

硕 士 学 位 论 文

基于位置信息的用户级网络中立性研究

**The Research on User-level Network Neutrality
Using Geographic Information**

作 者 姓 名：_____ 田 丰 _____

学 科、 专 业：_____ 软 件 工 程 _____

学 号：_____ 21417025 _____

指 导 教 师：_____

完 成 日 期：_____

大 连 理 工 大 学

Dalian University of Technology

大连理工大学学位论文独创性声明

作者郑重声明：所呈交的学位论文，是本人在导师的指导下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内容和致谢的地方外，本论文不包含其他个人或集体已经发表的研究成果，也不包含其他已申请学位或其他用途使用过的成果。与我一同工作的同志对本研究所做的贡献均已在论文中做了明确的说明并表示了谢意。

若有不实之处，本人愿意承担相关法律责任。

学位论文题目：_____

作者签名：_____ 日期：_____年____月____日

摘 要

随着“开放网络”以及“中立性”概念的提出，网络服务提供商理应为用户提供公平、高质量的网络服务。但是现阶段，处于利益，在网络资源分配以及网络使用中普遍存在着资源的垄断以及对用户的不公平对待，也就是中立性违反问题。

在网络资源分配中，面向二级频谱市场的频谱拍卖是保证中立性的最好方法。而在网络使用中，保证中立性首先要做的就是让用户可以独立进行网络诊断，面对问题网络时能确定发生问题的部分。针对第一个问题，本文设计了一个空间时间重用的频谱拍卖机制 RTDA。RTDA 在满足用户的频谱需求的同时通过赢家判定和定价机制保证了真实性，从而实现了中立性。本文使用非合作式博弈证明了 RTDA 的中立性的同时也利用用户分组提高了频谱利用率。针对第二个问题，本文提出了基于 HTTP 与路径松弛的用户级别路径诊断方案，HIDA。HIDA 通过发现合适的 Web 服务器，帮助用户利用现有资源进行基于网络层析技术的中立性敏感的网络诊断。HIDA 通过结合位置信息的启发式服务器发现策略，帮助用户尽可能快的发现可用的服务器。

仿真试验中，RTDA 实现高达 100% 的频谱利用率，HIDA 实现在理想情况下的 0% 误判率。本文进行了拓展性的测量与实验，讨论了 RTDA 中时间分片数量对单信道效用等参数的影响，同时揭示了网络服务提供商之间数据交换的特点，为 HIDA 的性能验证提供了实测的数据。实验结果显示，HIDA 的基于动态时间规整的路径匹配方法比传统的基于 IP 地址的方法提高了 127%。HIDA 的启发式服务器发现策略比经验化方法降低了最多 25% 的查询次数。

关键词：开放网络；中立性；网络服务提供商；网络资源分配；网络诊断

The Research on User-level Network Neutrality Using Geographic Information Abstract

With the proposing of “Open Network” and “Network Neutrality”, Internet service providers (ISPs) are supposed to provide users with network services with high quality equally. However, problems of recourse monopoly and unequal treatment towards users in resource allocation and network usage exist nowadays, which are also neutrality violation.

In network resources allocation, the best way of obtaining neutrality is spectrum auction of secondary spectrum market. To guarantee the neutrality in the usage of network, users should have a method that could help locate the fault point, to diagnose the network dependently when facing a problematic Internet path. To address the first problem, this paper proposes a spatially and temporally reusable auction mechanism named RTDA. RTDA uses new winner determine, pricing and user grouping methods to achieve truthfulness and spectrum reusing. This paper uses non-cooperative gaming to prove the neutrality in RTDA and uses buyer grouping to improve the usage rate of spectrum channels. To address the second problem, this paper proposes a user-level diagnostic approach based on HTTP and path relaxation, named HIDA. HIDA could discover available Web servers for user in implementing tomographic inference without deploying any vantage point. Meanwhile, HIDA uses geolocation-based heuristic discovery strategy to help users discover servers as soon as possible.

In simulation, RTDA could improve the spectrum usage rate up to 100% while HIDA achieves 0% false positive fault. In extensive measurement and experiments, this paper discusses the influence of number of time slots on particular features such as utility of channels. It also reveals new observations of interconnections between ISPs, and provides real data for evaluating the performance of HIDA. The results show that HIDA’s DTW-based matching method could improve the performance up to 127% compared with traditional IP-based method. The heuristic server discovery strategy could decrease the number of queries up to 25% than empirical method.

Key Words: Open Network; Neutrality; Internet Service Provider; Network Resources Allocation; Network Diagnosis

目 录

摘 要.....	I
Abstract	II
1 绪论.....	1
1.1 研究背景与意义.....	1
1.1.1 频谱分配中的中立性.....	1
1.1.2 网络诊断中的中立性.....	2
1.2 国内外研究现状.....	3
1.3 本文主要工作.....	4
1.4 本文结构安排.....	5
2 相关技术.....	7
2.1 相似性度量.....	7
2.1.1 时间序列.....	7
2.1.2 层析技术与动态时间规整.....	8
2.2 基于地理位置的网络优化.....	10
2.2.1 频谱感知干扰.....	10
2.2.2 网络映射与路由策略.....	11
3 基于中立性的频谱拍卖.....	13
3.1 问题抽象模型.....	13
3.1.1 参数定义.....	14
3.1.2 真实性.....	14
3.1.3 个体理性.....	14
3.1.4 预算平衡.....	15
3.1.5 频谱重用性.....	15
3.2 RTDA 拍卖机制.....	17
3.2.1 用户分组.....	18
3.2.2 赢家决定.....	19
3.2.3 定价.....	20
3.3 中立性证明.....	21
3.3.1 预算平衡.....	21
3.3.2 个体理性.....	21
3.3.3 真实性.....	22

3.4	RTDA 性能评价	25
3.4.1	实验配置	25
3.4.2	实验结果	26
4	基于中立性的网络诊断	30
4.1	问题模型抽象	30
4.1.1	基于层析技术的网络诊断	30
4.1.2	基于位置信息的网络映射	31
4.2	网络映射与路径松弛	33
4.2.1	自治域系统映射中的路径松弛	34
4.2.2	地理位置映射中的路径松弛	34
4.2.3	相关问题与解决方案	35
4.3	基于 DTW 的路径匹配	37
4.4	基于位置信息的路径诊断	39
4.4.1	启发式发现策略	39
4.4.2	基于层析的问题路径推断	40
5	面向用户的网络测量与验证实验	41
5.1	基于用户的测量工具	41
5.1.1	网络爬虫	42
5.1.2	IP 地址与地理位置	42
5.1.3	测量实验设置	43
5.1.4	测量与观测结果	44
5.2	HIDA 性能评价	48
5.2.1	地理覆盖率	49
5.2.2	路径匹配实验	50
5.2.3	服务器发现实验	51
5.2.4	层析诊断仿真实验	53
结 论	55
参 考 文 献	56
攻读硕士学位期间发表学术论文情况	59
致 谢	60
大连理工大学学位论文版权使用授权书	61

1 绪论

1.1 研究背景与意义

在网络信息时代,网络服务提供商(Internet Service Provider, ISP)在人类生活中扮演着越来越重要的角色。这些网络服务提供商拥有各种网络资源,例如频谱资源、网络带宽资源等等。随着美国联邦通讯委员会(Federal Communications Commission, FCC)对“开放网络”以及“中立性”^[1]概念的提出,网络服务提供商应该基于中立性为用户公平的提供高质量的网络服务,即在开放的网络中,用户可以在任何的时间连接到任何想到达的合法的地方。但是现阶段,出于利益的原因,存在着对网络资源的垄断以及在网络使用中对用户的不公平对待,也就是所谓的违反中立性的问题。同时由费恩伯格提出的“技术民主化”^[2]的理论也指出,用户最为参与者的利益应该是首先得到保证的。因此,如何在网络资源的分配以及使用过程中保证网络服务提供商的中立性就成为了一个非常重要的问题。

狭义上来讲,中立性是指在开放互联网环境下,网络服务提供商不能对用户的流量进行阻塞(Blocking)、节流(Throttling)以及用户差异性(Paid Prioritization)。然而从广义上来说,网络中立性是一种对于网络用户的公平自由的开放的原则。所以对于用户来说,与网络服务提供商有关的中立性问题可以定位在资源分配与网络诊断两个方面。

1.1.1 频谱分配中的中立性

首先,在网络资源分配中,中立性就代表着所有的用户需要网络资源的时候能够公平的获得自己需要的网络资源。而现阶段,网络资源分配问题中最重要的就是频谱资源的分配问题。在这个问题上,传统的频谱管理方式是通过将频谱直接分配给网络服务提供商进行独占式的使用。但是,当网络服务提供商获得这部分频谱以后,并不一定能真正的使用这部分信道,也就导致了用户在享用不可再生的频谱资源中不能被公平的对待,也就是违反了中立性。因此,为了保证资源分配中的公平性, FCC 提出了基于拍卖的频谱管理方式。

网络服务提供商与对频谱资源有需求的用户就构成了二级频谱市场(Secondary Spectrum Market, SSM)。在这个模型中,由主用户(Primary Users, PUs)也就是拥有频谱的网络服务提供商作为卖家,二级用户(Secondary Users, SUs)也就是需要频谱资源的普通用户作为买家。在二级频谱市场中,主用户通过出售自己并未使用的空闲频谱从而获得收益,而二级用户通过支付相应的费用从而获得自己所需要的频谱,从而既

满足了主用户的经济利益又满足了二级用户的对于无线频谱的迫切需求，从而能够得到双赢的结果。

同时，频谱作为一种有限的资源的同时也具有一些与传统资源商品不同的特性^[3]。首先，最主要的就是其重用性，即在信噪比足够小从而不产生冲突的情况下，或是在不同时间隙使用同一信道的情況下，多个用户可以获得同一个频谱使用权。由于许多频谱授权拥有者，也就是网络服务提供商的排他式使用，使得大量的空闲频谱没有得到有效的利用^[4]。现在的网络环境日益复杂，各式各样的用户的私有设备也希望能够在实际的使用过程中独立的使用私有的信道，比如个人的家庭网络等等。随着物联网技术的发展，频谱资源也就愈发紧张。白频谱信道已经在西方国家率先开始进行分配。如何在进行二级频谱市场中，保证所参与的用户能够被公平的对待，也就是是频谱资源再分配过程中的中立性，是一个非常重要的问题。因为在这个过程中，会存在恶意的网络服务提供商，通过改变频谱价格或者操纵用户的方式来影响整个分配的公平性，从而继续对频谱资源的垄断。

1.1.2 网络诊断中的中立性

用户在获得了网络资源，进行网络的使用过程中保证网络服务提供商的中立性也是非常重要的。而现阶段，用户使用最为广泛的就是 **Internet** 网络。但是对于用户来说，在网络出现问题的时候并不能很好的了解问题出现的位置以及原因，因为所有的诊断工作基本上都是由网络服务提供商参与的。这也就导致很多问题，尤其是涉及到网络中立性的问题。因此为了保证用户在网络使用过程中能够获得中立性，首先要做的就是能够实施基于用户层面的 **Internet** 网络诊断。

Internet 网络由许多网络服务提供商组成，同时，这些网络服务提供商通过互联互通的方式相互进行数据交换。每个网络服务提供商作为一个自治域系统（**Autonomous System, AS**），自己管理自己的网络。一旦有网络流量需要穿越两个或两个以上的网络服务提供商的网络，就需要专门实现互联互通的网络交换点（**Internet eXchange Point, IXP**）来进行数据交换。比如说，存在一个违反中立性的网络服务提供商，他可能会由于利益的驱使对来自于不同网络服务提供商的流量进行不同程度的区别对待。比如说，当存在一个位于自己网络的服务器被来自于不同网络服务提供商网络内的客户端访问的时候，该网络服务提供商就有可能提高自己网络内部的流量的优先级，甚至利用深度包检测等技术对其他网络服务提供商的流量进行截留。而这种差异化对待的方式也就导致了中立性的违反。

当面临某一条出现问题的网络路径的时候,网络服务提供商可以通过其私有的网络管理工具帮助他们定位具体的问题发生位置;然而,作为终端用户而言,他们也想很方便的知道什么网络的什么位置发生了问题,从而作为依据反馈给相应的网络服务提供商来改善自己的网络状况。同时在网络使用过程中,他们应该更关心的也是到底是网络的哪一个部分(例如,某一个网络服务提供商或者某一段网络服务提供商之间的互联互通)产生了问题。更重要的是,中立性有关的网络诊断要完全依靠用户层面的信息,因为没有任何一个网络服务提供商,会为验证自己的中立性提供相关的数据。因此,解决这些问题的最好方式就是进行用户层面的网络诊断。

1.2 国内外研究现状

在网络中立性相关研究领域,影响力最高的无疑是 FCC。对应于其开放网络的理念,针对于上文提到的两个问题, FCC 分别提出了频谱拍卖的网络资源分配方式^[5]以及在 Internet 中 ISP 的相关服务准则也就是狭义的网络中立性^[6]。

在对用户进行网络资源分配的过程中,最为客观、透明公平的频谱分配方式就是以市场为主导的频谱拍卖方式。比如 FCC 就在近年了举行了几次针对空拍频谱的分配拍卖。因此,通过拍卖的手段来管理空闲的频谱也就成为一种公平客观并且基于市场的有效手段。同时在拍卖中,保证中立性的方式就保证参与者出价的真实性^[7]。因此,在本文中,真实性等价于中立性。如何在充分满足用户的频谱需求的前提下,进行真实公平的频谱管理分配,并且尽可能的提高频谱资源的利用率是一个亟待解决的问题。其决定性因素是买家对频谱的估值所做出的出价,由价格的高低决定频谱的归属^[8]。现有的频谱拍卖大都着重于拍卖中收益的最大化,忽略了频谱重用率的提高。其中最基本的 McAfee^[9]与 VCG^[10]机制仅仅注重最基本的经济鲁棒性,忽略了频谱可重用的特性。WRAN^[11]也是实现了买家收益的最大化,但是没有考虑频谱的重用问题。TASG^[12]虽然考虑了频谱的重用性,但是在重用过程中并未考虑在二级用户使用的过程中可能会产生冲突的情况。在 VERITAS^[13]中,虽然兼顾了重用性与经济鲁棒性两方面,但是只是实现了面向买家的单向拍卖。U-Auction^[14]实现了空间的频谱重用与真实性,但是并没有考虑时间上的重用。目前为止,比较完善的双向拍卖框架是由 Zhou Xia 等人提出的 TRUST^[15]。TRUST 同时注重了真实性以及频谱重用性,因此在本文中也将其作为一个重要的基础以及比较的对象。

针对用户层面的网络诊断问题,之前的研究工作大都集中在本地局域网(Local Area Network, LAN),比如 Kanuparth 提出了使用的基于侦测方法的诊断方法^[16]。对应到骨干网络, Mahajan 等人提出了用户层面的路由器级别的诊断方法^[17]。他们通过记录不

同的 IP-ID 来检测相关的问题。然而，现阶段网络服务提供商对于私有路由器的管理愈发严格，从而导致路由器级别的诊断变得更加困难^[18]。除此之外，现阶段应用最广泛的 Traceroute 也是用户层面的测量与诊断所必备的工作。但是匿名路由器 (Anomalies)^[19]，超时 (Timeouts)^[20]以及别名 (Aliases) 都会产生较大的影响。Traceroute 主要是通过包含不同的 TTL 值的探测数据包来发现到达服务器的沿途的中途的端口，只要依靠的是监听来自每个端口在 TTL 到期之后发送的 ICMP 的应答包，从而描绘出从客户端到服务器的网络路径。然而，在实际使用过程中，由于上述出于安全的网络服务提供商对自己的路由器信息的隐藏，用户常常会面临例如图 1.2 中的超时问题。除此之外，负载均衡 (Load Balancing)^[19]导致的 Traceroute 导出的路径的环路问题也可能会影响整个诊断的结果。Coates 等人首次将网络层析技术 (Network Tomography Technology) 应用在 Internet 相关的测量与诊断中^[21]。网络层析技术利用统计和推断的方法对网络性能和拓扑结构进行推理^[22]。在最近的网络诊断研究中，CLINK^[23]为用户提供了一种基于端到端测量与网络层析技术的诊断方案，将网络诊断的问题转化为了简单的布尔判断问题。网络层析技术也在中立性的判定中得到了很好的应用，例如 System 4^[24]。但是涉及用户层面的端到端的测量是相对困难的。在现阶段，尤其在中国，比如著名的 Measurement Lab (M-Lab)^[25]以及 PlanetLab^[26]等在中国活跃的结点少之又少。同时，大部分针对网络中立性的工作都是与测量有关，比如基于 M-Lab 平台的 Neubot^[27]。所以，基于 HTTP 的端到端测量就变得非常实用了。所以为了获得实用性与更好的可接入性，可以使用基于 HTTP 的 Web 服务器来代替这些观测服务器。因为在网络层析中，可以作为观测点的服务器主要有两个特性。首先，从客户端到这些服务器的网络路径与被检测的网络路径应该至少有一处重合的部分；其次，用户可以在这些路径上实施端到端的网络测量。He 已经对层析技术中的监测点替换问题进行了讨论^[11]。并且基于 HTTP 的测量也可以反映网络中的问题，尤其是丢包率、延迟和拥塞^[28]。Fathom^[29]也为用户提供了基于 HTTP 的网络测量的平台；WPof^[30]同样提供了基于 Chromium 浏览器的网络测量方式。除此之外，利用地理信息对网络诊断进行增强的方法也是很多的，因为在骨干网路由中，地理信息的特征是非常明显的^[31]。

1.3 本文主要工作

在本文定位于基于中立性的频谱资源分配问题以及用户层面的网络诊断问题。在这两个从用户角度出发的问题中，本文分别设计和实现了基于中立性的双向频谱拍卖机制 (Reusable Truthful Double Auction Mechanism for Wireless Spectrum Management, RTDA)

以及单用户实施的基于 HTTP 与路径松弛的中立性敏感的用户级网络诊断方案 (ISP-level HTTP-based Path Diagnostic Approach with Neutrality-aware Relaxations, HIDA)。

首先, 针对网络资源分配中的中立性问题, 本文设计了一个基于频谱重用的频谱双向拍卖机制 RTDA, 在实现真实性, 也就是中立性的同时提高了频谱的利用率。不仅从用户角度实现了资源分配的中立性, 同时从网络服务提供商角度, 能够通过出售空闲频谱获得收益, 从资源利用率角度实现频谱资源的有效利用。

其次, 针对网络使用过程中的中立性问题, 本文提出了基于 HTTP 的中立性敏感的网络诊断方案 HIDA。在 HIDA 中本文使用了基于地理位置的路径松弛方法以及基于动态时间规整的路径匹配方法。HIDA 把网络路径划分为网络服务提供商级别的片段, 从而帮助用户发现可用的 Web 服务器用来进行层析推断。

本文主要做出了以下工作与贡献:

(1) 利用时间信息: 本文利用了时间序列的基本思想, 实现了频谱拍卖中频谱的时间维度重用, 以及提出了基于动态时间规整的地理位置网络映射下的网络路径匹配方法。

(2) 利用位置信息: 本文在 RTDA 与 HIDA 中均使用了地理位置信息, 分别实现了频谱拍卖中频谱的空间维度的重用、路径的松弛以及启发式的 Web 服务器的发现方法。

(3) 理论证明: 本文通过非合作博弈推导证明了 RTDA 可以保证的与中立性密切相关的参与者出价的真实性, 同时也利用了中立性相关的理论推导证明了 HIDA 中基于地理位置的路径松弛方法的正确性。

(4) 用户级别的测量工具以及互联互通特性: 本文提出了一个用户层面的轻量化的网络测量工具。利用该工具可以获得拥有高地理覆盖率的数据用来评价本文提出的方法。同时, 也在测量中发现了一些关于互联互通的重要特性与观测结果, 验证了本文的假设与方法。

(5) 验证实验: 通过仿真实验, 本文验证了 RTDA 与 HIDA 在实际使用中的性能表现, 主要考察了频谱利用率、误判率等重要的指标。RTDA 实现高达 100% 的频谱利用率, HIDA 可以实现在理想情况下的 0% 误判率。实验结果显示, HIDA 的基于动态时间规整的路径匹配方法比传统的基于 IP 地址的方法提高了 127% 的同时, 其启发式服务器发现策略比经验化方法降低了最多 25% 的查询次数。

1.4 本文结构安排

本文按照如下的方式进行组织:

第一章，讨论了中立性在用户与网络服务提供商之间所代表的重要意义，分别从广义以及狭义的两个角度给出了定义，并分别对网络资源分配以及网络使用过程中的中立性进行了讨论。

第二章，介绍了本文提出的频谱分配机制 **RTDA** 以及网络路径诊断方法 **HIDA** 所使用的相关技术。**RTDA** 与 **HIDA** 均使用了时间与空间维度的信息来对现有的工作进行优化，例如相似性度量中的动态时间规整方法以及基于地理位置的网络优化中的空间重用以及网络映射技术。

第四章，介绍了基于中立性的二级频谱拍卖机制 **RTDA**。通过基于博弈论的方法对 **RTDA** 中与中立性相关的参与者出价的真实性进行了证明，并通过仿真实验验证了 **RTDA** 在实际使用中的效果

第四章，介绍了基于中立性的用户级网络路径诊断方法 **HIDA**。其中包括了对基于动态时间规整的路径匹配方法的证明。

第五章，介绍了本文设计实现的客户端网路测量工具。通过该工具，进行了大量的实测实验，并得出了有关网络服务提供商以及 **Internet** 互联互通的观测结果，同时利用实测实验的结果对 **HIDA** 在实际使用过程中的表现进行了评价。

结论部分主要对本文进行总结，并且对未来相关的工作进行讨论与展望。

2 相关技术

本章将对在 RTDA 与 HIDA 中, 本文使用了时间序列相似性度量以及基于地理信息的网络优化的方法进行讨论。

2.1 相似性度量

无论是在二级频谱拍卖中的频谱资源分配中, 还是在基于层析技术的网络路径诊断中, 相似性度量都是非常重要的一部分。进行频谱重用的时候要选择合适的用户进行分组, 而拍卖者就需要判断哪些用户可以在分配到同一个分组中来共享同一个信道, 而时间维度的频谱需求可以看作是时间序列; 在网络层析诊断中, 需要判断对应的路径有哪些重叠的部分可以用来诊断, 而网络路径可以看作不同的节点组成的时间序列。因此, 本文使用了基于时间序列的方法来进行相似性度量, 分别用时间序列的匹配以及动态时间规整的思想, 进行用户信道需求的冲突检测以及路径诊断中的网络路径匹配。

2.1.1 时间序列

时间序列 (Time Series)^[32]是指将同一类型的参数数值按其出现的时间维度上的先后顺序排列而成的数列^[33], 因此在实际使用过程中可以简单的表示为

$$S = \{s_1 = (v_1, t_1), s_2 = (v_2, t_2), \dots, s_n = (v_n, t_n)\}. \quad (2.1)$$

在该序列中, $s_i = (v_i, t_i)$ 表示的是在该时间点的所记录得到的信息 v_i 。在广义的时间序列中, v_i 不仅局限于数值, 也可以是任意的离散符号或者用户自定义的数据结构等等。只要符合时间的递增顺序就可以被认为是时间序列, 例如在 RTDA 中, 本文用时间序列对频谱进行时间维度的划分。在网络诊断中, HIDA 将网络路径当作时间序列以进行路径的匹配。

衡量两个时间序列的相似度主要是通过对两者的距离进行描述, 从而对其进行更合理的分类。欧式距离 (Euclidean Distance) 是最常用的距离测量的标准与方法。两个等长的时间序列之间的欧式距离就是对每个对应的时间节点的距离进行简单的相加。这些距离值可以是连续的实数数据也可以是离散的数据, 通过计算两个时间序列的距离就可以对其形似度进行合理的度量。比如说 RTDA 中代表用户对时间片需求的布尔类型的时间序列。所有的请求是否发生冲突, 就可以看作这些时间序列能否进行匹配。而在网络诊断中, 网络路径同样是可以看作时间序列。层析技术中的网络路径之间的重合部分的判断就可以看作是相应的时间序列片段的相似性匹配。但是, 需要强调的是, 网络路径由于动态路由以及丢包率的存在, 用户获得的 Traceroute 结果可能会存在误差, 因此这里额外的就需要相应的模式匹配的方法来进行修正。

2.1.2 层析技术与动态时间规整

网络层析技术（Network Tomography Technique）是近年来网络诊断与中立性推断领域最为常用的一种技术。网络层析技术是一种利用医学图像中的透视概念，通过不同路径下的端到端测量以及路径间的重合与分离，能够对不同网络路径的共享链路集合的状态进行判断，从而实现覆盖整个网络中的链路的诊断与测量方案。网络层析技术已经广泛应用于复杂多跳骨干网络的测量与诊断工作中，并且占据越来越重要的位置。

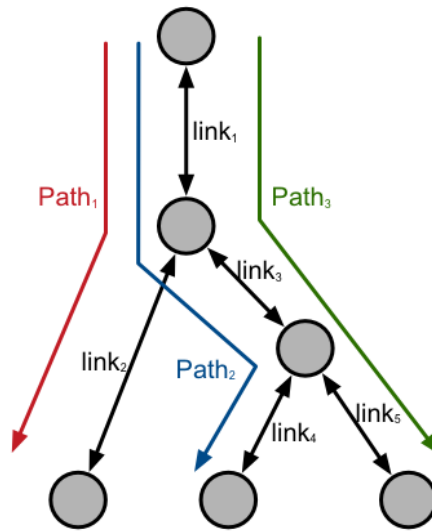


图 2.1 网络层析技术推断

Fig 2.1 Network Tomographic Technique Inference

如图 2.1 所示，是典型的网络层析技术在网络中立性判断与故障诊断中的应用。假设图中通过测量 $Path_1$ 、 $Path_2$ 与 $Path_3$ ，只要三者同时出现了问题，例如拥塞或者中立性违反，那么就可以断定最有可能发生问题的链路是 $link_1$ 。如果 $Path_1$ 是正常的、 $Path_2$ 与 $Path_3$ 却同时出现了问题，那么该拓扑中的故障点最有可能是 $link_3$ 。如上所述，通过不同角度对网络的描述结合起来反映出同一个网络问题的方式就是网络层析技术。

而如何判断这些路径是否重合就需要用到路径匹配的一些方法，其中就包括相似性都来那个。动态时间规整（Dynamic Time Wrapping, DTW）^[34]就是一种非常重要的时间序列的相似性度量技术。它可以帮助用户在两个非等长的时间序列中找到最佳的匹配距离。在这个过程中，时间序列可能在时间维度被进行缩放，以获得最佳的匹配。

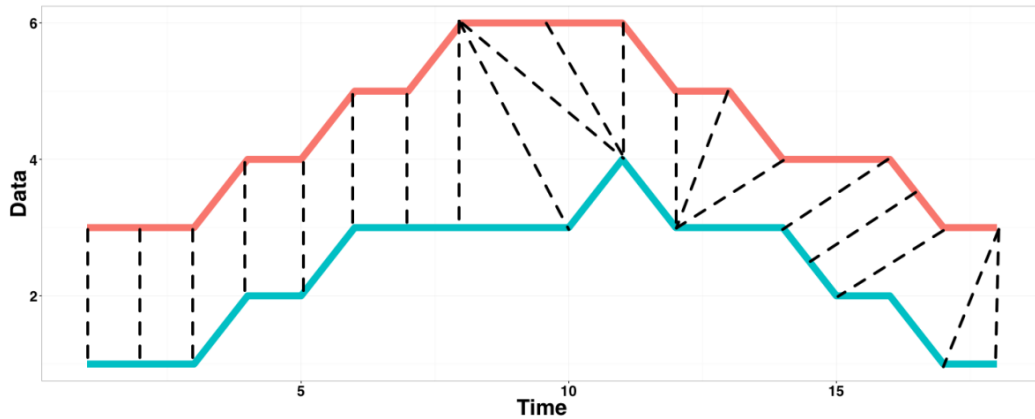


图 2.2 动态时间规整示例

Fig 2.2 Example of Dynamic Time Wrapping

如图 2.2 所示是 DTW 的一个常见的例子。纵向的每条虚线连接的是不同的时间序列中拥有相同的数据值的数据点。但是为了区别的更加清晰，在图中按照纵向将两个时间序列平移开，并用虚线进行连接。如果在图 2.2 中所示的两个时间序列的波动趋势是完全重合的话，那么相应的所有的虚线都应该是平行的。但是，由于时间上的波动，导致这些虚线出现了相交的部分。但是，实际情况时，原本应该完全匹配的两个时间序列因为时间上的波动，导致了误差的产生。因此，在时间维度进行相应的规整，然后对信号再进行匹配，就是动态时间规整。在动态时间规整中，最常用的方法就是动态规划。比如图 2.2 所示的情况，两个信号的波动趋势是相同的，但是在细节上并不是完全重合的，因此就需要要动态时间规整的方法来进行信号的模糊匹配。接下来，就要介绍具体如何进行时间维度上的调整，从而帮助用户能够使得两个信号能够匹配，同时还不违反基本的事实。

在实际的应用中，动态时间规整是非常有用的，其最基本的思想就是对信号在时间维度上进行一定程度的放缩，从而达到匹配的目的。例如在 HIDA 中，对于用户得到的网络路径来说，其实就是一个线性的时间序列，每个节点的 IP 地址就可以看作是相应的节点信息。网络的动态性，比如负载均衡以及匿名路由器的存在，导致即使是等价的路径也有可能会出现不匹配的状况。根据动态时间规整的思想，只要可以确定两条路径是相似的，那么从用户角度就可以进行一定程度的放缩，从而解决因为用户角度路径测量的误差导致的路径不匹配问题。比如在 HIDA 中，本文使用的就是地理位置信息来对路径进行松弛。同时在 RTDA 中，本文也使用了地理位置信息进行频谱的空间维度重用，所以接下来就对地理位置信息相关的网络优化问题进行讨论。

2.2 基于地理位置的网络优化

2.2.1 频谱感知干扰

在认知无线电（Cognitive Radio, CR）相关的频谱二次分配领域中，非常重要的一个频谱重用的因素就是地理位置信息^[35]。由于无线电信号在大气中传播会由于能量的衰减导致信号的衰落。因此，其传播范围也是一定的，并不能无限制的传播下去。根据这个特性，在不同地理位置的用户就可以在同一时间共享同一信道。因为其信号由于在空间维度上距离足够远，就无法进行相互干扰。信号与干扰加噪声比（Signal to Interference Plus Noise Ratio, SINR）则是用来反应可用信号与干扰信号之间的强度的比值，通常被称为“信噪比”。当某个区域的某个信号的信噪比足够低的时候，就可以认为该信号在这个区域不会对其他的信号造成影响，那么这个信道可以分配给多用户同时使用。而与信噪比密切相关的就是两个信号源之间的距离以及设备发射的功率。在路径衰弱无线电传播模型中，在距离发射点 $r(l)$ 的接收点的接受信道 l 的信号强度为 P_l/γ^α ，其中 α 为该路径上的损耗系数，是由当前的传播媒介所决定的一个常数。在传输中，只要保证信号的接受强度小于 β 倍的外界噪声，那么，在接收点就可以认为发送点的信号对其不造成干扰。两个不会产生干扰的情况下就会满足以下的条件：

$$SINR = \frac{P_l/\gamma^\alpha}{N + \sum_{l' \in S/\{l\}} P_{l'}/\gamma^{\alpha(l',l)}} \leq \beta. \quad (2.2)$$

其中， β 表示的是不产生干扰的最大的 SINR 值， N 是环境噪声， S 表示的是链路的集合。其中 $r(l)$ 就可以定义为发射点在这个信道上的干扰半径。

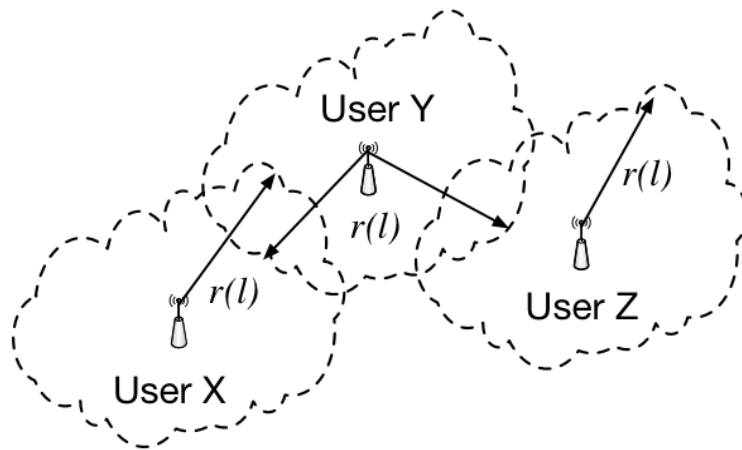


图 2.3 空间干扰模型

Fig 2.3 Spatial Interference Model

如图 2.3 所示, 用户 X 与用户 Z 均存在于用户 Y 的干扰半径之内, 并且是三个用户均使用相同的信道。同时, 用户 X 与用户 Z 均不在相互之间的干扰半径之内。因此, 用户 X 与用户 Z 就可在任意时间共享这个信道, 而用户 Y 则不行。

现有的技术, 能够根据用户的地理位置信息进行认知无线电中频谱资源再利用的优化。比如说, 基于用户地理位置的动态频谱规划以及位置辅助的网络规划^[35]等等。更重要的是, 在频谱资源分配的过程中, 基于地理位置可以让用户在相同的时间共享同一个信道。比如 TRUST^[15]就是第一个将地理位置信息考虑进去, 实现地理位置层面的频谱重用的拍卖机制。

2.2.2 网络映射与路由策略

网络映射 (Network Mapping) 技术是研究网络发展与变革的一个非常重要的方法。网络映射是指在不同的层面上进行聚集。如图 2.4 所示的一条网络路径为例, 网络映射可以是在自治域系统层面, 该层面也是最粗略的层面, 用来研究网络服务提供商之间的关系; 也可以是端口层面, 该层面是最细致的层面, 主要用来研究路由协议的相关特性。不同的网络映射的层面都是有其对应的相关的网络的不同的角度的, 因此每个层面的网络映射都是很重要的。同时, 对于网络路径来说, 其不同层面的网络映射中的投影就是路径松弛的方法。

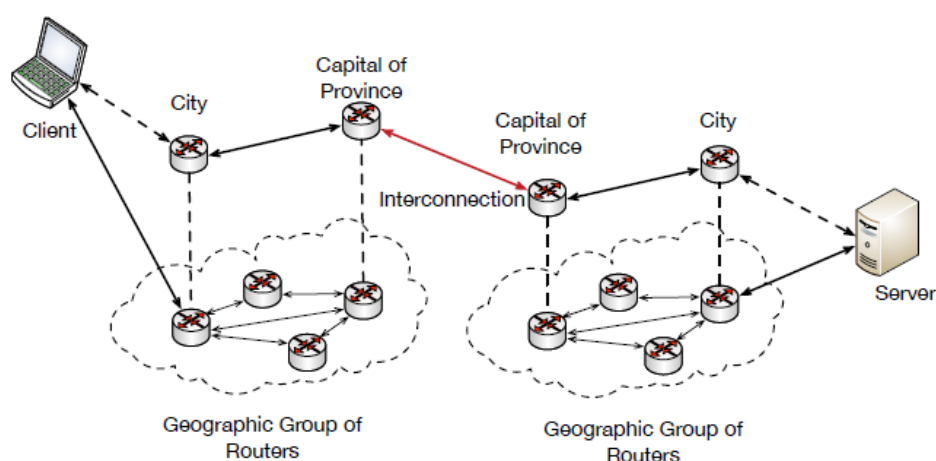


图 2.4 网络映射模型

Fig 2.3 Network Mapping Model

在有关骨干网的相关研究中, 地理位置层面的网络映射是非常重要的一个方法。网络服务提供商可能在同一个地区设置很多的服务器, 从而形成一个网络服务提供点 (Point of Presence, PoP)。同时, 将网络映射到 PoP 层面, 在一定程度上也可以认为

是地理位置层面，为用户提供了一个更好的路由器的聚集，同时能够最小程度的较少有效信息的丧失^[36]。使用地理位置层面的路由器聚集，除了可以帮助用户检测到非常重要的节点之外，也可以帮助用户理解包括网络路由策略在内的许多其他的其他网络特性。而且从实用性角度来说，地理位置层面的网络映射也是很必要的。比如说，通过地理位置信息，可以对 Web 相关的广告服务以及内容分发进行优化^[37]。

在自治域系统中，使用的是开放式最短路径优先（Open Shortest Path First, OSPF）为例的内部网关协议（Interior Gateway Protocol, IGP）。在每个自治域之间，运行的则是边界网关协议（Border Gateway Protocol, BGP）。同时，这些所有的协议一定是无环的、最短路径优先的。同时，这些路由协议在实际的使用过程中也是与地理位置信息是息息相关的。

理想情况下，所有的路由应该选择在地理上、网络层面都是最短的路径，从而提高数据传输的性能。因为，更少的地理位置的转换也就意味着更少的数据交换的过程。而且，现在也普遍的存在通过地理位置的方式对 BGP 协议进行优化的案例，比如将地理位置信息加入到 BGP 协议中^[38]。同时，也有研究者通过对网络的进行划分统计来探究地理位置信息与骨干网部署之间的关系^[39]。

更重要的是通过端到端的测量并且结合地理位置信息，就可以发现网络服务提供商在路由策略选择上的机制，比如选择最近出口点路由（Hot-Potato）策略以及选择最远出口点路由（Cold-Potato）策略。对于这两个路由策略的判断上，用户完全可以使用地理位置信息^[40]。在 Hot-Potato 路由策略中，网络服务提供商会尽可能快的将流量转发出去；相反，在最 Cold-Potato 路由策略中，网络服务提供商会尽可能久的持有流量。前者是为了尽可能减少网络服务提供商自己的负担；后者是为了尽可能保证对流量的控制权。而这两种不同的路由策略对网络中立性的影响是不同的。同时在 Cold-Potato 路由策略中，由于网络服务提供商会尽可能的持有数据流量，那么就导致发生在网络服务提供商之间的数据交换是非常的少的。因此，在网络路径中需要识别的互联互通部分也就比较少。这种情况下，是非常有利于网络服务提供商层面的网络诊断的。因为在层析技术中，相应的路径需要被划分的部分也是更少的。相反，Hot-Potato 路由策略的情况就需要进行仔细的判断，因为如果出现由于负载均衡导致的路径混乱的情况，Cold-Potato 路由策略就很容易被误判为 Hot-Potato 路由策略。因此，通过测量确定网络服务提供商的路由策略是非常重要的。在测量中，用户也可以根据网络服务提供商之间数据交换的次数以及数据交换发生的地理位置来判断其具体使用的是什么样的路由策略。

3 基于中立性的频谱拍卖

在本章节将首先讨论与中立性有关的第一个问题，就是网络资源分配过程中的中立性问题。本文提出了名为 RTDA 的基于拍卖的频谱分配机制，以解决频谱资源再分配过程中的关于参与者出家真实性的中立性问题以及频谱的利用率问题。RTDA 是一个网络服务提供商作为一级用户与普通用户作为二级用户共同参与的双向频谱拍卖机制，除了实现频谱资源分配中的中立性相关的真实性之外，同时还可以实现空间以及时间上的多维重用。在本章中，本文给出 RTDA 的相关设计细节以及基于博弈论的理论证明。最后将给出对 RTDA 表现进行评价的仿真实验结果。

3.1 问题抽象模型

在本文中，假设如图 3.1 所示的一个场景，有 M 个网络服务提供商作为卖家， N 个二级用户最为买家参与拍卖，并且存在一个中立的机构作为拍卖者，例如 FCC。并且所有的用户都可以根据其空间与时间维度的信息，在不相互干扰的情况下组成买家分组。除此之外，本文还假设这些信道都是同质的，例如 Wi-Fi 中的 11 个信道。

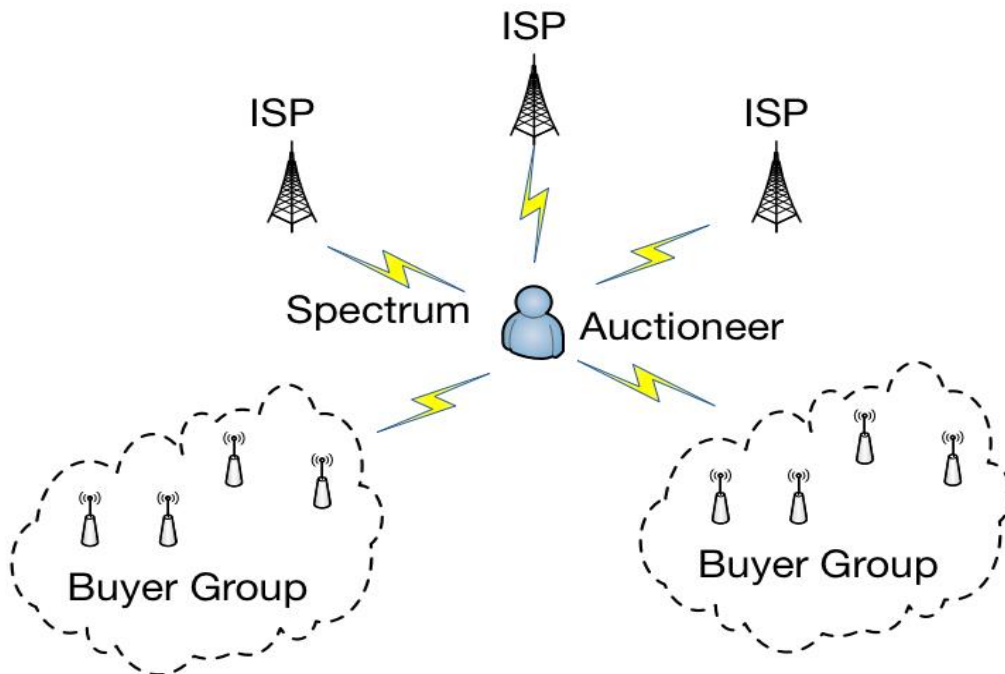


图 3.1 二级频谱拍卖模型

Fig 3.1 Secondary Spectrum Auction Model

3.1.1 参数定义

(1) 卖家参数

对于任意的 i 卖家来说, B_i^s 代表卖家 j 的出价, 即卖家 i 所能接受的卖出信道的最小价格; V_a^b 代表卖家 i 对于某个信道的真实估价; P_i^s 则代表卖家 i 在赢得拍卖后所要收取的价格即出清价格 (Clear Price); 因此, 对于一个卖家 i 来说, 如果卖家 i 赢得了某次拍卖, 则其在这次拍卖中所收获的效用则为 $U_i^s = P_i^s - V_i^s$, 否则, 卖家 i 因为在此次拍卖中获得 $U_i^s = 0$ 的效用。

(2) 买家参数

对于任意买家 j 来说, B_j^b 代表的是其对任意信道的出价, 同时也代表了买家 j 能够支付给卖家的最高价格; V_j^b 则代表买家 j 对于该信道的真实估价, 也就是买家 j 认为信道的实际的价值; P_j^b 则代表买家 i 在赢得拍卖后所要支付的出清价格

在本文中, T 被定义为每个信道所划分的时间片的总数。为了区分时间重用中的用户的报价, 对于任意的买家对于任意买家 j 来说, b_j^t 代表的是其对信道中的任意等长的时间片 t 的出价。 V_j^t 代表着这个时间片的真实的价值。同时, P_j^t 代表着买家 j 对时间片 t 所支付的费用。那么当用户竞拍成功的时候, 其效用为 $U_j^b = V_j^b - P_j^b$, 如果失败则为 0。

3.1.2 真实性

如果在一个双向拍卖中, 不存在任意一个卖家或者 i 买家 j , 在其他任意用户进行任意出价时, 通过不真实的出价, 即出价与真实估价不符, 即 $B_i^s \neq V_i^s, B_j^b \neq V_j^b$, 来获得更高的效用, 那么这个双向拍卖机制就是具有真实性的。因此只要将拍卖机制设计为, 理性的用户进行正常出价, 其出价一定满足真实性的要求, 那么该机制就一定是中立性的。为了进一步说明, 接下来要对个体理性进行定义, 也就是解释什么样的用户是理性的用户。

3.1.3 个体理性

在一次频谱拍卖中, 如果不存在任意一个获胜的买家支付了超过其出价的出清价格并且也不存在任意一个获胜的卖家被支付了低于其报价的出清价格, 那么这次拍卖就被认为是具有个体理性的性质。该性质可以由以下公式表达:

$$P_i^s \leq B_i^s, P_j^b \geq B_j^b, \forall i \in [1, M], j \in [1, N]. \quad (3.1)$$

个体理性保证了在真实出价的情况下, 没有参与者会因为真实出价而得到一个负的收益, 即效用, 以激励他们参与拍卖, 同时也是为了区别恶意破坏拍卖的中立性的用户。

个体理性的性质也是证明频谱拍卖机制设计的可信度的一个重要因素，同时也是证明其他重要性质成立的一个前提。

3.1.4 预算平衡

在本文中，拍卖的利润 φ 被定义为从买家收取的收入与支付给卖家的费用之间的差额。如果每次拍卖的利润 φ 都一定大于零，那么就表示该拍卖机制是预算平衡的。预算平衡是用来衡量每次频谱拍卖中，总的效用的最重要的标准，如果存在为负值的拍卖的收益 φ ，那么创造的社会价值就是负的，也就说明拍卖是没有意义的，从而也就没有网络服务提供商愿意进行中立性的频谱分配。如果对于任意信道的拍卖的过程中，这次拍卖的利润满足如下的公式，那么该拍卖算法就认为是符合预算平衡的：

$$\varphi = \sum_{j=1}^N P_j^b - \sum_{i=1}^M P_i^s \geq 0. \quad (3.2)$$

因为为了保证中立性，在拍卖的过程中双方的价格互相都是隐藏的，只有中立的拍卖机构能够获得，而拍卖者的收益也是在这两者之间产生。更重要的是通过保证拍卖者的收益不为负收益，从根本上保证了整个频谱拍卖机制的有效性，让整个频谱拍卖机制的总效用不为负值。也通过不为负的额外收益激励拍卖者组织参与频谱拍卖，从一方面也是激励相容的表现。

3.1.5 频谱重用性

(1) 基于位置信息的空间重用：对于买家来说，频谱的使用范围是有限的。在第二章中本文也提到无线信号的 SINR 对其信号强度是有决定性的影响的。也就是说，如果地理位置信息可以在频谱拍卖开始之前上报给拍卖者，拍卖者根据其地理冲突以及重用的程度，就可以通过多个买家使用同一个信道的方式实现频谱的重用。由于单纯基于空间的频谱重用，是可以忽略时间分段的影响的。不同频率的无线电资源存在着不同的环境特征，其不同信道的衰落也存在差异。但为了简化模型，在进行空间的频谱重用的频谱拍卖机制设计的过程中，本文假定对于任意的卖家 i 与买家 j 来说，所使用的都是统一的时间片的划分。在空间重用的过程，本文定义了买家的地理信息 D_j^b ，代表了在选定区域内用户 j 的位置，其具体的表示如下：

$$D_j^b = (x_j, y_j). \quad (3.3)$$

根据之前的 SINR 的定义，本文还定义最大的干扰半径 R ，代表了不会产生冲突的两个任意买家之间直线距离 $|D_{j_1}^b - D_{j_2}^b|$ 的最小值，并且当满足如下公式时，来表示任意分布的买家 j_1, j_2 之间不会产生冲突：

$$R \leq |D_{j_1}^b - D_{j_2}^b|. \quad (3.4)$$

但是，由于频谱的使用环境以及使用设备功率的差异，造成的不同买家设立的设备的干扰半径的差异性也是存在的。在使用过程中，每对用户的干扰情况都是单独计算的，而且 **RTDA** 允许用户存在不同的干扰半径。

(2) 基于时间序列的时间重用：在网络领域中，存在着不少进行时间重用的案例，比如无线网络协议中的时分复用等等。因此，在进行频谱重用的频谱拍卖中，时间自然作为一维重要的变量被考虑到频谱重用中。**RTDA** 将信道的时间分片设置为等长的布尔型数据的时间序列，不同用户之间的时间片的冲突都可以描述为时间序列的相似性度量问题，更确切的来说是时间序列的匹配问题。因此在进行时间重用设计时，在本文中有如下参数需要被设计考虑：

首先定义 t_j 为用户 j 所需要的时间片的数量。在 **RTDA** 中，本文用图 3.2 所示的时间序列来表示信道在时间维度的划分。

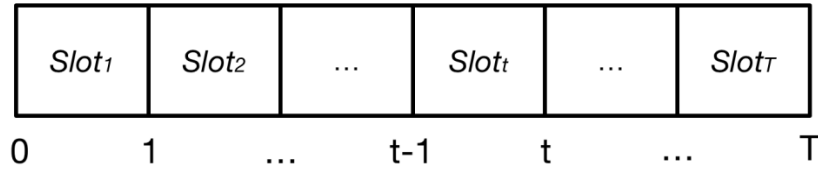


图 3.2 信道的时间序列

Fig 3.2 Time Series of the Channel

同时，不可能存在用户申请少于一个时间片，因此用公式 3.5 所示的不等式来表示一个用户所申请时间片的数量的限制。其中，**RTDA** 使用第一范式来代表时间序列中时间片的数量。

$$1 \leq \|t_j\|_1 \leq T, \forall j \in [1, N]. \quad (3.5)$$

在 **RTDA** 中，同时还定义了一个时间片矩阵 **TIME** 来计算每个信道被所有用户申请的时间序列的集合。如图 3.3 所示，每个信道对应一个矩阵，矩阵的每一行代表参与的买家用户，而列则代表划分好的时间序列。通过计算所有的时间序列矩阵 **TIME**，拍卖者就可以知道每个信道的时间维度的使用情况以及每个用户的信道分配情况。

对于每个信道来说，**RTDA** 都会维护一个相对应的矩阵 **TIME**，每个矩阵中都存在着每个用户的信道的时间维度需求的时间序列。因此，相应的，用户之间的时间片的冲突就可以转化为对时间序列集合矩阵的计算，从而就可以将问题转化为线性问题。在每个时间序列中，**RTDA** 将每个时间段的数据设置为布尔变量，然后通过范式计算的方式计算两个时间序列的欧式距离。只要在两个用户的时间片需求的时间序列能够完全不匹

配，也就意味着两个用户在时间维度上是不会产生冲突的。在不考虑两个用户的干扰半径的情况下，那么两个用户就是可以在相对应的时间内毫无干扰的使用同一信道的。即使两者存在干扰，那么由于时间维度的优先级应该是高于空间维度，所以两个用户同样可以共享该信道。

	$Slot_1$	$Slot_2$...	$Slot_t$...	$Slot_r$
b_1	TRUE	FALSE	...	FALSE	...	TRUE
b_2	FALSE	TRUE	...	TRUE	...	FALSE
...
b_j	TRUE	TRUE	...	TRUE	...	FALSE
...
b_J	TRUE	TRUE	...	FALSE	...	FALSE

图 3.3 TIME 矩阵

Fig 3.3 TIME Matrix

需要注意的是，如果该信道被时间重用，那么该信道将被分片，否则信道作为整体被出售，其效用并不会因为效用被进行时间上的划分而增加或减少。因为对于卖家来说，出售的是频谱的某段时间的完整使用权，所以进行时间片划分的效用 U_i' 、未进行时间片划分的效用 U_i ，满足如下的条件：

$$U_i' \equiv U_i. \quad (3.6)$$

3.2 RTDA 拍卖机制

如图 3.4 所示，RTDA 的整个频谱的分配过程主要包括以下三个部分，分别为用户分组（User Grouping）、赢家决定（Winner Determination）以及定价（Pricing）。其中买家分组是为了实现频谱的重用，而赢家决定与定价是为了保证分配过程的真实性与中立性。在买家分组中，本文使用了空间维度的干扰半径模型以及时间维度的时间序列模型，来对用户进行合理的分组规划。同时，按照基本的 McAfee 的设计机制，提出了全

新的基于时间空间重用情况下的赢家决定机制。最后，根据用户分组的情况，对分组中的每个用户以及每个参与拍卖的网络服务提供商进行出清价格的定价。在整个过程中，RTDA 实现了包括真实性、个体理性以及预算平衡在内的所有频谱拍卖过程中所要求的经济鲁棒性相关的特性，从而保证了频谱资源分配过程中的中立性。同时也通过买家分组对每个信道进行共享的方式，实现了信道的多维重用，从一定程度上解决了先阶段，频谱资源不足的问题。

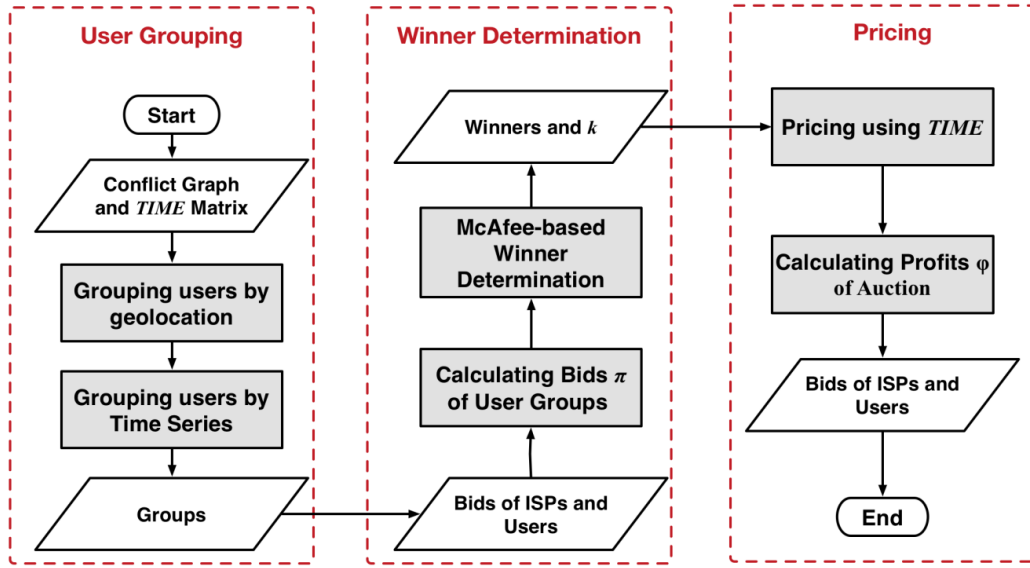


图 3.4 RTDA 工作流程图

Fig 3.4 RTDA Workflow

3.2.1 用户分组

在 RTDA 的用户分组过程中，买家用户上报的地理位置信息进行分组。与 TRUST 中的机制不同的是，在买家 j 上报个人地理位置信息 D_j^p 的同时，同时上报了自己所需要使用的时序 t_j 。首先，定义序列 G_1, G_2, \dots, G_L 为个 L 买家分组，买家的分组情况是根据空间与时间重用得到。RTDA 使用冲突图（Conflict Graph, CG）来表示用户之间在空间上的冲突。

用户作为图中的节点，两个用户之间存在着边，那么就代表着冲突。RTDA 使用基于贪婪算法重新设计的 OPT 分配算法来选择最大的用户分组，然后再根据用户所提交的 t_j 来重新计算分组。如表 3.1 中的用户分配算法与表 3.2 中的 OPT 算法所示，是涉及到的买家分配算法以及相应的 OPT 算法。在本文中， G_l 代表第 l 个二级买家用户分组。每个分

组中包含至少一个用户，并且每个分组中的用户一定可以保证在时间与空间维度都不会相互产生干扰。换句话说，每个分组内的所有用户都可以在允许的时间、空间范围内任意的使用自己通过拍卖获得的信道，而不需要担心不同设备对同一信道的干扰问题。

表 3.1 用户分组算法

Tab. 3.1 Users Grouping Algorithm

算法： 用户分组算法
输入： 冲突图 CG ， $TIME$ 矩阵，二级用户的数量 N
输出： 分组情况，分组数量 L
1: while CG 存在权值不为 N 的边 do
2: 调用 $OPT(CG, N)$ 函数得到不冲突点的集合
3: 将集合中的节点加入 G_l ，并计算 L
4: 将 G_l 中的节点在 CG 中相关边的权值设为 N
5: end while
6: return Groups, L

表 3.2 OPT 算法

Tab. 3.2 OPT Algorithm

算法： OPT 算法
输入： 冲突图 CG ，二级用户的数量 N
输出： 分组结果
1: If CG 存在权值不为 N 的边 then
2: 寻找途中度数最小的点，加入分组结果中
3: 将节点在 CG 中相关边的权值设为 N
5: 分组结果=分组结果 $\cap OPT(CG, N)$
6: else
7: return 分组结果
8: end if

3.2.2 赢家决定

保证真实性与中立性很重要的一个问题就是赢家决定问题（Winner Determination Problem, WDP）。对于任意的分组 G_l ，其分组的出价为 π_l ，而 π_l 也是由所有在分组的内的用户的出价来决定的，不仅如此，RTDA 也考虑到了时间重用中的条件，按照如下的方式来决定：

$$\pi_l = \min\{b_n^b | n \in G_l\} \cdot \sum_{n \in G_l} \|t_n\|_1. \quad (3.7)$$

然后得到分组的定价之后，RTDA 按照 McAfee 这个最基本的拍卖机制的方式来设计 WDP 问题的解决方案。首先，RTDA 将用户分组的报价按照降序进行排列如公式 3.8 所示：

$$\pi_1 \geq \pi_2 \geq \pi_3 \geq \dots \geq \pi_L. \quad (3.8)$$

然后再将所有卖家的报价按照升序的方式进行排列，获得相应的卖家的出价序列，如 3.9 所示：

$$B_1^s \leq B_2^s \leq B_3^s \leq \dots \leq B_M^s. \quad (3.9)$$

为了确定拍卖中的获胜的买家用户与卖家，将两个序列进行对应的比较，然后选择满足如下公式的 k 值：

$$k = \operatorname{argmax}_{l \leq \min\{L, M\}} \pi_l \geq B_l^s. \quad (3.10)$$

也就是说，按照双向拍卖的基础，要选择出两个序列中对应的两个 $k - 1$ 长度的序列，并且这两个子序列满足买家出价序列大于卖家报价序列。根据如上的设计，在本文中使用了表 3.3 中所示的赢家决定算法。

表 3.3 赢家决定算法

Tab. 3.3 Winner Determination Algorithm

算法： 赢家决定算法
输入： t_N, G, L, b_N, B_M
输出： 用户分组赢家，卖家赢家， k
1: for $l \in [1, L]$ do
2: 计算 $\pi_l = \min\{b_n^b n \in G_l\} \cdot \sum_{n \in G_l} \ t_n\ _1$
3: end for
4: 计算卖家出价序列 $B_1^s \leq B_2^s \leq B_3^s \leq \dots \leq B_M^s$
5: 计算买家出价序列 $\pi_1 \geq \pi_2 \geq \pi_3 \geq \dots \geq \pi_L$
6: 计算 $k = \operatorname{argmax}_{l \leq \min\{L, M\}} \pi_l \geq B_l^s$
7: 选出 $k - 1$ 个用户分组与 $k - 1$ 个卖家
8: return 用户分组，卖家分组， k

3.2.3 定价

在决定了拍卖过程中的赢家之后，RTDA 需要知道每个赢得拍卖的用户需要按照什么价格来进行支付。在这里，RTDA 分别将卖家与买家的实际支付的出清价格定义为 B_k^s

和 π_l ，也就是在 WDP 中两个序列比较的临界点的报价与出价。这样，每个卖家都要获得 B_k^s 的，而每个买家分组都要支付 π_l 。那么，具体到每个分组中的用户来说，就需要支付如下的费用：

$$P_j^b = \frac{\pi_k \cdot \|t_j\|_1}{\sum_{n \in G_l} \|t_n\|_1}, \forall j \in [1, N]. \quad (3.11)$$

将个体理性考虑在内，那么也就是说不存在任意的买家用户或者卖家网络服务提供商在这个过程中会获得负值的效用。与此同时，作为拍卖中心的执行者来说，其所获的收益，也就是本次拍卖所产生的收益为：

$$\varphi = (k - 1) \cdot (\pi_k - B_k^s). \quad (3.12)$$

3.3 中立性证明

通过对买家进行分组，RTDA 实现了频谱的时间、空间重用，接下来的这节内容将对频谱拍卖维持经济鲁棒性的所需的预算平衡、个体理性、真实性进行证明，保证频谱拍卖机制设计的有效性。首先，需要定义 P_j^b 为用户 j 的出清价格，然后用 p_j^b 来代表每个时间片的出清价格，从而证明在时间重用中 RTDA 的正确性。同时这两个价格还满足如下的等式：

$$P_j^b = p_j^b \cdot \|t_j\|_1, \forall j \in [1, N]. \quad (3.13)$$

3.3.1 预算平衡

定理 1： RTDA 具有预算平衡的性质，即 $\varphi \geq 0$ 。

证明： 在计算 k 的过程中，所选出的 k 满足了 $\pi_k \geq B_k^s$ ，并且若存在交易成功的话， $k \geq 1$ 恒成立。根据 φ 的表达公式 $\varphi = (k - 1) \cdot (\pi_k - B_k^s)$ ，也就可以容易 $\varphi \geq 0$ 是恒成立的，因此 RTDA 一定是具有预算平衡的性质。

3.3.2 个体理性

定理 2： RTDA 一定是个体理性的

证明： 证明拍卖机制是个体理性的，那么就是证明所有的参与者不会为了获得负的效用而参与拍卖。也就是要证明没有用户会愿意付出高于自己对信道的估价的报价来进行拍卖。而且，在拍卖失败的时候，买家或者卖家都会因为没有成交而不需要支付任何的费用，从而获得等于 0 的效用。因此，在拍卖不成功的时候，个体理性是存在的。

首先，要证明买家用户的个体理性。按照 WDP 中的设计，用户分组的报价按照降序进行排列。对于竞拍成功的 G_l 分组中的任意一个用户 j 来说，该分组的报价 $\pi_l \geq \pi_k$ 是恒成立的。那么对于该用户来说，其被收取的费用也就满足以下等式：

$$P_j^b = \frac{\pi_k \cdot \|t_j\|_1}{\sum_{n \in G_l} \|t_n\|_1} \leq \frac{\pi_l \cdot \|t_j\|_1}{\sum_{n \in G_l} \|t_n\|_1}, \forall j \in [1, N]. \quad (3.14)$$

同时按照买家用户分组的定义，对于 G_l 分组中任意的用户 j 来说，其出价也就一定满足公式 3.15 所示的条件：

$$B_j^b = b_j^b \cdot \|t_j\|_1 \geq \frac{\pi_l \cdot \|t_j\|_1}{\sum_{n \in G_l} \|t_n\|_1} \geq p_j^b \cdot \|t_j\|_1 = P_j^b. \quad (3.15)$$

并且，当 b_j^b 为该分组的最小出价时，那么 $B_j^b = \pi_l \cdot \|t_j\|_1 / \sum_{n \in G_l} \|t_n\|_1$ 是成立的。所以说 $B_j^b - P_j^b \geq 0$ 也是恒成立的。从而也就说明，买家用户个体理性也一定是存在的。

接下来是卖家的个体理性。与买家用户相似，卖家服务提供商的竞价全都按照升序排列，那么任意卖家 i 的出清价格 P_i^s 一定满足 $P_i^s = B_k^s \geq B_i^s$ 。也就是说，其效用大于等于零也是一定成立的。

综上所述，在 RTDA 的设计过程中，所有参与的二级用户作为买家已经网络服务提供商作为卖家一定都是个体理性的。

3.3.3 真实性

比较复杂的是真实性的证明。其证明就必须要考虑如表 3.4 所示四种情况，如果四种情况下真实性全部都满足的话，那么就可以认定设计的频谱拍卖机制是具有真实性的。由于卖家出价与买家出价是对称的，那么真实性的证明方式也就是类似，在本文中就以买家出价的证明为例。

表 3.4 出价与拍卖结果
Tab. 3.4 Bidding and Results of Auction

出价情况	情况 1	情况 2	情况 3	情况 4
不真实	失败	失败	成功	成功
真实	失败	成功	失败	成功

为了证明真实性，就需要证明两个特性，一个是 WDP 的单调性，另一个是定价方式与出价的不相关性。首先是证明 WDP 的单调性。

引理 1：对于给定剔除买家 j 的出价集合：

$$b_j^b, \dots, b_{j-1}^b, b_{j+1}^b, \dots, b_N^b. \quad (3.16)$$

以及卖家的出价集合 $\{B_i^s\}_{i=1}^M$ ，如果买家 j 能够通过出价 b_j^b 赢得拍卖，则其也能通过出价 $b_j^{b'} > b_j^b$ 赢得拍卖。

证明：当买家 j 出价为 b_j^b 时，其所在的分组出价为 π_l ；当其出价为 $b_j^{b'}$ 时，分组出价为 π_l' 。并且可以很容易看出当极限情况存在是，即买家 j 的出价始终为分组 G_l 中的最低价时，才能存在 $\pi_l' > \pi_l$ ，所以 $\pi_l' \geq \pi_l$ 恒成立。在 WDP 中， G_l 始终能够赢得拍卖，因此引理 1 成立。

引理 2：对于给定剔除卖家 i 的出价集合：

$$B_1^s, \dots, B_{i-1}^s, B_{i+1}^s, \dots, B_M^s. \quad (3.17)$$

以及买家的出价集合 $\{b_j^b\}_{j=1}^N$ ，如果卖家 i 能够通过出价 B_i^s 赢得拍卖，则其也能通过出价 $B_i^s > B_i^{s'}$ 赢得拍卖。

证明：（同引理 1 证明）。

接下来要通过证明引理 3、引理 4 来证明 RTDA 中定价方式与出价的不相关性。

引理 3：对于给定剔除买家 j 的出价集合：

$$b_j^b, \dots, b_{j-1}^b, b_{j+1}^b, \dots, b_N^b. \quad (3.18)$$

以及卖家的出价集合 $\{B_i^s\}_{i=1}^M$ ，如果买家 j 能够通过出价 b_j^b 赢得拍卖，也能通过出价 $b_j^{b'} > b_j^b$ 赢得拍卖，那么对应的出清价格 P_j^b 与 $P_j^{b'}$ 是相等的。

证明：当 $b_j^{b'} > b_j^b$ 时，在引理 5 的证明中已经进行了。因此只需证明 $b_j^{b'} \leq b_j^b$ 的情况即可。如果要拥有相同的出清价格 P_j^b ，则在不同的出价下，买家 j 都要能赢得拍卖。如果在出价为 $b_j^{b'} \leq b_j^b$ 的情况下都赢得了拍卖，则存在极限情况，买家 j 的出价始终为分组 G_l 中的最低单价。但无论如何，买家 j 的出价都只会影响买家分组 G_l 的出价，但不影响其分组大小，同时，每个赢家分组的出清价格都只会是 π_k ，买家 j 的出清价格 P_j^b 也不会因为其投标出价的改变而改变，则因此得证引理 7 成立。

引理 4：对于给定剔除卖家 i 的出价集合：

$$B_1^s, \dots, B_{i-1}^s, B_{i+1}^s, \dots, B_M^s. \quad (3.19)$$

以及买家的出价集合 $\{b_j^b\}_{j=1}^N$ ，如果卖家 i 能够通过出价 B_i^s 赢得拍卖，则其也能通过出价 $B_i^{s'} < B_i^s$ 赢得拍卖，那么对应的出清价格 P_i^s 与 $P_i^{s'}$ 是相等的

证明：（同引理 3 证明）。

证明了 WDP 的单调性以及定价与出价的不相关性之后，接下来就进入真实性的证明。若要证明真实性，就需要证明买家或者卖家不能通过投标不符合其真实估价的出价来提高自己的效用，也就意味着其出价的占优策略是依照真实估价进行出价，即 $B_j^b = V_j^s$ 与 $B_i^s = V_i^s$ 。

定理 3：RTDA 一定是具有真实性的，即 $B_j^b = V_j^s$ 与 $B_i^s = V_i^s$ 。

证明：对于表 3.4 中的四种情况，如果四种情况下真实性全部都满足的话，就可以认定本文所设计的频谱拍卖机制是具有真实性的。

(1) 情况 1：无论买家 j 通过真实或者不真实的出价都不会赢得信道，也就意味着无论怎样出价都不能赢得信道，其效用始终为零，是不证自明的。

(2) 情况 2：这种情况下只有买家 j 不真实出价 $b_j^{b'}$ 小于其真实出价 b_j^b ，并且由于不真实出价却在拍卖中失败了，获得效用为零，同时在赢得拍卖后其效用为正值，其不真实出价导致了更低的效用，证明在这种情况下具有真实性。

(3) 情况 3：只有买家 j 的不真实出价 $B_j^{b'} = b_j^{j'} \cdot \|t_j\|_1 > V_j^b = b_j^j \cdot \|t_j\|_1$ 时，才会发生这种情况。现在假定，当买家 j 出价为 $b_j^{b'}$ 时，其所在的分组出价为 π_l ；当其出价为 $b_j^{b'}$ 时，分组出价为 π_l' 。如果买家 j 的出价想要影响分组的出价的话，那么其真实总出价 B_j^b 必须为分组内最低出价，即 $\pi_l = b_j^b \cdot \sum_{n \in G_l} \|t_n\|_1$ ，如果通过提高自己的出价赢得了拍卖，易证得其赢得拍卖的分组的出清价格 P 一定满足 $\pi_l' \geq P \geq \pi_l$ 。因此，其结算之后的效用为：

$$V_j^b - \frac{P \cdot \|t_j\|_1}{\sum_{n \in G_l} \|t_n\|_1} = B_j^b - \frac{P \cdot \|t_j\|_1}{\sum_{n \in G_l} \|t_n\|_1}, B_j^b = \frac{\pi_l \cdot \|t_j\|_1}{\sum_{n \in G_l} \|t_n\|_1}. \quad (3.20)$$

因此，通过化简以及带入相关的变量，从而就可以得到如下的不等式，从而证得不等式成立：

$$\frac{\pi_l \cdot \|t_j\|_1}{\sum_{n \in G_l} \|t_n\|_1} - \frac{P \cdot \|t_j\|_1}{\sum_{n \in G_l} \|t_n\|_1} < 0, \forall j \in [1, N]. \quad (3.21)$$

由于其真实出价在拍卖中失败的效用为 0，而通过不真实出价获得效用为负值，所以根据个体理性，就可以知道在这种情况下，用户的出价时不存在的，那么也就证得在这种情况下，RTDA 是具有真实性的。

(4) 情况 4：在引理 3 的证明中已经证明了无论其出价真实与否，如果都赢得了信道，那么其出清价格是不变的，由于其真是估价的唯一性，因此得证其效用也是相同的，因此可以证明这种情况下具有真实性。

综上所述，对于以上四种情况 RTDA 都是可以证明是具有真实性的，而且这四种情况也将频谱分配过程中的所有报价情况涵盖了，根据以上结果，就可以得证 RTDA 对于买家来说具有真实性，相应的也就证明 RTDA 是具有真实性的，从而证明了 RTDA 在频谱分配过程中是具有中立性的。

因此, 本文通过非合作博弈的方式对用户的出价进行了分情况讨论, 从而证明了相关的经济鲁棒性进而证明了 RTDA 在网络频谱资源分配中的中立性, 从而保证了用户在请求网络资源的过程中能够得到公平的对待。

3.4 RTDA 性能评价

本节主要对 RTDA 在实际使用过程中的各项参数如时间片划分的数量以及信道的利用率等方面进行讨论, 主要通过两个仿真实验来进行描述。首先, 本文针对信道所进行分割的时间片的数量对 RTDA 所涉及的各项参数的影响, 最主要的频谱拍卖相关的收益以及单信道的效用。其次, 本文通过仿真实验来比较 RTDA 与 TRUST 在信道的利用率方面的提高, 证明在实际的使用过程中, RTDA 的时间维度的频谱重用是有效的。

3.4.1 实验配置

在仿真试验中, 首先需要对以下三个涉及用户需求的空间时间分布的参数进行选择以覆盖一般的情况:

(1) 报价分布: 本文在实验中仿照 TRUST 机制进行假设, 任意买家 j 的报价 b_j^b 是在 $(0, 1]$ 之间随机分布的, 并且其值被设定为 $b_j^b = B_j^b / \|t_j\|_1$ 。对于任意的卖家 i 的报价是在 $(0, 2]$ 中随机分布。

(2) 时间片: 在所有试验中, 每个信道会被分割成等长的时间序列, 而每个时间序列由时间片组成。每个买家会随机请求至少一个时间片。

(3) 冲突情况: 在本文中假设用户的使用场景时随机分布的组群拓扑。首先随机的分布 50% 的买家在给定区域的中心, 从而创建了一个中心热点。然后, 在剩余的空间内再随机分布剩余的用户。在这个过程中, 用户之间可能会存在着距离上的冲突, 因此使用上文中提到的 SINR 与冲突图模型来解决冲突。

在仿真实验中, 主要使用了以下两个参数来衡量拍卖机制的表现:

(1) 信道利用率 (Usage Rate): 本文用每个信道被二级用户成功使用的时间片数量该信道所被划分的时间片数量的比值来表示该信道的利用率。同时, 所有信道的平均的利用率就是平均信道利用率。

(2) 单信道效用 (Utility): 按照 TRUST 中的定义, 本文将每个信道的效用也就是同时使用某一条信道的二级用户的数量 (Number of Users) 定义为单信道效用, 主要用来标识信道在实际使用过程中所发挥的效用。

(3) 拍卖收益 (Profit): 在实验中, 本文还考察了与频谱分配中经济特性相关的拍卖收益。在之前的章节中已经证明 RTDA 是具有预算平衡的特性的。因此在试验中, 也将考察时间片数量对整个拍卖收益也就是拍卖者的收益情况的影响。

3.4.2 实验结果

仿真实验主要包含时间片实验以及利用率实验。在试验中，将要考察的是时间片数量对拍卖的相关的重要参数产生的影响，其中最重要的就是频谱利用率、单频谱的效用以及整个拍卖过程所产生的效用。本小结中将对这三个数值进行分别的讨论，并给出预估的比较合适的时间片数量。

(1) 时间片实验：在时间片实验中，主要考察的是信道所分割成的时间片的数量（Number of Time Slots）的多少对信道相关的各个参数的影响。在试验中，信道被划分为 2 到 50 个不同的时间片们，并且分别观察信道利用率（Usage Rate）、单信道的效用（Utility）以及整个基于拍卖的频谱分配过程中所产生的收益。在实验中，考察 2 到 50 个不同的时间片数量，针对每个时间片数量分别进行了 1000 次的实验。在实验结果中，使用盒图来统计每次实验的分布情况，并对不同时间片数量下的参数的均值进行考察，如图 3.5 至 3.7 所示。

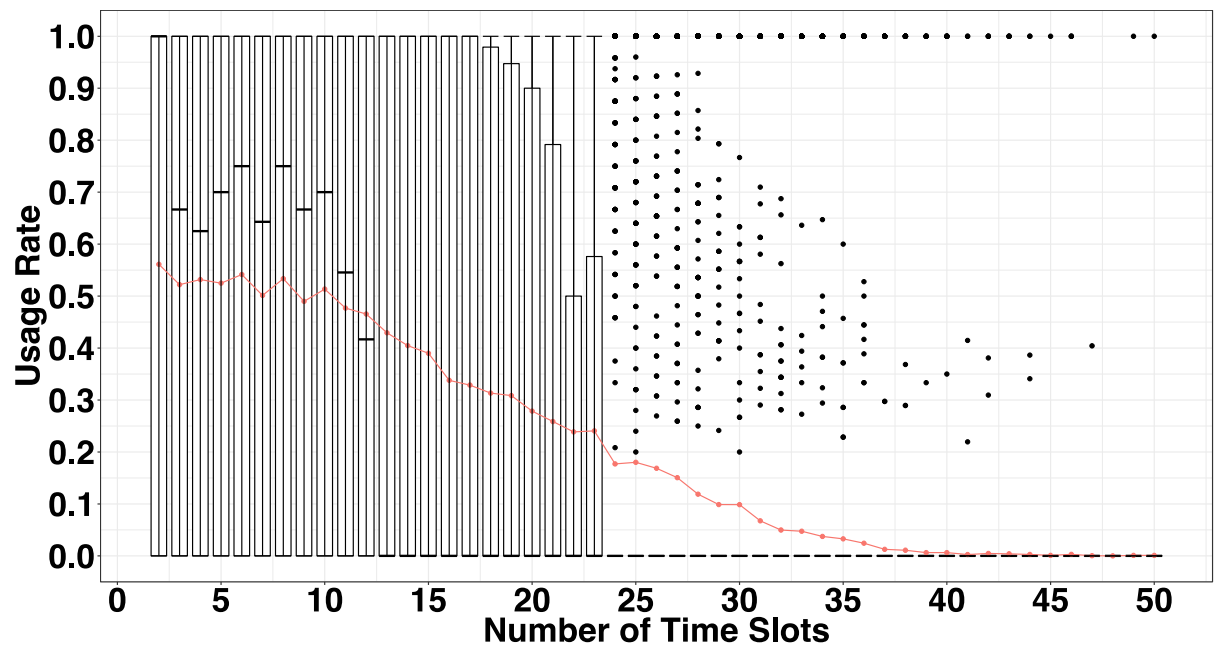


图 3.5 信道平均利用率

Fig 3.5 Average Usage Rate of Channel

如图 3.5 所示，平均的信道利用率（Usage Rate）随时间片数量变化的趋势是整体递减的。但是可以很清楚的发现，当把信道划分为 2 到 10 个时间片的时候是基本稳定的。如图所示，平均每个信道在时间维度上，被占用的情况大概是 50%左右，然而随着

时间片的增多,由于用户提交的时间片需求的增多,而且使用的是正态分布的随机请求,也就导致了冲突的增多。因此,平均的信道利用率是下降的。在实际生活中这个现象也是有迹可循的。比如说,在人们使用 WIFI 的繁忙时段,11 个信道的冲突是非常高的。但是在空闲时段,信道的利用率就较低了。而且可以看到,在 23 个时间片之后,实验结果基本上就是呈现两极分化的趋势,换言之,如果能够成功分配,基本上利用率就会接近 100%,否则就会分配失败。这个结果对于频谱资源以及网络服务提供商来说是有利的,因为这样既不会造成频谱资源的浪费也不会造成利益的损失。

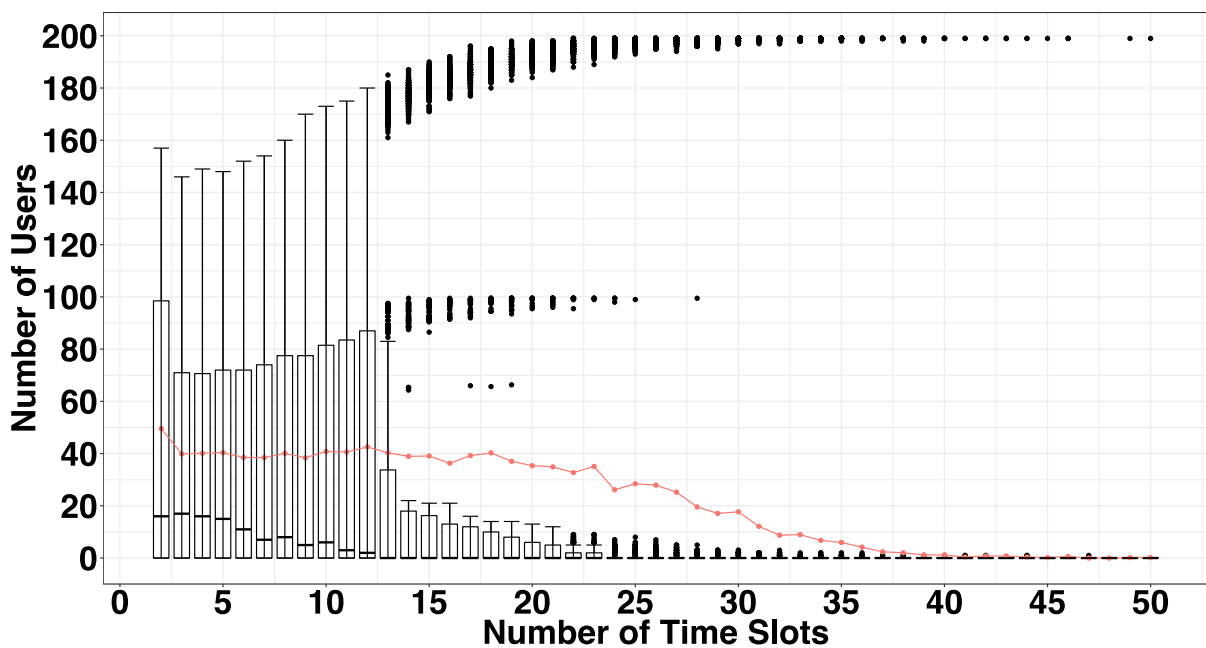


图 3.6 单信道平均效用

Fig 3.6 Utilization of Single Channel

如图 3.6 所示。与信道平均利用率相似,在 15 个分片之前,平均单信道的效用基本上是稳定的。在 2 到 25 个分片范围内,信道的效用是基本稳定的,也就是说,在这个范围内,用户正态分布的随机申请信道时间片,对 RTDA 的影响不大。但是随着分片数量的增多,单信道的在同一时间使用的用户也在减少。但是对于实际的分布情况来看,随着时间片数量的增多,频谱的利用率也会呈两极分化。这对于频谱资源来说无疑是有利的,这样未被分配的信道可以再次进行分配。因此,时间片数量的增多可以呈现出多样化的用户对时间维度的信道的需求,从而增加了其成功获得信道的概率。换言之,时间片数量的增多无论对信道还是用户来说,都是有积极作用的。

如图 3.7 所示，是在拍卖过程中，整个拍卖收益的情况随时间片的增多的变化无论从整体的平均趋势还是从不同时间片数量的分布趋势都是减少的。从图中红线可以看到，基本上其变化是符合一个线性的表现。但这个收益并不代表网络服务提供商作为卖家的收益，是整个拍卖产生的社会效用。因此，在 RTDA 的工作中，本文定位于保证中立性的情况下面向用户的资源分配，并且尽可能的提到频谱的利用率。所以说，对于总收益的降低是可以容忍的。因为在满足用户的资源需求的同时，对于收益肯定是有相对应的权衡的。而且，结合图 3.5 与图 3.6 中的结果，就可以看出，对于时间片的数量肯定是越多越有利于用户的获得自己需要的信道，同时信道的利用率也会大大提高。但是对于拍卖者来说，例如 FCC，其利益的降低也是明显的。但是 RTDA 可以保证的是，无论时间片的数量为多少，整个拍卖都一定是符合预算平衡的需求的。因此从信道和用户角度来看，时间片的选择还是尽可能多才有利于提高资源的利用率，实现用户层面的中立性需求。

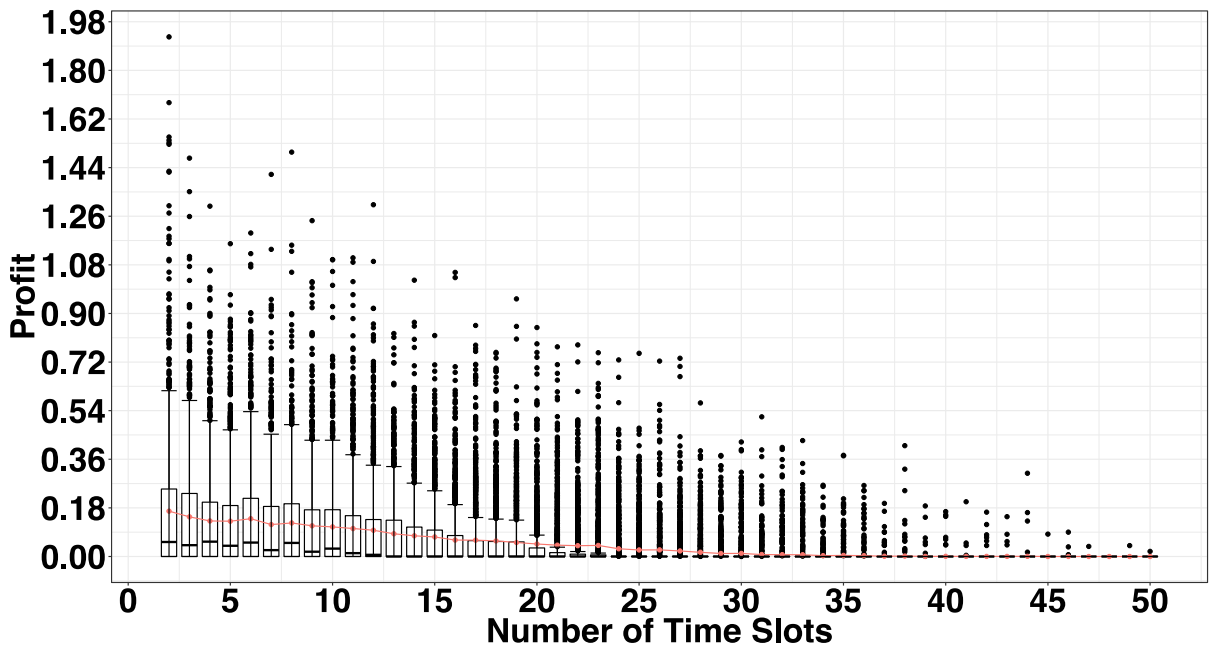


图 3.7 拍卖收益

Fig 3.7 Profit of Auctioneer

(2) 拍卖机制比较：如表 3.5 所示，相比较与其他的拍卖算法，RTDA 在保证以真实性为基础的公平性之外，还保证了实现了频谱在空间以及时间上的多维重用，从而提高了其利用率。

表 3.5 频谱拍卖机制比较

Tab. 3.5 Comparison of Auction Mechanisms

名称	频谱重用	真实性	预算平衡	个体理性
VCG、McAfee	未实现	实现	未实现	实现
U-Auction	空间	实现	实现	未实现
VERITA	空间	实现	实现	实现
TRUST	空间	实现	实现	实现
RTDA	空间、时间	实现	实现	实现

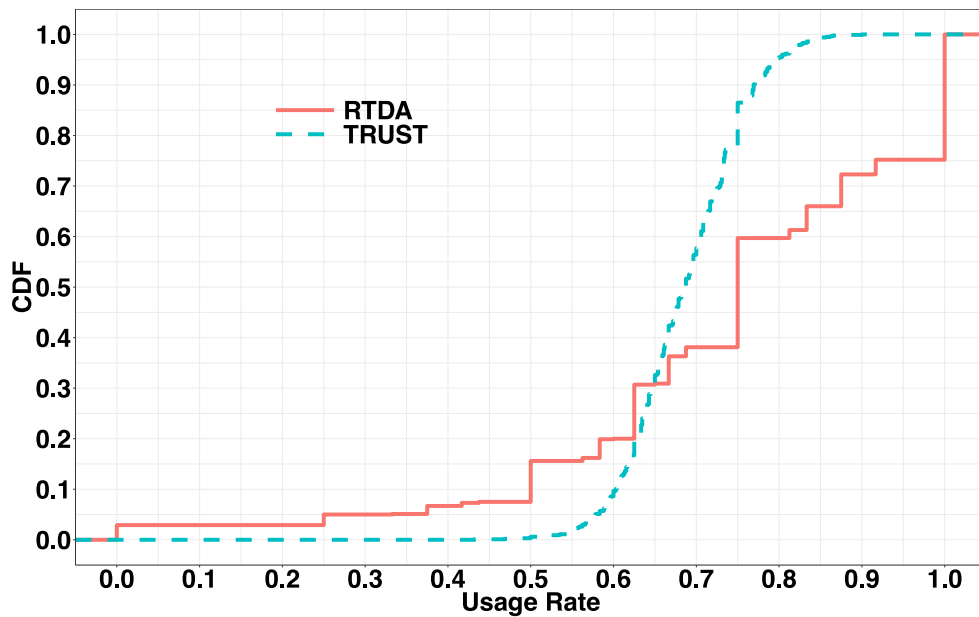


图 3.8 信道利用率比较实验

Fig 3.8 Usage Rate of Channels compared with TRUST

(3) 利用率对比实验：如图 3.8 所示，通过比较 RTDA 与 TRUST 对信道的利用率来比较证明 RTDA 的差异。在 50% 以上的信道利用率部分，TRUST 的信道利用率基本稳定在 70% 左右，呈现出正态分布的表现。但是 RTDA 很明显的在 80% 以上的信道利用率区域是有明显的优势的。虽然图中较低的利用率中 RTDA 也存在分布，而 TRUST 不存在。出现这样的问题，其主要原因就是较少的时间片分片数量导致的，因为按照这样的计算方式，RTDA 的时间片分片导致了 0% 利用率的分片的存在，从而导致了这种结果，因此并不能证明 RTDA 的表现一定低于 TRUST。但是高利用率的部分却可以证明 RTDA 的表现一定是优于 TRUST 的。

4 基于中立性的网络诊断

保证用户在网络使用过程中能够得到中立性，首先要做的就是能够完全依靠用户可以获得的信息进行网络诊断。没有任何的网络服务提供商愿意提供真实的数据来帮助用户对自己的中立性做出判断，因为这是与其自身利益相违背的。而通过用户层面获得的信息是有局限性的，比如出于安全原因，网络服务提供商会对内部的路由器进行匿名处理等等。但是与中立性相关的诊断的精度可以是粗粒度的，因为用户需要确定是哪一段网络发生了问题，而不是传统网络诊断中出问题的具体的链路。尤其是对于复杂的多跳骨干网来说，地理位置层面的映射或者自治域系统级别的映射就是足够的。因为一旦用户可以在这个层面上确定的找到问题点，那么就足够可以反馈给网络服务提供商来完善自己的网络。针对这个问题，本文提出了 **HIDA**，主要利用中立性相关的特点对用户通过 **Traceroute** 获得网络路径在不同的网络映射中进行松弛，也就是将端口级别的网络路径映射到不同级别的网络映射上，从而实现粗粒度的、对用户来说实用的中立性相关的诊断。

4.1 问题模型抽象

4.1.1 基于层析技术的网络诊断

为了区别于 **RTDA** 中的符号以及方便对网络使用过程中的中立性问题进行描述，本文使用了如表 4.1 中所示的符号，来表示相应的网路路径的有关变量：

表 4.1 HIDA 相关符号与解释
Tab. 4.1 Notations and Explanations of HIDA

符号	解释
$link$	一段任意的网络链路或分段
$segment$	一段任意的网络路径分段
$Path_{ij}$	从客户端 <i>i</i> 到服务器 <i>j</i> 的任意的网络路径
$\phi_{Path_{ij}}$	路径 $Path_{ij}$ 在端到端测量中的表现，例如丢包率等
λ	任意无环的 $link$ 序列
Π	Π_λ 为任意路径对 $Path_{il}$, $Path_{ij}$ 相交为 λ 的集合的幂集合
Λ	Λ_n 为 <i>n</i> 个 $link$ 组成的 λ 的子集

在层析诊断技术中，主要依靠的是端到端测量和路径的抽象化处理。给定了网络拓扑与端到端测量的丢包率，就可以很方便的建立其每个链接 $link$ 与路径 $Path_{ij}$ 之间的线性关系如 4.1 所示。

$$\log(\phi_{Path_{ij}}) = \sum_{k=1}^n \log(\phi_{link_k}), \forall link_k \in Path_{ij} \quad (4.1)$$

当把丢包率通过阈值分类法进行分类之后，那么就可以将线性问题转化为布尔问题。例如，当 $\phi_{Path_{ij}}$ 高于某个阈值的时候，就可以认为 $Path_{ij}$ 是有问题的，反之就是正常的。因此公式 4.1 就可以转化为：

$$y_{Path_{ij}} = \bigvee_{k=1}^n x_{link_k}, \forall link_k \in Path_{ij} \quad (4.2)$$

其中， $y_{Path_{ij}}$ 与 x_{link_k} 分别代表了 $Path_{ij}$ 与 $Path_{ij}$ 的路径状态， $\bigvee x_{link_k}$ 则代表了二进制或运算。因此在用户层面的网络诊断中，就可以将问题抽象成一个布尔问题，从而确定出问题的发生位置。如图 4.1 所示，就可以将原本的测量工作简单的转化为诊断方案，只要用户可以得到相应的拓扑信息与端到端测量。但是对于终端用户来说，如果想要在不依靠其他的手段独立的完成这部分工作的话，就需要首先要通过客户端工具例如 Traceroute 获得网络的拓扑信息，然后在利用现有的设施来进行相应的端到端测量。

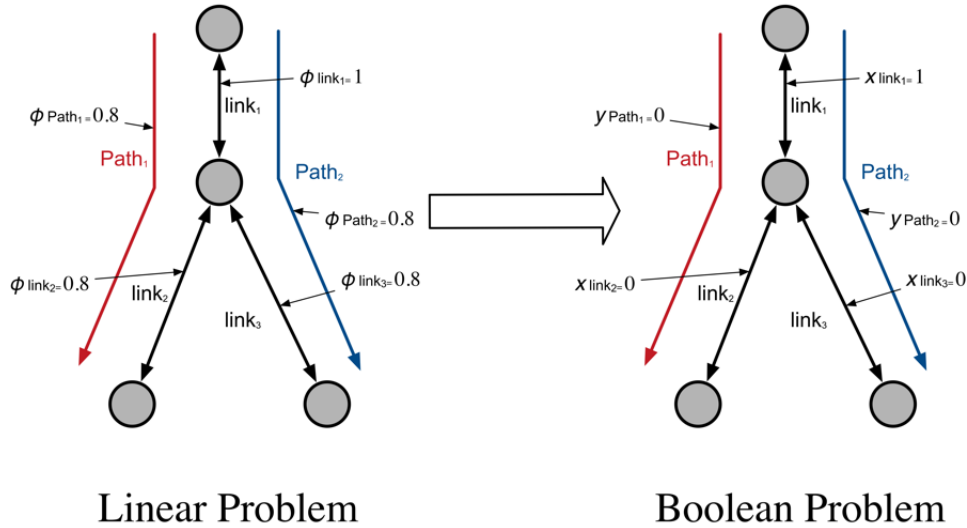


图 4.1 网络服务提供商级别的路径松弛

Fig. 4.1 ISP-level Relaxation

4.1.2 基于位置信息的网络映射

由于对于终端用户来说，最重要的不是判断具体的产生问题的链路，而是弄清楚具体出现问题网络服务提供商或者互联互通部分即可。所以说，利用路径松弛的方法，虽

然损失了一定的精确度，但是在用户层面的网络诊断中，是可以接受的。同时，网络服务提供商级别的路径松弛对于用户来说也是足够的。举例来说，即使实施了更高精确度的网络诊断例如路由器级别的诊断，用户也不能在路由器级别实施任何的可以改变网络状况的解决方案。同时，更高的精度代表着用户更高的实施难度。而这样的代价就代表着实用性的降低。反而，网络服务提供商级别的诊断，对于用户来说已经是可以接受的。由于诊断难度的降低，用户可以方便快捷的实施诊断，同时把诊断的结果反馈给网络服务提供商，这样被诊断处存在问题的网络服务提供商就可以利用自己更加精确的网络管理工具对网络进行诊断修复。所以说，损失一定的精确度却获得了更好的实用性，对于信息匮乏的终端用户来说是非常值得的。

基于以上的理论，用户层面的诊断就可以在一个粗粒度的精度上来进行，也就是路径松弛。通过松弛来获得一个精确度较低的路径，最主要的就是地理位置级别或网络服务提供商级别的路径。因此，在本文中主要使用了以下两种网络路径松弛的方式：

(1) 第一个层面，使用的是基于网络服务提供商级别的路径松弛。在网络服务提供商级别的路径松弛中，每个网络服务提供商以及每个互联互通部分都被分割成了一个路径 $segment$ 。如图 4.2 所示， $Path_{ij}$ 是经过两个网络服务提供商以及一个互联互通部分。从客户端到 City₃ 结点就可以被认为是一个 $segment$ 。

(2) 第二个层面，使用的是基于地理位置级别的路径松弛。每个 $segment$ 都划分为了许多地理位置级别的 $link$ 。并且，每个地理位置，也就是城市，作为一个结点，如图 4.2 所示， $Path_{ij}$ 就包含了 7 个 $link$ 以及 5 个地理位置级别的中间结点。

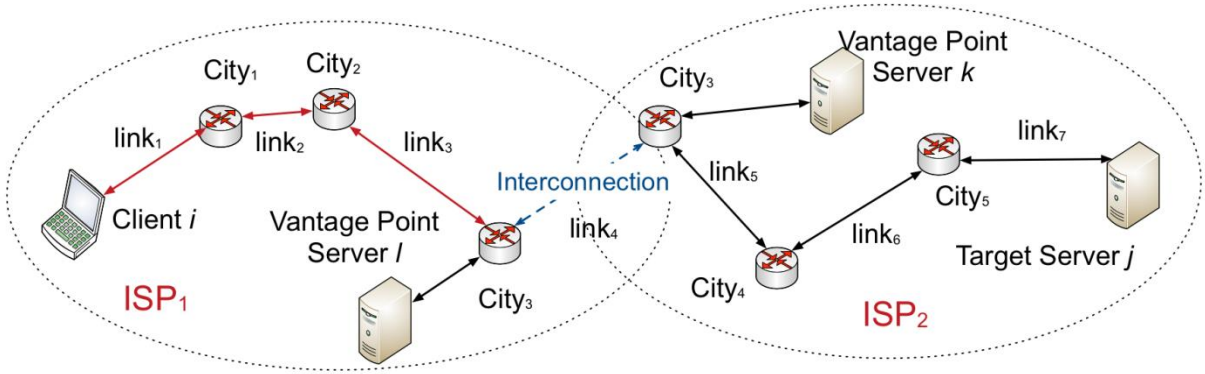


图 4.2 网络服务提供商级别的路径松弛

Fig. 4.2 ISP-level Relaxation

针对这个问题，本文提出了 HIDA，通过路径松弛的方法来解决 Traceroute 的相关问题，通过基于地理位置信息的启发式方法帮助用户发现可用的 Web 服务器，以便独立进行层析诊断工作。如图 4.3 所示，HIDA 与 RTDA 相类似，也包含三个主要部分。目标路径生成（Target Path Generation）会根据用户 Traceroute 的结果生成地理位置级别与网络服务提供商级别的网络映射中的目标问题路径。观测点选择（Vantage Point Selection）使用了地理位置信息与启发式的服务器发现策略为用户选择相应的可测量的 Web 服务器。在进行了端到端的测量之后，就可以有目的的进行故障点的层析推断（Tomography Inference），从而帮助用户确定发生问题的故障网络段。HIDA 会根据用户提供的需要进行诊断的目标地址，可以是 URL 也可以是 IP 地址，来选择合适的可用的 Web 服务器，然后推断出可能导致问题发生的网络段。

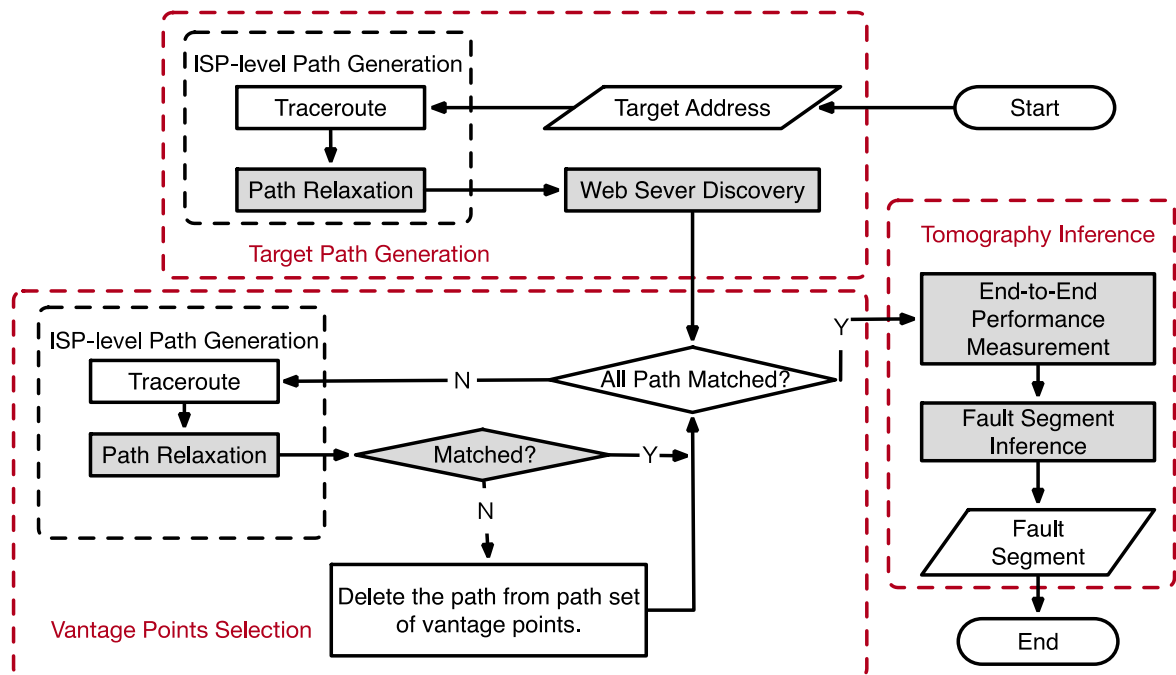


图 4.3 HIDA 工作流程图

Fig. 4.3 HIDA workflow

4.2 网络映射与路径松弛

在本文中，进行如上所述的路径松弛是有根据的。同时，HIDA 主要进行的是网络服务提供商级别以及地理位置级别的两个层面的路径松弛，主要是出于以下几个原因：

(1) 自治域系统中的划分：城域网（Metropolitan Area Network, MAN）在自治域系统中一般是一个管理的单元，也就是说网络服务提供商的网络在不考虑其他特定情况下是以城市为一个最小管理单位的。关于这一点，可以从中国互联网络信息中心（China Internet Network Information Center, CNNIC）提供的报告中得到相关的依据。

(2) 网络诊断的实用性：路由器级别的诊断是对于终端用户来说是不实用而且没有必要的。为了搞清楚出问题（主要是拥塞问题和连接性问题）的路径分段 $segment$ ，用户们也有理由把网络服务提供商通过松弛的方式区分开来。

(3) 别名（Aliases）问题的解决：一台网络设备，例如一台路由器是包含不止一个端口的，而每个端口可能都分配了一个地址。这也就造成了路由器级别诊断过程中的别名问题。但是，一台设备不可能存在于两个地理位置上，如果把地理位置作为一个结点的单位，那么也就是通过地理位置级别的路径松弛解决了别名问题。

4.2.1 自治域系统映射中的路径松弛

在网络服务提供商级别的网络映射中，如图 4.2 中路径 $Path_{ij}$ 就被划分为了三个网络服务提供商级别的 $segment$ ，并且分别包含了不同数量的地理位置级别网络映射中的 $link$ ：

(1) $Client_i - ISP_1$ ：该网络段涵盖了 ISP_1 的网络，包括了从客户端 $Client_i$ 到 ISP_1 的网络的出口路由器（Egress Router），包括了之前提到的 $link_1$ 、 $link_2$ 以及 $link_3$ 。

(2) $ISP_1 - ISP_2$ ：这段包括了从 ISP_1 的网络的出口路由器到 ISP_2 的网络的入口路由器（Ingress Router）。这段路径段主要表示的是网络服务提供商间的互联互通部分，同时如果这段网络出问题，也就是代表着在 IXP 处进行的数据交换发生了问题，也就是图中的 $link_4$ 。

(3) $ISP_2 - Server_j$ ：这段路径段主要表示的是从 ISP_2 的网络的入口路由器到服务器 $Server_j$ 的路径段。这段路径主要表示的是 ISP_2 的网络，其中包括 $link_5$ 、 $link_6$ 以及 $link_7$ 。

4.2.2 地理位置映射中的路径松弛

当把路径进行分段之后，用户就可以按照之前讨论的把路径 $Path_{ij}$ 中的各个网络服务提供商以及有关数据交换的互联互通部分标记了出来，这就完成了网络服务提供商级别的路径松弛。但是仅仅这样是不够的，因为还要在接下来寻找不同路径之间的重叠部分。除了之前提到的原因之外，为了辅助路径重合的判断，要进行地理位置级别的路径松弛。

每个地理位置，也就是城市被看成是一个路径中的结点。每个结点不再是有单独的 IP 地址，而是存在一个地址池，而地址池包含所有该地理位置结点的 IP 地址。同时，只有完全相邻的地理位置才会被视为统一结点。比如说，如果路径为城市 A 到城市 B 再到城市 A，那么在城市 A 的 IP 地址会被相应的分为两个结点。通过这种方式，路径就进行了地理位置级别的松弛。

4.2.3 相关问题与解决方案

虽然上文中提到的路径松弛方案，是很容易理解的，但是在实际的实施过程中还存在着许多的实际问题，比如 Traceroute 超时问题，负载均衡问题，互联互通部分的检测问题以及路径重合的查找等等。

(1) Traceroute 超时与匿名路由器：因为 Traceroute 使用的是多个端口的 ICMP 的返回信息叠加从而描绘出整体的网络路径的方式，随意每个端口出现的问题都会反应在路径上，影响用户的判断。

在实际生活中，网络服务提供商出于各种各样的安全问题，对路由器级别的信息加以保护。从而导致匿名路由器(Anonymous Router)的出现，也就意味着单纯由 Traceroute 所获得的信息构建出的网络路径是不完整的。在用户层面的网络诊断中，有两个问题是需要关注的：

① 单纯使用 IP 地址作为路径匹配的标准是实用性很差的，因为匿名路由器可能会导致路径的不一致，违反实际情况。

② 而在地理层面的路径匹配是相对使用的，因为在一定程度上解决了匿名路由器带来的影响。

(2) 数据交换的检测：在路径松弛的过程中，还需要准确的检测出所有的网络服务提供商之间的数据交换，才能准确的定位网络服务提供商的互联互通部分。在 HIDA 中，通过检测出 IP 地址所属的网络服务提供商来检测数据交换的发生。HIDA 使用的是自治域系统号码(Autonomous System Number, ASN)。因为 ASN 是各个网络服务提供商作为自治域系统的唯一标识。并且这些代码都是可以公开的查询到的。比如说 CNNIC 就向所有的用户公开开放了有关所有网络服务提供商的 ASN 与对应的 IP 地址段。因此，HIDA 就可以在避免任何较高权限的情况下，查询到 ASN 与对应的 IP 的变化，从而检测出相应的数据交换的发生。需要注意的是，在检测数据交换之前，必须要进行负载均衡在 Traceroute 中的相应的处理，否则，基于这种方法的数据交换的检测，会受到影响，并产生误差。

(3) 负载均衡：在本文中研究的方向是基于用户层面的网络诊断。而 Traceroute 是用户端使用最广泛，也是使用最方便的诊断工具之一。因此，为了使用户避免任何的更高权限的操作，本文使用 Traceroute 来获得用户视角的网络路径。但是，基于 Traceroute 的路径产生面临的最大的问题就是负载均衡带来的与实际路径的误差。

负载均衡，按照网络的整体情况来影响每个数据包在路由过程中所经过的端口，能够有效的解决拥塞、增加吞吐量^[41]。因此，在互联网中，存在这许多负载均衡结点以及等价路径。而这个实时改变的等价路径反而造成了网络诊断过程中的个别问题。

上文中提到，Traceroute 是基于发送不同的 TTL 的多个数据包，从而模拟出实际的某个流量所走的实际路径。但是在实际生活中，负载均衡问题是广泛存在的。

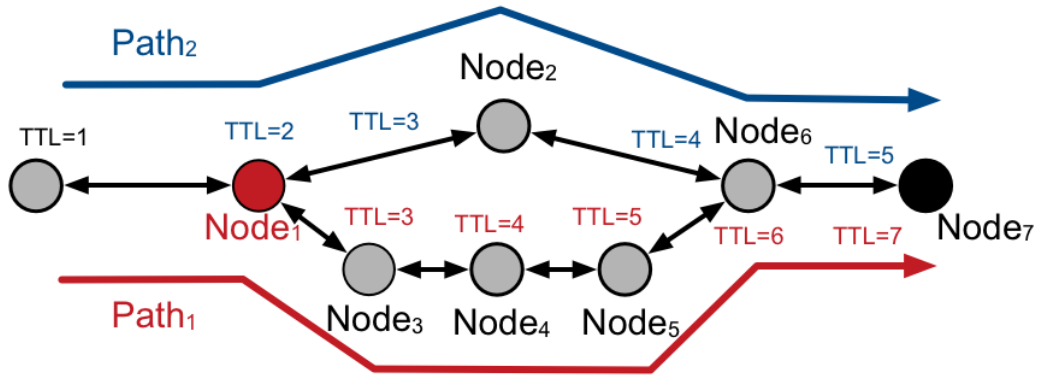


图 4.4 负载均衡问题

Fig. 4.4 Load Balancing Problem

如图 4.4 所示， $Node_1$ 、 $Node_2$ 、 $Node_3$ 、 $Node_4$ 、 $Node_5$ 以及 $Node_6$ 位于城市 A，而 $Node_7$ 位于城市 B。其中，假定 $Node_1$ 为负载均衡结点，也就是说数据流量会根据链路的拥塞情况分别选择 $Path_1$ 或 $Path_2$ 。如果数据流量经过 $Path_1$ 到达目的服务器，而由于 Traceroute 的工作机制，很有可能会存在侦测的数据包在 $Path_1$ 与 $Path_2$ 中来回转换。比如说，TTL 为 5 的侦测数据包走的是 $Path_2$ ，而 TTL 为 6 的数据包经过的是 $Path_1$ ，那么 $Node_7$ 就会在 Traceroute 的输出中出现在 $Node_6$ 之前。从而也就导致了原本在地理位置级别松弛方法中实际为城市 A 到城市 B 的路径变为了城市 A 到城市 B 再到城市 A。这也就是让用户在寻找路径重叠部分过程中，出现偏差。因为按照 HIDA 中的方法，中间间隔城市 B 的两个城市 A 结点，不会被作为同一个结点，而在路径重合中提到的地址池也就不会出现匹配。

在解决这个问题过程中, HIDA 使用了 OSPF 路由算法的特性来解决。首先, OSPF 路由协议首先都被设计为无环的。因为, 在实际测量过程中也发现, 有可能会出现不连续的两个相同的 IP 地址, 比如说, 如图 4.4 所示, 在 $Path_1$ 中 TTL 为 6 的 $Node_6$ 也有可能在 $Path_2$ 中 TTL 为 4, 也就说, 很有可能出现 TTL 为 6 与 4 的结点在 Traceroute 的输出中全都为 $Node_6$ 的地址, 而恰好, TTL 为 5 的结点为 $Node_7$ 。因为, 实际的路径中, 是不可能存在环路的, 否则违背了其设计原则。所以用户就可以把 $Node_6$ 在 Traceroute 中的位置提前, 因为每个结点 HIDA 都选择最早出现的位置, 就会得到一个相对符合拓扑顺序的结点。这样的话, 就可以在一定程度上解决负载均衡所带来的影响, 从而使用户能够获得一个比较稳定, 相对正确的网络路径。

4.3 基于 DTW 的路径匹配

在进行 Web 服务器发现之前, 首先要知道的是在路径松弛中, 什么样的路径以及相应的服务器是可以用来辅助进行层析推断的。首先测量的路径必须与目标路径存在重叠部分; 其次, 这个服务器一定是可以进行端到端测量的, 这个那么这个服务器才算是可用的服务器。由于大量的工作已经证明基于 HTTP 的 Web 服务器是可以测量的, 那么在本文中, HIDA 的主要工作就是要帮助用户发现与目标路径存在重叠部分的服务器。

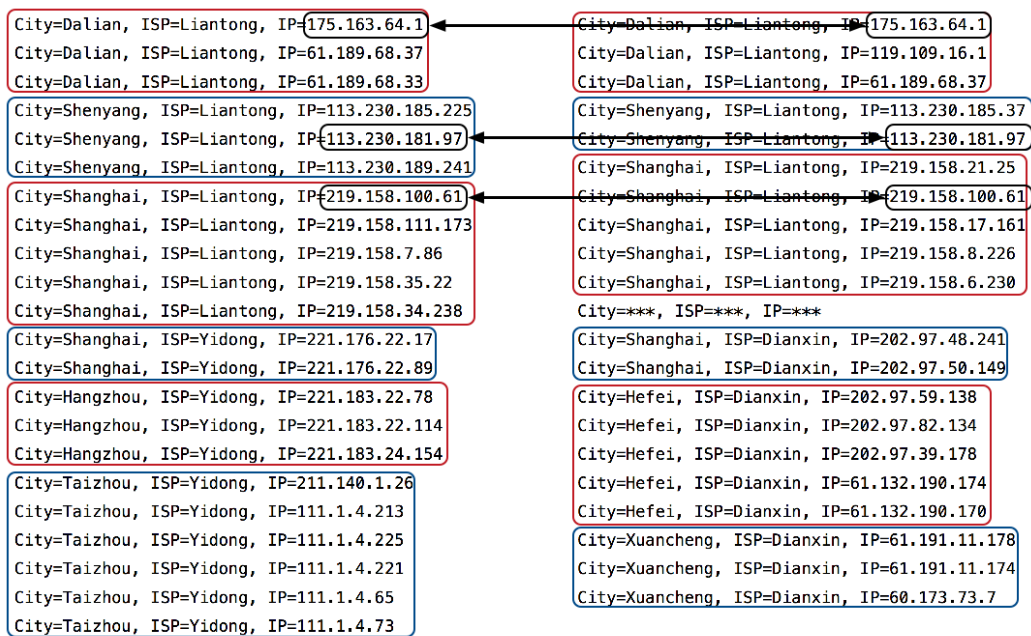


图 4.5 地理位置级别的路径匹配

Fig.4.5 Geolocation-level Path Matching

在传统的层析诊断中，未进行路径松弛的网络路径是用 IP 地址作为每个结点的标识的，从而相应的，也应该用 IP 地址来进行路径的匹配。而在路径松弛中的路径匹配，在每个结点 HIDA 维护了一个 IP 地址池保存该地理位置结点中的 IP 地址。如图 4.5 所示，HIDA 使用这些相应结点中的地址池中的地址作为每个结点匹配的依据。在地理位置级别网路映射中的路径匹配，HIDA 使用的是基于 DTW 思想的路径匹配方法。如图 4.5 所示，每个地理位置都被划分为了一个结点，而这些结点中都有一个地址池。具体的匹配的规则如下所示：

(1) 如果两条路径能够在相应的结点的地址池中，找到至少一个完全相同的 IP 地址，那么这两个结点就是匹配的。

(2) 如果两条路径，在分界点（也就是需要判断的路径端的终点）之前的所有结点是匹配的，那么这两条路径就是在相应的 *segment* 上是匹配的。

因为中立性要求，对于所有的对网络合法的流量都要平等对待，不能阻塞，截流或者有差异化。由于 HIDA 涉及的方法是针对网络服务提供商层面的诊断，就有必要将中立性相关的内容考虑进来，因为现有网络下，网络服务提供商会利用深度包检测等技术，做出一些违反网络中立性的行为，其中也就包括区别对待不同的网络服务提供商的流量。如图 2.2 中的 $Path_{il}$ 与 $Path_{ij}$ 之间的唯一区别就是其网络流量是否发生了网络服务提供商之间的数据交换（是否经过了互联互通的部分）。接下来要利用引理 1 与引理 2 以及上面提到的内容，对定理 1 进行证明。

定理 4: $Path_{il}$ 与 $Path_{ij}$ 在与中立性有关的地理位置级别的路径松弛中是等价的。

引理 5: 如果 $Path_{il}$ 与 $Path_{ij}$ 是等价的，那么他们一定存在相同的起点结点和终点结点。

引理 6: 如果 $Path_{il}$ 与 $Path_{ij}$ 在地理位置级别是等价的，那么他们在网络服务提供商级别映射中也一定是等价的。

证明: 在 DTW 中，有一个重要的原则就是信号的开始和结束部分一定要匹配。相应的，HIDA 在路径匹配中也是这么做的，在被匹配的每一段路径段中，一定存在相同的起点 IP 地址与相同的终点 IP 地址。因为一旦存在相同的起点与终点，那么这两两个路径段在中立性满足的情况下就一定是等价的。如图 4.5 所示，每个不等长的路径都被线性缩放为了等长的地理位置级别的路径，每个地理位置的地址池中可能包含不同数量的 IP 地址，但是采用递归的思想，只要每个对应的地理位置级别的路径段可以对应匹配，那么 HIDA 需要考察的网络服务提供商级别的 *segment* 就可以对应匹配。从而证明了机遇 DTW 的路径匹配得到的重合部分一定是匹配的，从而该方法是正确的。

(1) 在地理位置级别等价：根据引理 6，如果 $Path_{il}$ 与 $Path_{ij}$ 在地理位置级别是等价的，那么他们在网络服务提供商级别也就是等价，那么任意的流量都会被平等的对待，也就是说如果出现问题，一定是两条路径同时出现问题。从而可以证明在网络服务提供商级别的路径松弛中是等价的，定理 4 成立。

(2) 在地理位置级别不等价：根据引理 5， $Path_{il}$ 与 $Path_{ij}$ 是有可能在网络服务提供商级别等价的，如果不等价，那么这两条路径上，因为都是来自于基于 HTTP 的应用，那么如上所述，他们唯一的区别就是是否存在网络服务提供商间的数据交换。也就是说当检测到是经过互联互通部分的流量时，即使是存在相同的起点和终点，那么不同的流量也是会按照特定的路径进行传输。而这种情况，可以等价为网络服务提供商内部网络的两条等价两路外加出现问题的互联互通部分。也就是说，不等价性导致的问题，完全可以被认为是数据交换出现了问题，因为流量被违反了中立性原则，区别对待了。那么就证明，路径是等价的，定理 4 成立。

(3) 地理位置级别的匹配：之所以要使用地理位置级别的匹配，是因为，如果不是在中间结点中匹配地理位置，就不能保证流量是存在于同一个网络中。

综上所述，就可以证明定理 4 是成立的。本文还进一步假设，在网络服务提供商级别的路径中，只要在首尾结点能够进行地理位置结点的匹配，而中间结点只要进行更底层次的地理位置的匹配就可以视为路径实现了匹配。因为该方法可以用来与 HIDA 所使用方法进行比较，但是在实际使用中是无法满足要求的，因此就不进行相应的证明。

4.4 基于位置信息的路径诊断

根据 HIDA 的设计，本文提出了一个启发式的服务器发现策略。首先，假设已经目标路径 $Path_{ij}$ 被划分为了相应的路径段。用户就可根据相应的地理位置信息以及网络服务提供商的信息来相应的发现服务器。

4.4.1 启发式发现策略

例如图 4.2 所示的场景，如果要对路径 $Path_{ij}$ 进行诊断的话，HIDA 就会相应的选择位于地理位置 $City_3$ 并且使用 ISP_1 网络的服务器，并验证是否匹配。假设存在一个网络路径经过了中国联通与中国电信的网络，客户端位于大连市的中国联通网络中，服务器位于广州的中国电信中。对应这个路径的两个网络服务提供商的数据交换发生在上海市。因此，HIDA 就会优先选择地理位置位于上海市的 IP 地址对应的使用中国联通网络的服务器以及位于上海市的中国电信的网络的服务器来进行辅助的层析的判断。相当于 HIDA 对于服务器的搜索划定了范围，在一个规定的更小的范围内进行优先的搜索。因为，可以确定的是，能够使用的服务层析诊断来进行网络服务提供商级别的服务器一定

是位于上海并且分属于两个不同的网络的服务器。所以说，只要在这个范围内进行搜索，理论上就会对成功的可能性有很大的提高。

但是，如果按照这个策略没有找到相应的匹配的路径的话，那么 HIDA 就会使用经验化方法来进行匹配。在传统的层析技术中，一般的经验化方法就是通过搜索用户使用过的历史路径来进行匹配。也就是说，经验化的方法会通过暴力搜索的方式对用户所使用的网络路径进行搜索。但是可以发现，随着用户使用时间的增长，这种方式会不断的增加搜索的时间。在实验中，本文也实现了这个经验方法，与 HIDA 中使用的启发式的发现策略来进行比较。

4.4.2 基于层析的问题路径推断

为了验证在查找到合适的服务器之后，用户是否能够快速准确的定位到发生问题的 *segment*, HIDA 使用的是最简单的一种基于投票的推断算法。现有的算法例如 CLINK，能够更好的定位可能发生问题的 *segment*，但是这并不是本文所讨论的重点。与这些方法不同的是，在 HIDA 使用的是针对相应的问题的链路，然后选择相应的可以用来检测的服务器，然后再进行测量的方式。相反，传统的层析诊断的方法是先进行大规模的测量，然后再根据端到端的测量结果以及现有的路径的信息来进行层析的诊断。所以说，两者是有本质的区别的。如表 4.2 所示，是该算法的伪代码

表 4.2 问题路径段推断算法

Tab. 4.2 Problematic Path Segment Inference Algorithm

算法： 问题路径段推断算法
输入： 测量 $Path_{il}$ ，路径的端到端测量 $\phi_{Path_{ij}}$ ，目标路径 $Path_{ij}$
输出： 问题路径段 $link_p$
1: 投票集合 $\pi = \{0\}$
2: for $i \in Path_{il} \cap Path_{ij}$ do
3: if $\phi_{Path_{ij}}$ 发生了拥塞等问题 then
4: $\pi_i ++$
5: else
6: $\pi_i = 0$
7: end if
8: end for
9: return $argmax_{i \in Path_{il}} \pi_i$

5 面向用户的网络测量与验证实验

为了验证 HIDA 的功能以及假设, 本文还设计和实现了一个面向用户的轻量级的测量工具。本文还在不同的城市部署实施了相关的测量工作, 其中包括大连、秦皇岛、鞍山和盘锦等。在测量实验中, 不仅获得了用户层面的真实的测量数据, 还得到了一些新的关于网络服务提供商间网络的互联互通的观测结果以及地理位置信息与网络路径之间的关系的结果。这些观测结果不仅验证了 HIDA 的假设, 而且实测的实验数据也为今后相关的工作打下了基础。

5.1 基于用户的测量工具

如图 5.1 所示, 本文设计了一个用户级别的轻量级的测量工具。该工具可以很方便的部署在客户端机器上, 并且不需要高级用户权限。

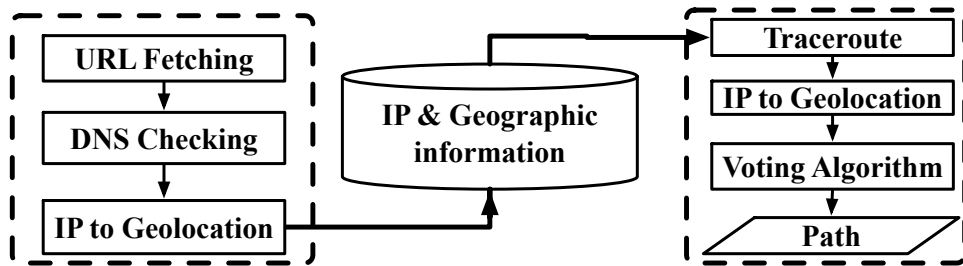


图 5.1 用户级别的测量

Fig. 5.1 User-level Measurement

该测量工具基于 Python 与 Traceroute 的用户级的测量工具, 不仅能够实现用户端轻量级的测量, 而且考虑到了各种各样的问题处理机制, 比如网络中断的恢复、基于 DNS 的实时地址解析等等。该工具可以实现在任意运营商网络下进行地理位置信息与网络路径相关的测量。

除此之外, 为了让用户在客户端能够获得尽可能多的可以使用的 Web 服务器的信息, 本文同样设计和实现了在 HIDA 以及测量工具中都使用的网络爬虫。然后使用 Traceroute 将到达选中的服务器网络路径推算出来, 并使用 HIDA 中的地理位置获取的方法以及上下文相关的纠正算法, 来获得这些网络路径与地理位置的关系, 并且将所有地理位置与互联互通部分标识出来, 并使用 R 语言进行可视化的数据分析, 得到相关的结果。

5.1.1 网络爬虫

在 HIDA 以及测量实验中, 让用户发现可以使用的服务器的前提是要尽可能多的发现服务器同时要覆盖更多的地理位置。因此, 本文设计实现了一个基于 Python 的网络爬虫, 来帮助用户抓取网络中的网络地址。

在该网络爬虫中, 使用了实时 DNS 查询的方法, 来确定具体连接的 IP 地址以及模拟用户的实际网络访问情况, 从而尽可能真实高效的收集可用服务器。在实际的使用过程中, 用户经常使用的导航网站就是一个比较好的获取服务器信息的资源。比如大学排名、政府门户、银行信息等等都包含了丰富的地理位置信息。因为在之前的实验中发现, 许多带有地理位置特点的网站比如省市政府的主页以及教育机构例如大学的主页大部分是与其所具有的地理位置特色相对应。例如说, 大连市政府门户网站的服务器就部署在大连市, 更重要的是, 大连市政府门户网站同样也代表了大连市这样一个地理位置。因此, 本文在实际的测量实验过程中主要选取了如表 5.1 所示的三种与地理位置密切相关的主页及服务器类型:

表 5.1 测量服务器选择
Tab. 5.1 Selection of Measured Web Servers

符号	解释
.gov	政府部门的主页及服务器
.edu	各类院校的主页及服务器
.com	部分的商业网站, 主要由各类地方银行的主页构成

5.1.2 IP 地址与地理位置

为了探究 Internet 与地理位置信息之间的关系, 需要进行 IP 地址与地理位置信息的转换。本文使用的是 IP 与地理位置数据库来完成, 比如 IPIP.net、MaxMind 等等。但是, 现阶段, 并没有一个百分之百完全准确的数据库来供用户使用^[42]。同时, IP 地址的动态转换, 也不允许用户使用单一的地理位置数据库来进行匹配。因此, 本文提出了一个如表 5.2 所示的基于投票的地理位置判定算法。在匹配过程中, 使用了基于上下文的方法, 也就是说如果在数据库中存在这个地理位置, 在上下文临近节点中如果出现了相同的地理位置, 那么应该优先选择这个地理位置。否则的话就会使用多个数据库中得票最多的地理位置。这么做的理由是:

(1) 不存在百分之百完全正确的数据库, 因此多个数据库存在的信息在用户角度看来都有可能是真实对应的地址。

(2) 如果按照上下文中的地理位置进行匹配, 在地理位置级别的路径松弛与网络映射中也不会影响整个路径的准确度。

关于网络服务提供商的识别问题, 使用了从 CNNIC 获取的 ASN 的信息来匹配相关的网络服务提供商。而且需要注意的是, 即使地理位置相同, 如果 IP 属于不同的网络服务提供商, 也必须被区分开来。

表 5.2 IP 地址与地理位置转换算法

Tab. 5.2 IP to geolocation Transferring Algorithm

算法: 上下文相关的 IP 与地理位置转换

输入: IP 地址 ip , 数据库集合 δ , 地址 ip 在路径中的上下文 c

输出: 地址 ip 对应的地理位置

```

1: 投票集合 $\pi = \{0\}$ 
2: 地理位置集合 $\lambda = \emptyset$ 
3: for  $db \in \delta$  do
4:   获得 $ip$ 在 $db$ 中对应的地理位置 $l$ 
5:    $\lambda = \{l\} \cup \lambda$ , 将 $l$ 加入到地理位置集合
6: end for
7: if  $c \in \lambda$  then
8:   return  $c$ 
9: else
6:   return  $\operatorname{argmax}_{l \in \lambda} \pi_l$ 

```

5.1.3 测量实验设置

在测量实验被部署在了中国的部分城市部署了。针对城市的特点, 主要分别在代表不同城市等级的几个城市进行了测量, 在本文中主要使用例如鞍山, 大连, 盘锦和秦皇岛等几个城市的数据来进行讨论。因为 Traceroute 是用户端获得网络路径最常用的手段, 本文使用 Traceroute 来进行用户级别的路径测量。在每个城市中, 至少测量 196 条网络路径已经涉及中国全部的省份以及主要城市。

在所测量的服务器的选择上, 本文也才采取了一定的策略。首先使用了之前网络爬虫中提到的三个类型的服务器以便在最短时间内能够最多的覆盖相应的地理位置。除此之外, 为了验证 HIDA 中关于网络服务提供商之间互联互通的假设, 本文还测量了 Alexa

上的中国大陆排名前 100 的网站。关于这些网站，本文主要关注访问这些网站路径是否包含网络服务提供商之间的数据交换。

5.1.4 测量与观测结果

观测实验的结果主要与网络服务提供商网络之间的数据交换与用户层面的网络测量有关。这些结果验证了 HIDA 的方法以及为之后的评价工作提供了数据。

(1) Cold-Potato 路由策略与 Hot-Potato 路由策略：如图 5.2 所示是位于不同地理位置的客户端到所选定服务器的网络路径。其中包括了从鞍山、大连、盘锦以及秦皇岛出发的网络路径。在同种不同的颜色代表着不同的网络服务提供商标，其中主要是对中国三大网络服务提供商中国联通、中国电信以及中国移动的测量。

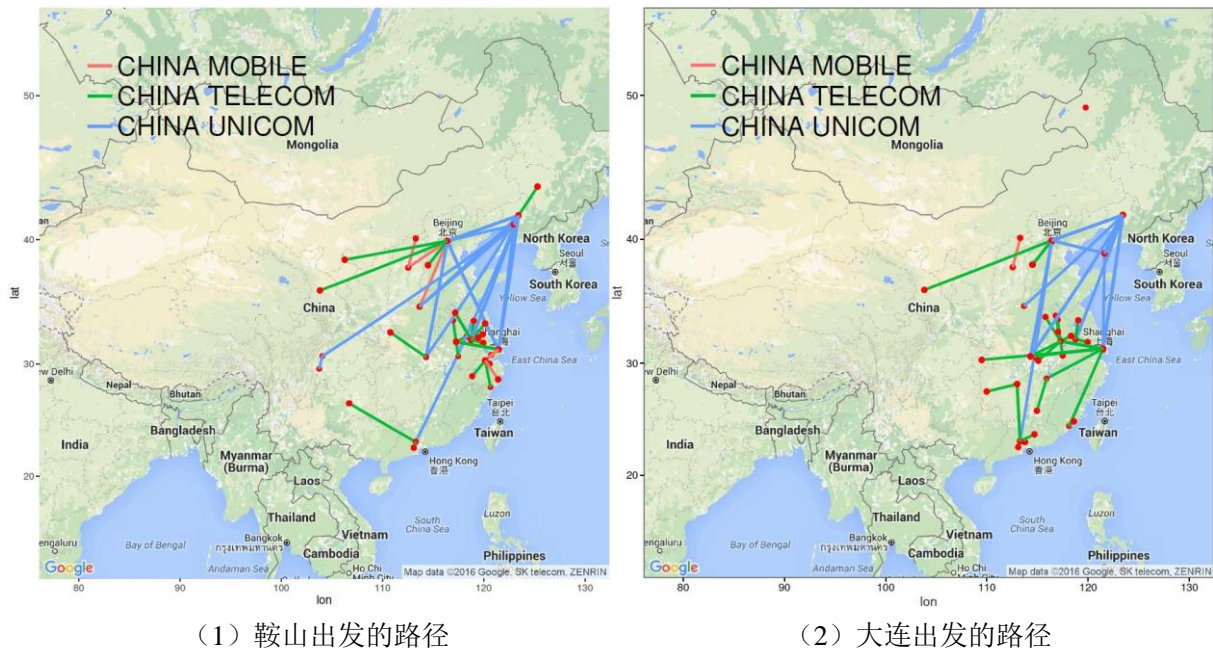


图 5.2 路由与地理位置信息

Fig. 5.2 Routing and Geographic Information

如图 5.2 所示，一旦某个网络服务提供商获得了对流量的控制权，它将尽可能长时间的拥有这些流量。根据测量的得到的结果，从用户角度来说，中国网络服务提供商主要实施的是 Cold-Potato 路由策略，也就是尽可能多的控制流量在自己的网络中。通过 Cold-Potato 路由策略，网络服务提供商对于网络流量有最大化的控制权，从而能够对网

络有更大的控制权。然而，Cold-Potato 同时也意味着在每条网络路径中，存在着更少的互联互通部分，这对于对数据交换比较敏感的 HIDA 来说是非常有利的。

如图 5.3 所示，是由位于客户端在大连市到服务器在深圳市的一条网络路径。在该路径中，网络服务提供商之间的数据交换发生在广州市。而在从 CNNIC 获得的网络组织方式来看，数据交换同时也可以发生在北京市，而且可以从图 5.3 中可以看出，流量数据在联通的网络中也经过了北京市的结点。因此就可以判断出，网络服务提供商所执行的是 Cold-Potato 路由策略。更少的互联互通部分也就意味着在网络服务提供商级别的路径松弛中，网络服务提供商级别的 *segment* 的数量能够尽可能的减少，从而使得 HIDA 需要更少的观测服务器就可以进行诊断。而如果是 Hot-Potato 路由策略的话，会导致网络服务提供商级别的路径松弛接近于路由器级别的路径，因为在不停的转换网络服务提供商。而就算是 Hot-Potato 路由中，地理位置级别的路径松弛也是可以使用的，因为就算是数据交换，也必须发生在同一城市的数据交换设备之间，所以不影响 HIDA 的设计，但是其表现应该是会收到影响的。



图 5.3 Cold-Potato 路由策略示例

Fig. 5.3 Example of Cold-Potato Routing Strategy

(2) 互联互通部分的数据交换：数据交换都发生在 Internet 交换点。而以中国为例，这些交换点全部位于中国的主要城市，比如说北京、上海、广州等。在实际测量过程中统计得到，从 Alexa 得到的网站中，中国大陆的前 100 名的商业网站中，53% 都是

位于北京、上海以及广州这三个最主要的交换点城市。也就是说，在用户的正常使用过程中，在这几个城市中，可以发现更多的商业服务器。从而，就存在更多的用来对目标路径进行诊断的候选的服务器。所以，通过本文设计的启发式的发现策略无法发现可以使用的 Web 服务器的时候，就会使用前文中提到的经验化方法来进行相应的补充。从而基本保证了用户能够发现至少一个可用的服务器。

由于现代的 Internet 网络对于内容分发网络（Content Distribution Network, CDN）大量使用，导致网络服务提供商之间的数据交换不是非常的频繁。但是需要注意的是，CDN 主要是为了数据体积较大的加载目标设计的。因此，对于常用的小数据量的文本文件，例如基于 HTML 的网页文件以及页内脚本还是需要进行源服务器的访问的。如图 5.4 所示，通过测量中国大陆访问流量前 100 的网站，可以看出，数据交换还是大量存在的。从而说明，HIDA 中进行网络服务提供商级别的路径松弛，从而进行中立性相关的诊断还是很有必要。



图 5.4 Cold Potato 路由策略

Fig. 5.4 Cold Potato Routing Strategy

（3）匿名路由与超时事件：在测量实验中，Taceroute 的最大 TTL 值被设置为了 20，然后由于匿名路由器导致的网络超时事件的出现频率从而计算出对层析技术的影响。因为基于层析的网络诊断，对于路径的重合程度是有要求的。如果所有的路径都没有重合，那么就无法进行层析诊断。所以说，匿名路由器与其导致的超时事件，对于用户角

度的路径发现甚至是网络诊断是有着较大的影响的。尤其是在用户层面的测量中，因为 Traceroute 是现阶段用户角度所能获得网络路径的唯一手段。

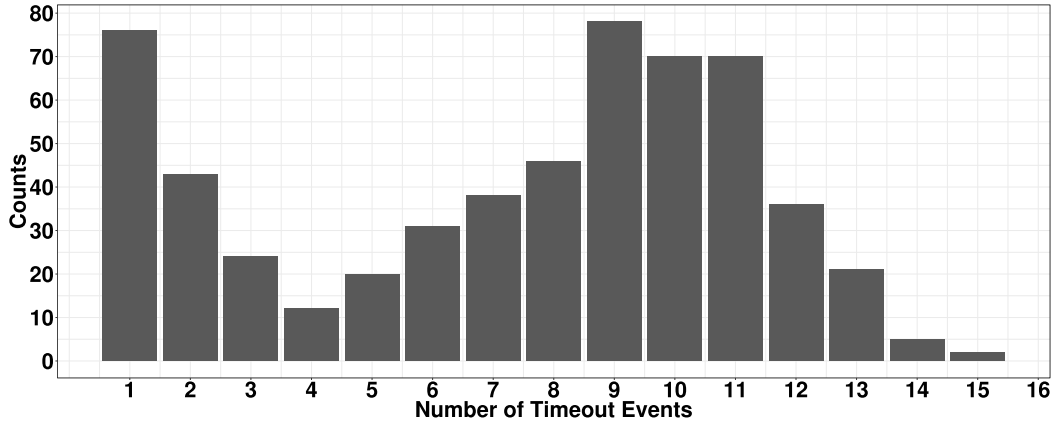


图 5.5 Traceroute 中的超时问题

Fig. 5.5 Timeout Issues In Traceroute

如图 5.5 所示，在实际的测量中每条路径都会存在超时时间。这个也是很好解释的，因为随着对于网络安全的要求的提高，越来越多的网络服务提供商隐藏了越来越多的路由器的信息，例如禁止 ICMP。所以说这种情况是普遍存在的。在每个城市测量的所有 196 条网络路径中，都出现了超时的问题，同事超时事件所占比例在 40% 至 50% 的这些网络路径数量占了绝大多数。对于这以结果，也从一个方面证明了本文所使用路径松弛方法是正确的。如果不使用一定的路径松弛，在一个粗粒度上进行诊断的话，用户是没有办法获得相应的准确的有效的结果的。

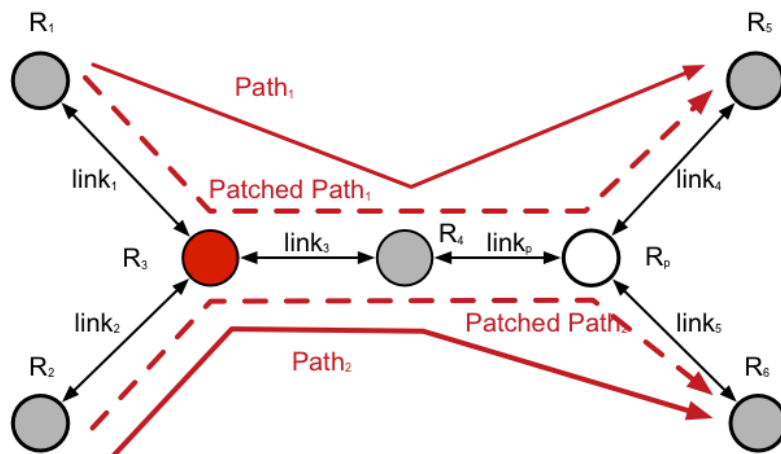


图 5.6 匿名路由器对中立性诊断影响

Fig. 5.6 Influence of Anonymous Router on Neutrality Inference

如图 5.6 所示,匿名路由器对中立性诊断也是有影响的。尤其是会导致例如 System 4^[24]的基于层析推断的检测技术的误差。

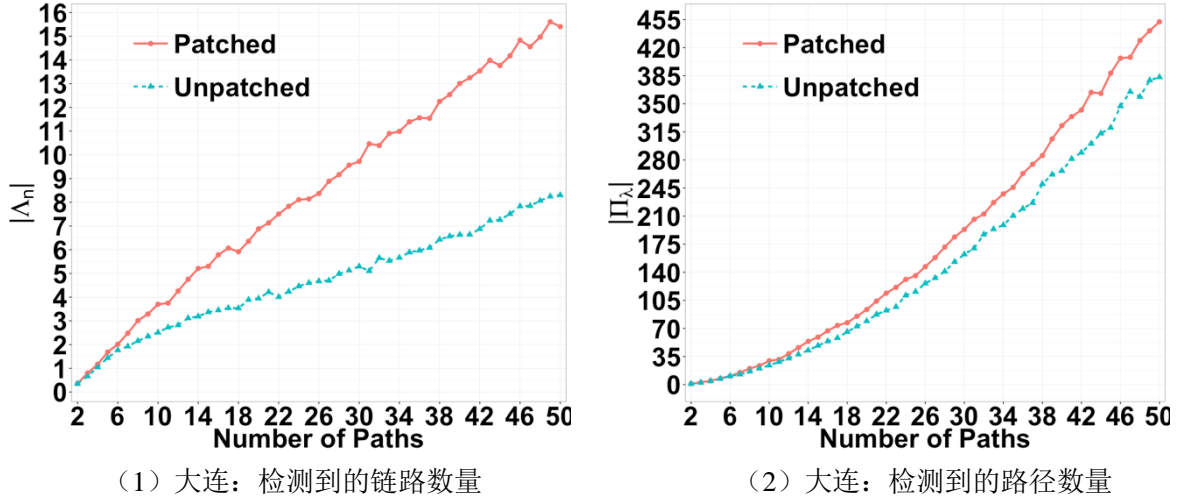


图 5.7 地理位置信息对中立性检测影响

Fig. 5.7 Influence of Geographic Information on Neutrality Inference

为了验证匿名路由器与超时事件对中立性相关的诊断的影响,本文实现了 System 4 的路径重叠算法,并通过地理位置补充的方法,对 System 4 的结果进行了比较。在试验中, Λ_n 与 Π_λ 恰好可以分别代表可以识别的路径的数量以及可以识别的链路的数量。如图 5.7 所示,使用地理位置信息对 System 4 进行补充之后, System 4 的中立性检测结果可以明显的高于未补充的情况。因此就可以证明,匿名路由器与超时事件对于基于侦测的中立性监测是影响非常大的。而 HIDA 使用的方法,既满足了用户角度与中立性相关的诊断的需求,而且还有有效的避免了这些问题。因此就可以得出结论,使用地理位置信息级别网络映射中的路径松弛进行用户层面中立性相关的监测是很实用有效的。

5.2 HIDA 性能评价

在本节中,通过测量实验中获得的实测数据,来验证 HIDA 在实际诊断过程中的表现。首先,需要讨论的就是 HIDA 与测量实验中测量服务器的选择策略所实现的地理位置级别的覆盖率。其次,通过对比基于 IP 地址的传统匹配方法,验证了 HIDA 中的基于 DTW 的路径匹配方法在路径匹配方面的提高。然后,本章讨论了所设计的启发式 Web 服务器的发现策略的表现。通过仿真实验,对 HIDA 在实际的诊断过程中的表现进行验证。并且在诊断率 (Detection Rate, DR) 与误判率 (False Positive Rate, FPR) 方

面对 HIDA 的功能进行了评价，因为 HIDA 的目标是能提供给用户可靠的依据来验证自己面临的网络问题所涉及到的网络服务提供商，因此误判率是保证其结果可靠性的最重要的因素，而不是传统基于层析的诊断所关注的诊断率。

5.2.1 地理覆盖率

为了验证所测量的目标服务的代表性是否满足需求，本文使用地理位置的覆盖率来评价实测实验中关于测量使用的 Web 服务器的选择策略。在实测实验中，如果所测量的服务器的地理位置覆盖率越高，就说明单一用户能够尽可能的覆盖更加广泛的区域。从另一个角度来说，在参与用户较少的情况下，或者说对于单一用户来说，所测量的数据也就更有效。如图 5.8 所示，如果仅仅使用之前提到的三种最基本的目标服务器，能够覆盖最多 27% 的城市。图中所示的是客户端位于中国联通（CHINA UNICOM）的网络下，并且测量的网络路径区分包含或者不包含网络服务提供商之间的数据交换，也就是说测量的也是在基本的服务器集合中，使用中国联通网络或者 CDN 以及需要访问位于其他网络服务提供商网络的两类不同的服务器。通过测量的这些服务器在地理位置级别的覆盖情况可以看出，在同一网络服务提供商网络下，基本能达到 15% 到 20% 的覆盖率。在不存在 CDN 影响的情况中，在中国联通的内部网络下的地理位置覆盖能够达到 20% 以上。而存在网络服务提供商之间互联互通的情况最高能达到 27%，这样就意味着在对互联互通部分进行诊断的时候是可以进行的。而且需要值得注意的是，包括中国主要城市在内的可能作为 IXP 的省会城市都是包含在内的，而且省市之间的比值为 9.17%，远低于测量中任何的地理位置覆盖率。

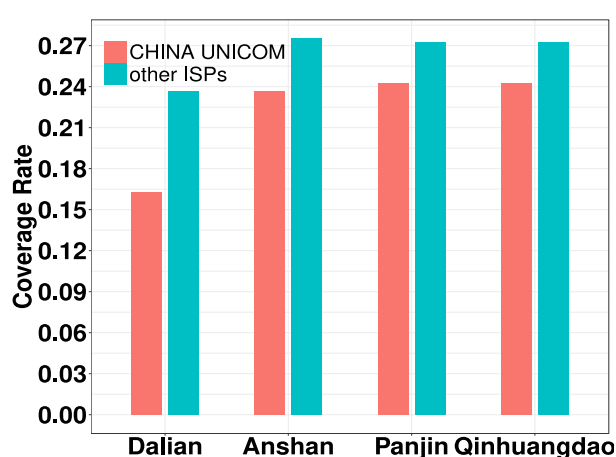


图 5.8 城市覆盖率

Fig. 5.8 Coverage Rate of Cities

如图 5.9 所示，在省一级地理位置的覆盖上，最少也能达到 60% 以上，而未覆盖的主要是西部内陆地区，这些地区并不存在互联互通设备，同样也就不存在数据交换的发生，所以对于网络服务提供商级别的诊断是没有影响的。而平均来说基本上 80% 以上的省级覆盖率也是很好的结果。

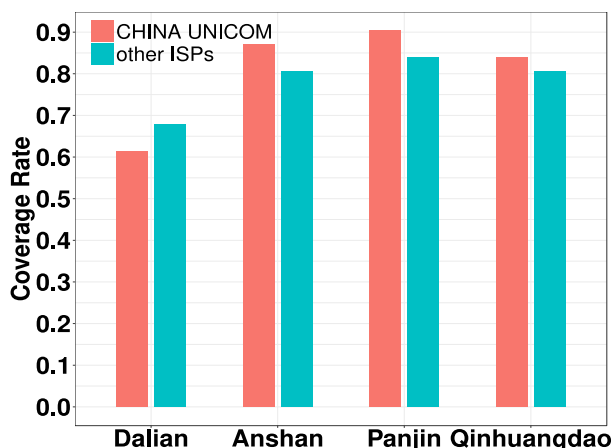


图 5.9 省覆盖率

Fig. 5.9 Coverage Rate of Provinces

需要值得注意的是，实验统计的仅仅是选定的三类网站的服务器。这样的结果对于网络服务提供商级别的网络诊断也是足够的，所以对用户来说也是由参考价值的。

这样的结果能够说明三个问题。首先，在现代网络中，CDN 逐渐的得到了很广泛的应用。但是，实验辨明传统服务器的存在也是普遍的。其次，15% 到 20% 的覆盖率基本上覆盖了中国的所有主要城市。而之前讨论过，数据交换的发生都是在主要城市的。而本文所涉及的网络服务提供商级别的诊断方法，在理想情况下，是可以很好的识别这些问题的。从而，在实际过程中，本文设计的服务器发现策略也是可以在相应的位置发现服务器的。最后，由于对于主要城市的完全覆盖，也就说明测量实验的结果是可靠的。

5.2.2 路径匹配实验

为了验证 HIDA 中使用的路径松弛方法，本文利用测量实验中得到的数据，设计并实验了路径松弛的实验，来验证 HIDA 与传统基于 IP 地址的方法的差异性。在实验中比较了 HIDA 中基于地理位置的路径松弛方法以及传统的单纯基于 IP 地址以及更加广泛的基于网络服务提供商级别的路径松弛方法。如图 5.10 所示，对 196 条跨网络服务提

供应商的路径，也就是存在数据交换的路径，与近客户端的未发生数据交换的路径进行路径匹配。

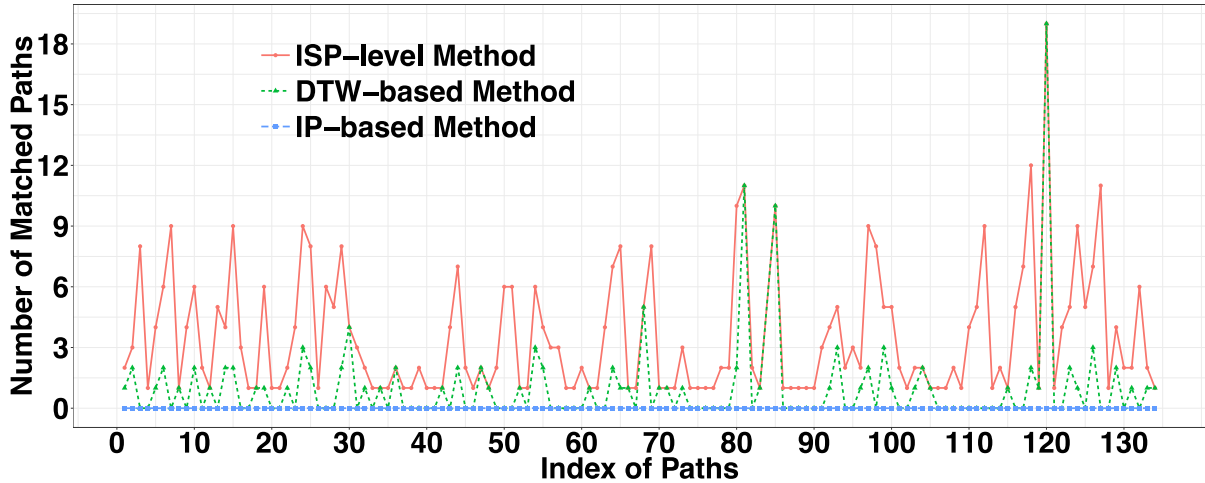


图 5.10 网络服务提供商级别路径松弛实验

Fig. 5.10 ISP-level Relaxation Experiment

可以看到，HIDA 的表现还是比较具有实用性的。纯粹基于 IP 地址的方法，在实际的操作过程中，是没有找到任何的重合路径的，因为在实际的测量中，即使是存在相同的终点 IP 地址，利用纯粹的 IP 地址匹配的方法，也是没有识别出来的。所以，HIDA 相较于该方法，肯定损失了一定的精度，但是对于用户来说，是非常有价值的，因为 HIDA 对于实用性来说是有提高的。同时，更重要的是，相比于网络服务提供商级别的路径松弛，HIDA 中的基于 DTW 的路径匹配方法更加准确，同时其表现也相差不大。

5.2.3 服务器发现实验

为了验证具体 HIDA 的启发式 Web 服务器发现策略的表现，使用测量实验得到的数据按照发现策略进行匹配。

(1) 测量数据

通过统计所有能够找到有重叠部分的路径所使用的查询的次数来对 HIDA 与基于历史路径的经验化方法来进行比较。所谓的经验化方法，就是用户在没有地理位置辅助的情况下，最一般的解决方案是在历史 Traceroute 获得的路径中查找可以用来进行辅助诊断的链路。其实，更一般的方法是随机的选择链路。但是随机的获取链路，对于实际使用来说，查询的次数会更多，所以在试验中用在历史路径中的方法来代替，只要相比经验化方法更有效，就说明 HIDA 所提出的方法是有提高的。

如图 5.11 所示，分别是在四个城市中，将测量的数据输入到 HIDA 得出的有关于服务器发现查询次数得到的结果。

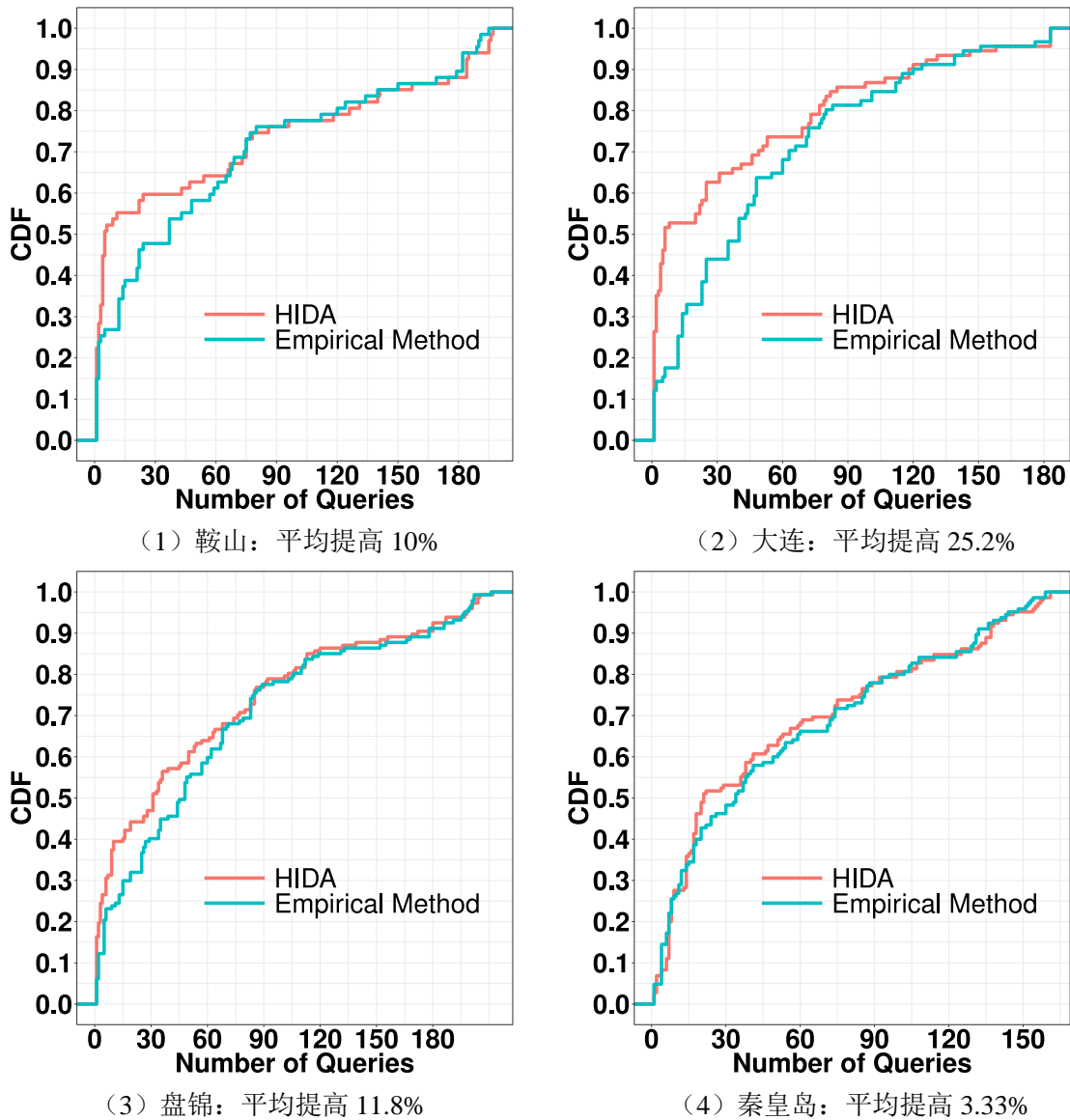


图 5.11 发现服务器所用查询次数

Fig. 5.11 The Number Of Queries Of Server Discovery

不难看出，在四个城市中，HIDA 都减少了查询的次数（Number of Queries）。而图 5.11（4）所示的秦皇岛市，在实际测量过程中，由于网络状况不佳，所以用户得到

的 Traceroute 结果也不理想。但是,在这种情况下,HIDA 依然相较于经验方法有所提高。

如图 5.11 (2) 中大连的情况所示,HIDA 可以实现高达 25% 的查询次数的降低。在这个网络环境中,通过测量发现,相关的网络参数还是比较稳定的。使用的是咖啡馆中的无线网络,网络延迟一直维持在 200ms 以下。而且是在中午时间,在相对嘈杂的环境中进行的实验。也就是说,该环境应该是正常用户所处的网络环境,是比较有代表性的。也同时说明,本文所使用的测量方式是能反映出实际网络环境的差异的。

如图 5.11 (1) 与图 5.11 (3) 所示的鞍山市与盘锦市的实验结果可以发现,这两个城市作为典型的中国二线城市,其表现也是相似的。相关的网络环境比较稳定,但是相比于大连市,网络状况不够优秀,会存在丢包与高延迟问题而从而导致的 Traceroute 的异常。而且这两个城市不是主要的网络节点,但是,HIDA 的服务器发现策略依然能够发挥功效。同时,为了保险起见,在用基于地理位置的方法无法发现可用的匹配的路径的时候,使用相同的经验化方法作为补充。所以说,如图 5.11 中所示四个城市在较多的查询次数的部分的分布情况,HIDA 与经验化方法是相似的。

5.2.4 层析诊断仿真实验

如表 5.3 所示,在仿真试验中,本文比较了相关的几个比较重要的基于层析的网络诊断方法。HIDA 与传统基于层析技术的诊断方式是有本质的区别的。如表 5.3 所示,HIDA 在误判率方面,是有最好的表现的,在本文涉及的所有的仿真实验中,HIDA 可以实现在理想情况下的 0% 的 FPR,而其他的测试方法,比如 CLINK 都会出现误判。但是,在检测率(DR)方面,HIDA 是存在着由其本质监测方式所带来的诊断方面的缺陷的。而且 HIDA 所使用的是有针对性的监测方式,而传统的层析诊断方式所使用的是基于数据的诊断方式。更重要的是,HIDA 更适合于中立性相关的诊断。

表 5.3 FPR 与 DR 比较
Tab. 5.3 Comparison of FPR and DR

层析方法	参与用户数	误判率	诊断率
HIDA	1	0%	50%
CLINK	≥ 2	0.7%	91.2%
SCFS	≥ 2	0.6%	62%
MCMC	≥ 2	1.3%	73.4%

首先,中立性相关的诊断要定位于确定的网络服务提供商;其次,针对于提供证据性的网络诊断,FPR 方面的结果要比 DR 方面的结果更加重要。因为在实际使用过程中,

一旦用户得到了相应的诊断结果，用户首先要保证的就是证据的准确性与唯一性。最后，网络服务提供商级别的问题的出现，通常也是单一的存在的。也就是说，HIDA 得到的结果能够很好的符合这些要求。

在仿真实验中，本文使用的是类似于如图 4.2 所示的树形网络拓扑图。因为在实际判断中，HIDA 有针对性的为目标链路建立起了相应的拓扑图，所以说相对于针对层析技术的研究。如图 5.12 所示，相比较于 CLINK 与 SCFS，HIDA 在理想条件下能够实现 0% 的误判率。而且需要注意的是，在实验中是进行随机的选取发生问题的路径段，同时发生问题的链路段的数量也是随机的。而在实际生活中，由于用户使用了网络服务提供商级别的路径松弛，所以说基本上产生问题的 *segment* 只有一个，而且最关键的是 HIDA 只会提供给用户在目标路径上的问题链路，那么就不会出现误判。

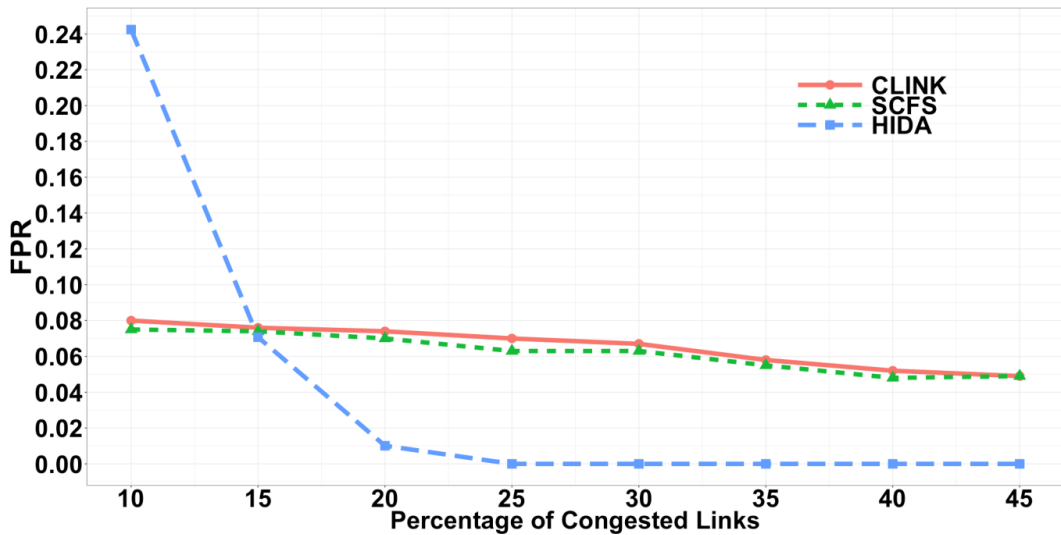


图 5.12 路径检测仿真实验

Fig. 5.12 Simulation Experiment of Path Diagnosis

结合监测率来看，是可能出现漏判的，但是误判的几率是相当低的。因为举例来说，在实测试验中，可以看出，Cold-Potato 路由策略下，基本上就只存在两个网络服务提供商的网络，那么在 HIDA 中就存在 5 个松弛过的网络链路，因此问题链路的比率（Percentage of Congested Links）就一定会大于 20%，所以 HIDA 在这种特定情况下的表现是优于其他方法的。这种差异的出现是由于监测方式的根本不同导致的，所以 HIDA 能够满足中立性相关的用户级别诊断的需求。

结 论

本文针对开放网络原则下的网络中立性，分别针对网络资源的分配以及网络的使用过程中的中立性进行研究。中立性要求以用户为中心，进行合理公平的分配，同时不允许出现由于用户差异性以及流量差异性导致的网络服务提供商内部或者网络服务提供商之间的阻断、节流以及差异性表现。

在网络资源分配相关的中立性中，本文提出了一个基于面向与网络服务提供商与二级用户的具有真实性的双向网络频谱拍卖机制 **RTDA**。**RTDA** 不同于传统的网络拍卖机制，在实现了与中立性相关的真实性的同时实现了时间与空间的多维重用。能够为用户的需求提供最大程度的满足，同时能够尽可能的提高有限的频谱资源的利用率。针对网络的使用过程中的保证中立性问题。本文提出了一个基于 **HTTP** 的中立性敏感的网络路径诊断方法 **HIDA**。**HIDA** 主要通过为用户发现广泛存在并且具有良好介入性的 **Web** 服务器，以便代替在基于网络层析技术的故障点推断中的测量观测点，从而让用户可以有目的性的进行层析诊断。**HIDA** 中包括了全新的基于 **DTW** 的路径匹配方法以及启发式的基于地理位置信息的服务器发现策略。

在本文中，分别使用非合作博弈以及数学推导的方式对 **RTDA** 与 **HIDA** 进行了理论证明以验证其正确性。同时使用仿真实验的方式对他们在实际使用中的性能进行了评价。无论是 **RTDA** 还是 **HIDA** 都相对于向前的方法有明显的提高，其中 **RTDA** 相较于 **TRUST** 在频谱利用率方面能够实现更高的频谱利用率的同时，也实现了其他的频谱拍卖机制未能实现的空间与时间的多维重用。在 **HIDA** 中提出的启发式服务器发现策略比经验化方法平均提高了 20%，同时基于 **DTW** 的匹配方法也相较于基于 **IP** 地址的匹配方法提高了 127%，而且 **HIDA** 可以实现在理想情况下的 0% 误判率。为了验证本文的工作与发现新的规律，本文还设计并实现了用户级的网络测量工具，并在多地部署了用户层面的面向网络服务提供商的测量工作。不仅发现了现阶段的网络服务提供商使用的路由策略，同时也对地理位置信息对中立性相关检测的影响进行了讨论。

在未来的工作中，将把 **RTDA** 与 **HIDA** 与结合众包（**Crowdsourcing**）技术结合起来，让更多的用户参与到开放网络相关的中立性的分配以及诊断中，从而从用户的角度，对网络服务提供商进行评价与监督，建立起和谐与用户友好的网络环境。

参 考 文 献

- [1] Choi J P, Jeon D S, Kim B C. Net neutrality, business models, and internet interconnection[J]. American Economic Journal: Microeconomics, 2015, 7(3): 104-141.
- [2] 朱春艳. 论费恩伯格的“技术民主化”理论[J]. 自然辩证法研究, 2008, (07): 33-37.
- [3] Huang J, Berry R A, Honig M L. Auction-based spectrum sharing[J]. Mobile Networks and Applications, 2006, 11(3): 405-418.
- [4] FCC Spectrum Policy Task Force. Report of the spectrum efficiency working group[EB]. (2002, 11) [2013, 02].
- [5] Bykowsky M M, Cull R J, Ledyard J O. Mutually destructive bidding: The FCC auction design problem[J]. Journal of Regulatory Economics, 2000, 17(3): 205-228.
- [6] Brenner D L, Maxwell W. The Network Neutrality and the Netflix Dispute: Upcoming Challenges for Content Providers in Europe and the United States[J]. Intellectual Property and Technology Law Journal, 2011, 23(3): 3.
- [7] Cramton P. Spectrum auction design[J]. Review of Industrial Organization, 2013, 42(2): 161-190.
- [8] Zhang Y, Lee C, Niyato D, et al. Auction approaches for resource allocation in wireless systems: A survey[J]. Communications Surveys and Tutorials, 2013, 15(3): 1020-1041.
- [9] Wurman P R, Wellman M P, Walsh W E. A parametrization of the auction design space[J]. Games and economic behavior, 2001, 35(1): 304-338.
- [10] Babaioff M, Nisan N. Concurrent auctions across the supply chain[C]. Proceedings 3rd ACM Conference on Electronic Commerce (EC-2001). Tampa, USA, 2001: 1-10.
- [11] T. He, L. Ma, A. Gkelias, et al. Robust monitor placement for network tomography in dynamic networks[C]. 35th Annual IEEE International Conference on Computer Communications (INFOCOM 2016), San Francisco, USA, 2016.
- [12] Lin P, Feng X, Zhang Q, et al. Groupon in the air: A three-stage auction framework for spectrum group-buying[C]. 32th Annual IEEE International Conference on Computer Communications (INFOCOM 2013), Turin, Italy, 2013.
- [13] Zhou X, Gandhi S, Suri S, et al. eBay in the sky: strategy-proof wireless spectrum auctions[C]. Proceedings of the 14th Annual International Conference on Mobile Computing and Networking (MOBICOM 2008), San Francisco, USA, 2008: 2-13.
- [14] 张林, 叶保留, 王钦辉. U-Auction: 面向真实利用率的频谱拍卖机制[J]. Journal of Frontiers of Computer Science and Technology, 2013, 7(12): 1073-10.

- [15] Zhou X, Zheng H. Trust: A general framework for truthful double spectrum auctions[C]. 28th IEEE International Conference on Computer Communications (INFOCOM 2009), Rio de Janeiro, Brazil, 2009: 999–1007.
- [16] Kanuparth P, Dovrolis C, Papagiannaki K, et al. Can user-level probing detect and diagnose common home-WLAN pathologies[J]. ACM SIGCOMM Computer Communication Review, 2012, 42(1): 7–15.
- [17] R. Mahajan, N. Spring, D. Wetherall, et al. User-level internet path diagnosis[J]. ACM SIGOPS Operating Systems Review, 2003, 37(5): 106–119.
- [18] M. H. Gunes, K. Sarac. Inferring subnets in router-level topology collection studies[C]. Proceedings of the 7th ACM SIGCOMM Internet Measurement Conference (IMC 2007), San Diego, USA, 2007, pp. 203 – 208.
- [19] B. Augustin, X. Cuvelier, B. Orgogozo, et al. Avoiding traceroute anomalies with Paris traceroute[C]. Proceedings of the 6th ACM SIGCOMM Internet Measurement Conference (IMC 2006), Rio de Janeiro, Brazil, 2006.
- [20] R. Padmanabhan, P. Owen, A. Schulman, et al. Timeouts: Beware surprisingly high delay[C]. Proceedings of the 2015 ACM Internet Measurement Conference (IMC 2015), Tokyo, Japan, 2015.
- [21] Coates A, Hero III A O, Nowak R, et al. Internet tomography[J]. IEEE Signal processing magazine, 2002, 19(3): 47–65.
- [22] Vardi Y. Network tomography: Estimating source-destination traffic intensities from link data[J]. Journal of the American Statistical Association, 1996, 91: 365–377.
- [23] H. X. Nguyen, P. Thiran. The boolean solution to the congested IP link location problem: Theory and practice. 26th IEEE International Conference on Computer Communications (INFOCOM 2007), Anchorage, USA, 2007.
- [24] Z. Zhang, O. Mara, K. Argyraki. Network neutrality inference. Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication (SIGCOMM 2015), London, United Kingdom, 2015.
- [25] Dovrolis C, Gummadi K, Kuzmanovic A, et al. Measurement lab: Overview and an invitation to the research community[J]. ACM SIGCOMM Computer Communication Review, 2010, 40(3): 53–56.
- [26] Fiuczynski M E. PlanetLab: overview, history, and future directions[J]. Operating Systems Review, 2006, 40(1): 6–10.
- [27] Basso S, Servetti A, Masala E, et al. Measuring DASH streaming performance from the end users perspective using neubot[C]. Proceedings of the 5th ACM Multimedia Systems Conference (MMSys 2014), Singapore, 2014.

- [28] Y. Liu, W. Tome, L. Zhang, et al. An end-to-end measurement of certificate revocation in the web' s pki. Proceedings of the 2015 ACM Internet Measurement Conference (IMC 2015), Tokyo, Japan, 2015.
- [29] M. Dhawan, J. Samuel, R. Teixeira, et al. Fathom: A browser-based network measurement platform. Proceedings of the 12th ACM SIGCOMM Internet Measurement Conference (IMC 2012), Boston, USA, 2012.
- [30] X. S. Wang, A. Balasubramanian, et al. Demystifying page load performance with WProf. Proceedings of the 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 2013), Lombard, USA, 2013.
- [31] Y. Tian, R. Dey, Y. Liu, et al. China' s internet: Topology mapping and geolocating. 31th IEEE International Conference on Computer Communications (INFOCOM 2012), Orlando, USA, 2012.
- [32] Hamilton J D. Time series analysis[M]. Princeton: Princeton university press, 1994.
- [33] Wiener N. Extrapolation, interpolation, and smoothing of stationary time series[M]. Cambridge: MIT press, 1949.
- [34] Salvador S, Chan P. Toward accurate dynamic time warping in linear time and space[J]. Intelligent Data Analysis, 2007, 11(5): 561-580.
- [35] Celebi H, Arslan H. Utilization of Location Information in Cognitive Wireless Networks[J]. IEEE Wireless Communications, 2007, 14(4): 6-13.
- [36] Shavitt Y, Zilberman N. A structural approach for PoP geo-location[J] Computer Networks, 2012, 56(3): 1029-1040.
- [37] Huffaker B, Fomenkov M. DRoP: DNS-based router positioning[J]. ACM SIGCOMM Computer Communication Review, 2014, 44(3): 5-13.
- [38] 胡俊. 一体化网络中基于地理位置信息的域间路由的设计与实现 [D]. 北京:北京交通大学, 2009.
- [39] Liberti L, Lavor C, Maculan N, et al. Euclidean distance geometry and applications[J]. Siam Review, 2014, 56(1): 3-69.
- [40] Subramanian L, Padmanabhan V N, Katz R H. Geographic Properties of Internet Routing[C]. Proceedings of the General Track: 2002 USENIX Annual Technical Conference, Monterey, USA, 2002.
- [41] Cao Z, Wang Z, Zegura E. Performance of hashing-based schemes for internet load balancing[C]. 19th IEEE International Conference on Computer Communications (INFOCOM 2000), Tel Aviv, Israel, 2000.
- [42] Poese I, Uhlig S, Kaafar M A, et al. IP geolocation databases: Unreliable?[J]. ACM SIGCOMM Computer Communication Review, 2011, 41(2): 53-56.

攻读硕士学位期间发表学术论文情况

- 1 RTDA: A novel reusable truthful double auction mechanism for wireless spectrum management. 第一作者. International Conference on Big Data Computing and Communications (BigCom), 2015. Springer 2015. pp 14-27. Taiyuan, China, 2015. EI 检索, EI 检索号: 20154701583397. (本硕士学位论文第三章)
- 2 A Novel Framework for Analysis of Global Network Neutrality Based on Packet Loss Rate. 第二作者. International Conference on Cloud Computing and Big Data (CCBD), 2015. IEEE 2016. pp 297-304. Shanghai, China, 2016. EI 检索, EI 检索号: 20162102424120. (本硕士学位论文第一章)
- 3 Patching Traceroute Using Geographic Information In Neutrality Inference Crowdsourcing. 第一作者. International Conference on Big Data Computing and Communications (BigCom), 2017. 在投. (本硕士学位论文第四、五章)

致 谢

三年时光匆匆而过，我也要即将度过我的研究生生涯。对于在这三年生活中给予我莫大帮助的人，我首先要表示我由衷的敬意与感谢。

首先，我要感谢我的导师，XXX 老师。在我进入实验室开始自己的研究工作之后，王雷老师一直帮助我、指导我，教会了我从调研到论文书写的很多细节方面的知识，让我逐步学会了如何进行科研工作。而且 XXX 老师宽容的胸怀以及耐心的指导，让我在许多迷茫的时候给予了我最大的支持与帮助。正是 XXX 老师将我从一个对科研一窍不通的大学生带入了科研的世界，并教会给了我能够受用一生的经验与知识。无论对我三年的研究生生活还是今后漫漫的人生路上来说，都是一笔不可多得的宝贵的财富。对此，我再次表示由衷的感谢与敬意。

我还要感谢谷歌的 XXX 博士，XXX 博士在我整个研究生生涯都为了提供了莫大的帮助。无论从科研方面还是人生规划方面，XXX 博士都为了指引了方向，让我在迷茫中能够快速起航。同时我也要感谢实验室的 XXX 老师、XXX 老师、XXX 老师，是你们在我最需要帮助的时候，无私的为我提供最宝贵的支持与建议。

其次，我要感谢我的朋友与同学，在整个毕业设计的准备过程中，是他们在我最需要帮助的时候为我提供了最无私的支持。我要感谢 WILNA 实验室中共同奋斗的各位。正是你们的出现，让我知道了团队的力量；正是你们的出现，让我的学习科研之路变得丰富多彩。在共同努力的每一个日日夜夜，是你们一直陪伴在我的身边；在共同进步的每一个分分秒秒，是你们给了我最大的动力与支持。无论今后我身在何方，你们都是我永远的朋友，永远的战友。

其次，我要感谢我的父母与家人，是他们从我的人生的开始就在我的身后毫无条件和保留的支持我、鼓励我，让我拥有莫大的自信与力量克服一切困难。从我蹒跚学步开始直到我长大成人，没有你们就没有我的今天。无论成功还是失败，你们总会报以微笑和支持。家人也是我奋斗不止勇往直前的动力源泉，你们对我的恩情不仅仅局限于一朝一夕，而是充斥在我人生中的点点滴滴。

最后，我还要对各位百忙之中为我审阅论文、参加我的答辩的各位专家、教授表示由衷的感谢！

大连理工大学学位论文版权使用授权书

本人完全了解学校有关学位论文知识产权的规定，在校攻读学位期间论文工作的知识产权属于大连理工大学，允许论文被查阅和借阅。学校有权保留论文并向国家有关部门或机构送交论文的复印件和电子版，可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印、或扫描等复制手段保存和汇编本学位论文。

学位论文题目：_____

作者签名：_____ 日期：_____年____月____日

导师签名：_____ 日期：_____年____月____日