Communication Review, 2007, 37(3):41–52

[10] Deering S. Metro-based addressing: A proposed addressing scheme for the ipv6 internet. Presentation, Xerox PARC, 1995.

[11] Rekhter Y, Lothberg P, Hinden R, et al. An ipv6 provider-based unicast address format. 1997.

[12] O'Dell M. Gse: An alternate addressing architecture for ipv6. draftietf-ipngwg-gseaddr-00. txt. Network Working Group, 1997.

[13] Hinden R M, Haberman B. Unique local ipv6 unicast addresses. 2005.

[14] Levy M, Pounsett M. A mechanism to allocate ipv6 blocks for bgp networks based on the networks as number. 2013.

[15] Li J, Veeraraghavan M, Reisslein M, et al. A less-is-more architecture (lima) for a future internet. Computer Communications Workshops (INFOCOM WKSHPS), 2012 IEEE Conference on. IEEE, 2012. 55–60

[16] Libbgpdump. https://github.com/woodrow/cidr-report_analysis/tree/master/libbgpdump/libbgpdump-1.4.99.13

[17] Prefer routing principle. http://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/13753-25.html

[18] Tier1 definition. https://en.wikipedia.org/wiki/Tier_1_network

[19] Caida as-rel info. http://data.caida.org/datasets/as-relationships/serial-1/

[20] Caida as-rel readme. http://data.caida.org/datasets/as-relationships/README.serial-1

[21] http://www.nongnu.org/quagga/

[22] Docker. http://docs.docker.com/

致　谢

非常感谢尹霞老师、王之梁老师、施新刚老师对我毕设工作的指导和帮助，特别感谢王之梁老师每周和我讨论毕设内容，不仅在理论上及时纠正我的错误，而且在实践的过程中帮助我解决实际问题，让我初步认识了做研究和解决问题的方法。

感谢姚姜源、吴丹、耿海军、王太红、郭迎亚、张晗、杨言等实验室同学们在我毕设遇到问题时，认真耐心地与我讨论，帮助我顺利完成该论文中的实验。

声　明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签　名：毛庆　日　期：2015.6.23

附录 A　外文资料的调研阅读报告或书面翻译

未来互联网中的Less-Is-More结构[15]

在未来互联网的发展中，人们提出了一种解决域间路由问题的新型寻址和路由设计，称之为Less-Is-More结构。该设计不同于最近提出的身份位置分离的解决方法，而是使用与位置无关的名称和与位置有关的地址。该设计需要结合两大相关政策，即stub必须是PA地址和stub级别的路由表信息不能扩散到全局路由表中。但政策的可行性还在研究中。该政策和设计的结合很大程度上导致全局路由表变小，但是也带来了四大挑战，即地址重编号（当stubs改变服务商），多宿主，移动性和流量工程。解决这些挑战的方法也有很多，比如使用多地址，基于端口的名字，LIMA概念的地址分解，特定传输协议（比如能够进行动态地址重配的SCTP协议），新的管理平台和控制平台程序。从基础的RIB数据分析可以确定LIMA结构可以将全局的路由表大小从现在的335K表项变成6815个表项。Stub的更新导致服务提供商的变化，维护LIMA结构的主要因素为服务提供商的改变，最近6个月平均每月有2450家服务商变化经过33k的stub。

I. 介绍

A．背景

2007年，互联网构架委员会的一篇报告[1]指出，全局的路由表由一级的互联网服务提供商维护，在默认的区域也同样适用的全局路由表，正在以一个逐渐增长的惊人的速率增长。与此同时，有人提出了一个相反的观点[2]，17%的指数年增长速率与内存技术的提高相一致。这个想法在现今为了节省能源，减少资源消耗的大趋势中，不能被接受。也就是说，在未来互联网中的设计，应该控制资本支出（路由器内存的花费）和操作支出（管理和控制）。

B．假设

LIMS的假设是实行新的地址分配政策，在这个政策中，stub必须是PA地址，同时stub级别的可达信息不能传播到全局的路由表中，这不需要一个有意

义的管理部门，如果传播到全局，网络使用效率将会很低。该政策很明显导致全局路由表更小，而且有更低的增长速度。但是这个政策带来了四个挑战：地址重编号，多宿主，流量工程，和移动性。解决这些挑战的方法正在发展完善中，文中的方法是将地址分解这个新颖的概念和提出的机制相结合。

C. 关键点

这篇论文提出名为LIMA的新型互联网结构，在这个结构中使用的数据包大小尽可能的最小，这给处理一些小概率的事件，比如地址重编号和连接链路失败，增加了负担。通过分析BGP RIB数据，我们发现LIMA结构有利于减小全局路由表的大小，而LIMA主要开销是随着stub变化服务提供商变化的数目的大小。

在浏览第二部分的相关工作之后，在第三部分描述了LIMA结构下的寻址和路由设计，第四部分解决LIMA结构带来的四个挑战，第五部分描述了LIMA的组件。第六部分对LIMA的优势和开销做了基本的分析。最后，第七部分总结了我们的工作。

II. 相关工作

我们把现在解决全局路由表存在问题的方法归为4类，如表1。表中右上角的方法使用了称为身份标识的第三个参数，该参数的加入使得不需要改变应用和传输协议，就可以减少全局路由表的大小。理想状况下，与位置无关的名字和与拓扑位置有关的地址足够了，因为应用中嵌套了IP 地址并且TCP不支持地址迁移。当使用与拓扑位置有关的地址进行全局路由时，需要第三个参数，身份标识。如果所有存在的地址都是与拓扑位置相关的地址，一旦stub改变了自己的服务商，如果没有身份标识，所有的主机和stub 的路由接口都需要重编号。这种位置身份分离的概念有以上的优势，但也有自己的劣势。比如每个数据平台的包需要更多的空间放置额外的指令，也需要一个管理身份和位置映射控制平台。比如数据平台应该包括NPTv6[9]和DRUID[12]的地址翻译和在LISP[5]和shim6[10]中的IP-in-IP隧道。从带宽的角度看，控制平台中身份和位置的映射看起来十分琐碎，它需要增加对于错误配置的故障排除操作花费。相反，LIMA的解决方法不需要身份，只使用名称和地址。

LIMA的层次化寻址解决概念和平铺式寻址概念，比如ROFL[13]，[14]，完全相反。层次化的寻址结构带来了很多问题，除了路径伸展、还有复杂管理系统、移动性和多宿主等等，都在这篇论文中有说明。

TABLE I: Classification of addressing and routing mechanisms

Routing Policy / Address Assignment	Stub reachability permitted in global routing tables	Stub reachability not permitted in global routing tables
Provider Independent (PI) addresses permitted for stubs	Today's Internet (IPv4 and IPv6)	eFIT [3], ILNP [4], LISP [5], HIP [6], MILSA [7], FARA [8], NPTv6 [9], Shim6 [10], TurfNet [11]
Only Provider Aggregatable (PA) addresses for stubs	None	LIMA (our strawman solution)

寻址路由分类机制

为什么我们的解决方案是“less-is-more”。 在第一部分B假设中提到的结合政策的LIMA和地址分解概念都能够用ipv6在网络层测试，因为ipv6 支持我们设计最关键的一个需求，那就是支持多地址，我们可以通过多地址可以找到接口。LIMA是“less-is-more”，因为从现今的ipv6的解决方法角度考虑，这种方法去掉了ARP和最大长度匹配；从LISP和NPTv6的角度考虑，不需要翻译和隧道的支持；从NDN[15]的角度考虑，NDN中的名字是基于每个包查找，NDN中使用160位地址的AIP协议[2]，但LIMA要求更短的固定长度分解地址，比如说32位。为了避免每一包有更复杂的处理行为，LIMA对于额外的管理会做一个处罚，但仅限于地址重编号和多宿主stub中连接链路失败这两个小概率事件。暂且不论网络层，很多方面比如应用层，套接字接口，传输层协议，DHCPv6，DNS和BGP都需要做相应的变化为了支持LIMA结构，来解决该结构带来的四大问题。这些需要的变化正在当今互联网中慢慢进行。

III. LIMA路由和寻址

LIMA主要是为了域间路由交流信息而设计的。具有额外的LIMA控制平台功能的Ipv6路由器将被使用在LIMA结构中。在该结构中，这些路由器的主要功能是作为域间路由器体现的，而不是域内路由器。在展示LIMA的寻址和路由之后，我们还讲述了两个域内stub网络的例子。

A. 寻址

简单来讲，LIMA的寻址是分层次的。这和ipv4/ipv6寻址有相似之处，需要一个全局路由前缀和一个接口身份标识，不同之处如下:

1. 使用自制系统号作为前缀
2. 给服务提供商分配的AS号必须是全局独一无二的AS号，而给stub分配的AS号是由它的服务提供商分配的本地服务商的AS号。这种方法不同于stub中可以在IP中使用PI地址。
3. 重新使用在域内路由网络中使用的地址作为接口身份标识，类似于ipv6地址可选选项中的MAC地址被扩展成EUI-64格式，然后被使用在interface-ID领域。

第一个概念在参考文献[2]也被提及过。现今ipv4网络中前缀数目大约是335k，远大于服务提供商的AS号数目大约6185；第二个概念也在别的工作中出现过，比如eFIT，区分用户网络和服务商网络。尽管第一部分提到过，eFIT[3]需要一个身份标识，而LIMA不需要。第三个概念和less-is-more的理念相一致，去掉了ARP和相关的安全威胁。虽然去掉了ARP，但是考虑到安全因素，我们提议在LIMA中使用DHCPv6，而不是SLAAC[16]。MAC地址有时候是很私密的，因为它可以暴露NIC服务商的名字。如果使用动态赋值MAC地址就可以避免泄露MAC 地址。为了避免欺骗，需要增加源地址过滤。

以上对LIMA寻址的描述，从IP寻址角度看，给读者提供了一个新的思路。基本的原理是把地址分成三个有区别的部分：全局独一无二的服务提供商AS号、本地服务商stub的AS 号、本地stub域内地址。这三个部分对应到ipv6的地址结构上，前4个比特是全局独一无二的服务商AS 号，接着4个比特是本地服务商stub的AS号，最后的8个比特是本地stub域内地址。L-DHCPv6和DNS需要进行修改来适应LIMA，L-DHPv6和L-DNS 被用来标识LIMA版本的协议。一个服务商可以有多个AS号，一个服务提供商可以给它的stub分配多个AS号。

我们面临的一个问题就是如何区分一个组织是服务提供商还是stub。比如内容传送网络服务提供商：谷歌，雅虎，微软和Akamai，这些服务提供商并不能提供一条网络专线用于网路传输，现在他们的域名和很多的stub域相关联。类似的，一些客户互联网服务提供商不提供转发服务，不管是从他们的顾客来的起源和终结。我们可以通过一个给定AS的域间连接数目来区分该AS是一个

服务提供商还是一个stub。这个问题在未来还可以继续研究。

B. LIMA路由

LIMA路由不同于下面写的IP路由。在IP路由中，一级路由表的信息包括PI和多宿主PA stub，路由需要实现最大长度匹配。在LIMA中，一级路由表没有任何关于stub的消息，也没有最大长度匹配。相反，在LIMA中，在服务提供商网络边界路由器中的分离路由表中维护的信息是，服务商AS号和stub AS号。快速的并行查表可以通过硬件实现。当一个数据报到达目的服务商网络时，stub AS号表决定该数据报从哪个边界路由器出去。Stub边界路由器有两个表，当出口数据报经过边界路由器时，边界路由器查询服务商AS号路由表，当入口数据报进入有多个stub AS号的stub时，边界路由器查询stub AS号路由表。

C. 域内stub网络举例

接下来我们将通过两个域内stub的例子来说明分解寻址的概念：一个扁平化以太交换网络和一个层次结构私有ipv4路由网络。在以太网案例中，IDA就是MAC地址。我们曾经提出给L-DHCPv6服务器赋动态的MAC地址，为了强度更大的资产管理。另外，在初始化的时候，L-DHCPv6 服务器要发送 服务提供商AS号，stub AS 号 对给终端。L-DHCPv6客户端在考虑地址分解的各个部分之后，在接口配置中创建他们的ipv6 地址。在单个以太接口的一个多宿主stub中的一台的主机将有一个单一的IDA（MAC地址）。用MAC地址联系多对 服务提供商AS号， stub AS 号产生多个ipv6地址。类似，在一个层次结构私有ipv4路由网络中的stub将会使用私有ipv4地址作为IDAs，写在ipv6地址的ID区域。

每一个主机接口对应一个权威的与IDA相关的名字。另外，一个名字对应一个主机，匹配多个接口对应的多个权威的名字。L-DNS服务器会存储所有终端名字与IDA的映射和一些简单的入口映射从组织名字映射到 服务提供商AS号， stub AS 号 对。一个完全符合规定的主机域名需要结合带有主机名的组织名字。L-DNS的查询和安全动态DNS更新支持分解地址结构，可以使用stub名字或者一个特殊的终端名字。

IV. 设想的四种挑战的解决方法地址重编号

A. 地址重分配

为了彻底实现全自动重新编号，LIMA采用参考资料[17]，[18]的机制，该机制可以分为三类：(i)主机相关，(ii)DNS相关，(iii)路由器相关。

主机相关。关键的特征包括：(a)多地址寻址,(b)基于端口的名字(NBS)[19](c)地址分解的LIMA概念(d)SCTP[20]和MPTCP[21] 的使用。多地址对带有停工的地址重编号至关重要，因为当在执行地址重编号的所有步骤中，stub能维护与旧服务提供商的连接信息1到2天。然后，要求应用只能使用域名，避免使用NBS在应用中缓存任何地址的情况，应用对于NBS都应该是很灵活的。应用只存储或者处理名字，NBS 层将名字翻译成地址。再者，地址重分配需要把新的 服务提供商 AS 号， stub AS 号对广播给stub里面的所有终端，L-DHCPv6客户端收到发过来的参数，结合没有变化的IDAs生成ipv6地址，配置接口。最后，因为大部分的TCP连接都是短暂的，而且与旧服务商的连接需要每过一段时间进行维护，以防DNS缓冲查询区的数据失效，所以基于地址的一些旧的服务器不再使用后，某些连接需要终止。如果是长期的TCP连接，支持动态地址重配的传输层协议发生类似的情况，就重新连接。为了支持这个结构，我们提出使用SCTP和MPTCP。

DNS相关。接下来，让我们研究一下DNS的更新和DNS的缓存。现在的DNS服务器对每一个域名都维护一个完整的IP记录。在LIMA 中，我们把数据库的每一条记录改为组织名（比如： Virginia.edu）对应一个或者多个 服务提供商AS号， stub AS 号 对，存在单个的记录匹配主机的名字和IDAs。这样的结构更容易应对服务提供商的变化。为了让缓存区DNS记录在其他stub和服务供应商中使用，stubs要维护和旧服务提供商之间的连接长达最大生存时间。

路由器相关。我们认为一个LIMA路由控制器需要运行(i)L-DHCPv6客户端，(ii)一个L-DNS客户端，(iii)拥有一个标识性的路由接口。LIMA的分解地址将采用发展比较完善的自动路由配置技术，比如Netconf[22]。隧道配置应用应该采用名称而不是IP地址。LIMA的分解地址也需要最新的防火墙过滤技术。

B. 多宿主

图一展示了在LIMA政策下，当stub收到来自每一个服务提供商的本地服务提供商stub的AS号通过其服务提供商构造的 服务提供商AS 号， stub AS 号映射对，比如A-2和B-1。如果stub和服务商A的连接断开了，考虑LIMA的路由政策的限制，一级互联网服务提供商不能到达A-2，结果没有数据包通过服务

商B到达A-2。对于这个问题，我们提出了一个解决方法，那就是在stub的边界路由器和服务商A的边界路由器之间建立一条通过服务商B的隧道（图1中的点虚线），然后我们就可以把这条隧道作为stub和服务商A之间的备用链路。理论上应该规定隧道的优先级来保护经过服务商B的连通链路。在BGP中MED值可以用来设置直接连接链路作为第一选择，将备用链路作为第二选择，当链路失败时使用备用链路。

为了数据报无中断转发，有三点要求。第一，当收到路由器发来的SNMP表明直接链路连接断开的信息，为了防止使用A-2地址进行新的连接，stub的错误管理系统告诉L-DHCPv6服务器，它会发出广播信息警示所有的终端停止使用A-2全局地址。第二，错误管理系统也应该通知L-DNS服务器（LIMA版本的DNS服务器），让它不要提供A-2地址的查询结果。第三，对于正在进行的连接，终端上面的L-DHCPv6客户端应该开始SCTP或者MPTCP动态地址重配。

C. 移动性

未来互联网设备中占据很大比例的可能是无线设备。其中，很大部分是移动，其中漫游比例可能较小。我们认为在LIMA层次化结构中，使用现今移动IP的方法来解决设备移动性的问题。

然而，为了减少路径扩展问题，我们建议结合动态DNS解决方法来增强移动IP的解决方法。现在有很多的方案提出使用安全动态DNS更新的结构来处理移动位置的管理，比如参考资料的[23],[24]。LIMA也使用了这样的方案。在LIMA中启用DNS服务器是非常有用的，因为这样当一个服务商改变的时候，stub的DNS服务器会告诉它的漫游设备，它本地发生了一些变化 服务提供商AS号， stub AS 号 。在LIMA中使用移动IP可以处理DNS缓冲池中地址初始化时的连接，此外，本地stub边界路由器对它所有的终端支持本地代理，对游客支持外地代理。

D. 流量工程

与服务提供商相关的流量工程。在LIMA中取消现在符合ASN标准的前缀和最大前缀匹配，还有其路由政策中不允许stub的信息传到全局路由表，这些都可以导致路径延伸。比如两个主干路由器，Internet2和ESnet 。这两个路由器在洛杉矶、西雅图、芝加哥、纽约、华盛顿相互连接。ESet有两个stub客户，一个在加利福尼亚（CA），另一个在纽约（NY）。如果stub的信息可以传播出去，

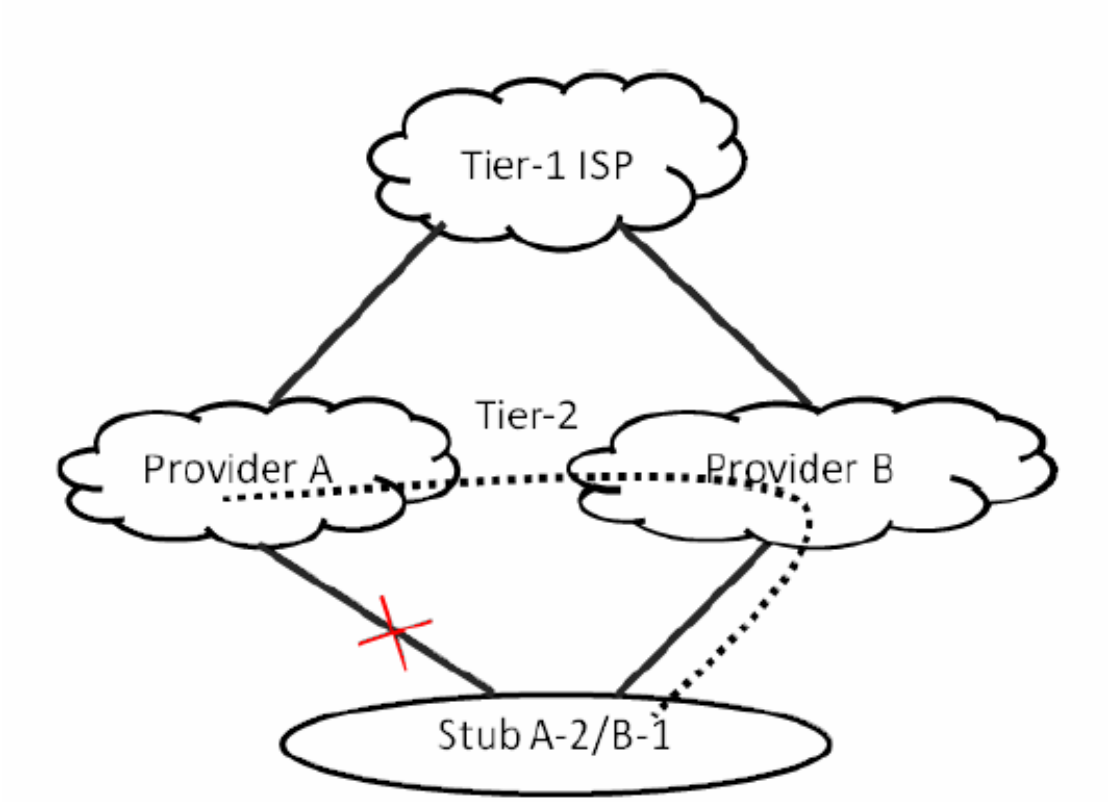

Fig. 1: LIMA multihoming

LIMA 多宿主结构图

ESet能把它的两个stub的更长前缀匹配通知给Internet2 。从Internet2中美国堪萨斯州的路由器发给ESnet中美国加利福尼亚的包，在Internet2 中将会向西朝着西雅图的路由器向前发包，如果这个包是发往美国纽约的路由器，那这个包将会往相反的方向朝着芝加哥路由器向前发包，前提条件是对于每一个路由器，在BGP更新时，都会收到来自不同路由器的信息，ESnet将会配置不同的MED值。但是，在LIMA中如果ESnet只能广播一个服务商AS号，这么有效的路由不能实现。

我们提出的解决方法是一个服务商有多个AS号，使得可以通过不同的AS号定位到这个服务商网络的不同地方。未来的工作中将会具体分析，为了实现好的交易，分配的AS号要尽可能的少，为了在低路径扩展值的情况下，不增加全局路由表的大小。

与Stub相关的流量工程。在当今的互联网中，为了平衡通过stub服务商的入口流量，一个多宿主的stub通过它的每一个服务商，有选择性地将更长前缀匹配的地址发给全局路由表。但在LIMA中，这是不可能的。我们提出了一个基于DNS和DHCP的方法来解决stub的流量工程问题。当应用程序请求选择地址的时候，stub权威的DNS服务器能整理多个地址将其返回给应用程序。对于出口流量，当告知stub的所有终端 服务提供商AS号， stub AS 号 对，L-DHCPv6客户端使用不同的次序，同时用NBS层选择地址。

V. LIMA组成部分

图二展示了由基于LIMA的边界ipv6路由器组成的stub网络的内部结构。使用NBS接口而不是TCP或者UDP接口修改应用。NBS为了调用套接字定义了一个新的本地类型AF_NAME。接收方的listen和accept调用，发送方的open调用都是用域名而不是IP地址。Read和Write系统调用是NBS套接口描述器的接口函数。源地址信息放置在发送给目的地址的第一个ipv6扩展包中，接收应用会用名字而不是IP地址缓存关于源的信息（比如：颁发许可证的服务器）。

除了为了支持LIMA的分解结构而修改DHCPv6，地址重编码或者发生连接线路失败时，需要进行强制广播操作。不能使用DHCPv6重配消息，因为这个消息是用来重配一个简单的客户端的，而我们需要一类消息是stub中对所有全局地址可达终端加入或者删除 服务提供商AS号， stub AS 号 对的命令。对

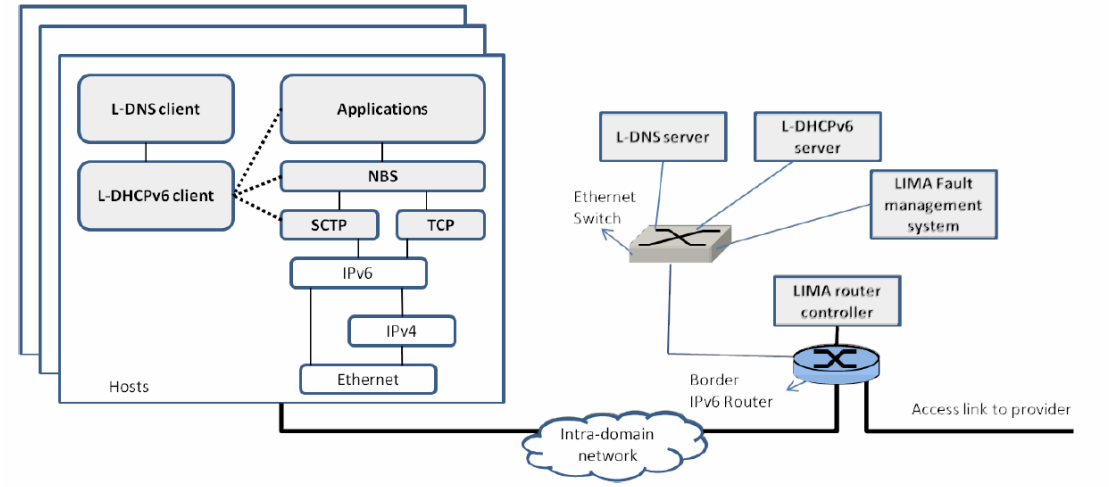

Fig. 2: LIMA stub architecture

LIMA stub结构图

于L_DHCPv6服务器发送stub边界路由器接口的IDAs（也就是当今网络中的网关地址）和DNS服务器的IDA（现今DNS 服务器的完整IP地址需要发给DHCP客户端）是非常高效的。因为LIMA比现在的网络更加依赖域名，预计来源于DNS客户端运行在终端上的安全动态DNS更新发生更加频繁，如图二。如果新的主机没有权威DNS更新要求的证书，L-DHCPv6服务器将会返回初始注册名字和IDA匹配。该行为要求在DHCPv6消息中增加名字。

需要对DNS进行简单地修改来适应地址分解结构。比如，原来数据库的结构应该修改成在第四部分A中提到的结构。一个自制系统地资源记录和对移动性的支持也应该加进去。在处理链路连接失败的时候，要求LIMA错误管理系统与L-DHCPv6和L-DNS服务器通过一个协议相连接，如图2。在初始化和服务商改变的过程中，LIMA路由控制器支持路由接口的地址重分配。

为了支持LIMA的地址重分配，预计BGP也会做相应的修改。因为stub的AS号是本地服务商下的AS号，所以当BGP的更新和stub的AS号相关，stub 边界路由器不仅要更新，stub的服务商边界路由器也要更新，而且服务商的AS号需要在服务商之间传播。

VI. 分析

评估LIMA需要几种不同的分析和原型结构，在这篇基础研究中，我们进行了两大分析。该部分A中验证使用LIMA降低全局路由表大小减少方面的优势，

B中概括了地址重分配的评估。

A. 路由数据分析

通过分析近十年的路由器RIB数据[25]，我们可以绘制AS总数、stub的AS数目和服务商AS数目的增长曲线，如图三。在LIMA中，全局路由表随着服务商AS的信号低速增长，现今共有6185 个服务商。和335K的前缀和17%的指数级年增长比率[2]形成反差。

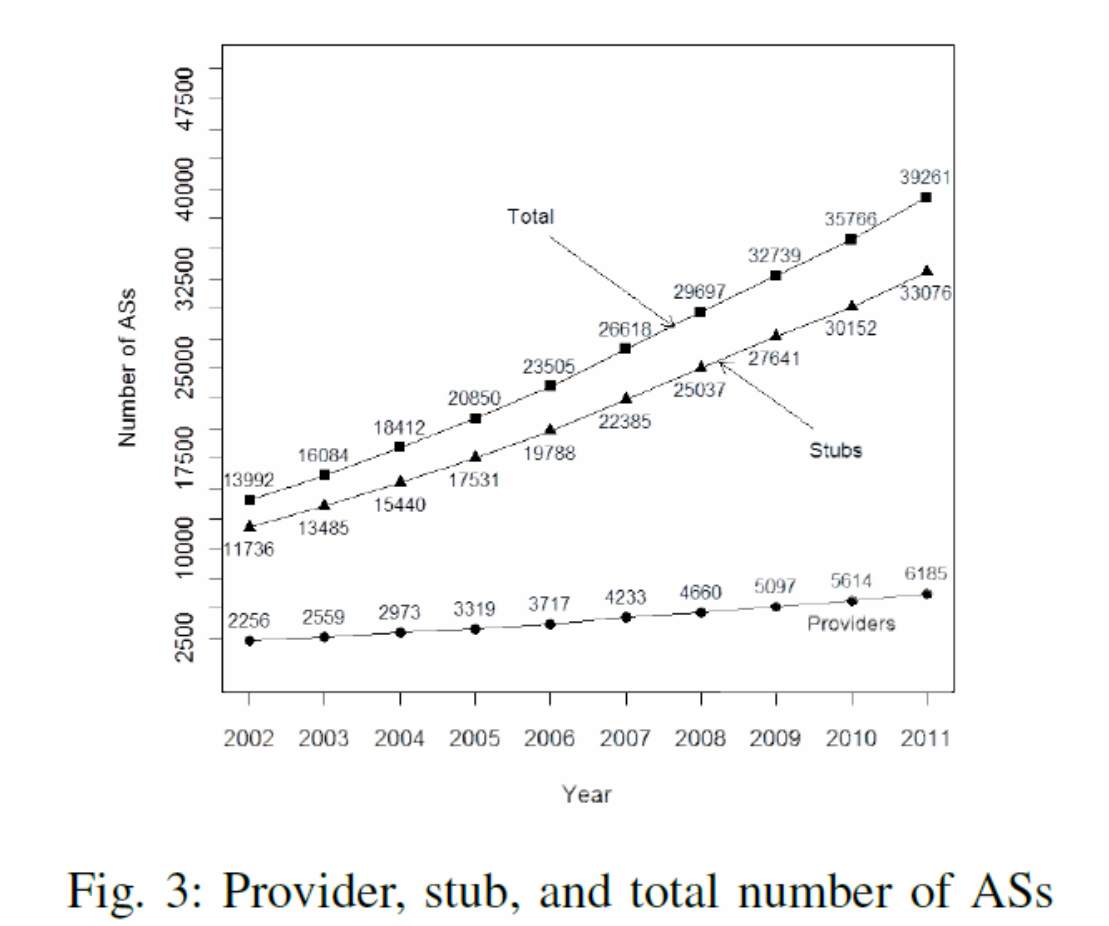

Fig. 3: Provider, stub, and total number of ASs

LIMA Provider Stub AS总数

B. 评估地址重编号

为了总结地址重分配花费代价的特征，我们分析了RIB数据为了确认stub增加或者删除服务商的频率，如表2。一些stub只增加或者减少服务商，但是增加或者减少服务商都能引起重编码操作，这两种情况都列在表2中。

TABLE II: Across all stubs (approx. 33K)

Month	#Provider additions	#Provider deletions
05/2011	1396	1049
06/2011	1408	1024
07/2011	1454	1112
08/2011	1435	1102
09/2011	1317	943
10/2011	1359	1092

服务商增加和减少数据

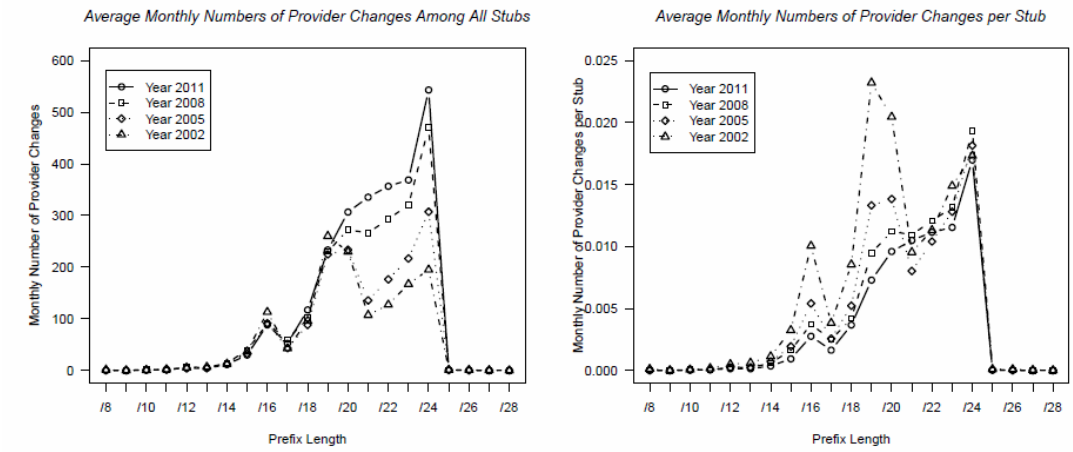

Fig. 4: Average per-month numbers

stub相关的每月服务商变化的平均数目

图四绘制了与stub相关的每月服务商变化的平均数目，与前缀块大小的功能类似。前缀块越多，引发问题的可能性越大，即使地址重分配流程是完全自动的。随着时间的推移，总数在增长，但是比率恒定。比如，2011年和2002年，对于/24类型的stub服务商平均每月变化的数量是543 和193。而2011年和2002年，stub的总数是29466和11250.因此，每个stub 的服务商平均每月变化数目是2011年0.018，2002年0.017 。

VII. 结论

文中呈现的是一个完全地址分配和路由政策的结合，可能在今天的管理机构不流行，但是在处理地址重分配、多宿主、移动性和流量工程的问题中很灵活。相关的政策建议在stub 上减少PI地址的使用，因为不允许stub级别的信息传到全局路由表中。在一个路由器比较多的stub中地址重分配在今天面临很大的挑战，需要有一些基础的改变，比如不允许在应用中使用IP地址而不是基于端口的名字，分解地址对于新的服务商AS 号要做强制广播。和今天的IP相比，LIMA除了是更新的位置-身份分离方案，还按照比例缩减了每个包的指令（取消了最大长度匹配）。但是它增加了控制和管理平台，仅仅处理小概率事件，比如服务商改变和连接链路断开。LIMA 既减少了操作花费（管理和启动消耗），还通过降低内存和流程开销减少了资本支出。

参考文献

[1] D. Meyer, L. Zhang, and K. Fall, “Report from the IAB workshop on routing and addressing,” Internet Engineering Task Force, RFC 4984, Sep. 2007. [Online]. Available: http://www.rfc-editor.org/rfc/rfc4984.txt

[2] D. G. Andersen, H. Balakrishnan, N. Feamster, T. Koponen, D. Moon, and S. Shenker, “Accountable internet protocol (AIP),” in Proc. of the ACM SIGCOMM, 2008, pp. 339–350.

[3] D. Massey, L. Wang, B. Zhang, and L. Zhang. Enabling future Internet innovations through transitwire (eFIT). [Online]. Available: http://www.netsfind.net/Funded/eFIT.php

[4] R. Atkinson, S. Bhatti, and S. Hailes, “Evolving the internet architecture through naming,” IEEE Journal on Selected Areas in Communications, vol. 28, no. 8, pp. 1319–1325, Oct. 2010.

[5] D. Farinacci, V. Fuller, D. Meyer, and D. Lewis, “Locator/ID Separation Protocol (LISP),” IETF Draft Version 16, Tech. Rep., Oct 2011.

[6] R. Moskowitz, P. Nikander, P. Jokela, and T. Henderson, “Host identity protocol,” Internet Engineering Task Force, RFC 5201, Apr. 2008. [Online]. Available: http://tools.ietf.org/html/rfc5201

[7] J. Pan, R. Jain, S. Paul, and S.-I. Chakchai, “MILSA: A new evolutionary archi-

tecture for scalability, mobility, and multihoming in the future internet," IEEE Journal on Selected Areas in Communications, vol. 28, no. 8, pp. 1344–1362, Oct. 2010.

[8] D. Clark, R. Braden, A. Falk, and V. Pingali, "FARA: Reorganizing the addressing architecture," in Proc. of ACM SIGCOMM Workshop on Future Directions in Network Architecture, 2003, pp. 313–321.

[9] M. Wasserman and F. Baker, "IPv6-to-IPv6 network prefix translation," Mar. 2011. [Online]. Available: http://tools.ietf.org/html/draft-mrw-nat66- 12

[10] C. de Launois and M. Bagnulo, "The paths toward IPv6 multihoming," IEEE Communications Surveys and Tutorials, vol. 8, no. 2, pp. 38–50, Second Quarter 2006.

[11] J. Pujol, S. Schmid, L. Eggert, and M. Brunner, "Scalability analysis of the TurfNet internetworking architecture," in Proc. of IEEE GlobeCom, Nov. 2007, pp. 1878–1883.

[12] J. Touch, I. Baldine, R. Dutta., G. Finn, B. Ford, S. Jordan, D. Massey, A. Matta, C. Papadopoulos, P. Reiher, and G. Rouskas, "A dynamic recursive unified internet design (DRUID)," Computer Networks, to appear 2011.

[13] M. Caesar, T. Condie, J. Kannan, K. Lakshminarayanan, I. Stoica, and S. Shenker, "ROFL: Routing on flat labels," in Proc. ACM SigComm, 2006.

[14] A. Singla, P. B. Godfrey, K. Fall, G. Iannaccone, and S. Ratnasamy, "Scalable routing on flat names," in Proceedings of ACM Co-NEXT, 2010, pp. 20:1–20:12. [Online]. Available: http://doi.acm.org/10.1145/1921168.1921195

[15] L. Zhang, D. Estrin, and J. B. et. al., "Named Data Networking (NDN) Project," Tech. Rep., 2010. [Online]. Available: http://www.nameddata. net/ndn-proj.pdf

[16] S. Thomson and T. Narten and T. Jinmei, "IPv6 stateless address autoconfiguration," IETF RFC 4862, Sep. 2007.

[17] B. Carpenter, R. Atkinson, and H. Flinck, "Renumbering still needs work," Internet Engineering Task Force, RFC 5887, May 2010. [Online]. Available: http://tools.ietf.org/html//rfc5887

[18] T. Chown, M. Thompson, A. Ford, S. Venaas, C. Schild, and C. Strauf, "Cookbook for IPv6 renumbering in SOHO and backbone network-

s," University of Southampton, Tech. Rep., 2005. [Online]. Available: http://www.6net.org/publications/deliverables/D3.6.1.pdf

[19] J. Ubillos, M. Xu, Z. Ming, and C. Vogt, "Namebased sockets architecture," Sep. 2010. [Online]. Available: http://tools.ietf.org/html/draft-ubillos-name-based-sockets-03

[20] P. Natarajan, F. Baker, P. Amer, and J. Leighton, "SCTP: What, Why, and How," IEEE Internet Computing, vol. 13, no. 5, pp. 81–85, sept.-oct. 2009.

[21] A. Ford, C. Raiciu, and M. Handley, "TCP extensions for multipath operation with multiple addresses," Oct. 2010. [Online]. Available: http://tools.ietf.org/html/draft-ietf-mptcp-multiaddressed-02

[22] R. Enns, M. Bjorklund, J. Schoenwaelder, and A. Bierman, "Network configuration protocol (netconf)," IETF RFC 6241, 2011.

[23] A. Ahmed, S. Reaz, M. Atiquzzaman, and S. Fu, "Performance of DNS as location manager," in IEEE Int. Conference on Electro Information Technology, May 2005, pp. 1–6.

[24] B. Yahya and J. Ben-Othman, "Achieving host mobility using DNS dynamic updating protocol," in Local Computer Networks, 2008. LCN 2008. 33rd IEEE Conference on, oct. 2008, pp. 634 –638.

[25] Route Views Project. [Online]. Available: http://www.routeviews.org/

综合论文训练记录表

学生姓名	王庆	学号	2011011239	班级	计 13
论文题目	基于 AS 编址的互联网可扩展路由机制的仿真和实现				

主要内容以及进度安排

为了解决 IPv6 网络中路由的可扩展问题，提出基于 AS 的新型编址方案，即将 AS 号嵌入到 IPv6 地址中。在此编址条件下，通过对基于 AS 号的 BGP 协议进行 SIMBGP 仿真评价、Quagga 系统实现和小环境试验床测试等工作，验证该新型编址方案对路由可扩展性的意义。

进度安排如下：

第 1 周-第 2 周：　阅读文献，准备开题

第 3 周-第 4 周：　熟悉 BGP 原理，熟悉仿真平台代码

第 5 周-第 8 周：　修改仿真平台，完成仿真实验

第 9 周-第 12 周：　使用 Quagga 软件路由器实现原型系统

第 13 周-第 16 周：搭建实际路由器环境，测试验证

指导教师签字： 张

考核组组长签字： 金勇

2015 年 3 月 12 日

中期考核意见

AS 采样点时候建议采得更全面一些，考虑 AS 的 Tier。基本按照进行。通过。

考核组组长签字： 裴丹

2015 年 4 月 23 日

指导教师评语	基于AS编址的CABA方案对互联网可扩展路由问题有所帮助。本论文对该方案进行了仿真评价和测试。结合FIB表分析和simBGP仿真进行了性能评价。在Quagga平台上进行了实现，在Docker平台上进行了线路集测试。达到了本科综合论文训练的要求。 指导教师签字：王之梁 2015 年 6 月 15 日
评阅教师评语	论文选题具有重要意义， 论文研究较为深入，达到综合论文训练要求。 评阅教师签字：[illegible] 2015 年 6 月 15 日
答辩小组评语	同意通过答辩。 答辩小组组长签字：[illegible] 2015 年 6 月 19 日

总成绩：92

教学负责人签字：[illegible]

2015 年 6 月 19 日