清 华 大 学

综 合 论 文 训 练

题目：基于 RPKI 的 BGP 路由数据验证与分析

系　　别：电子工程系

专　　业：电子信息科学与技术

姓　　名：许文杰

指导教师：李星　教授

2019 年 1 月 16 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：_____ 导师签名：_____ 日　期：_____

中文摘要

BGP 协议是当前互联网默认采用的域间路由协议。然而 BGP 路由协议存在许多安全隐患，其中之一是所谓的前缀或者子前缀攻击，也即一个自治域非法发布了一个其所不应当发布的前缀。前缀或者子前缀攻击可能会导致流量劫持或者窃听等安全威胁[1]。为了应对前缀或者子前缀劫持所带来的安全威胁，IETF 标准化了一套基于 RPKI 的 BGP 前缀源验证的措施[2]。本项目首次系统探讨并测量了基于 RPKI 的 BGP 源验证中存在的伪无效前缀的问题。为了刻画前缀的聚类结构，本项目建立了前缀聚类森林用于后续分析。本文首先将无效前缀划分为了无效自身前缀、无效客户前缀、无效迁移前缀和无效劫持前缀，然后分析指出了造成伪无效前缀判断的可能原因并综合多源信息设计方法测量了不同原因造成的伪无效前缀。基于对伪无效前缀的分析结果，本文提出了改善原有基于 RPKI 的 BGP 路由源验证机制的建议，并提供了相应的理论分析。最后，综合 ROAs 数据库、自治域关系图谱、前缀聚类结构、自治域路径结构和历史路由记录等多源信息，本项目首次实现了一个检测和发布基于 RPKI 对 BGP 路由表进行源验证时出现的伪无效前缀的系统，有望帮助网络运营者合理设定路由策略，避免由于信任 RPKI 的伪无效验证结果而造成“断网”现象。

本文的创新点主要有：

- 首次系统地指出并测量了基于 RPKI 的 BGP 路由前缀源验证方法中存在的伪无效性的问题，并且分析了其可能的原因
- 基于对伪无效前缀产生的原因分析，提出了改善原有验证机制的建议，包括引入（路径，前缀）对的验证方法和引入验证请求机制，并做了相应的理论分析
- 首次实现了一个检测和发布基于 RPKI 对 BGP 路由表项进行验证时出现的可能的伪无效前缀的系统，有望帮助网络运营者合理制定路由策略，避免由于信任 RPKI 的验证结果而出现“断网”现象

关键词：边界网关协议，资源公钥基础设施，路径（由）源验证，地址聚类

ABSTRACT

BGP is the default interdomain routing protocol in today's Internet. However, there are some potential security threats in BGP, among which is the prefix or subprefix hijacking. In prefix or subprefix hijacking, an AS illegally announces a prefix not owned by itself, possibly resulting in traffic hijacking and man-in-the-middle attack[1]. To deal with the potential security risk resulted from prefix or subprefix hijacking, IETF has standardized a BGP origin validation method based on RPKI. To characterize the aggregation structure of the prefixes, prefix aggregation forest is constructed for further analysis. This project measures and analyses the phenomena of false invalid BGP items in validation process systematically for the first time. Specifically, the invalid BGP prefixes are classified into invalid self prefix, invalid customer prefix, invalid transfer prefix or invalid hijacking prefix. Then the underlying reasons for false invalid BGP items are discussed and measured systematically using the method designed with multi-sourced information. Based on the measurement and analysis result, several improvement approaches are analyzed and proposed in this paper and relevant theoretical analysis is given. Finally, a false invalid BGP items detection and publication system utilizing the multi-sourced information of ROAs database, the relationship of ASes, the structure of prefix aggregation, the structure of AS path and the historical routing records is implemented and presented to help network operators design their routing policies better and avoid possible loss of connection to some IPs due to believing in RPKI's false invalid validation result. This paper's main innovative points include:

- Systematically points out and measures the false invalid prefixes and analyzes the underlying reasons in RPKI based BGP origin validation process for the first time.
- Proposes some improvement approaches for RPKI based BGP validation mechanism, including (path, prefix) validation and validation request and conducts relevant theoretical analysis.

- Implement a false invalid prefix detection and publication system in RPKI based BGP validation process, potentially helping the network operators make their routing policies better and avoiding loss of network connection due to believing in false invalid prefix.

Keywords: BGP; RPKI;ROV; address aggregation

目 录

主要符号对照表

AS	自治域（Autonomous System ）
BGP	边界网关协议（Border Gateway Protocol ）
IETF	互联网工程任务组（The Internet Engineering Task Force）
p_i	第 i 个前缀
RC	资源证书 (Resource Certificate)
RFC	Request For Comments 文件（记录了互联网相关标准等）
RIPE	全世界五个 RIR 之一，主要负责欧洲、中东和中亚部分地区的数字资源分配[3]
RIR	负责一个区域互联网数字资源分配的组织 (Regional Internet Registry)
ROA	路由源授权（Route Origin Authorization ）
ROV	路由源验证（Route Origin Validation ）
RPKI	资源公钥基础设施（Resource Public Key Infrastructure ）
T	树

第 1 章　引言

1.1　研究背景

1.1.1　BGP 的安全隐患和应对措施

互联网作为一种无中心的网际网，是由超过 50000 个自治域组成的。为了实现不同自治域之间的通信，需要一种协议来传播自治域之间的路由信息。当前互联网默认采用了 BGP 协议，也即边界网关协议来在自治域之间传播路由信息。然而，正如许多报道[4][5] 所指出的，BGP 协议存在路由劫持的安全威胁。BGP 路由劫持可能会导致用户无法正常访问网络服务以及“中间人”攻击[1] 等恶性后果。

BGP 路由劫持在今天的互联网上频繁地发生[6]，已经造成了大量的大面积“断网”的问题[7]。例如在著名的 AS7007 事件中，该 AS 宣称了大量非法的长前缀，造成了路由黑洞[8]。

目前为了应对 BGP 路由劫持的问题，尽本文作者所知，主要有两类方案。第一类是基于网络监测的方案，也即在互联网上布设多个监测点，通过被动和主动地测量、监测 BGP 路由表、网络连接等的异常现象来检测出路由劫持[9][10]。另一类方案是利用 RPKI 提供的 ROAs（一种将自治域和其可以合法宣布的地址前缀绑定在一起的验证数据）来对 BGP 路由表项的路径源和对应前缀的合法性进行验证[2]。然而，虽然理论上 RPKI 可以验证路由前缀的合法性，但是基于 RPKI 的验证措施目前仍在部署的早期，暴露出了各种问题[11]。

1.1.2　地址的良性解聚类与恶性解聚类

BGP 的安全隐患在很大程度上与地址的解聚类密切相关。由于网络层的最长地址匹配原则，当一个自治域发布一个更长前缀时，将更容易导致 BGP 路由劫持的发生。因此，对前缀聚类关系的分析将有助于分析 BGP 的安全性质。然而，互联网上的解聚类类型并不仅仅是出于劫持的恶性解聚类，也存在着出于各种商业竞争、网络连接和流量方面的考虑而产生的良性解聚类[12][13][14]。正确地区分解聚类现象是恶性的解聚类还是良性的解聚类对于 BGP 路由安全是至关重要的。

1.2 研究现状

一方面，关于地址前缀解聚类，Tian Bu 等人的开创性工作将地址的解聚类现象按照发生的原因划分为了多出口，负载均衡，聚类失败和地址碎片四类，并且测量了这四类现象发生的比例和随着时间变化的趋势。[12] 关于自治域之间的商业关系，Lixin Gao 等人利用不同自治域由于相互关系不同导致的路由策略不同得到了对自治域关系的推断算法[15]。Luckie 等人则进一步对互联自治域之间的商业关系进行推断并且给出了关系数据集[16]。

另一方面，当前，对于 BGP 路由劫持的检测方案主要分为两大类。第一类方案基于对 BGP 路由表、网络连接性等的测量和异常发现[9][10]。另外一类则基于 RPKI 进行 BGP 路由表项的源验证[2]。

然而，基于 RPKI 的 BGP 路由路径源验证方法并不是完美无缺的，目前为止，由于各种原因，基于 RPKI 的 BGP 路由源验证方法出现了各种问题。D. Cooper 等人指出了 RPKI 可能被一些表现不当的地址分配者利用来操纵某段前缀的可达性[17]。E Heilman 等人指出了 RPKI 缺少透明度的问题并提出了提高 RPKI 的透明度的方法[18]。Y. Gilad[19] 等人则指出了基于 RPKI 的路由路径源验证方法中 ROA 中所设置的最大长度本为了用户的便利，但却带来了安全上的问题。Y. Gilad 等人调查了路径源验证的采用率，发现了 RPKI 的采用率极其低[20] 并调研了可能的原因，他们同时指出只有当 RPKI 的采用率到达一定值之后，其安全上的优势才能充分体现出来[20]。在 Y. Gilad 等人给出的调研结果中，有超过三成的网络运营商之所以不采用 RPKI 是因为存在网络连接丢失的可能[20]。正如 Y. Gilad 等人指出的，之所以存在采用 RPKI 之后，网络连接丢失的问题，是因为存在误判的问题，也即将一原本合法的地址判断成非法的并进一步将这个路径遗弃[20]。

1.3 研究意义和目标

为了进一步推动 RPKI 的部署，有必要系统地对基于 RPKI 的 BGP 路由源验证中将合法的地址前缀误判为无效的情形进行系统的测量分析，并相应地提供一份白名单数据，从而降低因为误判而导致的合法地址无法正常访问的事件发生的可能性。然而，尽本文作者所知，目前学术界在这方面的工作是空白的。

基于目前学术界在 BGP 路由安全验证领域的研究现状，本项目试图：

- 分析 BGP 路由前缀的良性解聚类（也即非攻击的解聚类）和恶性解聚类并试图区分二者
- 利用基于 RPKI 的方法验证 BGP 路由表项
- 结合对于 BGP 路由表中前缀的聚类结构和自治域的关系结构，推断出伪无效的 BGP 路由表项
- 结合数据平面的主动测量的结果，进一步地推断伪无效的 BGP 路由表项
- 在路由验证机制层面，分析现存基于 RPKI 的验证方式存在的问题，并且提出相应的可能的推荐解决方案
- 实现一个检测和发布基于 RPKI 的 BGP 路由表项验证中伪无效前缀的系统

本项目的工作结果将有望：

1. 系统地揭示当前基于 RPKI 的 BGP 路由表项验证方法中存在的伪无效路由表项的存在、类型和深层次的原因
2. 提供一份基于 RPKI 的 BGP 路由表项验证方法中存在的伪无效路由表项的白名单，从而减少网络运营商关于可能失去正常网络服务访问连接的担忧，降低因此而产生的“断网”风险，进一步推动 RPKI 的部署
3. 在验证机制层面提出建议改善原有的基于 RPKI 的源验证机制

第 2 章　基本原理和概念简述

本章将简要解释本文将涉及到的主要的基本原理和基本概念。

2.1　自治域

今天的互联网是一种网际网，是由超过 50000 个相互独立，对外拥有统一的路由策略，有统一的管理的子系统组成，这样的互联网上的子系统称为自治域，本文以下简称为 AS。每一个自治域都有一个唯一的自治域号码来进行标记[21]。IANA，也即负责互联网数字资源分配的组织，会向 RIR 分配自治域号，然后由 RIR 来在自己所在地区分配自治域号给各个自治域。例如 CERNET 即被分配了自治域号 4538[21]。2007 年之前，互联网使用的自治域号码采用了 16 个比特，而在这之后自治域号码所采用的比特的数目被扩展到了原来的两倍[21]。

2.2　BGP 协议

为了实现在一个自治域内进行路由信息的传播，可以采用域内路由协议，而为了实现自治域之间路由信息的传递，则需要采用域间路由协议。当今互联网默认采用了 BGP 协议作为域间路由协议。如图2.1所示为 BGP 协议的主要工作原理示意图。

BGP 协议的目标是在自治域之间传递路由信息，以图2.1为例，AS3 某一个路由器下拥有一块地址 C，为了将这个信息传播到全网络。AS3 会首先将这个信息传播给与自己直接相邻的 AS2，AS2 则会进一步将这个信息传递給与自己相邻的 AS1，如此不断将到达地址块 C 的路由信息扩散到整个互联网。在 BGP 路由信息的扩散过程中，会维护一个到达目标地址的自治域路径，自治域路径的最后一跳即是相应前缀的源自治域，也即该块地址所在的自治域。每一个自治域在把路由信息发布给相邻自治域之前，都会在自治域路径上附加上自己所在的自治域号从而维护一条到达目标地址的自治域路径。同时伴随着路由信息的传播，也会维护一个下一跳的地址[22] 用于网络层的路由。

在域内路由协议中，由于整个自治域受到统一的控制，一般不存在其他影

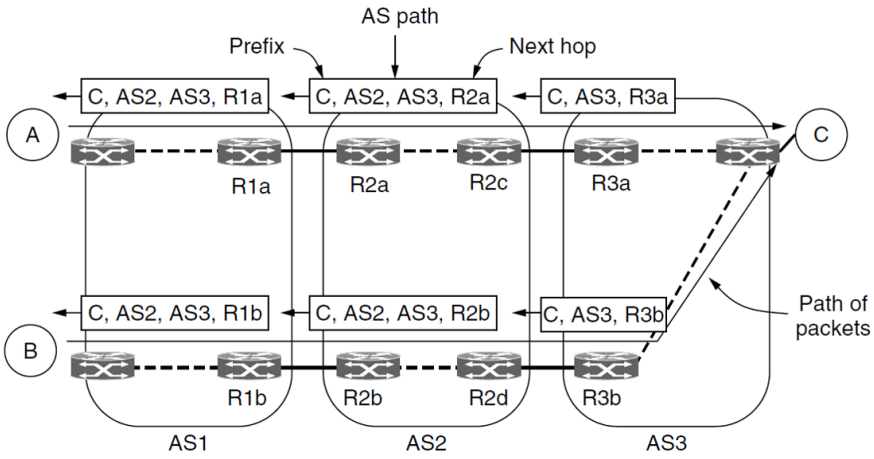


图 2.1　BGP 协议工作原理示意图[22]

响路由的干扰，路由的效率几乎是唯一需要考虑的因素。然而，由于互联网分布在整个世界上，各方面的因素非常复杂，在自治域之间进行路由的时候，自治域可能会有许多非效率层面的考虑，例如经济和安全上的考虑[22]。BGP 协议中维护的自治域路径恰好满足了这样的需求。

2.3　地址前缀聚类

所谓地址前缀，是指一个共享地址前若干位的 IP 地址集合。例如 1.2.3.0/24={1.2.3.0,1.2.3.1,1.2.3.2,1.2.3.4,⋯,1.2.3.255}，1.2.3.0/24 表示地址前 24 位与 1.2.3.0 相同的 IP 地址集合。所谓地址前缀的聚类指的是前缀长度更长的对应的地址子集并为了一个地址前缀长度更短的地址子集。如图2.2为地址前缀聚类的一个例子，在这个例子中，1.1.0.0/24 和 1.1.1.0/24 前缀聚类为了更短的前缀 1.1.0.0/23，这里就将 1.1.0.0/23 称为聚类过程中的父前缀，将 1.1.0.0/24 和 1.1.1.0/24 称为子前缀，将聚类过程的逆过程称为解聚类。

2.4　基于 RPKI 的 BGP 路由表项源验证机制

RPKI，也即所谓的资源公钥基础设施，是用于互联网路由数字资源验证的公钥密码框架[23]。RPKI 的功能之一是验证一个自治域是否有权限宣称是一个前缀的源。为了实现这一功能，目前 RPKI 提供了一种树状结构的签名机制。以图2.3为例，负责区域数字资源分配的 RIR 为下属 ISP 颁发了 RC（也即资源证

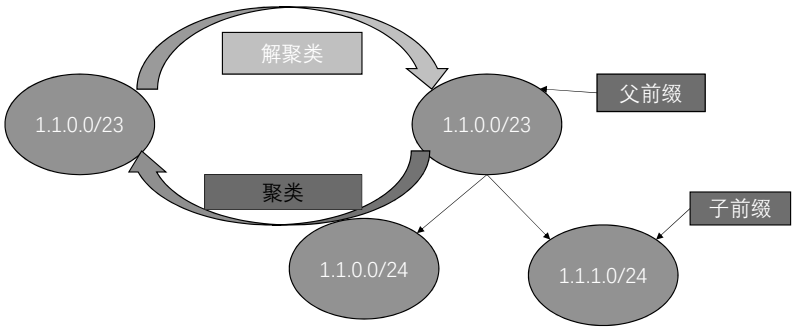


图 2.2　地址聚类与解聚类示意

书）用于证明其拥有一块地址块，进一步地，其又可以为下一级的 ISP 签署一块上述地址块对应的子地址块对应的 RC，同时为下属的自治域签署发布相应的地址块的 ROA[20]。

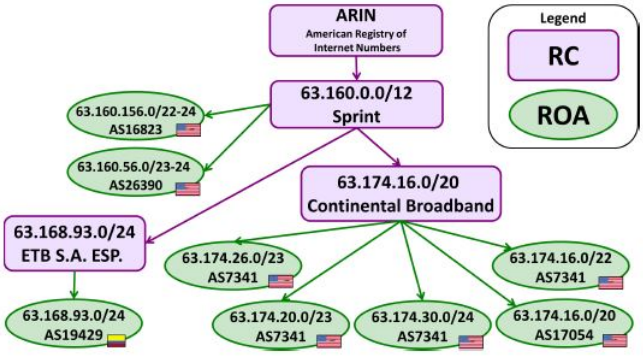


图 2.3　RPKI 的签名结构[20]

所谓 ROA，即路由源授权，它是一个被签了名的可验证的将自治域号和其

所能合法宣布的地址前缀绑定在一起的对象[24]。自治域可以通过一个分布式数据库来得到当前所签名的 ROAs 信息用以验证实际收集到的路由表[24]。

2.5 自治域商业关系

互联网是由自治域组成的，自治域与自治域之间不仅仅存在着拓扑上的连接关系，还存在着商业上的合作或者竞争的关系。L. Gao 的开创性工作将自治域之间的商业关系划分为了客户-供应商关系，对等关系和兄弟关系，并发展了启发式的自治域商业关系推断算法[25]。

第 3 章　路由数据的采集与简单统计分析

本章将要介绍 BGP 路由数据的收集和初步分析。

3.1　控制面 BGP 路由数据的收集

为了能够收集互联网上的 BGP 路由数据，选择 CERNET 拥有的 AS4538 作为对整个互联网路由信息的观测点，然后配置一台服务器，将该服务器划入一个私有自治域 AS65508，然后利用与 AS4538 的 Peer 连接收集到全网的 BGP 路由信息。如图3.1 为收集控制面 BGP 路由信息的拓扑图。

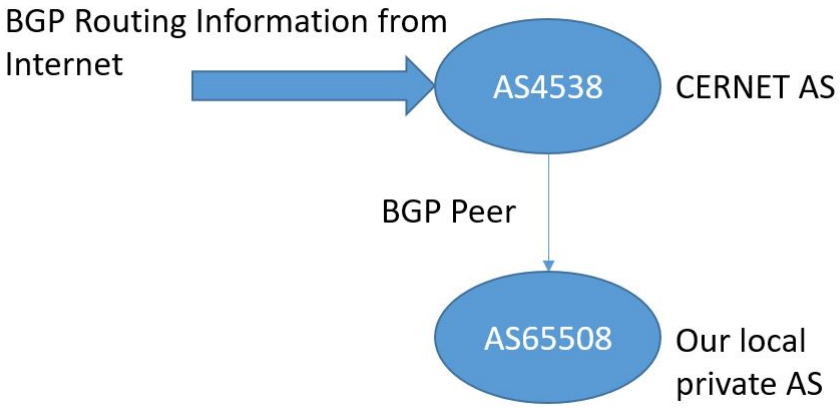


图 3.1　控制面 BGP 路由信息采集拓扑示意

利用 Quagga[26] 将服务器配置成为一个支持 BGP 协议的路由器，然后通过多跳 EBGP 来收集 AS4538 输出给 AS65508 的 BGP 路由信息，表3.1是收集 BGP 路由信息的服务器的基本信息。

表 3.1　收集 BGP 路由数据的服务器信息

服务器架构	x86_64
服务器操作系统	CentOs7
IPv4 地址	121.194.167.41
IPv6 地址	2001:250:3::6:1

在 BGP 协议中，每一个路由器会维护一个输入路由表，本地采用的路由表和输出路由表。自治域的边界路由器会收集来自相邻的非本自治域的路由器发出的路由更新的信息，然后将其存入输入路由表。在此基础上一个自治域会根据自身的路由策略，选择输入路由表中的路径作为本自治域采用的路由。最后自治域也会根据自身的路由输出策略和路由输出的对象选择相应的路由信息输出[27]。BGP 路由信息的收集可以分为两种模式，第一种是将当前的整个 BGP 路由表完全存储下来，第二种方式是将路由表的更新数据收集下来，如图3.2。

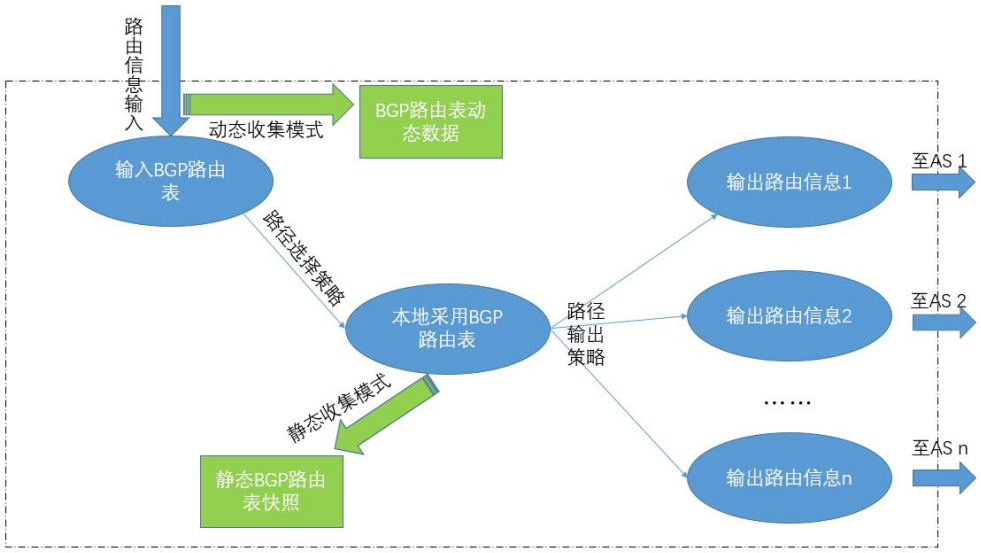


图 3.2　BGP 路由表的选择和输出过程以及路由表数据的收集模式

3.2　静态路由表数据

在本项目中，基本平均每天对当天路由表做一次快照存储（除服务器断电意外期间）。下面以 2018 年 2 月 28 日收集到的 BGP 路由表数据为例，对当前互联网的 BGP 路由数据做简要统计分析。该 BGP 路由表中共收集到 680000 多条 IPv4 路由表项，收集到 47000 多条 IPv6 路由表项。

首先考察地址前缀的分布，如图3.3为 IPv4 和 IPv6 中的地址前缀长度的分布。可以看到在 IPv4 中，/24 的地址超过了一半，进一步的测量表明这其中又有 50% 左右的/24 前缀来自于一个更短前缀的解聚类。如图3.4是解聚类类型的分布。在理想的状态下，BGP 路由表中的所有前缀对应的地址集合应该是两两

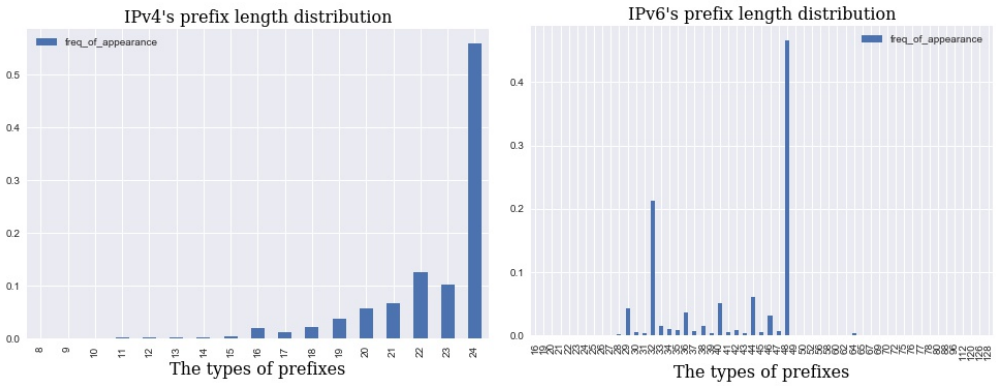


图 3.3　在 IPv4 和 IPv6 中的前缀长度的分布

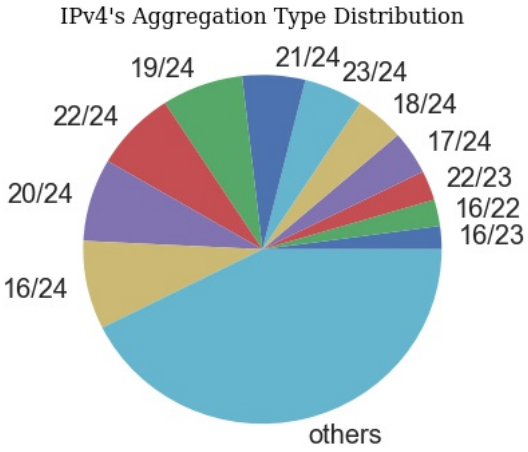


图 3.4　前缀聚类长度类型统计（IPv4）

不相交的，出现大量的解聚类前缀意味着对存储和计算资源的大量浪费，同时还可能存在着地址劫持的安全风险。

除了静态数据，长时间积累下来的历史 BGP 路由更新数据也展现出了丰富的动态特征，如图3.5为从 3 月 12 日 9 点到 4 月 9 日 9 点的 BGP 路由每 30 秒更新次数。

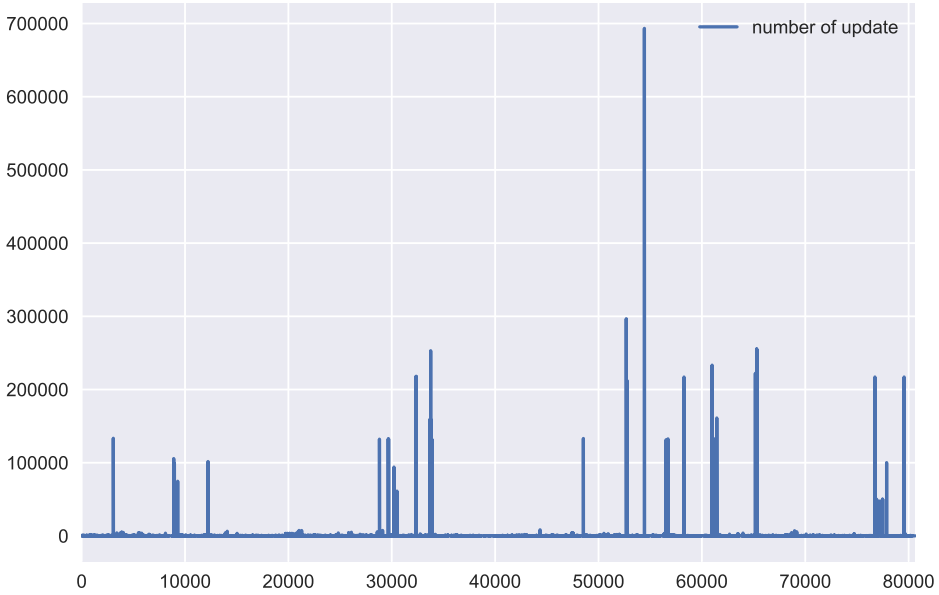


图 3.5　从 3 月 12 日 9 点到 4 月 9 日 9 点的 BGP 路由每 30 秒更新次数

第 4 章　前缀聚类森林的建立与前缀解聚类的分类

4.1　前缀聚类森林的建立

4.1.1　前缀聚类森林的定义

为了更好地表示地址聚类的结构，同时考虑到地址前缀之间的包含关系是一种偏序关系，可以选择用树的结构来表示地址聚类的结构。考虑到两个聚类前缀子集之间可能不相交，因此前缀聚类树会形成一个森林。如图4.1所示，为前缀聚类的示意图。为了本文之后的叙述方便，做如下形式化定义。

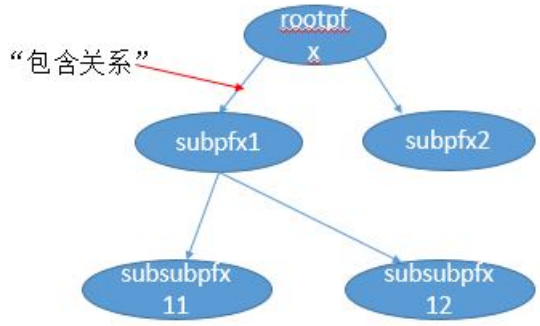


图 4.1　前缀聚类树示意

定义 1：　给定一个前缀的集合 $P = \{p_1, p_2, \cdots, p_n\}$，定义一个有向图 $G = (V, E)$，其中 $V = P = \{p_1, p_2, \cdots, p_n\}$，$(p_i, p_j) \in E$ 当且仅当 p_i 包含 p_j 并且若 p_k 包含 p_j 且 p_i 包含 p_k，则 $p_k = p_i$ 或者 $p_k = p_j$，称 G 为 P 的前缀聚类图。

定义 2：　给定一个前缀的集合 $P = \{p_1, p_2, \cdots, p_n\}$，称 p_i 为 P 的极大前缀，若不存在 $p_j \neq p_i$ 使得 p_j 包含 p_i，令 R 为 P 的极大前缀组成的集合，称 R 为 P 的极大前缀子集。

命题 1：　定义1中得到的图 G 是一个树的并集，即一个森林。

证明　对 n 归纳。$n = 1$ 时，显然成立。假设 $\forall P^{(n-1)}$ 满足 $|P^{(n-1)}| = n - 1$ 时，对应的前缀聚类图是一个森林，则当 $|P^{(n)}| = n$ 时，取一个 $P^{(n)}$ 的 $n - 1$ 元子集 P_0，

由归纳假设，P_0 对应的前缀聚类图 $G_0 = (V_0, E_0)$ 是树的并，设 P_0 对应的极大前缀子集为 R_0，设 $\{p\} = P^{(n)} - P_0$。设 S 为 p 在 $P^{(n)}$ 的子前缀集，Pa 为 p 在 $P^{(n)}$ 中的父前缀集，设 $P^{(n)}$ 对应的前缀聚类图为 $G^{(n)}$，下面分类讨论：

1. 若 $S = Pa = \emptyset$，则 $G^{(n)} = (V_0 \cup \{p\}, E_0)$ 依然是一个森林。
2. $Pa = \emptyset, S \neq \emptyset$，易见 $G^{(n)} = (V_0 \cup \{p\}, E_0 \cup \{(p, r) | r \in S \cap R_0\})$，易验证依然是一个森林。
3. $Pa \neq \emptyset, S = \emptyset$，易见存在叶子结点 le，使得 $G^{(n)} = (V_0 \cup \{p\}, E_0 \cup \{(le, p)\})$，依然是一个森林。
4. $Pa \neq \emptyset, S \neq \emptyset$，设 Pa 的极小前缀为 pa 和 S 中的极大前缀为 s，则 $G^{(n)} = (V_0 \cup \{p\}, E_0 \cup \{(pa, p), (p, s)\} - \{(pa, s)\})$ 依然是一个森林。

综上，$G^{(n)}$ 是一个森林。 □

4.1.2 前缀聚类森林的建立算法

如果直接将前缀逐个加入到前缀聚类森林中，则建立算法的时间复杂度将会是 $O(n^2)$，这在实际操作中是不可忍受的，下面按照时间复杂度更低的算法1建立前缀聚类森林。其中利用一个极大前缀和其所有的子前缀建立前缀聚类树的算法如算法2。[①]

新的前缀聚类森林的建立算法的时间复杂度主要来自排序，排序时间复杂度为 $O(n \log n)$，相比原来的 $O(n^2)$ 的时间复杂度得到了改善。

4.2 地址解聚类的分类

4.2.1 地址解聚类的原因分析

地址聚类森林上的每一条边都是前缀空间的一种冗余，既是对路由器存储和计算资源的消耗，也存在可能的地址劫持的安全威胁。Tian Bu 等人的开创性工作[12] 将地址前缀空间的膨胀归结为以下四项原因：

1. 多出口
2. 负载均衡
3. 聚类失败
4. 地址碎片

① $p_i[0], p_i[-1]$ 分别表示 p_i 前缀下的第一个和最后一个地址

Algorithm 1 前缀聚类森林建立算法

输入：前缀集合 $P = \{p_1, p_2, \cdots, p_n\}$　**输出**: 前缀集合 P 对应的前缀聚类森林 G

对 P 中前缀按照首地址排序，为了算法叙述方便，仍然将排序后的结果记为 $p_1, p_2, \cdots, p_n$

初始化极大前缀集合 $R \leftarrow \emptyset$, 非极大前缀集合 $S \leftarrow \emptyset$ 以及每个前缀的子前缀集合 $Sub(p_i) \leftarrow \emptyset, i \in \{1, 2, \cdots, n\}$

for $i \in \{1, 2, \cdots, n\}$ **do**
  if $p_i \in S$ **then**
    继续循环
  end if
  R 中加入 p_i
  for $j \in \{i-1, \cdots, 1\}$ **do**
    if $p_j[0] < p_i[0]$ **then**
      终止循环
    end if
    if p_i 包含 p_j **then**
      if $p_j \in R$ **then**
        将 p_j 从 R 中移除
      end if
      $Sub(p_i)$ 中加入 p_j
      S 中加入 p_i
    end if
  end for
  for $k \in \{i+1, \cdots, n\}$ **do**
    if $p_k[0] > p_i[-1]$ **then**
      终止循环
    end if
    if p_k 是 p_i 的子前缀 **then**
      $Sub(p_i)$ 中加入 p_k
      S 中加入 p_k
    end if
  end for
end for

31: **for** $p_i \in R$ **do**

32: 用 p_i 和 $Sub(p_i)$ 建立前缀聚类树 $T(p_i)$

33: **end for**

34: $G \leftarrow \cup_{p_i \in R} T(p_i)$

本文将对该四点原因做逐一解释。图4.2中有向直线段表示的是供应商-客户的商业关系，其中箭头所指的是客户，曲线表示的 BGP 路由信息传播的方向。

如图4.2(a)是所谓的多出口现象,AS2 拥有一块地址，并将该块地址中的一个子块分配给了自己的客户 AS1，然而 AS1 并不仅仅拥有 AS2 一个供应商，同时也拥有 AS3 作为供应商，这个时候 AS1 出于网络服务的稳定性等方面的考虑，可能会再通过 AS3 传播路由信息。由于 AS3 自身发布的地址无法与 AS1 发布的地址相聚合，在 AS4 的路由表中就会发生解聚类现象[12]。

如图4.2(b)是所谓的负载均衡现象，AS1 除了对外宣称一段父前缀外，为了使外部到达自身的流量可以尽量均衡地通过多条路径到达自身，进而充分利用网络连接，会将父前缀下的两个子前缀通过不同的网络路径对外宣布出去。从而使得 AS4 的路由表中发生前缀解聚类的现象[12]。

如图4.2(c)是所谓的聚类失败的现象，AS1 用完全一致的路由策略对外宣称了两段本可以合并的前缀，此时虽然 AS4 的前缀聚类森林中虽然不会出现相应的边，但是会存在两个自治域路径完全一致的前缀可以聚合为一个前缀，而不会损害 BGP 的路由策略的执行[12]。

如图4.2(d)是所谓的地址碎片的现象。在理想的情况下，为了满足一个自治域的需求，应当将尽可能连续的地址块分配给同一个自治域，使该自治域尽可能只需要发布一个前缀。然而在现实的互联网中，一个自治域可能会用完全相同的策略来发布两个不能聚合的前缀。在历史发展过程中，AS1 可能首先获得了一块地址，但之后可能又获得了一块不能与最初得到的前缀相聚合的地址前缀，进而用相同的路由策略来对外发布两块不能聚合的地址前缀[12]。

除了上述良性的（也即不造成安全威胁）的解聚类情形外，还存在错误配置或者恶性攻击造成的恶性的解聚类，可能会造成前缀劫持。

Algorithm 2 前缀聚类树的建立算法

输入：极大前缀 p 和其子前缀集 S　**输出**：p 和 S 所生成的前缀聚类树 T

初始化 $T \leftarrow (V, E)$，$V \leftarrow \{p\}$，$E \leftarrow \emptyset$

for $p_i \in S$ **do**

  $p_0 \leftarrow p$

  flag$\leftarrow$ 0

  while flag==0 **do**

    next$\leftarrow$ 0

    for $p_j \in \{p_0$ 在 T 中的子结点$\}$ **do**

      if p_i 包含 p_j **then**

        将 p_i 加入 V

        将 (p_0, p_j) 从 E 删除，将 (p_0, p_i) 和 (p_i, p_j) 加入 E

        flag$\leftarrow$ 1

        结束循环

      end if

      if p_j 包含 p_i **then**

        $p_0 \leftarrow p_j$

        next$\leftarrow$1

        结束循环

      end if

    end for

    if next==1 **then**

      继续

    else

      if flag==0 **then**

        将 p_i 加入 V，将 (p_0, p_i) 加入 E

      end if

    end if

  end while

end for

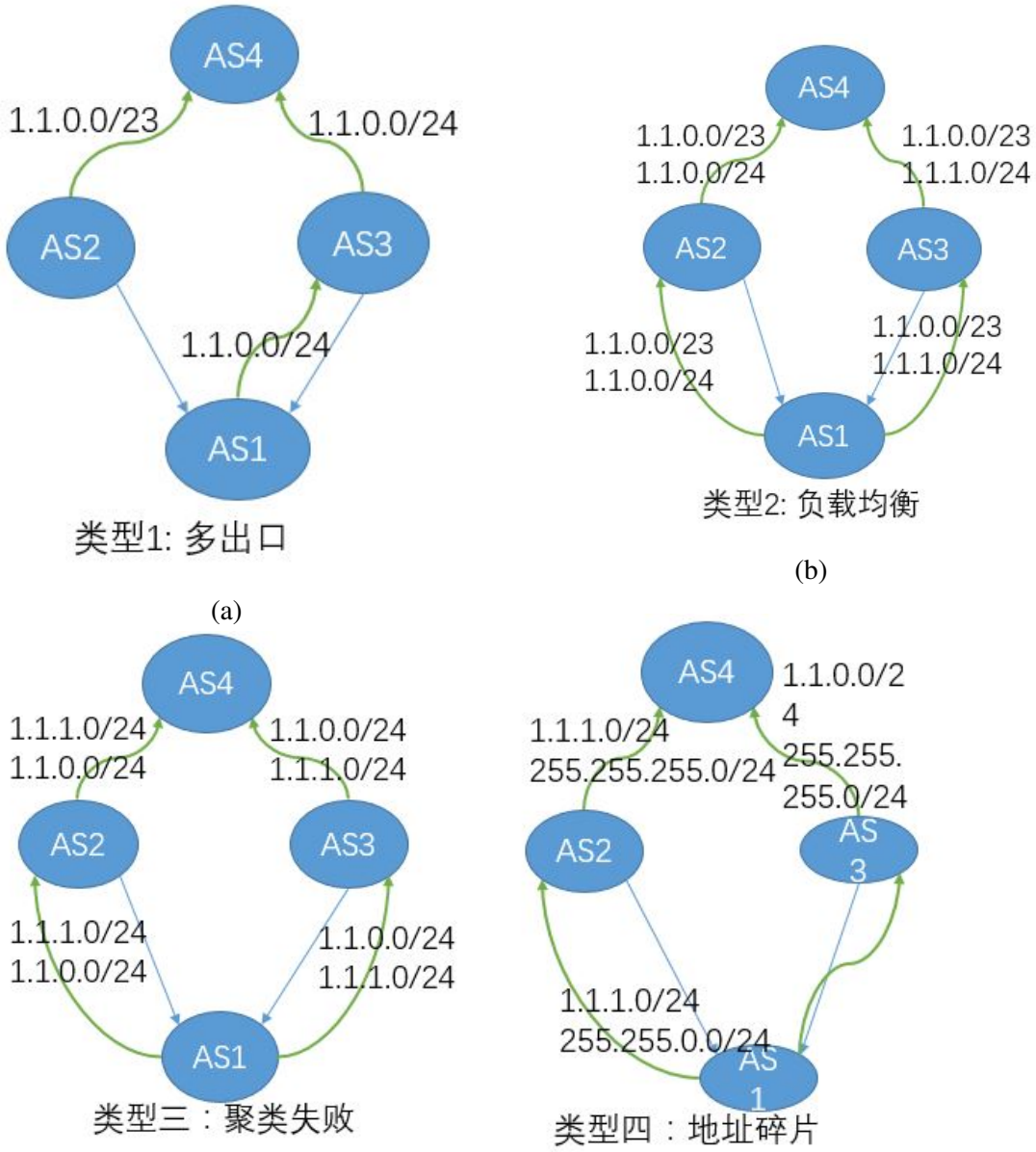


图 4.2　前缀解聚类的类型

4.2.2　基于前缀聚类森林的地址解聚类分类方法

Tian Bu 等人的工作[12] 测量了多出口、负载均衡、聚类失败和地址碎片等对路由表膨胀的贡献和变化趋势。然而，随着互联网的持续发展，新的解聚类的模式可能会出现，为了进一步深入探讨解聚类的类型及其成因，本文提出了一种新的基于前缀聚类森林的地址解聚类类型的分析测量方法。

首先分别考虑地址聚类森林中的每一个树，对于树结构中的每一条边 e，都

意味着一种前缀的解聚类，本文接下来将探讨每一条边 e 对应的地址解聚类类型。本文为多出口、负载均衡、聚类失败和地址碎片设计了相应的判定规则。

如图4.3(a)，当解聚类边 e 对应的父前缀和子前缀对应的源 AS 恰好是供应商客户关系，并且子前缀的源 AS 还存在其他的供应商时，则判断该解聚类为多出口造成的。

如图4.3(b)，当解聚类边 e 对应的父前缀和子前缀对应的源 AS 相同，并且子前缀有共享源 AS 但 AS 路径不同的父结点或者兄弟结点时，则判断该解聚类为负载均衡造成的。

如图4.3(c)，当两个无父结点的前缀可以聚合成一个更短长度的前缀且共享 AS 路径时或者一个父结点和一个子结点且不是负载均衡的案例时，则认为是该解聚类为聚类失败造成的。

如图4.3(d)，当两个无父结点的前缀不能聚合成一个更短长度的前缀且共享 AS 路径时，则认为是一个地址碎片化的案例。

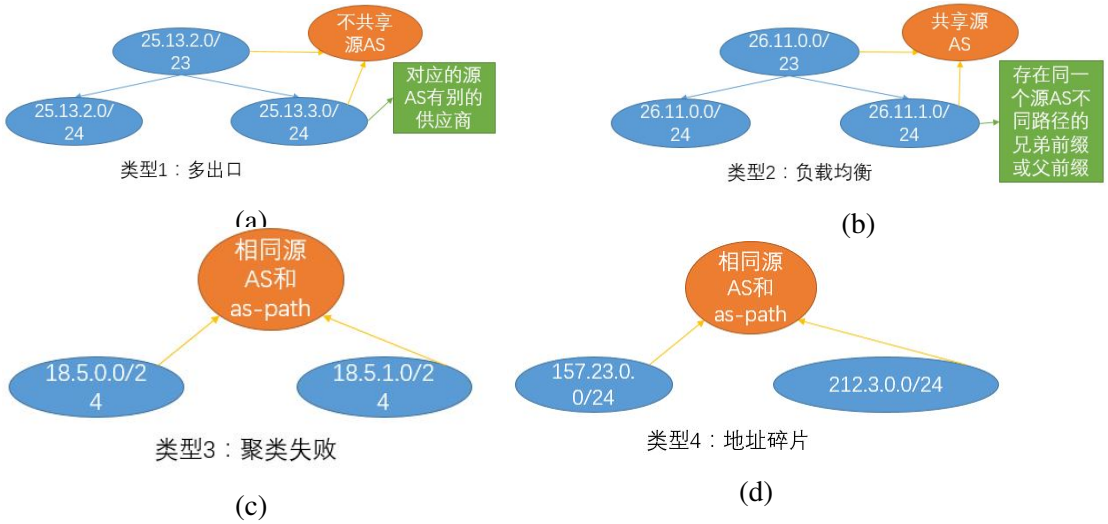


图 4.3　前缀解聚类的四种类型判定法则

4.2.3　自治域商业关系数据集

在具体阐述基于前缀聚类森林的地址聚类分类算法和结果之前，首先说明相关的自治域商业关系数据集。本次采用的数据集来自 Luckie 等人的工作[16]，他们在 CAIDA 官网[28] 上公布了自治域商业关系的测量结果。该数据集共记录了超过 290000 条的自治域商业关系。

4.2.4 基于前缀聚类森林的地址前缀解聚类分类算法

在 Tian Bu 等人提出的算法[12] 解聚类分类算法的基础上，利用前缀聚类森林，按照算法3来对地址前缀的解聚类进行分类。

Algorithm 3 基于前缀聚类森林的地址解聚类分类算法

输入：前缀聚类森林 $G = (V, E)$　**输出**：各类解聚类类型的统计

初始化多出口、负载均衡、聚类失败、地址碎片和其他原因造成的地址解聚类类型数量为 0

for $e \in E$ **do**

　按照4.2.2中设定的规则赋予 e 相应的解聚类类型

　相应的地址解聚类类型数量加 1

end for

按照 AS 路径将 V 的极大前缀子集划分为 $\{S_1, \cdots, S_m\}$[12]

for $k \in \{1, \cdots, m\}$ **do**

　对 S_k 中的前缀进行排序和聚类，循环操作至无法再聚集, 将聚类后前缀集设为 R_k[12]

　聚类失败数加 $|S_k| - |R_k|$，地址碎片数加 $|R_k| - 1$[12]

end for

4.2.5 基于前缀聚类森林的地址前缀解聚类分类结果

利用聚类森林和算法3，采用 2018 年 5 月 16 日收集到的 BGP 静态路由表为例，可以得到表4.1所示的解聚类类型的统计。

表 4.1　不同类型的前缀解聚类分类结果统计（IPv4）

前缀解聚类类型	数量
多出口	33427
负载均衡	140813
聚类失败	229920
前缀碎片	162145
其他	60793

注意到，有 60793 条解聚类类型是无法归入 Tian Bu 等人划分出的四类经典

解聚类类型[12]。为了进一步地分析这超过 60000 条的“其他”解聚类类型，探讨其可能的安全威胁，再对这 60793 条解聚类类型进行进一步的详细分类。

表 4.2 “其他”解聚类类型的进一步详细分类

源 AS 是否相同	源 AS 是否供应商-客户关系	父前缀源 AS 是否在子前缀 AS 路径	数目
否	否	是	6363
否	是	是	13623
否	否	否	40807

根据以上测量结果，可以有以下讨论：

1. 地址分配过程中，存在大量单出口但依然不对地址聚类的现象
2. 源 AS 不同且没有供应商-客户关系的解聚类现象可能是由地址的迁移导致的，也有可能是真实的子前缀劫持
3. 上述前缀的安全性需要更进一步的验证

第 5 章　基于 RPKI 对 BGP 路由表项源的验证

为了进一步地验证 BGP 路由表项的合法性，本章将讨论用基于 RPKI 的方法对 BGP 路由表项进行合法性验证。本章将指出基于 RPKI 对 BGP 路由表项进行验证的方法存在伪无效性的问题。然后本章将进一步对无效前缀进行分类，并指出事实上这些无效前缀中的大部分是一种良性解聚类的表现。最后本章将分析伪无效性的危害。

5.1　RPKI 的验证机制和 ROAs 数据来源

5.1.1　RPKI 的验证机制

正如本文基本概念和基本原理一章所述，RPKI 提供了一个分布式数据库存储 ROAs 用于进行 BGP 路由表项源验证。ROA 可以视为一种将自治域号和其所能合法发布的前缀进行绑定的可信记录。ROA 的格式为：自治域号码、合法前缀、最大长度和信任锚[2]。其中最大长度是指自治域在合法前缀下可以发布的子前缀的最大长度。信任锚是指 ROA 发布的树状结构中信任链的根。按照 RFC6811 中叙述的 ROA 验证路由表项的规则[2]，首先做如下定义：

定义 3：　称一个 ROA 表项覆盖了一个 BGP 表项，若该 ROA 表项中的地址前缀的前缀部分不长于 BGP 路由表项中的前缀部分，且 ROA 表项中的前缀部分和 BGP 路由表项中的相应部分是相互重合的[2]，也即 ROA 表项中前缀对应的地址集包含了 BGP 表项中前缀对应的地址集。

定义 4：　称一个 ROA 表项和一个 BGP 路由表项是匹配的，若该 ROA 表项覆盖了该 BGP 路由表项，并且 ROA 表项对应的 AS 号和 BGP 路由表项中的源 AS 号相同，并且 BGP 路由表项中前缀的长度不长于 ROA 表项中设定的最大长度[2]。

用 ROA 数据验证 BGP 路由表项采用以下验证规则:[2]

1. 若路由表项中的某一项没有被任何 ROA 表项覆盖，则验证结果为**未知**
2. 若路由表项中的某一项被某个 ROA 表项覆盖，但没有被匹配，则验证结果为**无效**
3. 若路由表项中的某一项被某个 ROA 表项匹配，则验证结果为**有效**

5.1.2 ROA 表数据源和基于此的 BGP 路由表项验证算法

本项目使用 RIPE-NCC 提供的 rpki-validator[29] 收集分布在全球的 ROA 数据库中的 ROA 数据，目前已经发布的 ROA 数据共有超过 51000 多条。

设 ROA 表中的表项数目为 m，如果直接对于大小为 n 的 BGP 路由表中的每一个路由表项都直接遍历 ROA 表进行验证，则时间复杂度将会是 $O(mn)$，实验结果表明，这样的验证方式在时间上是极其浪费的，无法在数据收集一节中介绍到的服务器上在一个小时内给出验证结果。本文提出了一种时间复杂度更低的 BGP 路由表验证算法，如算法4。该算法的复杂度来自于排序和对队列的移动，总的复杂度为 $O(m+n+n\log n+m\log m)$，较直接验证有了大幅改善。

5.1.3 BGP 路由表项的验证结果

以 2018 年 5 月 16 日收集到的静态 BGP 路由表数据和 ROA 数据表为例，验证得到统计结果如表5.1。

表 5.1 用 ROA 数据验证 BGP 路由表项合法性的结果统计

验证结果	条数	百分比
未知	635412	90.87%
无效	4949	0.71%
有效	58931	8.43%

由表中的统计结果可以看出当前基于 RPKI 的 BGP 路由表项 AS 源验证的方法存在如下问题：

1. 90% 左右的 BGP 路由表项得不到验证，可见当前 ROA 的数量远远没有达到能够保卫全球网络中大部分前缀安全的要求。
2. 发现了超过 4000 条的无效前缀，虽然 BGP 路由劫持的新闻时有报道，但是同时出现那么大范围的路由劫持可能性是很小的，本文接下来将重点对这些无效前缀做进一步的分析。

5.2 无效前缀是否真的“无效”？

上一节的结果表明，基于 RPKI 的 BGP 路由表项源验证方法得到的验证结果中存在超过 4000 条的无效前缀，然而考虑到现实网络中不太可能在同一时间

Algorithm 4 基于 RPKI 中的 ROA 数据进行 BGP 路由表验证的算法

输入：BGP 路由表 $\{b_1, b_2, \cdots, b_n\}$，ROA 数据列表 $\{r_1, r-2 \cdots, r_m\}$　**输出**：对 BGP 路由表表项的验证结果

初始化未知前缀集 U$\leftarrow \emptyset$，无效前缀集 InV$\leftarrow \emptyset$，有效前缀集 V$\leftarrow \emptyset$

对 BGP 路由表按照前缀对应地址集合的首地址进行排序，将结果从小到大仍然记为 $b_1, b_2, \cdots, b_n$，按照算法1的方法取出 ROA 数据列表中前缀为极大前缀的表项，记为 $q_1, q_2, \cdots, q_k$

初始化 $i \leftarrow n$，$j \leftarrow k$

while $i > 0$ **do**

  if b_i 前缀首地址小于 q_j 的前缀首地址 **then**

    if $j > 0$ **then**

      $j \leftarrow j - 1$ 并继续循环

    else

      for l 从 i 到 1 **do**

        U 中加入 b_i

      end for

      结束 WHILE 循环

    end if

  else

    if b_i 被 q_j 包含 **then**

      利用被 q_j 和 q_j 覆盖的 ROA 表项验证 b_i，并据验证结果将 b_i 加入 InV 或 V

    else

      将 b_i 加入 U

    end if

    $i \leftarrow i - 1$

  end if

end while

出现这样大范围的地址劫持，因此基于 RPKI 的无效验证结果是值得怀疑的。

5.2.1 一个伪无效性的个案分析

Orange S.A.（前法国电信）是一家多国通信服务供应商[30]，结合 CAIDA 提供的 AS 到组织的映射关系[31]，可以发现其控制了 AS3215 和 AS28708，在 ROA 数据库中这两个 AS 对应了两条 ROA 记录，如表5.2

表 5.2 Orange S.A. 所控制的 AS3215 和 AS28708 所对应的 ROA

自治域号	前缀	最大长度	信任锚
3215	194.3.0.0/17	24	RIPE
28708	80.10.0.0/19	24	RIPE

用基于 RPKI 的 BGP 路由表表项源验证方法可以检测到由于表5.2中的相应 ROA 覆盖所导致的无效 BGP 路由表项的源 AS 和前缀对，如表5.3。

表 5.3 Orange S.A. 控制的 AS3215 和 AS28708 对应的 ROA 覆盖所导致的无效 BGP 路由源 AS 和前缀对

自治域号	前缀
34444	194.3.118.0/24
8376	80.10.24.0/22
8376	80.10.20.0/22
8376	80.10.16.0/20
8376	80.10.16.0/22
8376	80.10.8.0/21

为了核实表5.3中的无效前缀是否真的是 AS34444 和 AS8376 的错误配置或者恶意攻击，本文作者向 Orange S.A. 的相关技术人员邮件反映了无效前缀的问题, 对方技术人员就上述问题做出如下回应：

1. AS34444 是对方的一个客户自治域，Orange S.A. 尚未在 RPKI 的体系中为该 AS 宣布的 194.3.118.0/24 前缀生成相应的 ROA，导致相应的 BGP 路由表项被判断为无效。
2. AS8376 对应的无效前缀是由 ROA 数据库中的记录缺失导致的，对方已经在数据库中补全了相应的缺失记录。

从对方的回复来看，被判断为无效的 BGP 路由前缀事实上是合法的前缀。从这个案例中可以总结出基于 RPKI 的 BGP 路由验证机制的问题：

1. ROA 的签署和被写入数据库与地址前缀分配之间的不同步可能会造成某段前缀在某一个时期被误判为“无效”。
2. 网络操作者为一个 AS 生成了 ROA，然后将一块子前缀分配给某个客户，而没有为客户生成相应的 ROA，客户可能通过别的出口再宣称该块地址，这个过程可能导致该前缀被误判为“无效”。

5.3 无效前缀的进一步分类

上一节的个案揭示了伪无效性在基于 RPKI 的 BGP 路由表项验证方法中的存在，为了更深入地分析无效前缀，本文将无效前缀划分为以下四类：

1. 无效客户前缀
2. 无效自身前缀
3. 无效迁移前缀
4. 无效劫持前缀

下面通过表5.4对上述四种无效前缀进行定义。

表 5.4 四类无效前缀的定义

无效前缀类型	源 AS 是否与 ROA 的 AS 相同	ROA 的 AS 是否与源 AS 为供应商-客户关系	是否为从 ROA 的 AS 迁移到源 AS
无效客户前缀	否	是	/
无效自身前缀	是	/	/
无效迁移前缀	否	否	是
无效劫持前缀	否	否	否

利用 CAIDA 提供的自治域商业关系数据[32]，可以建立自治域的商业关系图谱，如图5.1，用单向边表示供应商-客户关系（其中箭头指向客户），双向边则表示对等关系。

依然以 2018 年 5 月 16 日收集到的 BGP 静态路由表数据和 ROA 数据列表为例，5.1.3节已经给出了利用 ROA 数据验证 BGP 静态路由表数据得到的结果，在此基础上，对其中的无效前缀再次划分为本节定义的四类无效前缀中。

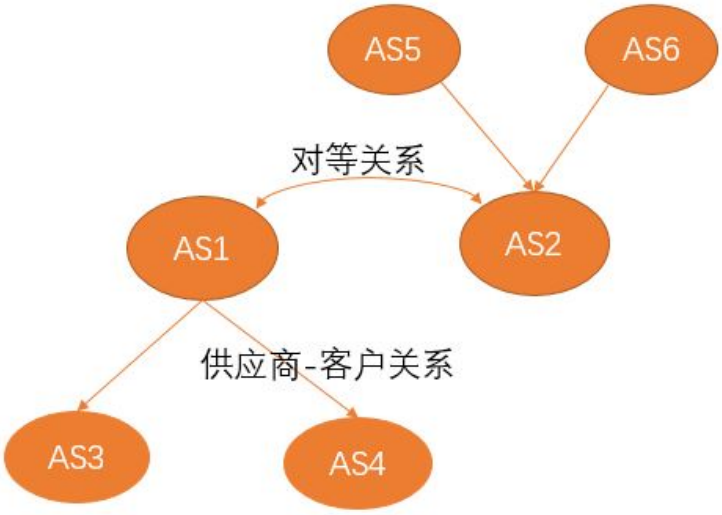


图 5.1　自治域商业关系图谱示意

在分类之前，首先需要对 BGP 静态路由表数据做一些预处理，按照规定，自治域不可以使用 16 位自治域号的首位和尾位以及 32 位自治域号的尾位，以及按照 RFC5398[33] 的规定，16 位自治域号中的 64496 到 64511 和 32 位自治域号中的 65536 到 65551 是保留给文件使用的, 又按照 RFC6996[34] 的规定，16 位自治域号的 64512 到 65534 和 32 位自治域号的某一段被保留为私用[21]。在 BGP 路由表数据中可以发现少量源 AS 号为以上 AS 的现象，这可能是由内部自治域号泄露所导致的，在预处理过程中需要将这一部分 BGP 路由表项滤除。

预处理之后将 BGP 无效路由表项分为表5.4中定义的四种，最终得到的分类结果的统计如表5.5。

表 5.5　无效 BGP 路由表项的进一步分类结果统计

无效前缀类型	数量	百分比
无效客户前缀	913	18.45%
无效自身前缀	2560	51.73%
无效迁移前缀或无效劫持前缀	1476	29.82%

5.4　伪无效性的可能成因

正如5.2.1节中所指出的，为客户分配地址之后没有及时发布 ROA 而造成的 BGP 路由信息宣布与 ROA 的不同步性可能会导致伪无效前缀。下面给出本文对

于伪无效 BGP 表项的严格定义：

定义 5：若某时刻利用 ROA 列表 rl 验证 BGP 表项 b 的结果为无效，但实际上通过 b 所提供的自治域路径可以正常访问到相应的合法 IP 地址，则称该 BGP 路由表项 b 相对于 ROA 列表 rl 是伪无效的。

本文指出一些原本良性的前缀解聚类会导致基于 RPKI 的 BGP 路由源验证方法得到伪无效前缀的判断并系统总结出了以下几类可能造成伪无效性的场景。

负载均衡 如图5.2(a)所示为负载均衡所导致的伪无效前缀，AS1 为了做负载均衡的目的，对外宣称了超过最大合法长度的前缀，从而导致了 AS4 利用 AS1 所对应的 ROA 进行路由表项验证时产生了伪无效表项。

多出口 如图5.2(b)所示，AS2 下的客户 AS1 被分配了一段自身拥有前缀的子前缀，AS1 将该段子前缀通过另外一个出口 AS3 传递给 AS4，AS4 利用 AS2 对应的 ROA 进行验证可以得出 AS1 发布的前缀的无效性，但事实上这段 AS1 宣称的地址是有效的，也即多出口导致了伪无效前缀的产生。

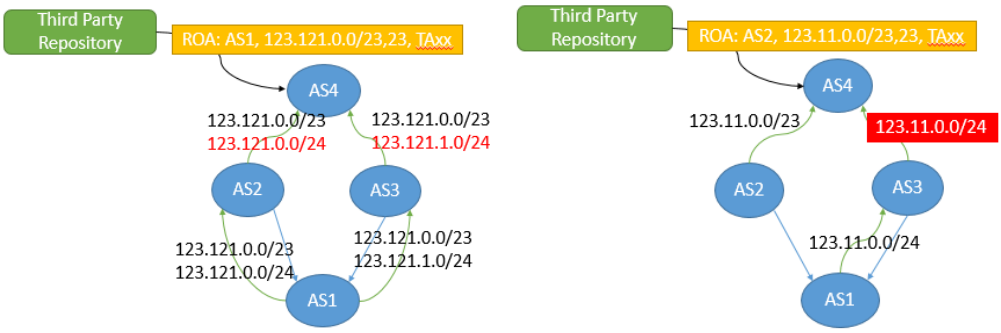
单出口中地址聚类失败 单出口中供应商 AS 应该将分配给客户 AS 的子前缀与自身 AS 宣称的父前缀相聚类，但现实情形下，出于吸引流量等方面的考虑，供应商 AS 可能保留客户 AS 宣称的长前缀，从而由自身对应的短前缀 ROA 覆盖导致客户长前缀被判断为无效导致伪无效前缀的产生[20]。

本自治域聚类失败 另一种可能是本自治域宣称的路由策略相同的前缀没有成功的聚类，导致其长度超过了 ROA 中规定的最大长度，进而发生了伪无效前缀的判断。

地址迁移 如图5.3为一个地址迁移导致伪无效前缀的场景，AS2 下拥有一块有相应 ROA 保护的前缀，处于某种原因，例如 IP 用户的地理位置变更、自治域之间进行了地址买卖交易，该前缀的一段子前缀从 ROA 对应的 AS 中迁出而迁入了另一个 AS，而该段子前缀没有相应的验证 ROA，从而导致伪无效前缀的判断。

5.5 伪无效性的危害

Y. Gilad 等人调研了 100 多个网络的实际运营者，结果显示超过 5% 的运营者会直接丢弃用 ROA 验证得到的无效前缀，另外还有超过 10% 的运营者会



(a) 负载均衡导致的伪无效前缀示例　　(b) 多出口导致的伪无效前缀示例

图 5.2　伪无效前缀产生场景举例

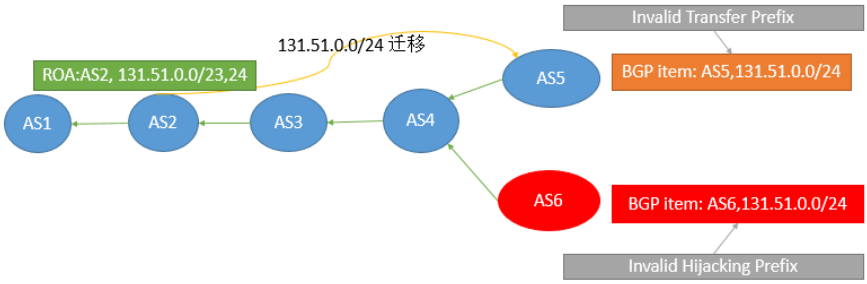


图 5.3　伪无效前缀产生场景（续）

给 ROA 验证得到的无效前缀赋予更低的偏好[20]。同时 Y. Gilad 等人还发现超过 30% 的运营商不采用基于 RPKI 的 BGP 路由验证方式的原因是可能的失去合法 IP 的连接[20]。

结合 Y. Gilad 的调研结果，可以看到作为一种虚警现象，伪无效前缀的存在将产生以下危害：

1. 大大降低了 RPKI 的可信任度，从而影响到 RPKI 的采用率，进而影响到 RPKI 的整体部署；
2. 实际采用 ROA 对 BGP 路由前缀进行验证并采取赋予无效 BGP 路由表项更低偏好的自治域将不能正常顺利的实现负载均衡、多出口等功能；
3. 直接丢弃无效前缀的自治域可能会因伪无效前缀丢失到特定合法 IP 地址的连接。

事实上，结合 Y. Gilad 等人的调研结果[20]，可以发现在一定程度上，存在图5.4中所示的恶性循环，由于 RPKI 的部分部署，一部分合法的 BGP 路由表项没有得到相应的 ROA 保护，导致了伪无效前缀的产生，信任伪无效前缀的“无效”性可能会带来对正常路由策略（良性解聚类）的干扰、甚至导致“断网”，一

部分网络运营者出于这方面的担忧会对 RPKI 采用保守的态度，制约了 RPKI 的实际采用率，低采用率又将降低签发 ROA 的安全收益，当签发 ROA 的安全收益不足以超过相应成本时，网络运营者就会失去签发 ROA 的动机[20]。而这些反过来，ROA 的低签发率又会继续导致伪无效前缀的存在。

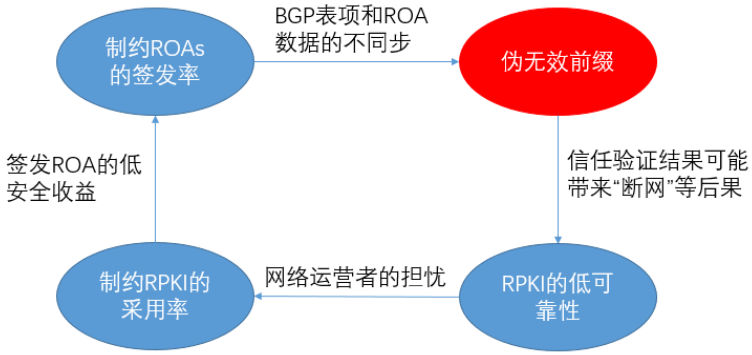


图 5.4　RPKI 部署过程中的恶性循环

作为一种应对措施，对伪无效前缀的检测和发布将有效地帮助自治域的操作者设定自己的路由策略，避免因丢弃伪无效前缀的路由表项而造成损失，从而推动 RPKI 的部署。

5.6　伪无效前缀的检测与分类

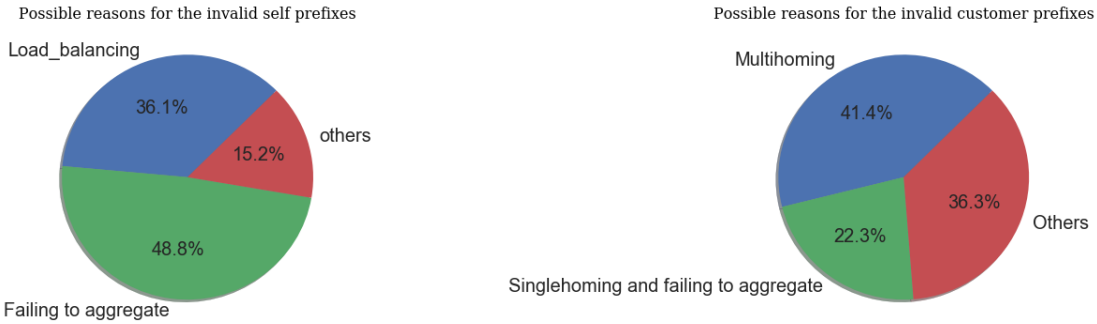
5.6.1　基于前缀聚类森林的检测方法

4.2.5节给出了前缀聚类森林中每一条边对应的解聚类类型，利用无效前缀对应前缀聚类森林中结点及其父结点的解聚类类型可以检测伪无效前缀并依据其成因对其进行分类。下面给出本文对伪无效前缀的检测规则：

1. 若无效前缀与致其无效的 ROA 前缀共享源 AS，在前缀聚类森林中存在父结点，且父结点到子结点的解聚类类型为负载均衡，则将该无效前缀分类为负载均衡导致的伪无效前缀；
2. 若无效前缀与致其无效的 ROA 前缀对应 AS 是客户-供应商关系，在前缀聚类森林中存在父结点，且父结点到子结点的解聚类类型为多出口，则将该无效前缀分类为多出口导致的伪无效前缀；
3. 若无效前缀与致其无效的 ROA 前缀对应源 AS 是客户-供应商关系，且

ROA 前缀对应的源 AS 是唯一无效前缀对应源 AS 的唯一供应商，则判断为单出口聚类失败导致的伪无效前缀。

继续以 2018 年 5 月 16 日收集到的 BGP 静态路由表数据和 ROA 数据为例，得到的对无效自身前缀的形成原因如图5.5(a)，得到的对无效客户前缀的形成原因如图5.5(b)。



(a) 无效自身前缀的形成原因　　(b) 无效客户前缀的形成原因

图 5.5　无效前缀的形成原因

如图5.6所示为无效自身前缀在不同自治域中的分布，可以看到近 80% 的无效自身前缀是由排名前 50 的自治域造成的，这也就意味着只要互联网上少量宣称了比自身 ROA 所规定前缀最长长度更长的自治域申请将 ROA 的最长长度调长或者签署新的 ROA 即可大幅度减少无效自身前缀的数量。

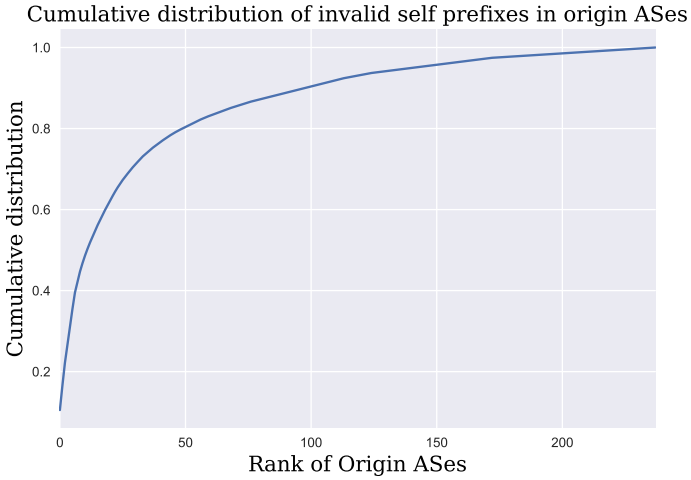


图 5.6　无效自身前缀在源 AS 中的累积分布

5.6.2 数据平面的应用测量

上一节探讨了利用前缀聚类森林来检测和分类伪无效前缀，但是对于无效迁移前缀和恶意的无效劫持前缀，很难通过控制平面的数据分析来区分二者。为此，本节将通过数据平面的应用测量来区分二者。正如 X. Shi[10] 等人指出的，如果一台路由器采用了（子）前缀劫持者所发布的路由信息，那么这台路由器通常无法与该段（子）前缀下的 IP 建立正常连接。本文进一步得到以下基于直观的判断：

- 通常而言，地址迁移之后会在相应的 IP 地址上建立起正常的网络应用，比如 http 服务器、邮件服务器、https 服务器等；
- 前缀被劫持之后，相应的网络服务将无法再正常地访问。

基于以上直观的理解，本文设计了一种检测由于地址迁移而造成的伪无效前缀的方法，步骤如下：

1. 利用 zmap[35] 等端口扫描工具来扫描事先选定的端口；
2. 用端口扫描的结果作为该段前缀的 IP 应用特征；
3. 利用前缀的 IP 应用特征区分无效前缀是地址迁移造成的伪无效前缀还是（子）前缀劫持。

本文采用如下简单的特征来检测伪无效前缀，设无效前缀长度为 l，则当式 (5-1) 满足时，即将该无效前缀判断为伪无效前缀。

$$\frac{\sum_{i\in\text{扫描端口集}}\text{端口}i\text{返回的响应数}}{\text{扫描的端口总数}\times 2^{32-l}} > \text{阈值} \tag{5-1}$$

其中上式左边可以理解为端口扫描的平均命中比率，当端口扫描的命中率比较高时，说明该段前缀下的 IP 上已经建立起了丰富的应用，因而更有可能是一种地址迁移的表现。当选择阈值为 0 时，扫描端口为 80 端口（对应 http 服务），可以从 1476 例无效迁移前缀或者无效劫持前缀中检测出 843 例发现了正常应用的由于地址迁移而导致的伪无效前缀。

进一步地，加入表5.6所示的常见端口[36] 进行扫描。

如图5.7所示为将无响应的 IP 去除后在 IPv4 中迁移无效前缀或者劫持无效前缀的各端口平均扫描命中率的分布。设定阈值为 0.001，可以得到 866 例伪无效前缀，比只扫描 80 端口和选择 0 为阈值多出了 23 例。

为了能够说明上述基于数据平面测量结果的可信性，利用 CAIDA 提供的

表 5.6　IPv4 端口扫描选用的常见端口及其对应的应用

端口号	相应的服务
20	FTP 数据传输
22	ssh 登录，文件传输等
23	无加密文本通信
25	SMTP
37	时间协议
53	DNS
179	BGP
443	HTTPS
547	DHCPv6 服务器

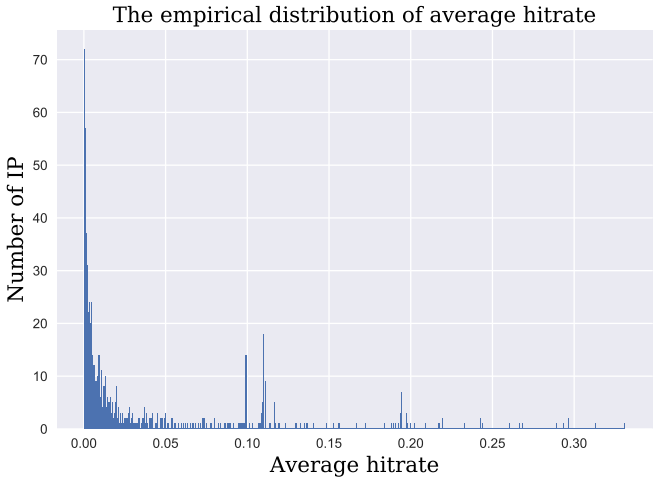


图 5.7　无效迁移前缀或者无效劫持前缀 IPv4 端口扫描平均命中率的分布

AS 到组织的映射关系可以发现 866 例伪无效前缀中事实上有 47 例来自于同一个组织，也就是说无效前缀的发布 AS 和致其无效的 ROA 的持有 AS 事实上在同一个组织的控制之下。进一步地，可以做出无效前缀的发布 AS 的组织和致其无效的 ROA 的持有 AS 的组织不同时相应的后者的累积分布如图5.8，可以看到 75% 左右的通过数据平面的测量得到的无效迁移前缀事实上是由所有相关组织中的 25% 左右的组织所造成的。进一步地对这些少量组织进行调查，可以发现其其实是网络服务的供应商，也即极有可能是这些供应商将自身持有的部分 IP 分配给了自己的客户，但是相应地 ROA 却没有给客户签署，进而导致了地址迁移之后伪无效前缀的产生。

为了进一步验证利用端口扫描结果判断得到的地址迁移所造成的伪无效前缀的判断，利用 ipwhois 提供的相关注册信息[37]，可以进一步发现在检测出的 866 例由地址迁移造成的伪无效前缀中，事实上有 742 例包含了相应的注册信息，这进一步说明了基于数据平面的应用端口扫描测量方法得到的结果的可信性。

另一方面，也可以发现有 457 例的迁移之后的前缀没有给出任何端口扫描的响应，这一部分将被加入疑似劫持列表。

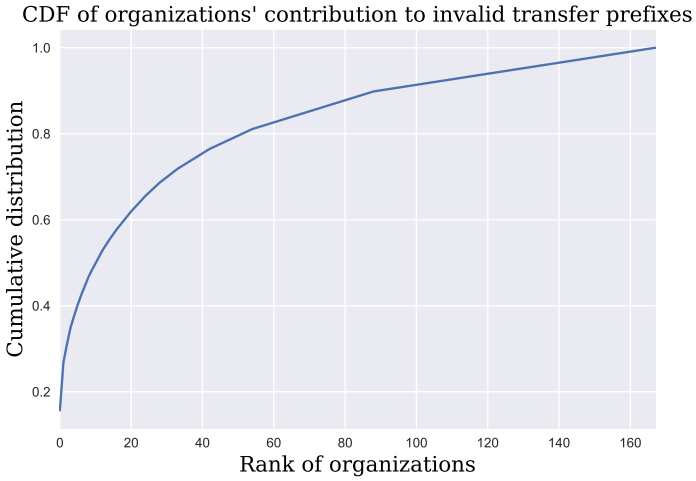


图 5.8　无效迁移前缀迁出 AS 在组织上的分布

5.6.3 时间维度上的路由源稳定性分析

通常而言，在互联网上长时间稳定存在的路由说明该路由的可用性和可靠性，劫持路由往往在一段时间后会被网络运营者发现，因而可以通过对路由的历史更新记录来判断一个路由是否是可信的。

分析表明，由 RPKI 验证以及端口扫描分析得到无效迁移前缀对应的（前缀，路由源）对事实上早在 2 月 28 日的路由数据中就已经存在了，这些前缀对应的路由表项事实上十分稳定的路由表项，这从另一个角度再次佐证了上文得到的伪无效性的判断。如表5.7所示为利用 5 月 16 日的路由数据得到的各类伪无效（前缀，源自治域）对从 2 月 28 日的路由表中开始的稳定存在情况。可以看到对于其他类型的（伪无效前缀，路由源）对，大部分也都是在时间维度上表现得非常稳定，也从另一个角度佐证了伪无效性的判断。

表 5.7　5 月 16 日的路由数据得到的各类伪无效（前缀，源自治域）对在 2 月 28 日的路由表中的存在情况

伪无效前缀成因	表项数目	在 2 月 28 日的路由表中相应(前缀, 路由源) 对即已开始稳定存在的数目	在 2 月 28 的路由表中相应（前缀, 路由源)对即已开始稳定存在的数目所占百分比
负载均衡	923	770	83.42%
本自治域聚类失败	703	684	97.30%
多出口	378	355	93.92%
单出口下与客户前缀聚类失败	204	177	86.76%
地址迁移	866	738	85.22%

第 6 章　RPKI 验证机制的可能改进措施的探讨

如前文所述，基于 RPKI 的 BGP 路由源验证机制存在着由于 ROA 相对地址使用者在自治域中切换的滞后性以及自治域前缀发布策略的灵活性导致的伪无效前缀的问题，这些问题暗示着当前基于 RPKI 的 BGP 路由验证机制层面的问题。因此从 BGP 路由合法性验证机制层面探讨可能的改进措施是非常有必要的。例如，Y. Gilad 等人探讨了基于 RPKI 的 BGP 路由源验证机制中 ROA 记录中的最大长度对 RPKI 的危害并建议网络操作者不要使用最大长度项[19]。本文接下来将基于前文中对伪无效前缀等的测量与分析在验证机制层面提出改进措施。

6.1　从验证（源，前缀）对到验证（路径，前缀）对

基于 RPKI 的 BGP 路由前缀源验证机制基于如图6.1所示的信任模型。本文

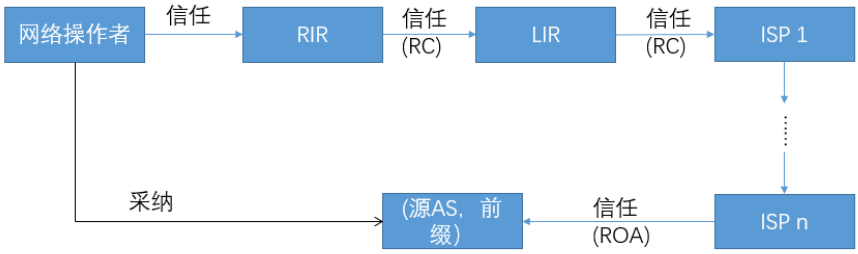


图 6.1　基于 RPKI 的 BGP 路由源验证机制的信任模型

一个关键的观察是当一个自治域采纳了一条 BGP 路由表项，也就意味着该自治域信任该项路由表项的合法性。本文接下来的讨论将基于基本假设1。

假设 1：　ROA 的持有 AS 不会采纳并对外输出劫持该 ROA 所覆盖前缀的 BGP 路由表项。

为了说明以上假设的合理性，分以下两种情况讨论。

1. 劫持前缀实际上仍然为 ROA 的持有 AS 所控制，此时 AS 会优先选择内部的合法前缀。

2. 劫持前缀已经不被 ROA 的持有 AS 所控制，因为 AS 事实上是由相应的网络运营者控制的，网络运营者应当具有判断自身控制的 AS 对应 ROA 覆盖的前缀是否为劫持的充分的信息，因此 ROA 的持有 AS 有充分的能力甄别劫持前缀，进而对劫持前缀不予采纳和输出。

本文提出了一种改进基于 RPKI 的 BGP 路由验证的 AS 路径验证方法，其判定规则如下。

未知 若 BGP 路由前缀没有被任何 ROA 前缀覆盖，则验证结果判断为未知；

源有效 应用源验证方法可以得到有效的结果；

路径有效 若存在 ROA 记录，使得其前缀覆盖了 BGP 路由前缀，且该 ROA 记录所对应的 AS 在 BGP 路由前缀所对应的自治域路径上，则判定为路径有效；

无效 非未知，且非路径有效，则判断为无效。

基于假设1和以上设计的 AS 路路径验证方法，可以得到命题2。

命题 2：在假设1成立的条件下，基于 RPKI 的路径验证方法相较源验证方法不会降低前缀劫持的检测率。

证明 假设 BGP *劫持*前缀对应的自治域路径为 $a_1, a_2, a_3, \cdots, a_n$，并且基于 RPKI 的源验证方法验证得到的该路由表项是无效的，则 $\forall a_i, i \in \{1, 2, 3, \cdots, n\}$，都不存在相应的 ROA 覆盖该 BGP 前缀，否则即违背了假设1，故该 BGP 路由表项不是路径有效的，进而是无效的。故路径验证方法可以检测出源验证方法能检测出的所有劫持前缀，因此基于路径验证的方法相较源验证方法不会降低前缀劫持的检测率。 □

显然地，可以得到命题3和命题4。

命题 3：基于 RPKI 的路径验证机制可以避免单出口聚类失败和部分地址迁移造成的伪无效前缀。

命题 4：源有效的 BGP 路由表项一定是路径有效的。

从命题2，3和4可以看到，在假设1成立的条件下，基于 RPKI 的路径验证机制相比基于 RPKI 的源验证机制可以在不降低劫持检测率的条件下降低虚警率。

依然以 2018 年 5 月 16 日收集到的 BGP 静态路由表和 ROA 记录数据为例，得到的无效前缀中路径有效的前缀占比如图6.2。可以看到接近 70% 的无效 BGP

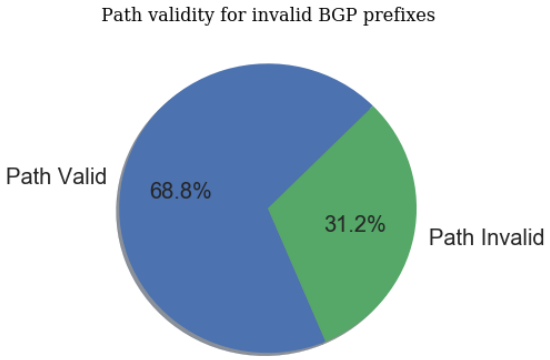


图 6.2　无效前缀的路径有效性

前缀其实是路径有效的，并且其中 91.63% 的路径有效的源无效前缀是在 2 月 28 日的路由表中就已经稳定存在的，说明了路径有效概念的合理性和对虚警率的大幅降低。

6.2　验证请求机制的引入

5.6.2节讨论了利用数据平面应用测量的特征来对 BGP 路由前缀来进行分类的方法。本节探讨在验证机制层面对原有的 BGP 路由验证机制进行补充。基于 RPKI 的 BGP 路由源验证机制需要依靠第三方提供的验证证书来进行验证，路由层面和证书层面的不同步性和不一致性可能会导致伪无效性。上一节中提到的（路径，前缀）对的验证机制可以避免由于单出口聚类失败和部分地址迁移所造成的虚警。但是，在如图6.3[①]所示的场景中，AS3 和 AS4 由于缺少关于地址迁移的信息，无法判断 AS5 和 AS6 发布的无效 BGP 路由信息是由地址迁移造成的还是由非法前缀劫持造成的，因而在接收到相应的 BGP 路由表项之后，无论是丢弃还是采纳，都存在“断网”的风险。

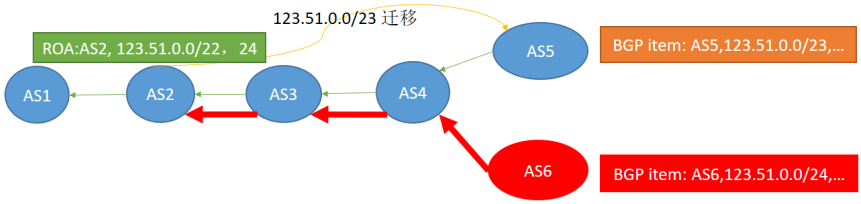


图 6.3　地址迁移和劫持难以区分的场景

① 本图中带箭头的线表示 BGP 路由表项的传播方向，虚线表示 IP 前缀的迁移方向。

假设 AS3 和 AS4 采用了保守的策略，即采纳了 AS5 和 AS6 发布的 BGP 路由表项，并将其传播到了 AS2。此处本文引入一个关键性的反向传播策略：

反向传播策略 在原有的基于 RPKI 的验证机制中，ROA 的持有者在接收到 BGP 路由表项之后，可以检查自身持有前缀是否有被劫持，如判断被劫持，则可以反向发起 ROA 的验证请求。

引入反向传播策略之后，对于 AS5 发布的迁移之后的前缀，由于 AS2 对于迁移地址有充分信息，因而将不会就此 BGP 路由表项反向发起 ROA 的验证请求，而对于攻击者 AS6 发布的劫持前缀，当 AS2 听到了 AS6 发布劫持前缀之后，则会反向地发起 ROA 的验证请求来帮助 AS3 和 AS4 区分地址迁移和前缀劫持，如图6.4所示。

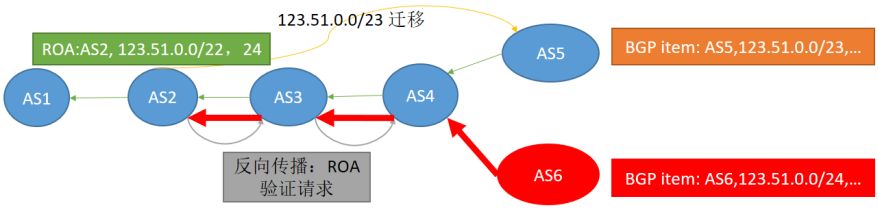


图 6.4　反向传播策略示意

验证请求的具体工作机制如下：

1. 每一个自治域的边界路由器可以利用某前缀下的独有 IP 沿着路由信息传播的反向路径发送验证请求的数据包
2. 数据包中包括了验证请求的发起 AS 号和请求验证的对象前缀
3. 参与验证请求机制的自治域收到验证请求的包之后，将会利用 ROA 数据信息，结合请求 AS 和请求对象前缀的信息，决定是否丢弃相关路由表项并进一步发起反向验证请求

验证请求机制的基本思想在于利用自治域之间的相互协作来检测（子）前缀劫持。下面指出验证请求机制的引入将具有以下特点：

1. 验证请求机制可以让自治域主动维护自身的前缀免于劫持，同时也有利于保障路由安全，因而可以给自治域提供采用该机制的动机。
2. 通过类似于 Traceroute[38] 的机制，当验证请求的发起者向疑似劫持的 IP 发送包含验证请求内容的包时，可以避免因为反向传播路径上尚未采用该机制的自治域的阻断导致的无法将验证请求传播到整个路径上，因而该机制

即便只是部分地部署，采用的自治域也可以得到安全上的收益，这也就意味着该机制是可以渐进式部署的。

3. 验证请求机制的引入将可以帮助采用该机制的自治域区分子前缀迁移和子前缀劫持。

第 7 章　IPv6 情形下的相应分析结果

随着 IPv4 地址的逐渐枯竭，IPv6 将成为互联网的未来。本章将阐述前文的测量方法应用到 IPv6 的结果，并重点分析 IPv6 和 IPv4 中测量方法或结果有所不同的部分。

7.1　IPv6 静态路由表的简单统计分析

如图7.1所示，为 IPv6 前缀的地址解聚类父/子前缀的分布，结合图3.3可以发现类似于 IPv4 中的关于地址聚类的结论，此处不再赘述。只不过相应的 IPv6 主要解聚类成了/48 的前缀。

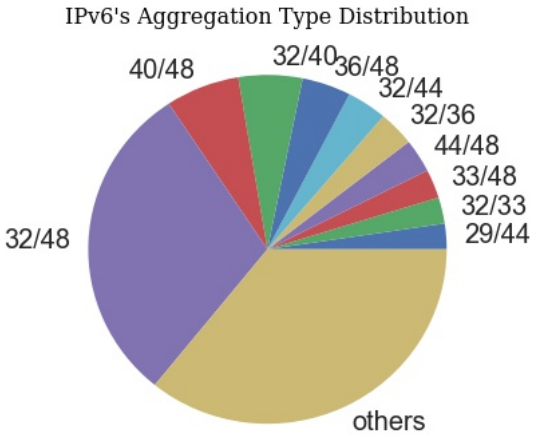


图 7.1　前缀聚类长度的统计（IPv6）

7.2　基于前缀聚类森林的地址前缀解聚类分类结果

和 IPv4 的情形一样，采用 2018 年 5 月 16 日收集到的 BGP 静态路由表，如表7.1所示为地址解聚类类型的分类结果。

和 IPv4 中的相应测量结果相比，IPv6 中的测量结果展现出了类似的规律。聚类失败和前缀碎片仍然是地址膨胀的主要成因。和 IPv4 的情形一样，可以发

表 7.1　不同类型的前缀解聚类分类结果统计（IPv6）

前缀解聚类类型	数量	百分比
多出口	2024	5.69%
负载均衡	4906	13.80%
聚类失败	15959	44.90%
前缀碎片	8455	23.79%
其他	4198	11.81%

现在 IPv6 中也存在不能归入到典型的四类解聚类类型的情形，并且占比较大。

采用和 IPv4 情形下相同的方法，对其他类的解聚类类型进一步地进行分类，可以得到如表7.2中所示的分类结果。可以看到在 IPv6 中同样存在大量源 AS 不

表 7.2　“其他”解聚类类型的进一步详细分类 (IPv6)

源 AS 是否相同	源 AS 是否供应商-客户关系	父前缀源 AS 是否在子前缀 AS 路径	数目
否	否	是	1145
否	是	是	1087
否	否	否	1966

同，且没有供应商-客户关系，父前缀源 AS 又不在子前缀自治域路径上的情形，这一类 BGP 路由表项是存在很大的安全隐患的。

7.3　BGP 路由表项源验证的结果

和 IPv4 中相同，用 2018 年 5 月 16 日收集到的 BGP 静态路由表数据和 ROAs 记录数据用相同方法进行验证分析，可以得到验证结果的统计如表 7.3。

表 7.3　基于 RPKI 的 BGP 路由源验证结果 (IPv6)

验证结果	数量	比例
未知	44596	87.02%
无效	347	0.68%
有效	6304	12.30%

从统计结果来看，基于 RPKI 的 BGP 路由源验证方法在 IPv6 中同样存在着以下问题：

1. 87.02% 的 IPv6 的前缀是没有得到 ROA 的保护，也即 RPKI 在 IPv6 中的部署事实上也是非常不足的。
2. 在 IPv6 的验证结果中出现了 347 例的无效路由前缀，和 IPv4 中一样，在同一时间出现那么多例的 IPv6 地址前缀劫持可能性是很小的，因此这部分验证结果也是值得怀疑的。

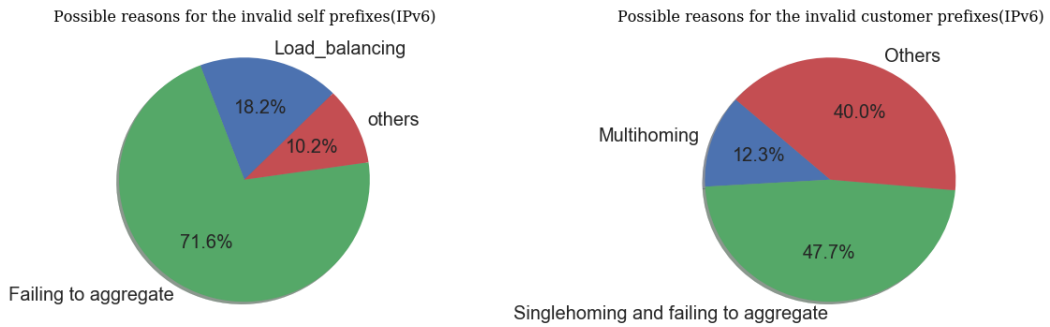
7.4 无效前缀的进一步分类以及相应的成因分析

利用和 IPv4 中相同的规则和算法将无效前缀划分为无效客户前缀、无效自身前缀、无效迁移前缀和无效劫持前缀。得到的分类结果如表7.4。

表 7.4 无效前缀的进一步分类结果 (IPv6)

前缀无效类型	数量	比例
无效客户前缀	65	18.73%
无效自身前缀	217	62.54%
无效迁移前缀或无效劫持前缀	65	18.73%

进一步考察每一类无效前缀的产生原因，可以得到如图7.2(a)和图7.2(b)所示的



(a) 无效自身前缀的形成原因 (IPv6)　　(b) 无效客户前缀的形成原因 (IPv6)

图 7.2 无效前缀的形成原因

7.5 IPv6 中对无效迁移前缀和无效劫持前缀的区分

IPv6 相对于 IPv4 的特点是其地址空间非常地庞大，考虑对一个“/64”的前缀下的地址进行全网段扫描，简单计算可知即便扫描带宽达到 Gbps 量级，扫描所需要的时间仍然是年量级的[39]。按照 X. Shi 等人的思想[10]，为了解决这个问题，可以利用 Alexa 提供的活跃域名对应的 IP[40]，对于某一个 IPv6 前缀，通过寻找仅仅被该前缀所覆盖的 IP，利用该 IP 对应的端口扫描结果来判断是无效迁移前缀还是无效劫持前缀。

由于找到的活跃 IP 地址极度地稀疏，目前利用这种方法检测得到的伪无效前缀的数量较为稀少。

7.6 IPv6 中伪无效前缀的时间维度稳定性

与 IPv4 中情形相类同，可以得到 IPv6 中的伪无效 (前缀，路由源）对在 2 月 28 日的路由表中的存在情况如表7.5。这里由于通过端口扫描可以检测到的无效迁移前缀的数目比较少，所以没有列出，另外的伪无效前缀大多数在时间维度上表现出了长时间的存在，这和 IPv4 中的结果是相似的。

表 7.5　5 月 16 日的路由数据得到的各类伪无效（前缀，源自治域）对在 2 月 28 日的路由表中的存在情况（IPv6）

伪无效前缀成因	表项数目	在 2 月 28 日的路由表中相应(前缀, 路由源) 对即已稳定存在的数目	在 2 月 28 的路由表中相应（前缀, 路由源)对即已稳定存在的数目所占百分比
负载均衡	41	36	87.80%
本自治域聚类失败	153	142	92.81%
多出口	8	7	87.5%
单出口下与客户前缀聚类失败	31	26	83.87%

第 8 章　基于 RPKI 的 BGP 源验证机制中伪无效性的检测和发布系统

第6章探讨了对基于 RPKI 的 BGP 路由源验证方法在验证机制层面降低虚警率的改善措施。然而基于 RPKI 的 BGP 路由源验证方法已经被标准化[2]，要想在验证机制层面进行修改并部署新的验证机制将会花费很长的时间，因而在短期内无法依赖 RPKI 来解决基于 RPKI 的 BGP 路由源验证方法中的虚警问题。一个过渡性的解决方案是提供一个开放的虚警列表，该虚警列表将帮助网络运营者确定自己的路由策略，必要的时候根据虚警列表设定相应的白名单，从而避免因为信任基于 RPKI 验证的结果而出现“断网”现象。

如图8.1所示为基于 RPKI 的 BGP 源验证机制中伪无效性的检测和发布系统的整体架构。

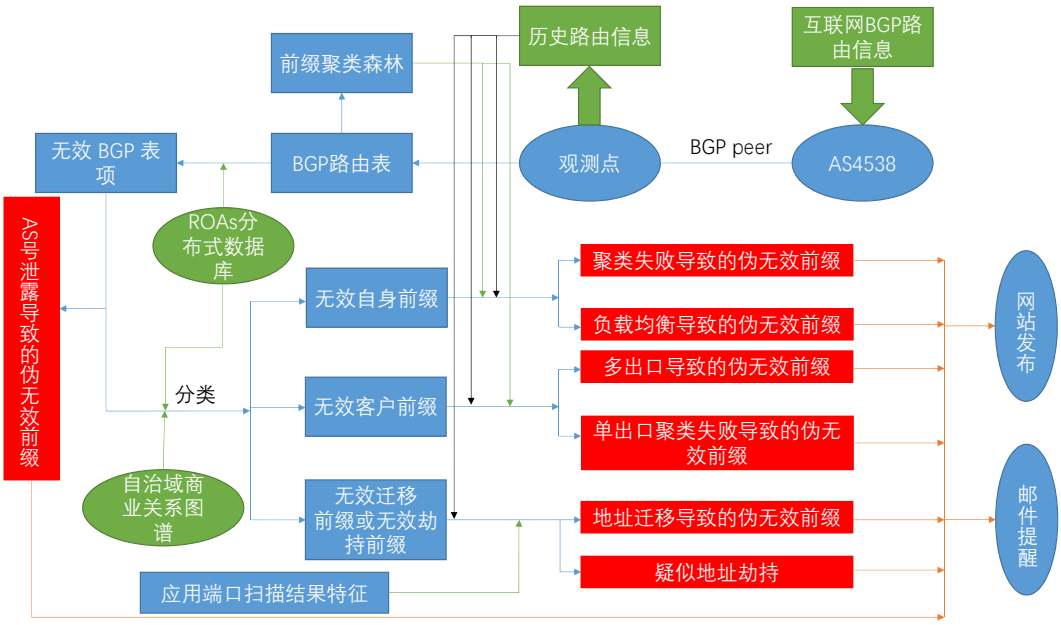


图 8.1　基于 RPKI 的 BGP 源验证机制中伪无效性的检测和发布系统

从 BGP 路由表数据的采集到检测到伪无效 BGP 前缀的整个过程已经在前文详述。下文将简单说明网站的功能。本网站的目的在于发布可能的基于 RPKI 的 BGP 路由验证机制中发现的伪无效前缀并指出其原因。

网站采用了 flask 提供的微框架，并采用 html 编写网页，提供了相应的由各种原因造成的伪无效前缀的列表以及相关总结。网站可以通过 202.38.101.13:5000 进行访问。

第 9 章　结论

本文作者首先利用基于排序的算法以较低的时间复杂度建立了 BGP **前缀聚类森林**，并基于此系统地测量了 BGP 路由前缀解聚类现状，将路由前缀解聚类现象分类为经典的多出口、负载均衡、聚类失败和地址碎片化四种类型[12]。本文作者指出，随着互联网的发展，地址解聚类类型已经不能简单的分为以上四种解聚类类型，而是出现了新的现象，也即**地址迁移**现象。

为了更进一步地分析 BGP 路由前缀解聚类是恶性的前缀劫持还是良性的正常解聚类，本文作者利用基于 RPKI（资源公钥基础设施）的方法利用 ROAs 来对 BGP 路由表项进行验证。本文作者发现基于 ROAs 的 BGP 路由表项源验证方法存在以下若干问题：

1. ROAs 数量的严重不足导致大部分的 IPv4 和 IPv6 前缀是无法得到安全上的保护的
2. ROAs 与实际的自治域使用前缀情形的不同步性以及良性解聚类的存在会导致伪无效前缀的产生

进一步地，为了分析伪无效前缀产生的原因，本文作者首先结合自治域商业关系图谱，将所有用基于 RPKI 的 BGP 路由源验证方法得到的无效前缀分类为**无效自身前缀**、**无效客户前缀**、**无效迁移前缀**和**无效劫持前缀**（在这一步，无法将后两种分离）。在此基础上，本文作者再进一步地结合前缀聚类森林和自治域商业关系图谱，发现无效自身前缀中的大部分事实上是由负载均衡和聚类失败所导致的，而无效客户前缀中的大部分其实是由多出口以及单出口条件下供应商父前缀和客户子前缀聚类失败导致的，并且本文作者给出了具体的测量比例。

为了能够区分**无效迁移前缀**和**无效劫持前缀**，本文作者提出了利用应用端口扫描特征和时间维度稳定性来进行区分的方法，并且发现 60% 左右的无效迁移前缀或者无效劫持前缀事实上是存在正常的网络应用的并且对应的（前缀，路径源）表现出了长时间的稳定性。这部分无效前缀极有可能是地址迁移导致的，事实上，ipwhois[37] 的注册信息佐证了 85% 左右的由地址迁移造成的前缀伪无效性的判断。

IPv6 是互联网的未来，但是目前关于 RPKI 在 IPv6 中的应用情况是缺少系统评估和测量的。本文作者将在 IPv4 的相关工作中建立起来的算法和方法应用到了 IPv6 中，并且获得了和在 IPv4 中类似的结果。

基于 RPKI 的 BGP 路由源验证方法中出现了伪无效性的问题，这在一定程度上是验证机制本身的问题，因此本文作者在验证机制层面提出以下两点建议：

1. 引入路径有效的概念，即验证（路径，前缀）对，若存在一个 ROA 记录覆盖了路由前缀，并且该 ROA 记录对应的 AS 在自治域路径上，则认为是路径有效的
2. 引入验证请求机制，也即 ROA 的持有者向其他 AS 发出对 BGP 前缀的验证请求

对于（路径，前缀）对的方法，本文作者指出在一个比较弱的假设下，该机制相对于（源，前缀）验证方法可以在不降低劫持检测率的条件下降低虚警率。对于验证请求机制，本文作者指出该机制可以部分地解决区分迁移前缀和劫持前缀的问题。

考虑到基于 RPKI 的 BGP 路由源验证机制已经被标准化[2]，在验证机制层面做改变需要时间，作为一种过渡性的解决伪无效前缀的方案，本文作者首次实现了一个基于 RPKI 的 BGP 路由源验证中伪无效前缀的检测和发布系统，利用本文作者的网站帮助网络运营者制定路由策略，减少伪无效前缀可能带来的影响正常路由策略甚至“断网”的危害。伪无效前缀的判断结果正在与网络运营商联系确认当中，目前为止没有收到对相关判断的否认回复。

以下几个方面是未来可以继续做的：

将 https 和路由验证相结合 https 是用来保证域名的安全性的，可以考虑将路由的相关信息嵌入到前缀中，然后将产生的 IP 地址映射到一个域名，利用 https 提供的域名验证机制来保证域名的合法性，进而保证路由的合法性。

BGP 路由更新的动态特征分析 本文主要分析了静态 BGP 路由表，没有挖掘 BGP 路由更新中的动态特征，通过动态特征来检测伪无效前缀是一个值得探索的方向。

由于毕业论文训练时间有限以及本文作者能力和视野有限，本文难免会有疏漏和不严谨之处，希望相关领域专家指正。

插图索引

表格索引

公式索引

参考文献

[1] Faruk A B M, Omar. Bgp security vulnerabilities analysis[J]. Work in Progress, draft-ietf-idr-bgp-vuln-00.txt, 2006.

[2] Mohapatra P, Scudder J, Ward D, et al. Bgp prefix origin validation[M]. [S.l.: s.n.], 2013.

[3] Wikipedia contributors. Ripe — Wikipedia, the free encyclopedia[EB/OL]. 2018. https://en.wikipedia.org/w/index.php?title=RIPE&oldid=840050034.

[4] Julian Z. In the news: A bgp hijacking technical post-mortem[M]. [S.l.: s.n.].

[5] Siddiqui A. What happened? the amazon route 53 bgp hijack to take over ethereum cryptocurrency wallets[M]. [S.l.: s.n.].

[6] Toonk A. Bgpmon[M]. [S.l.: s.n.].

[7] Wikipedia contributors. Bgp hijacking — Wikipedia, the free encyclopedia[EB/OL]. 2018. https://en.wikipedia.org/w/index.php?title=BGP_hijacking&oldid=838121683.

[8] Wikipedia contributors. As 7007 incident — Wikipedia, the free encyclopedia[EB/OL]. 2017. https://en.wikipedia.org/w/index.php?title=AS_7007_incident&oldid=815586186.

[9] [M]. [S.l.: s.n.].

[10] Shi X, Xiang Y, Wang Z, et al. Detecting prefix hijackings in the internet with argus[C]// Internet Measurement Conference 2012, Textbfcommunication Contribution Award. [S.l.: s.n.], 2012: 15–28.

[11] Geng G, Fu Y, Lee X D, et al. Rpki deployment considerations: Problem analysis and alternative solutions[M]. [S.l.: s.n.], 2016.

[12] Bu T, Gao L, Towsley D. On characterizing bgp routing table growth[J]. Computer Networks, 2004, 45(1): 45–54.

[13] Meng X, Xu Z, Zhang B, et al. Ipv4 address allocation and the bgp routing table evolution [J]. Acm Sigcomm Computer Communication Review, 2005, 35(1): 71–80.

[14] Gagliano R, Grampin E, Baliosian J, et al. Understanding ipv4 prefix de-aggregation: Challenges for routing scalability[C]//Ifip/ieee International Symposium on Integrated Network Management-Workshops. [S.l.: s.n.], 2009: 107–112.

[15] Gao L. On inferring autonomous system relationships in the internet[J]. Networking IEEE/ACM Transactions on, 2000, 9(6): 733–745.

[16] Luckie M, Huffaker B, Dhamdhere A, et al. As relationships, customer cones, and validation [C]//Conference on Internet Measurement Conference. [S.l.: s.n.], 2013: 243–256.

[17] Cooper D, Heilman E, Brogle K, et al. On the risk of misbehaving rpki authorities[M]. [S.l.: s.n.], 2013: 1–7

[18] Heilman E, Cooper D, Reyzin L, et al. From the consent of the routed: improving the transparency of the rpki[C]//ACM Conference on SIGCOMM. [S.l.: s.n.], 2014: 51–62.

[19] Gilad Y, Sagga O, Goldberg S. Maxlength considered harmful to the rpki[C]//The International Conference. [S.l.: s.n.], 2017: 101–107.

[20] Gilad Y, Cohen A, Herzberg A, et al. Are we there yet? on rpki's deployment and security [C]//NDSS. [S.l.: s.n.], 2017.

[21] Wikipedia contributors. Autonomous system (internet) — Wikipedia, the free encyclopedia [EB/OL]. 2018. https://en.wikipedia.org/w/index.php?title=Autonomous_system_(Internet)&oldid=841857414.

[22] Tanenbaum A, Wetherall D. Computer networks, 5th edition[M]. [S.l.: s.n.], 2010.

[23] Wikipedia contributors. Resource public key infrastructure — Wikipedia, the free encyclopedia[EB/OL]. 2018. https://en.wikipedia.org/w/index.php?title=Resource_Public_Key_Infrastructure&oldid=836550079.

[24] Lepinski M, Kent S. Rfc 6480 - an infrastructure to support secure internet routing[M]. [S.l.: s.n.], 2012.

[25] Gao L. On inferring autonomous system relationships in the internet[C]//Global Telecommunications Conference, 2000. GLOBECOM '00. IEEE. [S.l.: s.n.], 2002: 387–396 vol.1.

[26] Wikipedia contributors. Quagga (software) — Wikipedia, the free encyclopedia[EB/OL]. 2017. https://en.wikipedia.org/w/index.php?title=Quagga_(software)&oldid=795521111.

[27] Li T, Rekhter Y. Rfc 1654: A border gateway protocol 4 (bgp-4)[J]. Rfc Cisco Systems Ibm T.j.watson Research, 1994, 19(8): 193–199.

[28] Center for applied internet data analysis[EB/OL]. 2018. http://www.caida.org/home/.

[29] RIPE-NCC. Ripe ncc rpki validator 2.24 (updated 9 january 2018)[M]. [S.l.: s.n.], 2018.

[30] Wikipedia contributors. Orange s.a. — Wikipedia, the free encyclopedia[EB/OL]. 2018. https://en.wikipedia.org/w/index.php?title=Orange_S.A.&oldid=841418069.

[31] The caida ucsd as to organization mapping dataset[EB/OL]. 2018. http://www.caida.org/data/as_organizations.xml.

[32] The caida as relationships dataset[EB/OL]. 2018. http://www.caida.org/data/as-relationships/.

[33] Huston G G. Autonomous system (as) number reservation for documentation use", rfc 5398 [M]. [S.l.: s.n.].

[34] Mitchell J. Autonomous system (as) reservation for private use[J]. RFC, 2013, 6996: 1–4.

[35] Adrian D, Durumeric Z, Singh G, et al. Zippier zmap: internet-wide scanning at 10 gbps [C]//Usenix Conference on Offensive Technologies. [S.l.: s.n.], 2014: 8–8.

[36] Wikipedia contributors. List of tcp and udp port numbers — Wikipedia, the free encyclopedia [EB/OL]. 2018. https://en.wikipedia.org/w/index.php?title=List_of_TCP_and_UDP_port_numbers&oldid=843168735.

[37] ipwhois 1.0.0[EB/OL]. 2018. https://pypi.org/project/ipwhois/.

[38] Wikipedia contributors. Traceroute — Wikipedia, the free encyclopedia[EB/OL]. 2018. https://en.wikipedia.org/w/index.php?title=Traceroute&oldid=842533822.

[39] Gasser O, Scheitle Q, Gebhard S, et al. Scanning the ipv6 internet: Towards a comprehensive hitlist[M]. [S.l.: s.n.], 2016.

[40] Index of /i8-ipv6-hitlist/open/alexa1m/[EB/OL]. 2018. https://alctatraz.net.in.tum.de/i8-ipv6-hitlist/open/alexa1m/.

致　谢

感谢指导教师李星教授的悉心指导和勉励，没有李星老师在大方向上的指导、帮助和反馈，我是无法完成这篇论文的。

感谢实验室常得量学长在数据收集、网站搭建等方面提供的帮助，和他的讨论让我获益匪浅。感谢实验室王文鑫学长在服务器搭建、数据收集等方面的帮助，没有他的帮助，我难以解决很多技术问题。感谢实验室苏晨学长的热情帮助，他也帮助我解决了一些技术问题。

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：_____ 日　期：_____

附录 A　调研阅读报告（英文）

Report of Literature Survey for Final Year Project: BGP Routing Data Analysis

Background

BGP(Border Gateway Protocol) is a protocol in network layer used between different ASes(Autonomous Systems). Unlike intradomain protocols, whose objective is only to move packets from the source to the destination as efficiently as possible, BGP allows many kinds of routing policies to be enforced in the interAS traffic. Those routing policies can be highly individual and may involve political, security, or economic considerations. The primary function of a BGP speaking system is to exchange network reachability information with other BGP systems. A specific routing policy is implemented by deciding what traffic can flow over which of the links between ASes. BGP is highly vulnerable to misconfiguration and malicious attack. This final year project aims to detect anomaly and potential security threat through the analysis of BGP routing data using techniques including statistical inference and machine learning, etc. The significance of this project lies in that with powerful BGP data analysis algorithm, we can collect and analyze real time BGP routing data, forecast potential security threats and take actions before harm is done to our network.

How BGP works

Roughly speaking, BGP works as follows**??**:

1. One autonomous system may have multiple BGP speakers. And these BGP speakers listen to its peer BGP speakers from other ASes for network reachability information. Network reachability information is exchanged in the format of UPDATE message which contains a set of destinations that can be reached through the speaker and a path of ASes that the reachability information traverses.
2. The autonomous system will maintain three routing information bases: namely,

the Adj-RIBs-in, the Loc-RIB and the Adj-RIBs-out. The BGP speakers hear network reachability information from external BGP speakers and store the information in its Adj-RIBs-in. The network reachability information is exchaged between the internal BGP speakers of the same AS. Then on AS level, a Decision Process is invoked to decide which routes should be utilized to reach different sets of destinations. The routes ulilized are stored in Loc-RIB. The BGP speaker also maintains Adj-RIBs-out which contains the routes it advertises to specific peer BGP speaker.

3. Packets are forwarded according to the routing table in Loc-RIB.

Types of security threats concerning BGP

The security issues of BGP arise from three fundamental vulnerabilities**??**:

1. BGP has no internal mechanisms that provide strong protection of the integrity, freshness and peer entity authenticity of the messages in peer to peer BGP communications.
2. No mechanism has been specified to validate the authority of an AS to announce NLRI information.
3. No mechanism has been specified within BGP to ensure the authenticity of AS path attributes announced by an AS.

There are multiple types of security issues**??**:

Starvation Data traffic destined for a node is forwarded to a part of the network that can not deliver it.

Network Congestion More data traffic is forwarded through some portion of the network than would otherwise need to carry the network.

Blackholing Large amounts of traffic are directed to be forwarded through one router that can not handle the increased level of traffic and drops many/most/all packets.

Delay Data traffic destined for a node is forward along a path that is in some way inferior to the path it would otherwise take.

Looping Data traffic is forwarded along a path that loops so that the data is never delivered.

Eavesdrop Data traffic is forwarded through a router or a network that would otherwise not see the data, thus affording an opportunity to see the data.

Partition Some portion of the network believes that it is partitioned from the rest of the network, while, in fact, it is not.

Cut Some portion of the network believes that it has no route to a network to which, in fact, it is connected.

etc.

Some solutions to the security issues

Several approaches are proposed or implemented to enhance the security of BGP including the following:

Signature Approach

Sign the originating AS, predecessor information and AS path for authentication.

Filtering Approach

Rely on a registry to verify the AS path and the NLRI originating AS.

Information Sharing Approach

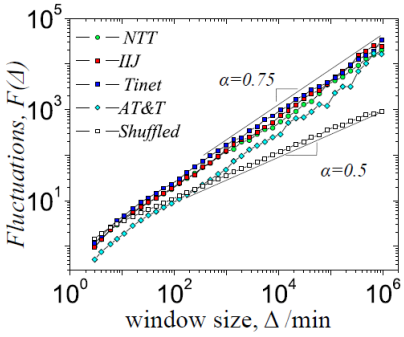
This approach proposes different ASes exchange routing information on demand and get a more complete view of the network. Through routing information exchanging and cooperation, the ASes may detect malicious attack more easily.

Summary of Existing Work on Analysis of BGP Routing Data

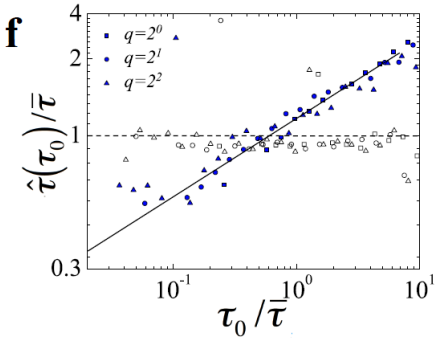
Statistical Signal Processing of BGP update data

Some work focus on using statistical signal processing techniques to analyse the update time series of BGP. For example, Maksim Kitsak, et al. utilize methodologies like DFA(Detrended Fluctuation Analysis), auto-correlation and spectrum analysis to detect the long-range correlation and memory in the update data of BGP**??**. As shown

in figureA-1(a), the DFA result shows the existence of long-range correlation in BGP update data. As shown in figure A-1(b), there is positive correlation between memory in BGP update process.



(a) DFA result of BGP update data



(b) The correlation between two consecutive time intervals of BGP update events

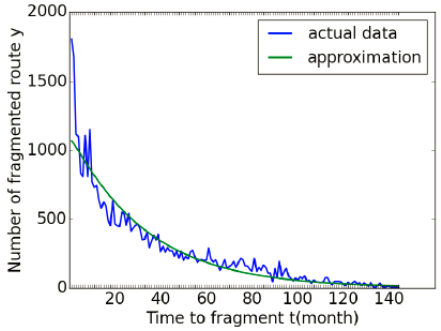
图 A-1 Statisical signal processing result of BGP update data**??**

Statistical Inference Based on Dynamic Data of BGP Routing

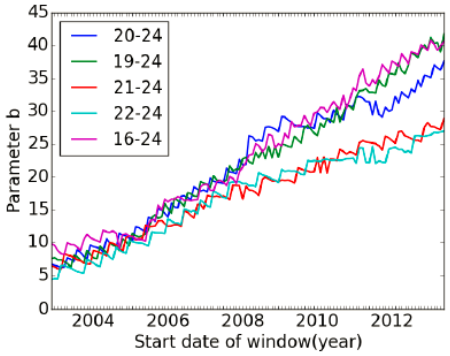
Some other work model the BGP update process as a stochastic process and do parameter estimation based on time interval data. For example, Motomu Utsumi, et al. provide a method to do statistical pattern prediction based on time window and parameter estimation**??**. As shown in figure A-2(a), the time interval between the appearance of an IP address prefix and the fragmentation of it approximately satisfy Gamma distribution and thus we can do parameter estimation based on that. And as shown in figure A-2(b), the scaling parameter which characterizes the length of the interval between the appearance of a 20 prefix and 24 prefix fragmentation, evolve almost linearly with some fluctuation and based on that we may do prediction.

Theoretical analysis based on graph theory

Graph is a powerful mathematical tool to model the connection and relationship between different autonomous systems. For example, Raj Kumar Rajendran, et al. provides a method to detect non-standard routing policies based on the analysis of common path of different AS paths**??**. Sobrinho, J., et al. a distributed address filtering and aggregation algorithm DRAGON based on the analysis of graph theoretical BGP



(b) 20-24 *fragmented*



(a) Parameter estimation of time interval of the IP address prefix fragment event

(b) The evolution of the scaling parameter of the Gamma model

图 A-2 Example of statistical inference in BGP update time series**??**

routes selection process**??**. Hitesh Ballani, et al. provides a method to detect prefix hijacking based on the analysis of BGP address prefix propagation process and routes selection policy**??**.

Perspective from network economics

BGP is designed for interdomain routing and in interdomain routing, ASes may provide transition service for other ASes. So BGP is closely related to economics. Much work has been done to analyze BGP routing data from an economic perspective. For example, Subramanian, et al. develop a heuristic algorithm to infer the business relationship between different ASes based on combinatorial optimization**??**. And Alex Fabrikant, et al. model the oscillation of BGP routing data as a Nash equilibrium**??**.

AS Relationship Inference

Since routing policy is highly individual, we may utilize a large amount of BGP routing data to infer the policies used by different ASes. For example, Xenofontas Dimitropoulos, et al. **??** provides a novel heuristic algorithm based on graph theory and weight function to infer the relationship between different ASes.

www.cidr-report.org

www.cidr-report.org is a website providing report on status summary, aggregation summary, last week's changes, more specifics and possible bogons of the Internet. The reports cover both IPv4 and IPv6. There are abundant results concerning the aggregation, dynamic characteristics and bogon routes from this website. The results from this website can be used to validate our own results.

Argus

Argus is a prefix detection system and is made up of three modules: the Anomaly Monitoring Module(AMM), the Live-IP Retriving Module(LRM) and Hijacking Identification Module(HIM).

RPKI

A potential alternative to BGP: Google's Espresso

While doing literature survey, I also find Google's Espresso, which is a potential alternative to BGP. In the final year project, I may also look for insight from Google's Espresso system.

*

References

[1] Y. Rekhter and T. Li, "A border gateway protocol 4 (bgp-4)," A Border Gateway Protocol 4 (BGP-4), 1994.

[2] S. Murphy, "Bgp security vulnerabilities analysis," 2006.

[3] M. Kitsak, A. Elmokashfi, S. Havlin, and D. Krioukov, "Long-range cor- relations and memory in the dynamics of internet interdomain routing," PloS one, vol. 10, no. 11, p. e0141481, 2015.

[4] M. Utsumi, H. Asai, and H. Esaki, "Spatio-temporal modeling of bgp rout- ing table evolution," in Proceedings of the 12th International Conference on Future Internet Technologies. ACM, 2017, p. 8.

[5] R. K. Rajendran, D. Rubenstein, and M. Wasserman, "A theoretical method for bgp policy verification," EE Department, Columbia University Technical Report, New York, NY, 2004.

[6] J. L. Sobrinho, L. Vanbever, F. Le, and J. Rexford, "Distributed route aggregation on the global network," in Proceedings of the 10th ACM In- ternational on Conference on emerging Networking Experiments and Tech- nologies. ACM, 2014, pp. 161–172. 5

[7] H. Ballani, P. Francis, and X. Zhang, "A study of prefix hijacking and interception in the internet," SIGCOMM Comput. Commun. Rev., vol. 37, no. 4, pp. 265–276, Aug. 2007. [Online]. Available: http://doi.acm.org/10.1145/1282427.1282411

[8] L. Subramanian, S. Agarwal, J. Rexford, and R. H. Katz, "Char- acterizing the internet hierarchy from multiple vantage points," in Proceedings.Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 2, 2002, pp. 618–627 vol.2.

[9] A. Fabrikant and C. H. Papadimitriou, "The complexity of game dynamics: Bgp oscillations, sink equilibria, and beyond," in Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, ser. SODA ' 08. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2008, pp. 844–853. [Online]. Available: http://dl.acm.org/citation.cfm?id=1347082.1347175

[10] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, k. claffy, and G. Riley, "As relationships: Inference and validation," SIGCOMM Comput. Commun. Rev., vol. 37, no. 1, pp. 29–40, Jan. 2007. [Online]. Available: http://doi.acm.org/10.1145/1198255.1198259