

## 网络断层扫描:理论与算法\*

李惠康, 高艺, 董玮, 陈纯

(浙江大学 计算机科学与技术学院, 浙江 杭州 310027)

通讯作者: 董玮, E-mail: dongw@zju.edu.cn



**摘要:** 网络测量为网络设计者与管理者提供网络内部细粒度的运行状态信息,是对网络进行高效管理与优化的基础.网络断层扫描是网络测量领域的一个研究热点,是一种端到端的网络测量方法.与传统网络内部测量方法不同,网络断层扫描利用端到端的测量信息计算和推断网络内部性能和状态,从而实现与网络组成和协议无关的网络测量,具有较低的测量开销.对近年来国内外学者在网络断层扫描研究领域取得的成果进行了系统的总结.首先介绍了网络断层扫描的基本模型,并指出了影响网络断层扫描性能的3个重要因素:监测节点部署、测量路径构造和测量数据分析;接着,依次归纳了这3个方面的研究进展和研究成果;随后分析了已有网络断层扫描方法在实际应用中存在的缺陷,并给出了应对这些核心缺陷的理论和关键算法;最后,基于现有研究成果讨论了网络断层扫描的发展趋势和进一步的研究方向.

**关键词:** 网络断层扫描;网络测量;性能指标;监测节点;链路识别性;测量数据

**中图法分类号:** TP393

中文引用格式: 李惠康,高艺,董玮,陈纯.网络断层扫描:理论与算法.软件学报,2021,32(2):475-495. <http://www.jos.org.cn/1000-9825/6134.htm>

英文引用格式: Li HK, Gao Y, Dong W, Chen C. Network tomography: Theory and algorithm. Ruan Jian Xue Bao/Journal of Software, 2021,32(2):475-495 (in Chinese). <http://www.jos.org.cn/1000-9825/6134.htm>

## Network Tomography: Theory and Algorithm

LI Hui-Kang, GAO Yi, DONG Wei, CHEN Chun

(College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China)

**Abstract:** Network measurement provides the network designers and managers with fine-grained information on the operational statuses of the network and is the basis for efficient network management and optimization. Network tomography is a hot topic in the field of network measurement and is an end-to-end approach for network measurement. Unlike the traditional internal approaches for network measurement, network tomography uses the end-to-end measurements to infer the internal network performance and network states, thereby incurring low overhead to achieve the network measurement that is independent of the network composition and the network protocols. This paper systematically summarizes the representative research works about network tomography in the past few years. First, the basic model of network tomography is given and three key factors that impact the performance of network tomography are identified: the monitoring node placement, the measurement path construction, and the measurement data analysis. Then, the related works are reviewed on these three factors separately. In particular, the major limitations of existing network tomography methods in practical applications are explored, and the efficient solutions proposed in recent years are introduced. Lastly, some challenges and future research directions are discussed in the field of network tomography based on existing research works.

**Key words:** network tomography; network measurement; performance metrics; monitoring node; link identifiability; measurement data

\* 基金项目: 国家自然科学基金(61872437, 61772465); 浙江省自然科学基金(LR19F020001)

Foundation item: National Natural Science Foundation of China (61872437, 61772465); Zhejiang Provincial Natural Science Foundation (LR19F020001)

收稿时间: 2020-05-30; 修改时间: 2020-07-22; 采用时间: 2020-08-19; jos 在线出版时间: 2020-09-10

随着互联网的飞速发展,网络在组成结构、数据流量和商业用途等方面变得越来越复杂.如何认知这些网络,对理解网络时代下用户行为有着重要意义.同时,这也是有效管理网络、保证可靠网络服务的基础.网络测量是指遵照一定的方法和技术,利用软硬件工具来测试、验证及表征网络性能和状态的一系列活动.网络测量可以是对网络性能的测量,如延时、丢包率、网络流量<sup>[1]</sup>,也包括对网络拓扑<sup>[2]</sup>、路由信息等其他网络特征的测量.它是理解和认识网络行为的重要手段,在网络建模、网络安全、网络管理和优化等许多领域都有广泛应用<sup>[3]</sup>.国内外学术界和研究机构对网络测量技术进行了大量的研究,总体上,这些网络测量技术一般采用分布式测量结构,在网络内部的相关节点上,通过测量代理采集有关测量数据,如报文丢失率、延迟和流量等,这些测量数据将汇集到一个中心节点上,通过建立适当的数学模型,对测量数据进行分析 and 计算,从而实现了对网络系统的性能评估<sup>[4]</sup>.这种网络测量技术属于网络内部测量,与网络体系结构和网络协议密切相关,并且需要网络内部相关节点的密切协作,具有较高的测量精确度.但也存在一些缺陷,主要表现在如下两个方面:(1) 网络测量依赖于特定的网络协议(如 TCP/IP 协议、SNMP 协议),无法实现与网络结构和协议无关的测量;(2) 测量依赖于自治系统内部节点之间的协作,出于网络安全和商业利益等原因,有些自治系统并不对外开放,难以实现内部节点的协作和信息交流,无法保证测量准确性.

为此,近些年来,国际上许多学术机构都在寻求其他途径来研究网络的整体性能及其相互影响,将已成功应用于医学、地震学、地质勘探等领域的成熟理论和方法应用于通信网络领域,衍生出了网络断层扫描(network tomography)技术<sup>[5]</sup>.它是根据网络外部(即网络边界)的测量信息来分析和推断网络内部的性能及状态,是一种在没有网络节点协作的条件下,通过主动发包探测或被动收集网络内部有用信息的新技术,结合数理统计方法,能够有效地推测出网络链路上的 QoS 指标,如延迟分布、丢包率、吞吐率和流量等.

目前,网络断层扫描的研究主要集中于两个方面:一是测量数据的收集,主要研究如何收集网络内部相关的有用信息;二是测量数据的分析,主要研究如何根据收集的数据来获取网络内部的性能指标和运行状态.本文在简要介绍网络断层扫描基本概念和形式化表述的基础上,从测量数据收集和测量数据分析两个主要方面综述了网络断层扫描的研究现状与关键技术.随后分析了已有网络断层扫描方法在实际应用中的核心缺陷,并系统介绍了针对这些核心缺陷近年来提出的最新理论及算法.最后,基于现有研究成果对网络断层扫描技术的发展趋势和下一步的研究方向进行了讨论.

## 1 网络断层扫描基础

### 1.1 网络断层扫描的基本概念

网络断层扫描通过网络边界上进行端到端(end-to-end)的测量,来推断网络内部状态和通信行为.为此,基于网络断层扫描的测量方法包含 3 个主要步骤:① 网络模型构建阶段,明确网络拓扑结构和网络性能模型;② 测量数据收集阶段,通过网络节点或边界上部署具有监测能力的装置(即监测节点),并采用主动测量或被动测量的方式,从监测节点上收集端到端的网络测量数据;③ 测量数据分析阶段,根据网络性能模型建立性能推测方法,运用统计学理论对测量数据进行分析处理,从而推算出网络内部性能指标.

网络断层扫描技术对网络性能的推算依赖于监测节点的测量数据值,测量方式的不同,会影响到推算结果的准确性.根据测量过程中是否向被测网络中注入额外的数据包,网络测量方式可分为主动测量和被动测量<sup>[6]</sup>.其中,主动测量通过向网络中发送特定的探测数据包,并通过分析探测数据包在网络中发生的特性变化,推算得到网络内部性能和行为参数,比如链路时延和丢包率等性能指标;被动测量通过网络中部署监测节点(或数据包捕获器),收集网络节点或链路上的数据包信息,被动地监测网络,获知网络行为状况.被动测量方式不必向网络发送探测数据包,不会占用额外的网络资源.然而,被动测量要依赖所测量网络节点上的负载情况及测量链路上的通信流量,从而使得其可用性较差,实现复杂性高,难以保证测量的准确性.

当前,网络性能测量成为了网络断层扫描的一个主要研究目标.通过对网络时延、丢包率和吞吐量等性能指标进行数据采集和统计,分析网络特定范围或端到端路径的连通性及有效性,以便于发现网络瓶颈,优化网络配置,保障网络服务质量.此外,通过网络断层扫描测量得到的数据,可用于网络故障的定位,从而有助于网络维

护管理.

## 1.2 网络断层扫描的形式化表述

为了便于网络断层扫描方法的研究与设计,通常需要对网络模型、测量目标和测量方法进行形式化的表述.下面将简要介绍网络断层扫描在这几方面常见的形式化表述方式.

首先,网络模型是网络断层扫描的基础,主要指在实现网络断层扫描技术时所采用的拓扑结构.为进行简洁的描述,可用  $G=(V,L)$  表示被测网络的拓扑结构.其中,  $V$  为网络节点集合,  $L$  为网络链路集合.每个节点表示一个计算机终端(terminal)、路由器(router)或者一个包括多个计算机/路由器的子网(subnetwork).任意两个节点间的连接称为路径(path),每条路径由一条或多条链路组成,任意两个节点之间的直接连接(不需要中间节点)称为链路(link),链路可以是单向的也可以是双向的.每条链路可以是一条物理链路,也可以是几条物理链路的抽象.对于如图 1 所示的网络拓扑,其节点集合  $V=\{v_1, v_2, v_3, v_4\}$ , 链路集合  $L=\{l_1, l_2, l_3, l_4, l_5, l_6\}$ , 节点  $v_1$  与  $v_3$  之间存在 5 条不经过重复节点的路径:  $l_5, l_2l_3, l_1l_4, l_2l_6l_4, l_1l_6l_3$ .

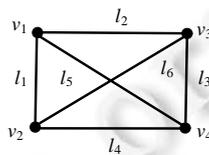


Fig.1 A sample network topology

图 1 一个简单网络拓扑

为了获得端到端的测量数据,网络断层扫描方法需要在网络部分节点上部署监测节点.用集合  $M=\{m_1, m_2, \dots, m_\mu\}$  表示网络部署的监测节点,在这些监测节点上可以发送和接收探测数据包.相应地,网络中其余的节点称为普通节点,用集合  $N(N=V \setminus M)$  表示.依赖于网络采用的路由策略,监测节点通过沿着某些测量路径发送探测数据包的形式,推算出网络节点及链路的性能指标和运行状态.在给定网络拓扑  $G$  和监测节点  $M$  的情况下,用集合  $P$  代表监测节点间的一组测量路径.需要注意的是,不同的路由策略可能会产生不同的测量路径集合.例如,当网络所有节点都运行了源路由机制(source routing<sup>[7]</sup>)时,监测节点能够严格地控制和指定每一个探测数据包的路由路径.因此,在源路由机制下,测量路径  $P$  可以包含监测节点间任意形式的路径.相反,当探测数据包的路由策略完全由网络正常运行的路由协议决定时,测量路径  $P$  只能包含监测节点间一组特定形式的路径,如最短路径.

对于一个含有  $n$  条链路的网络  $G$  (即  $n=|L|$ ),用  $\{l_i\}_{i=1}^n$  表示  $G$  中的链路集合,以及用  $\mathbf{x}=(x_1, \dots, x_n)^T$  表示链路性能指标向量,其中,  $x_i$  表示链路  $l_i$  的性能.给定监测节点间的一组测量路径  $\{p_i\}_{i=1}^{\gamma}$ ,用  $\mathbf{c}=(c_1, \dots, c_\gamma)^T$  表示测量路径性能指标向量,其中,  $c_i$  表示路径  $p_i$  的性能.在很多网络应用中,链路性能指标是可以累加的(additive).例如,多条链路上的时延是单条链路时延之和;多条链路的原始丢包率虽然为单条链路丢包率的乘积,但在经过对数函数  $(\log(\cdot))$  转换后,也可以表示成单条链路丢包率累加的形式.对于可累加的链路性能指标,一条测量路径的性能值等于这条路径上所有链路性能值的总和.因此,我们可以用一个线性方程组的方式将已知的路径测量数据与未知的链路性能指标关联起来:

$$\mathbf{R} \cdot \mathbf{x} = \mathbf{c} \quad (1)$$

其中,  $\mathbf{R}=(R_{ij})$  为一个  $\gamma \times n$  阶的测量矩阵(measurement matrix),并且测量矩阵  $\mathbf{R}$  的元素值要么为 0,要么为 1,即  $R_{ij} \in \{0, 1\}$ .具体地,当测量路径  $p_i$  经过链路  $l_j$  时,  $R_{ij}=1$ ; 否则,  $R_{ij}=0$ . 因此,在数据分析阶段,网络断层扫描的目标是在给定  $\mathbf{R}$  和  $\mathbf{c}$  的情况下,通过求解线性方程组(1),获得链路性能  $\mathbf{x}$  的值.

当链路  $l_i$  的性能指标  $x_i$  在线性方程组(1)中有唯一解时,称链路  $l_i$  为可识别的(identifiable),否则称  $l_i$  为不可识别的(unidentifiable).另外,当网络  $G$  中所有链路都是可识别的,称网络  $G$  为完全可识别的(completely identifiable); 否则,当网络  $G$  中只有部分链路是可识别的,称网络  $G$  为部分可识别的(partially identifiable).基于线性代数理论,只有当方程组(1)的测量矩阵  $\mathbf{R}$  列满秩时(即  $\text{rank}(\mathbf{R})=n$ ),网络  $G$  才能是完全可识别的.即:为了确切

地计算出网络中  $n$  链路的性能  $\mathbf{x}$ ,网络断层扫描必须要获得监测节点间  $n$  条线性无关的测量路径的性能数据.

进一步地,当网络测量的目标是判断链路通信是否正常(即链路测量指标只有正常和失效两种状态)时,我们将用到布尔网络断层扫描方法(Boolean network tomography).与常规网络断层扫描思想类似,布尔网络断层扫描通过测量监测节点间一组路径的状态,推断出网络内部链路的状态.形式化地,用  $\mathbf{x}=(x_1, \dots, x_n)^T$  表示链路状态向量以及  $\mathbf{c}=(c_1, \dots, c_p)^T$  表示测量路径状态向量.与常规网络断层扫描不同的是, $\mathbf{x}$  和  $\mathbf{c}$  为布尔向量,即元素  $x_i$  和  $c_i$  只有 0 和 1 两种取值.当  $x_i=1$ (或  $c_i=1$ )时,表示链路  $l_i$ (或路径  $p_i$ )通信发生失效;反之,当  $x_i=0$ (或  $c_i=0$ )时,表示链路  $l_i$ (或路径  $p_i$ )通信正常.考虑到对于一条路径  $p_i$ ,只有当  $p_i$  包含失效链路时, $p_i$  才会发生通信失效,可以用如下方式将已知的测量路径状态信息与未知的链路状态信息关联起来:

$$\mathbf{R} \odot \mathbf{x} = \mathbf{c} \tag{2}$$

其中, $\mathbf{R}=(R_{ij})$ 为  $\gamma \times n$  阶的测量矩阵(measurement matrix),并且元素  $R_{ij} \in \{0,1\}$ ." $\odot$ "为布尔矩阵的乘积运算,即  $c_i = \bigcup_{j=1}^n (R_{ij} \cap x_j)$ .布尔网络断层扫描的目标是在给定  $\mathbf{R}$  和  $\mathbf{c}$  的情况下,通过求解布尔方程组(2),获得链路状态  $\mathbf{x}$ .同理,当链路  $l_i$  的状态  $x_i$  可以由方程组(2)唯一确定时,称链路  $l_i$  为可识别的(identifiable).

需要说明的是:虽然方程组(1)和方程组(2)表述的是对链路性能和状态的测量,但网络中节点的性能和状态可以等价转换为与其关联链路的性能和状态,所以网络断层扫描方法仍然可以用方程组(1)和方程组(2)的形式来求解网络节点的性能和状态.

### 1.3 网络断层扫描的简单例子

图 2 所示为一个包含 7 个节点和 11 条链路(链路  $l_1 \sim l_{11}$ )的网络拓扑,其中, $m_1, m_2$  和  $m_3$  为部署的 3 个监测节点.为了测量出所有链路的时延,网络断层扫描方法在 3 个监测节点之间构造了 11 条测量路径  $p_1 \sim p_{11}$ .其中,1 条为  $m_1$  与  $m_2$  间的测量路径,7 条为  $m_1$  与  $m_3$  间的测量路径,3 条为  $m_2$  与  $m_3$  间的测量路径.

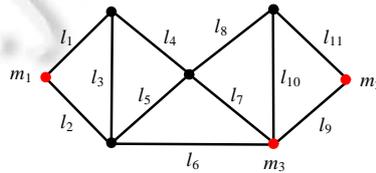


Fig.2 A network topology with three monitors:  $m_1, m_2$  and  $m_3$

图 2 部署 3 个监测节点( $m_1, m_2$  和  $m_3$ )的网络拓扑

用向量  $\mathbf{x}=(x_1, \dots, x_{11})^T$  表示链路  $l_1 \sim l_{11}$  的时延,用向量  $\mathbf{c}=(c_1, \dots, c_{11})^T$  表示测量路径  $p_1 \sim p_{11}$  的时延,从而可以得到如下形式的测量矩阵  $\mathbf{R}$ :

$$\begin{aligned} & m_1 \rightarrow m_2 : \\ & p_1 : l_1 l_4 l_8 l_{11} \\ & m_1 \rightarrow m_3 : \\ & p_2 : l_1 l_4 l_7 \\ & p_3 : l_2 l_6 \\ & p_4 : l_2 l_3 l_4 l_7 \\ & p_5 : l_2 l_5 l_8 l_{10} \\ & p_6 : l_2 l_5 l_7 \\ & p_7 : l_1 l_3 l_6 \\ & p_8 : l_1 l_4 l_5 l_6 \\ & m_2 \rightarrow m_3 : \\ & p_9 : l_9 \\ & p_{10} : l_{10} l_{11} \\ & p_{11} : l_6 l_3 l_8 l_{11} \end{aligned} \Rightarrow \mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

对于线性方程组  $Rx=c$ , 由于测量矩阵  $R$  是列满秩的(即  $rank(R)=11$ ), 所以链路时延  $x$  有唯一解, 即  $x=R^{-1}c$ . 在图 2 中, 尽管监测节点间还可以选取其他形式的测量路径(如  $l_2, l_5, l_7, l_9$ ), 但这些额外的测量路径都是与已有测量路径( $p_1 \sim p_{11}$ )线性相关的, 即, 其他形式的路径都可以用已有路径( $p_1 \sim p_{11}$ )线性表示. 因此, 从链路性能测量的角度来说,  $p_1 \sim p_{11}$  是一组最优的测量路径. 此外, 当我们任意移除一个监测节点(如  $m_2$ )后, 用剩余的 7 条测量路径( $p_2 \sim p_8$ )构成的测量矩阵将不是列满秩的, 即剩余的 7 条测量路径将不能保证所有链路的可识别性. 因此, 从链路性能测量的角度来说,  $m_1, m_2$  和  $m_3$  是一种最优的监测节点部署方式.

从图 2 的例子中可以看出, 监测节点部署和测量路径构造(或选取)对网络断层扫描方法的有效性有着重要的影响. 下一节我们将详细综述国内外在网络断层扫描监测节点部署和测量路径构造等方面的研究进展.

## 2 相关工作

近些年, 网络断层扫描作为网络测量领域的一个研究热点, 得到了国内外学者的广泛关注. 总体上, 网络断层扫描包含网络模型构建、测量数据收集与测量数据分析这 3 个主要步骤. 其中, 测量数据收集包含监测节点部署、测量路径构造与选取. 本节从监测节点部署、测量路径构造与选取、测量数据分析与推断这 3 个方面对近些年国内外代表性的研究工作进行归纳分析, 其中也包含对各研究工作网络模型的阐述.

### 2.1 网络监测节点部署

如何获取网络内部的有用信息, 是网络断层扫描研究的重要组成部分. 当前, 网络断层扫描的测量对象一般可分为网络状态参数和网络性能指标. 其中, 网络状态参数包含网络节点的配置信息和网络链路的状态信息, 网络性能指标包括链路吞吐量、链路丢包率和链路时延等. 这些指标反映网络的实时运行状态. 根据测量对象的不同, 网络监测节点部署可以分为传统网络断层扫描监测节点部署和布尔网络断层扫描监测节点部署.

#### 2.1.1 传统网络断层扫描监测节点部署

传统网络断层扫描技术主要针对网络性能指标的测量, 即: 基于监测节点间端到端的测量数据, 推断网络内部链路性能指标. 由于网络端到端测量数据的收集依赖于网络模型(包括网络拓扑、网络性能模型)和路由策略等的设定, 当前的研究工作考虑了网络断层扫描在不同网络模型和路由策略下的监测节点部署问题.

就网络拓扑结构而言, 通常可分为有向拓扑和无向拓扑两种类型: 在有向网络拓扑中, 两节点间链路(如  $v_1, v_2$ )性能在其不同通信方向上(如  $v_1 \rightarrow v_2$  和  $v_2 \rightarrow v_1$ )具有不同的表现; 相反, 在无向网络拓扑中, 两节点间链路性能在其不同通信方向上有相同的表现. 对于有向网络拓扑的链路性能测量, Xia 等人<sup>[8]</sup>证明了除非网络中所有的节点都是监测节点, 否则并不是所有链路的性能指标都是可测的. Gurewitz 等人<sup>[9]</sup>进一步证明了: 当只使用环路形式的测量路径(measurement cycle)时, 即使在网络中每个节点都是监测节点的情况下, 仍然不能准确测量出所有链路的性能指标. 对于无向网络拓扑的链路性能测量, 文献[10-13]提出了基于探测包往返时间 RTT(round-trip time)逐跳式(hop-by-hop)地测量链路时延的方法. 该方法需要网络所有节点都运行了网际控制报文协议(ICMP). 然而在有些实际的网络中, 并不是所有的节点都允许使用 ICMP 协议. 此外, 受因特网路由协议以及 ICMP 协议的限制, 探测包的往返路径可能是不对称的, 从而难以保证其测量结果的准确性. 由于逐跳测量方法在应用上存在的缺陷, 很多工作研究了端到端(end-to-end)测量方法的可行性.

文献[11, 12]研究了基于网络默认路由策略的链路性能测量问题, 并证明了部署最少数量监测节点来获得网络中所有链路的性能指标是 NP 难问题. 文献[13]进一步证明了: 即使在部分节点可以控制其本地路由策略的情况下, 部署最少数量监测节点来获得网络中所有链路性能指标的问题仍是 NP 难的. 文献[14]研究了在给定网络路由策略时的链路性能测量问题, 并提出了一个有效的近似计算方法, 以基于网络节点的连接信息估算出一个自治系统内部链路的性能指标. Gopalan 等人<sup>[15]</sup>研究了当探测包路径可控的情况下, 基于网络断层扫描技术的链路性能测量问题. Gopalan 等人<sup>[15]</sup>还在分析链路性能可测性(link identifiability)、网络拓扑和监测节点数量及位置三者之间关系的基础上, 提出了一种最优的监测节点部署算法, 并通过使用环路形式的测量路径进行探测包的发送和收集, 从而计算出网络所有链路的性能指标. 由于很多网络禁止兜圈子路由的方式, 基于环路形式测量路径的测量方法在应用中难以实现. 因此, 基于非环路形式测量路径的链路性能识别问题受到了越来越多

的关注。Ma 等人<sup>[16]</sup>推导了在仅仅使用非环路形式测量路径的情况下,网络中所有链路都可测的网络拓扑结构特征,并设计了一个最优的监测节点部署算法 MMP(minimum monitor placement)<sup>[17]</sup>。基于对拓扑连通性的划分,MMP 迭代地在每个连通分支上部署必要的监测节点,从而可以在线性时间内最小化监测节点的整体数量。这些早期工作为网络断层扫描技术的演变与改进提供了基础和指导。

现有的大多数监测节点部署算法旨在实现整个网络的可识别性,需要精确测量所有链路的性能指标。为了实现这一目标,即便最优的算法也可能需要部署数目不少的监测节点,如将 ISP 网络中,64%的节点部署为监测节点<sup>[17]</sup>。因此,部分网络可识别性(partial identifiability)问题受到了越来越多的关注。基于探测包可控路由的假设,文献[18]提出了一个有效的判别算法 DIL-2M,以快速地识别出一个部署有两个监测节点的网络中所有可识别的链路。基于该算法,文献[19]提出了一个近似最优的监测节点部署算法 GMMP,通过优化对给定数目监测节点的部署,从而最大化网络中可测链路的数量。另一方面,注意到有些实际应用往往只需要获取网络某些链路的性能,如处在网络关键位置上的链路。

为了提高网络断层扫描的灵活性,Dong 等人<sup>[20,21]</sup>提出了优先链路(preferential link)的概念,将网络中需要测量的链路指定为优先链路。为了降低优先链路测量的复杂度及减小监测节点的部署范围,Dong 等人首先提出了一种网络拓扑裁剪算法 Scalpel<sup>[20]</sup>。Scalpel 包含两个主要阶段:第 1 个阶段,从原拓扑中裁掉不含优先链路的 2-点连通分支(biconnected component)(如图 3(a)所示);第 2 个阶段,从第 1 阶段裁剪后的图中裁掉一些边缘的 3-点连通分支(triconnected component)和环分支(cycle)。需要注意的是:在剩余的拓扑图中部署监测节点,可能无法达到在原拓扑图中部署的效果。为了不影响监测节点部署的性能,Scalpel 会从每一个被裁剪的分支中选取一个辅助节点(如图 3(b)所示)。在使用 Scalpel 算法对原拓扑图  $G$  进行裁剪后,得到一个剩余的拓扑图  $G_r$  及辅助节点集合  $H$ 。以图  $G_r$  和节点集合  $H$  为输入,Dong 等人进一步设计了最优监测节点部署算法 OMA(optimal monitor assignment)<sup>[21]</sup>,通过从图  $G_r$  和辅助节点集合  $H$  中选取最少数量的节点作为监测节点,实现对原网络拓扑  $G$  中所有优先链路的性能指标的测量。上述这些旨在实现部分网络可识别性的监测节点部署算法有助于服务提供商和管理员以较小的成本满足网络测量需求,提高了网络断层扫描算法的灵活性。

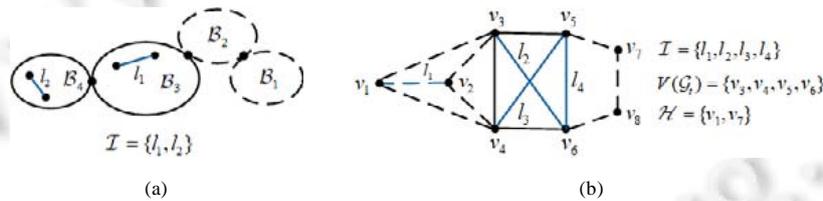


Fig.3 Illustrative example for the two stages of topology trimming algorithm

图 3 网络拓扑裁剪算法包含的两个阶段的说明示例

### 2.1.2 布尔网络断层扫描监测节点部署

当服务提供商和管理员想要测量的是网络状态参数时,要用到布尔形式的网络断层扫描技术。其中,网络内部运行状态可以用有限个零散数值表示,例如,用“0”表示链路通信正常,用“1”表示链路通信故障。

网络拥塞已成为网络传输性能和扩展性的重要影响因素,及时准确地定位网络拥塞,有助于网络性能调优和协议设计。基于网络中同时发生拥塞的链路数目较少的假设和给定的端到端路径测量信息,文献[22–24]使用压缩感知技术来识别因特网中发生拥塞的链路。在算法的实现中,通过引入一个用于判断链路通信正常与否的阈值,从而将链路性能小于阈值的数值转化为零。通过压缩感知技术,可以有效地恢复出链路性能大于阈值的数值,即识别网络中拥塞的链路。对于链路通信故障的定位问题,文献[25–28]利用概率统计方法从网络中找出一组最少数量的链路故障,从而能够完全解释测量的端到端的路径状态信息。文献[29,30]基于端到端的测量信息,使用贝叶斯方法计算网络中每一条链路发生故障的概率。

上述研究工作对网络允许的链路故障数目具有较严格的要求,在大规模网络场景中,难以保证其有效性。针对特定形式的通信模型(如客户-服务器),其网络拓扑往往相对固定。Duffield 等人<sup>[31]</sup>提出了一种简捷有效的链

路故障定位算法,利用客户端与服务器之间路径的相似性,能够基于少数端到端测量信息,定位网络故障链路.进一步地,Ahuja 等人<sup>[32]</sup>形式化分析了网络拓扑结构、监测节点部署和链路故障识别三者的关系,并推导了在使用包含环路的测量路径时,链路故障的可识别性条件,并提出了一个最优监测节点部署算法,以准确定位出任意  $k$  条链路的通信故障.另一方面,有些工作关注于对网络节点通信故障的监测和定位.文献[33–35]研究了在不同探测包路由策略(任意可控路由、可控无环路由和不可控路由)下,网络节点通信故障的定位问题.具体地,文献[33,34]提出了有效的算法来计算对于给定的网络拓扑和监测节点部署,在一个特定节点集合中,最多可以同时发生故障的节点数目,以及给定允许发生故障的节点数目时,可以明确推算出状态的最大节点集合.针对采用客户-服务器通信模型的网络,He 等人<sup>[36]</sup>提出了有效的服务器节点部署算法,通过测量服务器节点与客户端之间的路径信息,从而最大化可定位节点故障的数量.该算法有利于拓展布尔网络断层扫描的适用范围.更为一般地,Bartolini 等人<sup>[37]</sup>推导了在给定测量路径数目、路由策略和网络拓扑形式(如树状和网状拓扑)等一般性信息的情况下,可以定位的发生通信故障节点的最大数量.该理论结果可以为网络拓扑设计和路由选择等操作提供参考和指导.

## 2.2 测量路径构造与选取

在网络断层扫描中,端到端测量数据的收集取决于使用的测量路径形式,不同的测量路径很可能产生不同的测量数据,从而可能得到不同的测量结果.近些年,很多工作从几何代数的角度研究了测量路径的构造和选取问题.当测量路径可以包含环路时,Gopalan 等人<sup>[38]</sup>提出了一种基于给定网络监测节点的线性无关(linearly independent)测量路径构造方法,并形式化证明了算法构造的测量路径是监测节点之间存在的一组最大数量的线性无关路径.该算法为基于环路测量路径的网络断层扫描技术(如最少监测节点部署<sup>[18]</sup>)的实际应用奠定了基础.当测量路径不允许包含环路时,Wing 等人<sup>[39]</sup>推导了不同网络(有向和无向拓扑网络)中线性无关路径数目的界限值.

另一方面,Ma 等人<sup>[40]</sup>提出了一种具有多项式时间复杂度的无环测量路径构造算法 STPC,基于网络中部署的监测节点构造出一组最大数量的线性无关且不含环路的测量路径,实现对所有链路性能的测量.该算法也为基于无环测量路径的网络断层扫描(如最少监测节点部署<sup>[17]</sup>)的实际应用奠定了基础.为了简化监测节点间测量路径的构造,STPC 首先对网络监测节点进行合并操作.具体地,在网络拓扑中,STPC 引入一个虚拟监测节点,并分别用一条虚拟链路将原有的每个监测节点与虚拟监测节点连接起来.由于原拓扑链路(及路径)与新拓扑链路(及路径)之间存在着——对应的关系,监测节点合并操作不会影响网络链路的可识别性,从而可以将原监测节点间无环测量路径的构造问题转换为虚拟监测节点上有环测量路径的构造问题.鉴于生成树不含回路的特性,STPC 接着在新拓扑图中以虚拟监测节点为根节点,构造了 3 棵独立的生成树(如图 4 所示).在每个网络节点上,通过这 3 棵生成树的两两组合,可以得到 3 条从该网络节点到虚拟监测节点的测量路径.最后,再将虚拟监测节点上的测量路径转换为实际监测节点间的测量路径.此外,将测量路径数据表示为线性方程组的形式后,生成树各节点上测量路径的嵌套结构使得对方程组的求解不需要采用传统高斯消元法的方式,即显示反转方程组系数矩阵.因此,STPC 能够确保在线性时间内计算出所有网络链路的性能指标.需要说明的是:STPC 在构造测量路径时需要网络中存在足够多的监测节点,以确保对所有链路性能指标的推算.这一前提条件可能会给网络管理员带来不小的测量成本,从而导致 STPC 在实际应用时面临局限性.

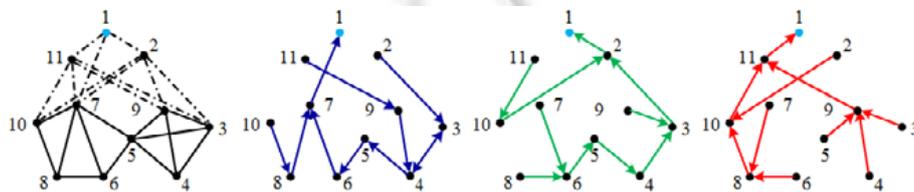


Fig.4 Illustrative example with a network topology and three independent spanning trees rooted at node 1

图 4 网络拓扑及其以节点 1 为根节点的 3 棵独立生成树的说明示例

测量路径的使用数量,很大程度上决定了探测包流量的大小<sup>[41-44]</sup>.为降低测量负载对网络正常通信的影响,有些工作关注于对测量路径的选取及测量路径数目的减少.针对网络路径性能测量问题,Chen 等人<sup>[45]</sup>提出了一个在给定监测节点部署的测量路径选取方法,通过选取并测量一定数量线性无关的路径,从而可以计算出其他路径的性能指标.为了说明测量路径选取方法的可扩展性,Chen 等人<sup>[45]</sup>还推导了在任意规模的网络中,线性无关路径选取数目的界限值,并将该界限值表达成网络链路数目的线性函数形式.该结果为路径选取方法在大规模网络中的应用提供了坚实的理论依据.注意到网络路径大多都是线性相关的,文献[46,47]基于给定的一组路径测量信息,利用线性代数方法计算出其他未被测量的子路径上的性能指标.针对网络链路性能测量问题,Tang 等人<sup>[48]</sup>基于网络拓扑和路径相关性信息设计了一种有效的路径选取方法,通过收集监测节点间少量路径的性能数据,推算出网络内部细粒度的链路性能界限值.

此外,针对网络中指定链路(即优先链路)的性能识别问题,基于实现网络部分可识别性的最优监测节点部署算法(如 OMA<sup>[21]</sup>),Zheng 等人<sup>[49]</sup>设计了一种测量路径选取方法 PathSelection.为了尽可能地减少测量路径的数目,PathSelection 首先从给定监测节点间的候选路径中确定出冗余的测量路径:可以被其他路径代替的路径以及对优先链路性能识别没有帮助的路径.然后,PathSelection 基于二分图模型(bipartite model)分析每一条优先链路识别与路径组合的关系,并迭代式地从二分图中找出一组可以识别所有优先链路且数目最少的路径作为最终使用的测量路径.在给定候选路径的情况下,PathSelection 虽然可以有效地减少测量路径数目,但在大规模网络中,要从给定的候选路径中找出所有可以识别优先链路性能的路径组合,并在二分图中确定一组数量最少的路径,可能会造成很大的时间开销.因此,如何从候选路径中快速地找到一组对链路性能识别最有帮助的测量路径,仍然是提高网络断层扫描有效性的一个关键问题.

### 2.3 测量数据分析与推断

在网络断层扫描技术中,监测节点按照一定规律(如周期性)或一定模式(如泊松分布)沿着测量路径发送探测数据包,将可以收集到一组(或一系列)端到端的测量数据.如何基于收集的测量数据准确地推断网络内部性能和状态对测量方法的有效性显得尤为重要.本节将简要介绍网络断层扫描常用的数据分析与推断方法.

根据网络性能参数特征,现有网络断层扫描的数据分析方法大致可以分为统计学习方法和代数计算方法.其中,统计学习方法针对的是性能参数在探测包收集过程中会变化的测量数据,而代数计算方法针对的是性能参数在探测包收集过程中相对稳定的测量数据.

基于统计学习的数据分析方法将路径性能和状态作为一个整体,通过端到端测量数据收集一个测量样本,并结合网络拓扑模型和网络性能模型推测出链路性能或通信状态的概率分布.其基本思想是:利用第 1.2 节公式(1)或公式(2)所示模型构造一个分布函数  $f(\mathbf{c};\mathbf{x})$  来描述网络性能指标或通信状态特性的依赖关系.如果存在完整观测样本,就可以用参数估计的方法确定  $\mathbf{x}$ .然而在网络断层扫描中,通常只能观测到部分样本数据,这使得数据推断变得比较复杂,需用统计学习的方法来估计那部分不确定的、观测不到的数据<sup>[50]</sup>.常见的基于统计学习的测量数据分析过程如图 5 所示.网络性能指标或通信状态对应于一个未知向量  $\mathbf{x}$ ,并在参数空间定义一个点,意味着已确定的推理估计目标.端到端测量数据  $\mathbf{c}$  是观测空间的一些点,每个测量数据都在这个空间产生一个特有的点.函数  $f(\mathbf{c};\mathbf{x})$  表示这些空间之间连接的统计量,实际上是一组概率密度函数.通过这些观测数据,推理方法就能利用估计规则  $\mathbf{x}'=g(\mathbf{c})$  来揭示测量对象  $\mathbf{x}$  的值.目前,网络断层扫描主要的统计学习方法包含极大似然估计方法(maximum likelihood estimation,简称 MLE)<sup>[51,52]</sup>、期望最大化(expectation maximization,简称 EM)方法<sup>[53]</sup>和贝叶斯(Bayesian)估计方法<sup>[54]</sup>等.

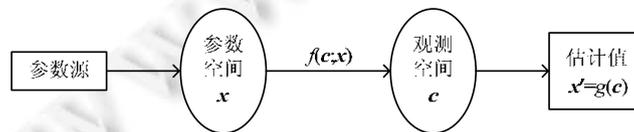


Fig.5 Statistical learning procedures for measurement data analysis

图 5 测量数据分析的统计学习过程

为测量所有链路性能的概率分布,在测量数据收集阶段,文献[55,56]利用多播的方式进行网络探测包的发送和接收,并提出了相应的多播树构造算法,证明了这种测量方式具有覆盖范围广、测量开销低的优势.在测量数据分析阶段,Chen 等人<sup>[57]</sup>应用傅里叶变换的技术对给定的路径性能数据做预处理后,结合广义矩估计的方法推断出单条链路性能参数的概率分布.注意到测量数据的质量可能影响到数据分析准确性,与传统给定路径测量信息来推断链路性能的方法不同的是,He 等人<sup>[58]</sup>基于费希尔信息(Fisher information)理论设计了一种探测包分配框架,用于计算在给定一组测量路径和给定探测包总数目的情况下,单条测量路径上探测包的发送数量,以使得收集到的路径测量信息可以最大化链路性能估计的准确率.同时,文献[58]将该探测包分配框架应用到链路时延测量和链路丢包率测量的实际问题中,基于框架收集到的路径测量信息和优化算法(D-最优化和 A-最优化),准确地估算出了网络链路时延和丢包率的概率分布情况.针对网络链路通信故障的定位问题,文献[25–28]利用极大似然估计方法从网络中找出一组最少数目的链路故障,从而能够完全解释收集的端到端路径状态信息.文献[29,30]基于给定的端到端的测量数据,利用贝叶斯学习方法计算网络中每一条链路故障的概率.

基于代数计算的数据分析方法的基本思想是:将链路性能指标或通信状态刻画为一个未知常量,通过收集一组端到端路径的测量数据,结合网络拓扑模型和网络性能模型推测出链路的性能或状态.对于可累加的网络性能指标(如链路时延),一条端到端路径的性能指标等于路径上所有链路性能指标的总和.因此,代数计算方法可以将收集的路径测量数据表述成一个线性方程组的形式(如公式(1)所示),然后应用线性代数理论求解形式化的线性方程组,从而获得有唯一解(即唯一确定)的链路性能指标值.目前,网络断层扫描主要的代数计算方法有高斯消元法<sup>[18,19]</sup>、LU 分解法<sup>[22,23]</sup>、逆矩阵法<sup>[15,16]</sup>和增广矩阵法<sup>[49]</sup>等.基于高斯消元法的数据分析具有较好的通用性,在实际网络断层扫描中应用得较多,但其计算复杂度较大,在可扩展性方面有局限性.相反,基于 LU 分解法和逆矩阵法的数据分析具有相对稳定的计算开销,但方法的使用需要满足一定的前提条件,如 LU 分解法需方程组系数矩阵(公式(1)中的  $\mathbf{R}$ )可分解成两个三角矩阵的乘积,逆矩阵法需方程组系数矩阵(第 1.2 节公式(1)中的  $\mathbf{R}$ )是满秩可逆的.基于增广矩阵的数据分析也具有较好的通用性,适用于需要测量不可识别(即无唯一解)的链路性能指标范围的场景.

### 3 应用挑战与关键技术

表 1 总结了近些年网络断层扫描方法的相关研究成果.

概况地说,文献[11,12,15,17]从探测数据包路由策略的角度研究了在实现网络完全可识别性时的监测节点部署问题.其中,文献[11,12]基于网络默认路由协议的链路性能测量问题,并证明在基于网络默认路由(即不可控路由)策略下最少监测节点的部署是一个 NP 难问题.随后,文献[15,17]分别从测量路径是否允许包含环路的角度提出了基于可控路由策略的监测节点部署理论和线性时间复杂度算法,从而为网络断层扫描技术的演变和改进奠定了基础.受网络测量成本的限制,文献[18–21]从实现网络部分可识别性的角度解决了监测节点的部署问题,确保可以用有限的成本满足服务提供商和网络管理员的测量需求,从而提高了网络断层扫描算法的灵活性.另一方面,网络内部运行状态也是网络测量的重要组成部分.文献[32–34,36]从探测数据包路由策略的角度研究了对特定数目链路(或节点)失效定位的监测节点部署问题,从而进一步拓展了网络断层扫描技术的适用范围.此外,文献[38,40,45,49]从测量对象和路由策略的角度研究了测量路径的构造和选取问题,并提出了相应的多项式时间复杂度算法.这些工作为网络断层扫描技术在各种形式网络中的实际应用提供了必要的基础.

从表 1 中可以看出:当前已有许多网络断层扫描的理论和方法,但这些方法的设计在测量目标(或测量对象)、网络模型和测量成本等方面都有较为明确且严格的规定(或假设).这些规定假设虽然有利于方法的演变与验证,但也给方法的实际应用带来了局限性.

总体上,现有的网络断层扫描方法在性能测量粒度(及精度)、网络拓扑结构和网络通信模型这 3 个核心方面都带有较强的假设和要求(见表 2),这影响了网络断层扫描方法在实际应用中的有效性和通用性,甚至使得方法在有些网络场景中难以被使用.总体上,设计一个通用、高效及鲁棒的网络断层扫描方法面临以下 3 项关键挑战:测量成本非常受限、网络拓扑的动态性以及网络通信的易失效性.近两年,越来越多的研究者开始关注已

有网络断层扫描方法的局限性,并提出了相应理论和关键技术,以有效应对这些挑战.本节归纳了在应对上述挑战的一些代表性研究工作及其关键技术.

**Table 1** Summary of network tomography methods

表 1 网络断层扫描方法总结

文献	研究目标	网络模型	测量路径	问题难度	算法复杂度
Bejerano 等人 <sup>[11,12]</sup>	部署最少数目监测节点实现对网络所有链路性能测量	静态、有失效	不可控	NP 难问题	$O( V ^3)$
Gopalan 等人 <sup>[15]</sup>	部署最少数目监测节点实现对网络所有链路性能测量	静态、无失效	可控、有环路	P 问题	$O( L )$
MMP <sup>[17]</sup>	部署最少数目监测节点实现对网络所有链路性能测量	静态、无失效	可控、无环路	P 问题	$O( V + L )$
OMP <sup>[18]</sup>	部署 2 个监测节点实现对网络可识别性链路数目的最大化	静态、无失效	可控、无环路	P 问题	$O( V ^3+ V ^2L)$
GMMP <sup>[19]</sup>	部署 $\kappa(\kappa \geq 2)$ 个监测节点实现对可识别性链路数目的最大化	静态、无失效	可控、无环路	NP 完全问题	$O( V ^3+ V ^2L)(\kappa=2)$ $O( V ^4+ V ^3L)(\kappa \geq 3)$
Scalpel <sup>[20]</sup> OMA <sup>[21]</sup>	部署最少数目监测节点实现对所有优先链路性能测量	静态、无失效	可控、无环路	P 问题	$O( V + L )$
Ahuja 等人 <sup>[32]</sup>	部署最少数目监测节点实现对 $k(k \geq 1)$ 条链路失效的定位	静态、有失效	可控、有环路	P 问题	$O( L )(k=1)$ $O(k V^4 )(k \geq 2)$
Ma 等人 <sup>[33,34]</sup>	部署最少数目监测节点实现对 $k(k \geq 1)$ 个节点失效的定位	静态、有失效	可控、无环路	P 问题	$O( V + L )(k=1)$ $O( V ^3 L )(k \geq 2)$
He 等人 <sup>[36]</sup>	部署最少数目监测节点实现对单个节点失效的定位	静态、有失效	不可控	NP 难问题	$O( V ^3+ V  L )$
CLIC <sup>[38]</sup>	在给定的监测节点间构造一组测量路径实现对最大数目链路性能测量	静态、无失效	可控、有环路	P 问题	$O( V ^2)$
STPC <sup>[40]</sup>	在给定的监测节点间构造一组测量路径实现对所有链路性能测量	静态、无失效	可控、无环路	P 问题	$O( V  L )$
Chen 等人 <sup>[45]</sup>	从给定( $r$ 条)路径中选取一组最少数目( $k$ 条)的测量路径实现对所有其他路径性能测量	动态、有失效	不可控	P 问题	$O(rk^2)$
Zheng 等人 <sup>[49]</sup>	从给定( $r$ 条)路径中选取一组最少数目的测量路径实现对所有优先链路性能测量	静态、无失效	不可控	NP 难问题	$O(r^3)$

**Table 2** Research status and key application challenges

表 2 现状及应用挑战

现状	应用挑战
测量粒度单一固定	测量成本非常受限
网络拓扑静止不变	网络拓扑动态变化
网络通信稳定可靠	网络通信容易失效

### 3.1 粒度可调网络断层扫描理论与技术

#### • 挑战

现有的网络断层扫描方法大多着重于实现对网络中所有链路性能指标确切值的测量,即测量粒度(及精度)单一且固定.这些方法存在两个主要缺陷:一是受网络拓扑和路由策略等的限制,测量网络所有链路性能的目标在很多应用场景中难以实现;二是测量链路性能指标的确切值往往会带来很大的开销,例如需要部署很多监测节点和发送大量探测数据包.为此,提出粒度(及精度)可调的网络断层扫描理论,并设计一种粒度(及精度)可调的网络断层扫描算法以有效平衡性能指标测量的精度与测量的开销,显得很有必要.

#### • 关键技术

在很多实际应用中,服务提供商与网络管理员通常关注的是网络内部性能是否符合与用户签订的服务水平协议(service level agreement,简称 SLA)的要求,如网络丢包率是否处在 SLA 规定的范围内<sup>[59]</sup>.因此,链路性能指标的界限值将足以满足服务提供商与网络管理员的性能测量需求.鉴于此,Feng 等人<sup>[60]</sup>提出了一种有效的网络断层扫描方法,基于给定的监测节点和端到端路径测量数据,推算出网络中所有链路性能指标的界限值(包括

性能上界和性能下界).简单地,对于可识别的链路,其性能指标值可由与端到端测量数据对应的线性方程组唯一确定,所以可识别链路的界限值为同一个数,即,其性能界限的上界值等于下界值.对于不可识别的链路,已有网络断层扫描方法不会提供任何关于其性能指标的信息,而文献[60]的方法可以求解出其性能指标的界限值,且此时其上界值不等于下界值.

以图 6 为例,网络中有两个节点( $v_2$  和  $v_3$ )部署为了监测节点.假设网络中 10 条链路( $l_1 \sim l_{10}$ )上的时延分别为  $\{x_1=7, x_2=1, x_3=5, x_4=4, x_5=6, x_6=8, x_7=13, x_8=10, x_9=3, x_{10}=3\}$ .需要说明的是,服务提供商与网络管理员一开始并不知道这些链路的时延,只有通过收集两个监测节点间端到端路径的时延数据,才能推算出单条链路的时延.因为网络中只有  $v_2$  和  $v_3$  为监测节点,基于链路可识别性条件<sup>[16,18]</sup>,已有的网络断层扫描方法只能测量出链路  $l_1$  和  $l_6$  的时延,即只有链路  $l_1$  和  $l_6$  是可识别的.而对于网络中不可识别链路( $l_2 \sim l_5$  和  $l_7 \sim l_{10}$ )的时延,已有的网络断层扫描方法不能提供任何信息.不同的是,通过使用文献[60]的网络断层扫描方法,网络管理员可以计算出不可识别链路上的时延范围: $B(x_2)=[0,7], B(x_3)=[2,9], B(x_4)=[3,10], B(x_7)=[0,27], B(x_8)=[0,27], B(x_9)=[0,7], B(x_{10})=[0,7]$ .

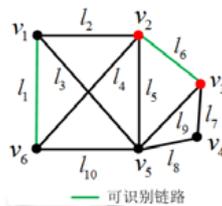


Fig.6 Illustrative example of network tomography for inferring performance metric bounds

图 6 测量性能指标界限值的网络断层扫描说明示例

通过对推算的链路性能指标界限值区间长度(bound interval length)赋上不同的约束,我们可以实现一种粒度可调的网络断层扫描.特别地,当要求链路性能指标界限值的区间长度为 0 时(即要获得链路性能指标的确切值),文献[60]的方法转化为已有网络断层扫描方法.因此,文献[60]中的网络断层扫描方法是已有的网络断层扫描方法在测量目标上的一种推广演变.相比已有的旨在实现对网络所有链路性能指标确切值的方法,虽然文献[60]的网络断层扫描方法通过推算链路性能指标界限值的方式能够在很大程度上减少测量的开销,但推算网络所有链路性能指标界限值的复杂度与网络规模有很大关系,在大规模网络中容易造成指数级时间复杂度.

然而有些场景并不需要测量所有链路性能指标界限值,相反,服务提供商与网络管理员只需要获取一部分网络链路的性能指标界限值.比如在因特网中,被用户反映通信不正常的链路要比其他链路重要得多<sup>[61]</sup>;在无线传感网中,位于基站(或网关)附近的链路由于承载的数据流量更大,通常成为限制整个网路系统性能的瓶颈<sup>[62]</sup>.通过测量网络中这些优先链路(或目标链路)的性能指标界限值,可以进一步减少测量开销.更重要的是,还可以降低测量复杂度,有利于提高方法在大规模网络中的有效性.为此,我们研究了基于断层扫描的网络优先链路(preferential link 或 interesting link)性能指标界限值的计算理论及测量技术<sup>[63]</sup>.不同于文献[60]需推算网络所有链路性能指标界限值的要求,我们的方法关注于对服务提供商和网络管理员在网络中任意指定的一组优先链路性能指标界限值的高效测量.当指定网络所有链路为优先链路时,优先链路性能指标界限值的求解问题将转化为网络所有链路性能指标界限值的求解问题.因此,我们提出的网络断层扫描算法<sup>[63]</sup>相比文献[60]的网络断层扫描方法更加灵活和普适,具有更大的应用范围.下面我们简要介绍实现对优先链路性能测量粒度可调的网络断层扫描理论及技术<sup>[63]</sup>的核心思想.

考虑到服务提供商与网络管理员在不同阶段可能具有不同的测量需求,如:在网络运行阶段,关注的是如何通过已有监测节点有效地计算出网络中优先链路性能的界限值;在网络扩建或升级阶段,关注的是如何部署更多数目的监测节点以缩小计算的优先链路性能界限的距离.为此,我们在文献[63]中同时解决了以下两个关键问题.

- 优先链路性能指标界限值计算:在一个给定监测节点和优先链路的网络中,如何有效地推算出这些优

先链路性能指标的最紧界限值,即最紧的上界值和最紧的下界值?

- 监测节点部署:在一个给定初始监测节点和优先链路的网络中,如何通过部署额外的监测节点以进一步缩小优先链路性能界限区间的长度(即上界值和下界值的差值)?

对于优先链路性能指标界限值的计算问题,基于对由端到端测量构造的线性方程组的可解情况及解空间分析,我们提出了网络优先链路性能指标界限值的计算理论,其中包含多种降低优先链路性能界限值计算复杂度的策略.由线性代数理论可知:方程组中不可解的变量可以分为自由变量和非自由变量两种类型,其中每个非自由变量可以表示成若干个自由变量的线性表达形式.然而,在一个没有唯一解的方程组中,通常有非常多种自由变量的组合方式.而每种自由变量组合方式都对应着一种方程组解的形式,在每种方程组解的形式中,都可以获得一个优先链路性能指标界限值.因此,在优先链路性能指标界限值的计算问题中,一个关键的挑战是如何确定出可以导致优先链路性能指标界限值最紧的一种自由变量组合方式.通过枚举所有自由变量组合方式并比较每种组合方式下优先链路性能指标界限值的方法,容易造成指数级的时间复杂度.注意到:在线性方程组中,对于一个特定变量(即目标变量),只有与其线性相关的变量才能影响到目标变量的求解.为此,在优先链路性能指标界限值的求解过程中,我们只选择与优先链路性能变量线性相关的变量作为自由变量,从而可以很大程度上缩小自由变量的选取范围,降低优先链路性能界限值计算的时间复杂度.

对于监测节点部署问题,基于对链路可识别性与优先链路性能界限值松紧程度关系的分析,我们设计了一个高效的监测节点部署算法 NMPI<sup>[63]</sup>.在网络原有监测节点的基础上,部署一个新的监测节点,从而最大程度上减小优先链路性能界限区间的长度.首先,NMPI 的设计基于一个重要的观察:在一个包含一组优先链路的网络中,对于新监测节点的部署,不可识别的优先链路性能指标界限区间长度只会受由新监测节点部署增加的不可识别链路的影响,从而相同的链路可识别性也会造成相同的优先链路性能指标界限区间长度.因此,在监测节点部署过程中,对网络链路可识别性变化的分析成为了一个关键的问题.

由于网络可能部署有多个监测节点,而链路可识别性与单个监测节点之间存在着复杂的关系,为了简化监测节点部署对网络链路可识别性变化的分析,NMPI 对网络拓扑进行扩展操作.具体地,对于一个已有  $\kappa(\kappa \geq 2)$  个监测节点的网络  $G$ ,引入两个虚拟监测节点(virtual monitor)  $m'_1$  和  $m'_2$ ,同时,在实际监测节点(actual monitor)和虚拟监测节点间增加  $2\kappa$  条虚拟链路,在两个虚拟监测节点之间也增加一条虚拟链路(virtual link),将新生成的拓扑称为网络扩展图  $G_{ex}$ (如图 7 所示).由于原网络拓扑实际监测节点之间的端到端测量都可以转化为虚拟监测节点  $m'_1$  和  $m'_2$  之间的端到端测量,并且对于网络链路的测量,  $m'_1$  和  $m'_2$  之间的端到端测量相比实际监测节点之间的端到端测量不会提供额外的信息,所以原网络拓扑  $G$  的链路可识别性与扩展图  $G_{ex}$  的链路可识别性是一样的,从而 NMPI 可以关注于仅包含两个监测节点的扩展图  $G_{ex}$  链路可识别性的变化.随后,NMPI 通过对新监测节点部署与网络扩展图  $G_{ex}$  结构变化关系的分析,实现了新监测节点在网络扩展图  $G_{ex}$  及原拓扑  $G$  中的最优部署.

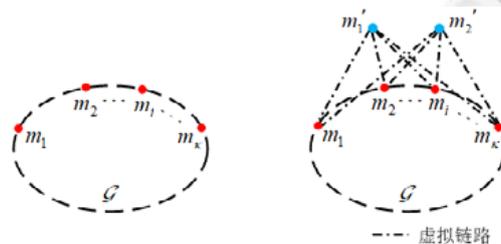


Fig.7 Illustration of topology extension operation

图 7 拓扑图扩展操作示例

此外,还有一些工作从测量对象的角度研究了粒度可调的网络断层扫描理论与技术.注意到在实际应用中,网络用户通常关心的是所用网络服务的性能,例如从客户端向服务器发送请求到客户端收到回应的时延,Yang 等人<sup>[64]</sup>解决了网络任意节点间端到端通信路径性能的测量问题,并提出了一种最优的监测节点部署算法 MPIP,通过收集监测节点间一组端到端测量数据,以实现任意指定路径(即优先路径,interested path)性能指标的推

算.为便于网络监测节点的部署,文献[64]首先研究了基于断层扫描的网络优先路径性能的可识别性理论,形式化表述了拓扑结构、监测节点和路径可识别性这三者的关系,并解决了给定监测节点部署时,网络可识别路径的判断问题.基于网络路径可识别性条件,MPIP 将网络拓扑划分成多个 2-点连通和 3-点连通分支,根据各连通分支中优先路径特性,迭代地在各连通分支中部署必要的监测节点,从而用最少的监测节点实现对所有优先路径性能指标的计算.需要说明的是,上述这些粒度(及精度)可调网络断层扫描方法<sup>[60,63,64]</sup>在实现过程中,使用的是不含环路的端到端测量路径,这有利于提高方法在实际网络中的可用性.

### 3.2 动态拓扑网络断层扫描理论与技术

- 挑战

随着物联网、软件定义网络等新型网络形态与网络技术的兴起,动态的网络路由拓扑越来越成为这些新型网络的重要特点<sup>[65]</sup>.例如在典型的物联网场景中,动态路由协议(如 RPL<sup>[66]</sup>、低功耗蓝牙 mesh 网络路由等)被广泛用于应对无线信道、节点位置等的动态性,从而使得网络路由拓扑不断发生变化.而在软件定义数据中心网络中,流量分割(traffic splitting)<sup>[67]</sup>技术及重路由(re-routing)<sup>[68]</sup>技术被广泛用于负载均衡,使得同一个数据流的数据包可以通过不同的路由到达接收端,从而导致了网络路由拓扑的动态性.现有网络断层扫描技术<sup>[16,17,19,21]</sup>的典型工作模式是基于已知的静态网络拓扑,部署一定数量的监测节点并获取监测节点间探测包的测量数据,然后结合网络拓扑信息建立关于内部运行状态的方程组,最后通过求解方程组完成网络测量的目标.然而,当网络路由拓扑动态变化时,某一时刻存在的路由路径下一时刻可能就不存在了,而某一时刻必要的监测节点在另一时刻可能是冗余的.因此,如何在动态拓扑的条件下进行精确、高效的网络断层扫描,是一个非常重要的问题.

- 关键技术

在动态拓扑网络中,由于链路(或节点)的动态性,在某一个时刻的拓扑中部署的监测节点可能并不适用于另一个时刻拓扑中链路的性能测量.一种简捷的动态拓扑网络测量方法是:当网络拓扑变化后,利用基于静态拓扑的网络断层扫描方法<sup>[21,25]</sup>在新生成的网络拓扑中重新部署一遍监测节点.这种应急式(reactively)的部署方法虽然可以实现对多个网络拓扑的性能测量,但频繁地更换监测节点部署位置一方面给网络管理员带来了不小的成本开销,另一方面也不利于网络测量任务的平稳进行.因此,在网络拓扑变化前,先验式(proactively)地实现对链路性能的测量显得尤为重要.

He 等人<sup>[69]</sup>首次研究了动态拓扑网络链路性能的测量问题,并提出了一种有效的网络断层扫描方法,通过在网络规划初期,一次性部署一定数目的监测节点,以实现对网络所有可能拓扑中所有链路的性能测量.然而,要想获得动态拓扑网络中所有链路的性能,即使最优的监测节点部署算法也可能需要使用大量监测节点(如网络 90% 以上的节点需为监测节点<sup>[69]</sup>),这将会造成很大的测量成本.另一方面,很多时候,服务提供商与网络管理员只关注网络某些链路的性能指标,例如处在网络关键位置上的链路性能指标.通过测量网络中优先链路的性能指标,在满足应用需求的同时,也可以很大程度上降低测量成本.为此,我们研究了面向动态拓扑的优先链路测量理论及技术,其中包含多种高效及鲁棒的网络断层扫描方法<sup>[70]</sup>,通过在动态拓扑网络中部署尽量少的监测节点,以实现对所有可能拓扑中优先链路的性能测量.不同于文献[69]计算网络所有链路性能指标的要求,我们的方法旨在用更少的成本计算出网络中任意指定的一组优先链路的性能指标.当指定网络所有链路为优先链路时,对优先链路性能指标的测量问题将转化为对网络所有链路性能指标的测量问题.因此,我们提出的网络断层扫描方法<sup>[70]</sup>是文献[69]中网络断层扫描方法的泛化,具有更好的测量灵活性.下面我们简要介绍实现对动态拓扑网络中优先链路性能有效测量的网络断层扫描方法<sup>[70]</sup>的核心思想.

网络断层扫描方法的设计需要网络拓扑信息作为基础,而尽早地获取动态网络在各个时刻的拓扑信息,有助于监测节点的部署.因此,我们首先研究了基于网络拓扑预测的优先链路可识别性理论,从而为动态拓扑网络监测节点的部署提供指导.由于动态网络拓扑结构的周期性和可预测性(例如行车路线和运行时刻固定的公交车网络、睡眠/工作模式周期性交替的低功耗传感网),可以利用拓扑时空相关性及过去一段时间的网络拓扑,预测未来一段时间内的网络拓扑.另一方面,我们设计了一种通用及简洁的时变网络模型来刻画动态拓扑网络的在各个时刻的拓扑结构.简单地说,我们将网络测量周期(或生命周期)划分成若干个等长的时隙,然后将网络

在各个时隙上的拓扑结构用如下序列的形式来表述:

$$\{G_1, G_2, \dots, G_t, \dots, G_T\} \tag{4}$$

其中  $G_t$  表示网络在第  $t$  个时隙上的拓扑.

图 8 所示为一个具有 4 个节点(A~D)的动态拓扑网络在 3 个不同时刻上的拓扑结构.

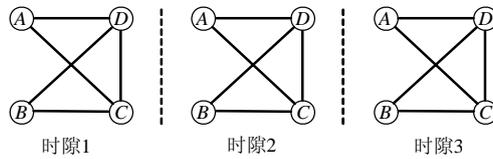


Fig.8 Illustrative example for topologies in different slots of a dynamic network

图 8 动态网络在不同时刻的拓扑结构示例

确定了动态网络在各个不同时刻上的拓扑后,我们提出了动态拓扑网络中优先链路的可识别性理论,解决如何一次性部署尽量少的监测节点以保证对网络所有拓扑中优先链路的性能测量问题.具体地,考虑到监测节点可能部署位置的数目将随着网络规模呈指数增长,我们分析优先链路的可识别性与监测节点部署位置的关系,通过裁剪掉对优先链路性能测量无影响的部分节点,缩小监测节点部署的选择范围,从而加快网络断层扫描的实现过程.

随后,我们解决了动态拓扑网络的监测节点部署问题.注意到现有工作<sup>[20,21]</sup>已实现对静态拓扑网络中优先链路性能测量的监测节点部署算法,存在一种比较直接的动态拓扑网络监测节点部署方式:在网络拓扑序列(如公式(4)所示)中依次应用静态拓扑的监测节点部署算法,获得网络各时隙上的监测节点部署,再将所有时隙上监测节点部署的并集作为动态拓扑网络最终的监测节点部署结果.这种监测节点部署方式虽然可以保证各网络拓扑中优先链路的可识别性,但可能会部署上冗余的监测节点,给服务提供商与网络管理员造成了不必要的开销.以图 9 所示的动态网络为例,在该网络中,节点  $v_2$  和节点  $v_6$  在时隙 1 的拓扑  $G_1$  中连通,在时隙 2 的拓扑  $G_2$  中断开.在网络两个时隙的拓扑中,假设网络管理员只需要测量链路  $v_1v_4$  和链路  $v_1v_5$  的性能.在拓扑  $G_1$  中,为了获得这两条优先链路的性能,基于文献[21]中的部署算法,我们可以从  $\{v_2, v_3, v_6, v_7, v_8\}$  中任选两个节点部署监测节点,所以  $M_1=\{v_6, v_7\}$  为一种可行的部署方式;而在拓扑  $G_2$  中,选  $v_2$  和  $v_3$  作为监测节点,可以测量出优先链路  $v_1v_4$  和  $v_1v_5$  的性能,即  $M_2=\{v_2, v_3\}$ .因此,最终我们在该动态网络中部署的监测节点为  $M_1 \cup M_2 = \{v_2, v_3, v_6, v_7\}$ ,即需要 4 个监测节点.由于  $M_2$  本身也适用于拓扑  $G_1$  中优先链路性能的测量,这种对不同时刻网络拓扑单独处理的方式部署了两个冗余的监测节点,即  $v_6$  和  $v_7$ .

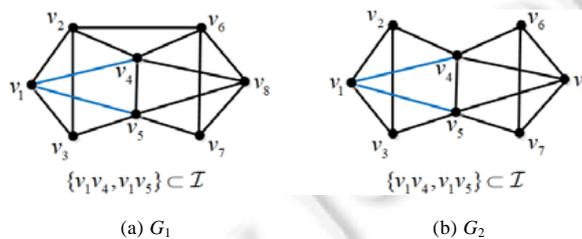


Fig.9 Illustrative example for monitoring node placement in a dynamic network

图 9 动态拓扑网络中监测节点的部署示例

为此,在监测节点的部署过程中,我们综合考虑了动态网络各时隙拓扑中监测节点的部署需求,并用如下约束条件的方式刻画单个拓扑的监测节点部署:

$$\{(S_i, n_i) | i=1, 2, \dots\} \tag{5}$$

其中,  $S_i$  为网络节点集合,表示在单个拓扑中可供选择的监测节点部署位置;  $n_i$  表示从  $S_i$  中至少需要选择的节点

数目.

得到动态网络每个拓扑中监测节点部署的约束条件后,我们将监测节点的部署问题规划成最小碰撞集问题(minimum hitting set problem,简称 min-HSP).利用最优化理论,可以从整个网络中选取最少数目的节点作为监测节点,以满足各约束条件.相应地,我们能够用最少的监测节点来保证动态网络中优先链路的可识别性.

### 3.3 通信失效网络断层扫描理论与技术

#### • 挑战

现有的很多关于网络断层扫描技术的研究工作尝试回答网络拓扑与链路可识别性具有何种关系这一问题<sup>[15,16,18,19]</sup>,想要了解对于给定拓扑,如何部署监测节点测量出网络中所有(或部分)链路的性能指标.但是这些工作假设在测量过程中网络通信都是完全可靠的,未考虑到网络通信失效的情况.然而在实际应用中,通信失效在多种形式的网络中是较为常见的<sup>[71]</sup>,如链路故障、节点变更和信道冲突等.因此,提出通信失效网络断层扫描理论并设计鲁棒的网络断层扫描算法,以保障在网络发生通信失效时测量任务的顺利进行,也是一个亟待解决的问题.

#### • 关键技术

对于特定的监测节点部署,当发生通信失效(如链路失效)时,其能观测到的路径是不同的,包含失效链路的路径将不能被观测到,从而可能使原来可识别的链路变得不可识别.以图 10 中链路  $l_6$  的测量为例,在该网络拓扑中,节点  $v_2$  被部署为监测节点,通过测量 3 条起止于  $v_2$  的路径性能数据( $p_1:l_2l_1l_6$ , $p_2:l_3l_4l_6$ , $p_3:l_2l_1l_4l_3$ ),可以推算出链路  $l_6$  的性能(如时延).用  $x_i$  表示链路  $l_i$  的性能, $c_i$  表示路径  $p_i$  的性能,则 3 条测量路径的性能与链路性能可以用图 10 中的线性方程组关联起来,其中每条测量路径分别对应于方程组中的一个线性方程.通过求解该线性方程组,可以计算出链路  $l_6$  的性能指标值  $c_6=(c_1+c_2-c_3)/2$ ,即链路  $l_6$  在监测节点  $v_2$  下是可识别的.然而,当链路  $l_2$  发生通信失效时,测量路径  $p_1:l_2l_1l_6$  和  $p_3:l_2l_1l_4l_3$  由于包含  $l_2$  将变得不可用,方程组将只包含方程②的信息,此时链路  $l_6$  将变得不可识别.因此,现有的监测节点部署方法无法有效应对链路失效对网络性能测量的影响.值得注意的是,有些链路在网络中任意其他  $k$  条链路失效的情况下仍然是可识别的.例如图 10 中  $l_3$  在网络其他任意 1 条链路失效时,仍可以通过其他测量路径的端到端数据推算出其性能指标值.如果我们能够了解怎样的拓扑特征可以使得链路具有这种性质,将有助于指导监测节点的部署以及网络拓扑的设计,实现鲁棒的网络测量.此外,不同的监测节点部署也可能导致不同的链路可识别性.如果将监测节点部署在  $v_5$ ,那么无论有无链路失效发生,网络中所有的链路都将变得不可识别.

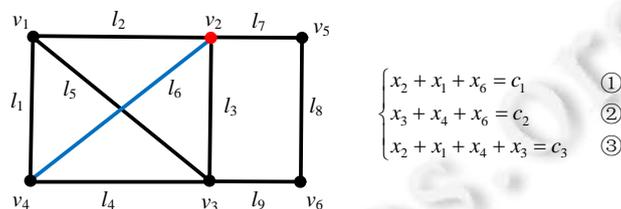


Fig.10 Example to illustrate the impact of link failures on network tomography

图 10 链路失效对网络断层扫描的影响示例

针对这种情况,我们在文献[72,73]中研究了网络失效下的链路可识别性理论,包含一条链路在其他任意  $k(k \geq 0)$  条链路失效情况下仍然可以识别的充分必要条件,并根据链路可识别性理论设计了监测节点的部署算法,解决了如何在有链路失效的情况下部署监测节点,以进行鲁棒的网络断层扫描.首先,为了量化链路失效对链路性能识别性的影响,我们提出了  $k$ -可识别性( $k$ -identifiability)的概念.

- 一条链路(非失效链路)被称为  $k$ -可识别的( $k$ -identifiable),当且仅当该链路在网络中任意  $k$  条链路失效时仍是可识别的;
- 一个网络被称为  $k$ -可识别的,当且仅当网络中所有链路都是  $k$ -可识别的.

需要说明的是:当  $k=0$  时, $k$ -可识别表示当网络无通信失效时链路是可识别的(即已有网络断层扫描方法的可识别性),此时也记为 0-可识别。 $k$ -可识别的链路一定是 $(k-1)$ 可识别的, $k$  越大,表示该链路对于其他链路失效的容忍度越大。可以看出:如果我们能得到链路  $k$ -可识别和网络拓扑结构的关系,将帮助网络管理者有效地部署监测节点,以使得网络中  $k$ -可识别的链路数目尽可能多。同时,这也能帮助网络设计者了解如何设计网络能够使得网络链路为  $k$ -可识别的,实现鲁棒的网络断层扫描<sup>[74]</sup>。

具体地,我们在文献[72,73]中通过回答以下两个密切相关的问题,来帮助网络设计者和网络管理员达到上述目标。

- 对于一个给定监测节点部署的网络,哪些链路是  $k$ -可识别的?
- 对于一个给定监测节点数目的网络,如何部署这些监测节点,以使网络中  $k$ -可识别的链路数目最大?

针对  $k$ -可识别性链路的判断问题,对于一个有  $n$  条链路的网络来说,网络中  $k$  条链路失效的场景有  $C_n^k$  种,通过枚举每一种失效场景去判断一条链路是否是  $k$ -可识别的是不可行的。为此,我们选择通过对网络拓扑进行适当划分来研究链路  $k$ -可识别性在监测节点部署和拓扑结构方面的判定理论。具体地,我们给出并证明了在不考虑链路失效( $k=0$ )时,链路可识别(0-可识别)的充分必要条件(见表 3)。在此基础上,进一步给出并证明了在考虑链路失效( $k>0$ )时,链路可识别( $k$ -可识别)的充分必要条件(见表 3)。基于这些证明的链路  $k$ -可识别性条件,我们可以基于网络拓扑结构和监测节点部署位置,在多项式时间内判断出一条链路是否是  $k$ -可识别的,而不需要枚举所有链路失效情况,显著降低了  $k$ -可识别性链路判定的复杂度。

**Table 3** Necessary and sufficient conditions for  $k$ -identifiability of links<sup>[72,73]</sup>

**表 3** 链路  $k$ -可识别性的充要条件<sup>[72,73]</sup>

可识别性	充要条件
0-可识别性( $k=0$ )	在只部署一个监测节点的网络中,一条链路是 0-可识别的当且仅当这条链路在网络的 $T_m$ , Type 1 或 Type 3 连通分支中
$k$ -可识别性( $k>0$ )	在只部署一个监测节点的网络中,一条链路是 $k$ 可识别的当且仅当这条链路在监测节点所在 $(k+2)$ -边连通分支里的 $(k+3)$ -边连通分支中

对于特定数目( $\kappa$ 个)监测节点的部署问题,一种比较直接的方法是:根据链路  $k$  可识别性的判定条件,在网络中枚举所有监测节点部署位置,并比较每种部署情形下  $k$  可识别链路的数目,最后输出使得  $k$  可识别链路数目最大的监测节点部署方式。考虑到网络中存在  $C_n^k$  种可能的监测节点部署方式,这种枚举方法的复杂度将随着网络规模呈指数级增长,不具有可扩展性。相反,我们在文献[72,73]中形式化地表述了特定数目监测节点的部署问题,并说明了这是一个 NP 完全问题。随后,我们提出了一种贪心算法 MPK 来解决这个问题。MPK 的核心思想是:因为在网络的 $(k+3)$ -边连通分支中部署一个监测节点必然能使其中的链路都是  $k$ -可识别的(见表 3),所以 MPK 首先在(网络的每个 $(k+3)$ -边连通分支中随机地选择一个节点,构成监测节点部署的候选集合。接着,MPK 枚举了第 1 个监测节点部署的所有位置,这是因为第 1 个监测节点不同的部署往往会给算法最后的结果带来较大的影响。当第 1 个监测节点的位置选定后,MPK 将一个地一个地从候选集合中选定剩余的 $\kappa-1$  个监测节点。每次选择中,MPK 都将选择使得当前  $k$ -可识别的链路数目最多的位置部署监测节点。需要说明的是:通过此算法,我们可以逐步增加监测节点的数目,直到网络中所有链路都为  $k$ -可识别时,此时的监测节点部署为开销最小的部署方式。

虽然文献[72,73]中的监测节点部署能够确保链路在网络失效情况下的  $k$ -可识别性,但  $k$ -可识别性概念针对的是网络随机(或任意的)链路失效场景,同时,监测节点部署过程也将网络链路失效刻画为不可预测的。这种悲观式链路失效刻画方式容易生成次优的监测节点部署结果,即部署冗余的监测节点。然而在很多网络场景中,基于网络通信模式特征和现有拓扑预测模型<sup>[75,76]</sup>,服务提供商和网络管理员能够一定程度上预测出未来一段时间内的网络失效情况,即存在一部分可预测的网络失效。为此,我们在文献[77]中进一步提出了一种能够同时应对可预测和不可预测链路失效的网络断层扫描方法,基于对网络链路失效的预测,在保证链路  $k$ -可识别性的同时,进一步减少监测节点部署数量,降低测量的成本。为了满足网络管理员在测量成本与实现复杂性上的不同

需求,我们还提出了多种监测节点部署算法<sup>[77]</sup>(如简单取并集法、增量部署法和综合部署法等),以达到监测节点部署成本和时间复杂度之间的平衡.

#### 4 总结和未来工作

与传统网络内部测量方法不同,网络断层扫描作为一种新型的网络外部测量方法,能够在没有中间节点协作的条件下,根据网络边缘的测量数据推测出网络内部性能指标和运行状态,从而实现与网络组成及协议无关的网络测量.本文首先给出了网络断层扫描的基本概念和形式化表述,并指出了影响网络断层扫描性能的 3 个重要因素:监测节点部署、测量路径构建和测量数据分析.接着,对近年来学术界在这 3 个影响因素方面的代表性研究工作进行了归纳分析.其中,针对网络状态识别问题,本文还介绍了布尔形式网络断层扫描在监测节点部署方面的研究成果.随后,本文分析了已有网络断层扫描工作在实际应用中的核心缺陷,并总结了针对这些核心缺陷近年来提出的最新理论及算法.

目前,网络断层扫描的发展和應用仍然是网络测量领域研究的热点,存在大量开放性問題有待更加深入的研究.以下列举出几个未来可能的发展方向.

- (1) 通用有效的网络断层扫描性能模型.当前,学术界已经提出了很多网络断层扫描算法,如 MMP<sup>[16]</sup>, OMA<sup>[21]</sup>, MPPI<sup>[64]</sup>和 MAPLink<sup>[70]</sup>等.这些网络断层扫描算法依赖于不同的网络模型,具有不同的数据收集和数据分析方式.此外,这些算法在测量开销、测量精度(及粒度)、实现复杂度等方面也有不同的表现.因此,设计一个通用有效的网络断层扫描性能模型,以准确刻画测量可用资源(带宽资源、计算资源、存储资源)与测量算法性能之间的关系,将有助于服务提供商与网络管理员在相同条件下(如给定测量开销)对比不同方法的优劣,从而便于根据当前网络的特性和测量的需求选择最合适的网络断层扫描算法,实现测量开销和测量精度之间的平衡.
- (2) 更鲁棒的测量算法和分析方法.网络断层扫描技术的应用,离不开测量数据的收集和分析.测量数据收集的质量,直接影响测量结果的性能.由于采用间接测量的方式,网络断层扫描测量数据的收集容易受到许多实际因素的影响,如链路拥塞状况、节点存储状况和网络路由等.而其中任意一个因素的变化,将很可能使得传统网络断层扫描方法变得不可用.虽然近几年已有部分工作<sup>[69,70,72,77]</sup>考虑了鲁棒断层扫描问题,但大多针对的是特定因素的影响,如节点位置更新和链路失效等.如何考虑更加全面及实际的网络设置,根据其特征进一步扩展和改进网络断层扫描算法,仍然是一个需要思考的问题.此外,受环境干扰、网络流量等影响,监测节点上收集的测量数据可能带有一些误差.虽然已有统计学习方法可以在一定程度上规避测量数据的噪声,但其往往具有较大的复杂度,难以保证测量的实时性.因此,未来需要在目前研究基础上探索更加鲁棒的网络测量算法和更加通用的数据分析方法.
- (3) 网络测量信息挖掘和智能分析.针对大规模网络的性能测量,是近年来网络测量领域的研究热点之一<sup>[74]</sup>.目前,实现对大规模网络的性能评估,需要长时间持续不断地测量才能获得全面而准确的数据.因此,网络断层扫描还要面对大规模测量可能带来的海量数据和数据时效性的问题<sup>[3]</sup>.另外,通过对网络断层扫描测量信息的跟踪与统计分析,有助于服务提供商与网络管理员迅速、直观地判断网络性能的变化趋势,并及时定位网络瓶颈和故障的位置,在网络性能恶化之前予以解决,保证网络服务的可用性和有效性.因此,实现网络测量信息的智能分析和预测将有助于网络的管理及优化,具有广泛的研究意义和应用价值.

可以预见:一种高效、鲁棒及智能的网络断层扫描技术将可以提升我们了解网络运行机理的能力,支撑网络性能的持续优化.

#### References:

- [1] Zhou AP, Cheng G, Guo XJ. High-speed network traffic measurement method. Ruan Jian Xue Bao/Journal of Software, 2014,25(1): 135-153 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4445.htm> [doi: 10.13328/j.cnki.jos.004445]

- [2] Yang H, Li CL, Sun Z, *et al.* A survey of Internet topology data acquisition methods. *Communications Technology*, 2019,52(8): 1811–1817 (in Chinese with English abstract).
- [3] Hu ZG, Tian CQ, Du L, Guan XQ, Cao F. Current research and future perspective on IP network performance measurement. *Ruan Jian Xue Bao/Journal of Software*, 2017,28(1):105–134 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5127.htm> [doi: 10.13328/j.cnki.jos.005127]
- [4] Zhang R, Jin YH, Yang T, Rong ZZ. Intelligent selection algorithm of measurement nodes in distributed network measurement. *Computer Science*, 2015,42(9):70–78 (in Chinese with English abstract). [doi: 10.11896/j.issn.1002-137X.2015.9.015]
- [5] Vardi Y. Network tomography: Estimating source-destination traffic intensities from link data. *Journal of the American Statistical Association*, 1996,91(433):365–377.
- [6] Zhang HL, Fang BX, Hu MZ, Jiang Y, Zhan CY, Zhang SF. A survey on Internet measurement and analysis. *Ruan Jian Xue Bao/Journal of Software*, 2003,14(1):110–116 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/110.htm>
- [7] Sunshine CA. Source routing in computer networks. In: *Proc. of the ACM SIGCOMM Computer Communication Review*. ACM, 1977. 29–33.
- [8] Xia Y, Tse D. Inference of link delay in communication networks. *IEEE Journal of Selected Areas in Communications*, 2006, 24(12):2235–2248.
- [9] Gurewitz O, Sidi M. Estimating one-way delays from cyclic-path delay measurements. In: *Proc. of the IEEE INFOCOM*. IEEE, 2001. 1–7.
- [10] Downey AB. Using pathchar to estimate internet link characteristics. In: *Proc. of the ACM SIGCOMM*. ACM, 1999. 241–250.
- [11] Kumar R, Kaur J. Practical beacon placement for link monitoring using network tomography. *IEEE Journal on Selected Areas in Communications*, 2006,24(12):2196–2209.
- [12] Bejerano Y, Rastogi R. Robust monitoring of link delays and faults in IP networks. In: *Proc. of the IEEE INFOCOM*. IEEE, 2003. 1–11.
- [13] Horton JD, Ortiz AL. On the number of distributed measurement points for network tomography. In: *Proc. of the ACM IMC*. ACM, 2003. 204–209.
- [14] Mahajan R, Spring NT, Wetherall D, Anderson TE. Inferring link weights using end-to-end measurements. In: *Proc. of the ACM Internet Measurement Workshop*. ACM, 2002. 1–6.
- [15] Gopalan A, Ramasubramanian S. On identifying additive link metrics using linearly independent cycles and paths. *IEEE/ACM Trans. on Networking*, 2011,20(3):906–916.
- [16] Ma L, He T, Leung KK, Swami A, Towsley D. Inferring link metrics from end-to-end path measurements: Identifiability and monitor placement. *IEEE/ACM Trans. on Networking*, 2014,22(4):1351–1368.
- [17] Ma L, He T, Leung KK, Swami A, Towsley D. Identifiability of link metrics based on end-to-end path measurements. In: *Proc. of the ACM IMC*. ACM, 2013. 391–404.
- [18] Ma L, He T, Leung KK, Swami A, Towsley D. Link identifiability in communication networks with two monitors. In: *Proc. of the IEEE GLOBECOM*. IEEE, 2013. 1513–1518.
- [19] Ma L, He T, Leung KK, Swami A, Towsley D. Monitor placement for maximal identifiability in network tomography. In: *Proc. of the IEEE INFOCOM*. IEEE, 2014. 1447–1455.
- [20] Gao Y, Dong W, Wu WB, Chen C, Li XY, Bu JJ. Scalpel: Scalable preferential link tomography based on graph trimming. *IEEE/ACM Trans. on Networking*, 2016,24(3):1392–1403.
- [21] Dong W, Gao Y, Wu WB, Bu JJ, Chen C, Li XY. Optimal monitor assignment for preferential link tomography in communication networks. *IEEE/ACM Trans. on Networking*, 2017,25(1):210–223.
- [22] Zhang Y, Roughan M, Willinger W, Qiu LL. Spatio-temporal compressive sensing and internet traffic matrices. In: *Proc. of the ACM SIGCOMM*. ACM, 2009. 1–12.
- [23] Firooz M, Roy S. Network tomography via compressed sensing. In: *Proc. of the IEEE GLOBECOM*. IEEE, 2010. 1–5.
- [24] Xu WY, Mallada E, Tang A. Compressive sensing over graphs. In: *Proc. of the IEEE INFOCOM*. IEEE, 2011. 1–9.
- [25] Kompella RR, Yates J, Greenberg AG, Snoeren AC. Detection and localization of network black holes. In: *Proc. of the IEEE INFOCOM*. IEEE, 2007. 2180–2188.

- [26] Zeng HY, Kazemian P, Varghese G, McKeown N. Automatic test packet generation. In: Proc. of the ACM CoNEXT. ACM, 2012. 241–252.
- [27] Dhamdhere A, Teixeira R, Dovrolis C, Diot C. NetDiagnoser: Troubleshooting network unreachabilities using end-to-end probes and routing data. In: Proc. of the ACM CoNEXT. ACM, 2007. 1–12.
- [28] Huang YY, Feamster N, Teixeira R. Practical issues with using network tomography for fault diagnosis. ACM SIGCOMM Computer Communication Review, 2008,38(5):53–57.
- [29] Nguyen HX, Thiran P. The Boolean solution to the congested IP link location problem: Theory and practice. In: Proc. of the IEEE INFOCOM. IEEE, 2007. 1–9.
- [30] Ghita D, Karakus C, Argyraki K, Thiran P. Shifting network tomography toward a practical goal. In: Proc. of the ACM CoNEXT. ACM, 2011. 1–12.
- [31] Duffield N. Simple network performance tomography. In: Proc. of the ACM IMC. ACM, 2003. 1–6.
- [32] Ahuja S, Ramasubramanian S, Krunz M. SRLG failure localization in all-optical networks using monitoring cycles and paths. In: Proc. of the IEEE INFOCOM. IEEE, 2008. 700–708.
- [33] Ma L, He T, Swami A, Towsley D, Leung KK, Lowe J. Node failure localization via network tomography. In: Proc. the ACM IMC. ACM, 2014. 195–208.
- [34] Ma L, He T, Swami A, Towsley D, Leung KK. On optimal monitor placement for localizing node failures via network tomography. Performance Evaluation, 2015,91:16–37.
- [35] Ma L, He T, Swami A, Towsley D, Leung KK. Network capability in localizing node failures via end-to-end path measurements. IEEE/ACM Trans. on Networking, 2017,25(1):434–450.
- [36] He T, Bartolini N, Khamfroush H, Kim I, Ma L, Porta TL. Service placement for detecting and localizing failures using end-to-end observations. In: Proc. of the IEEE ICDCS. IEEE, 2016. 560–569.
- [37] Bartolini N, He T, Khamfroush H. Fundamental limits of failure identifiability by Boolean network tomography. In: Proc. of the IEEE INFOCOM. IEEE, 2017. 1–9.
- [38] Gopalan A, Ramasubramanian S. On the maximum number of linearly independent cycles and paths in a network. IEEE/ACM Trans. on Networking, 2014,22(5):1373–1388.
- [39] Wing O, Kim W. The path matrix and its realizability. IRE Trans. on Circuit Theory, 1959,6(3):267–272.
- [40] Ma L, He T, Leung K, Towsley D, Swami A. Efficient identification of additive link metrics via network tomography. In: Proc. of the IEEE ICDCS. IEEE, 2013. 1–10.
- [41] Li F, Thottan M. End-to-end service quality measurement using source-routed probes. In: Proc. of the IEEE INFOCOM. IEEE, 2006. 1–12.
- [42] Nguyen HX, Thiran P. Active measurement for multiple link failures diagnosis in IP networks. In: Proc. of Passive and Active Measurement. 2004. 1–10.
- [43] Agrawal S, Naidu KVM, Rastogi R. Diagnosing link-level anomalies using passive probes. In: Proc. of the IEEE INFOCOM. IEEE, 2007. 1757–1765.
- [44] Barford P, Duffield N, Ron A, Sommers J. Network performance anomaly detection and localization. In: Proc. of the IEEE INFOCOM. IEEE, 2009. 1377–1385.
- [45] Chen Y, Bindel D, Song HH, Katz RH. Algebra-based scalable overlay network monitoring: Algorithms, evaluation, and applications. IEEE/ACM Trans. on Networking, 2007,15(5):1084–1097.
- [46] Zhao Y, Chen Y, Bindel D. Towards unbiased end-to-end network diagnosis. In: Proc. of the ACM SIGCOMM. ACM, 2006. 219–230.
- [47] Shavitt Y, Sun XD, Wool A, Yener B. Computing the unmeasured: An algebraic approach to internet mapping. IEEE Journal on Selected Areas in Communications, 2004,22(1):67–78.
- [48] Tang CP, McKinley PK. On the cost-quality tradeoff in topology-aware overlay path probing. In: Proc. of the IEEE ICNP. IEEE, 2003. 1–12.
- [49] Zheng Q, Cao GH. Minimizing probing cost and achieving identifiability in probe-based network link monitoring. IEEE Trans. on Computers, 2011,62(3):510–523.

- [50] Li GS, Cai WD. Research and development of network tomography. *Measurement and Control Technology*, 2008,27(2):1–4 (in Chinese with English abstract). [doi: 10.3969/j.issn.1000-8829.2008.02.001]
- [51] Coceres R, Duffield N, Horowitz J, Towsley D, Bu T. Multicast-based inference of network-internal characteristics: Accuracy of packet loss estimation. In: *Proc. of the IEEE INFOCOM*. IEEE, 1999. 371–379.
- [52] Coates M, Castro R, Nowak R. Maximum likelihood network topology identification from edge-based unicast measurements. *Performance Evaluation Review*, 2002,30(1):11–20.
- [53] Shih MF, Hero AO. Unicast-based inference of network link delay distributions with finite mixture models. *IEEE Trans. on Signal Processing*, 2003,51(8):2219–2228.
- [54] Zhang JZ. Origin-destination network tomography with Bayesian inversion approach. In: *Proc. of the IEEE/ACM WI*. IEEE, 2006. 1–4.
- [55] Lawrence E, Michailidis G, Nair VN, Xi BW. Network tomography: A review and recent developments. *Frontiers in Statistics*, 2006,54:345–366.
- [56] Bu T, Duffield N, Presti FL, Towsley D. Network tomography on general topologies. In: *Proc. of the ACM SIGMETRICS*. ACM, 2002. 1–10.
- [57] Chen AY, Cao J, Bu T. Network tomography: Identifiability and Fourier domain estimation. In: *Proc. of the IEEE INFOCOM*. IEEE, 2007. 1875–1883.
- [58] He T, Liu C, Swami A, Towsley D, Salonidis T, Bejan AI, Yu P. Fisher information-based experiment design for network tomography. In: *Proc. of the ACM SIGMETRICS*. ACM, 2015. 389–402.
- [59] Trienekens JJM, Bouman JJ, Zwan MVD. Specification of service level agreements: Problems, principles and practices. *Software Quality Journal*, 2004,12:43–57.
- [60] Feng CY, Wang LN, Wu K, Wang JP. Bound-based network tomography with additive metrics. In: *Proc. of the IEEE INFOCOM*. IEEE, 2019. 316–324.
- [61] Wu J, Wang Y, Mao ZM, Shin KG. Internet routing resilience to failures: Analysis and implications. In: *Proc. of the CoNEXT*. ACM, 2007. 1–12.
- [62] Liu YH, He Y, Li M, Wang JL, Liu KB, Li XY. Does wireless sensor network scale? A measurement study on greenorbs. *IEEE Trans. on Parallel and Distributed Systems*, 2013,24(10):1983–1993.
- [63] Li HK, Gao Y, Dong W, Chen C. Bound-based network tomography for inferring interesting link metrics. In: *Proc. of the IEEE INFOCOM*. IEEE, 2020. 1–10.
- [64] Yang RW, Feng CY, Wang LN, Wu WW, Wu K, Wang JP, Xu YL. On the optimal monitor placement for inferring additive metrics of interested paths. In: *Proc. of the IEEE INFOCOM*. IEEE, 2018. 2141–2149.
- [65] Zhang H, Cai ZP, Li Y. An overview of software-defined network measurement technologies. *Scientia Sinica Informationis*, 2018, 48(3):293–314 (in Chinese with English abstract). [doi: 10.1360/N112017-00203]
- [66] Vasseur JP, Agarwal N, Hui J, Shelby Z, Bertrand P, Chauvenet C. RPL: The IP routing protocol designed for low power and lossy networks. *Internet Protocol for Smart Objects (IPSO) Alliance*, 2011. 1–20. <http://www.cse.chalmers.se/edu/year/2019/course/DAT300/PAPERS/rpl.pdf>
- [67] Kang NX, Ghobadi M, Reumann J, Shraer A, Rexford J. Efficient traffic splitting on commodity switches. In: *Proc. of the ACM CoNEXT*. ACM, 2015. 1–13.
- [68] Yang BH, Liu JD, Shenker S, Li J, Zheng K. Keep forwarding: Towards  $k$ -link failure resilient routing. In: *Proc. of the IEEE INFOCOM*. IEEE, 2014. 1617–1625.
- [69] He T, Gkeliass A, Ma L, Leung KK, Swami A, Don T. Robust and efficient monitor placement for network tomography in dynamic networks. *IEEE/ACM Trans. on Networking*, 2017,25(3):1732–1745.
- [70] Li HK, Gao Y, Dong W, Chen C. Preferential link tomography in dynamic networks. *IEEE/ACM Trans. on Networking*, 2019,27(5): 1801–1814.
- [71] Turner D, Levchenko K, Snoeren AC, Savage S. California fault lines: Understanding the causes and impact of network failures. In: *Proc. of the ACM SIGCOMM*. ACM, 2010. 1–12.

- [72] Ren W, Dong W. Robust network tomography:  $k$ -identifiability and monitor assignment. In: Proc. of the IEEE INFOCOM. IEEE, 2014. 1–9.
- [73] Ren W. Study and design of a monitor assignment algorithm based on network tomography [MS. Thesis]. Hangzhou: Zhejiang University, 2017 (in Chinese with English abstract).
- [74] Han HS, Qiu LL, Zhang Y. NetQuest: A flexible framework for large-scale network measurement. IEEE/ACM Trans. on Networking, 2009,17(1):106–119.
- [75] Zhao M, Wang WY. Analyzing topology dynamics in ad hoc networks using a smooth mobility model. In: Proc. of the IEEE WCNC. IEEE, 2007. 3279–3284.
- [76] Gu Y, He T. Dynamic switching-based data forwarding for low duty-cycle wireless sensor networks. IEEE Trans. on Mobile Computing, 2011,10(12):1741–1754.
- [77] Li HK, Gao Y, Dong W, Chen C. Taming both predictable and unpredictable link failures for network tomography. IEEE/ACM Trans. on Networking, 2018,26(3):1460–1473.

#### 附中文参考文献:

- [1] 周爱平,程光,郭晓军.高速网络流量测量方法.软件学报,2014,25(1):135–153. <http://www.jos.org.cn/1000-9825/4445.htm> [doi: 10.13328/j.cnki.jos.004445]
- [2] 杨慧,李春林,孙治,等.互联网拓扑数据获取方法研究综述.通信技术,2019,52(8):1811–1817.
- [3] 胡治国,田春岐,杜亮,关晓蕾,曹峰.IP网络性能测量研究现状和进展.软件学报,2017,28(1):105–134. <http://www.jos.org.cn/1000-9825/5127.htm> [doi: 10.13328/j.cnki.jos.005127]
- [4] 张荣,金跃辉,杨谈,荣自瞻.分布式网络测量中测量节点的智能选择算法.计算机科学,2015,42(9):70–78. [doi: 10.11896/j.issn.1002-137X.2015.9.015]
- [6] 张宏莉,方滨兴,胡铭曾,姜誉,詹春艳,张树峰.Internet 测量与分析综述.软件学报,2003,14(1):110–116. <http://www.jos.org.cn/1000-9825/14/110.htm>
- [50] 李贵山,蔡皖东.网络断层扫描技术的研究与发展.测控技术,2008,27(2):1–4. [doi: 10.3969/j.issn.1000-8829.2008.02.001]
- [65] 张恒,蔡志平,李阳.SDN网络测量技术综述.中国科学:信息科学,2018,48(3):293–314. [doi: 10.1360/N112017-00203]
- [73] 任伟.基于网络断层扫描的监测节点部署算法的研究与设计[硕士学位论文].杭州:浙江大学,2017.



李惠康(1993—),男,博士,主要研究领域为网络测量,物联网系统与网络,无线与移动计算.



高艺(1987—),男,博士,副教授,博士生导师,CCF 专业会员,主要研究领域为物联网与传感网,边缘计算,无线与移动计算,软件定义网络.



董玮(1982—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为物联网系统与网络,无线与移动计算,智能感知技术,边缘计算.



陈纯(1955—),男,博士,教授,博士生导师,CCF 会士,主要研究领域为计算机软件与理论,大数据实时智能处理技术.