

测量调度的建模和分析

蒋序平, 陈 鸣

(解放军理工大学指挥自动化学院计算机教研室, 南京 210007)



摘 要: 通过研究一个大型网络测量系统中测量任务的调度问题, 建立了形式化的测量调度模型, 规范地定义了测量任务分布、资源耗费和资源分配等概念。基于该模型, 分析了大规模测量调度的分解、测量调度的拓扑约束和资源约束、以及测量任务排队等问题, 为网络测量系统的用户或管理员较好地规划组件配置、资源分配和测量任务分布奠定了基础。

关键词: 网络测量; 测量调度; 建模

中图分类号: TP391

文献标识码: A

Modeling and Analysis of Measurement Scheduling

JIANG Xu-Ping, CHEN Ming

(Institute of Command Automation, PLA University of Science and Technology, Nanjing 210007, China)

Abstract: The scheduling of measurements in a large-scale network measurement system is analysed in this paper. A formal measurement scheduling model and its key concepts (the deployment of measurements, resource cost, resource allocation, etc.) are defined. Based on this model, the large-scale measurement scheduling can be divided into some small independent ones. Furthermore, some related problems such as the topology and resource constraint of scheduling and the queuing of measurements are also discussed. This model facilitates the configuration of components, resource allocation and the deployment of measurements.

Keywords: network measurement; measurement scheduling; modeling

1 引言

测量调度是网络测量系统研究中的一个关键问题。

“宽带网监视和测量系统”^[1,2]是一个对网络运行状况进行监视和测量的软件平台, 它由多个监测中心和分布在各个测量点的探针组成, 如图 1 所示。为获得网络运行状况, 由监测中心发起调用, 在探针上执行的所有相关功能被称为测量任务。测量调度要解决的问题是哪些测量中心调用哪些测量任务在哪些探针上执行。

大型网络测量系统包括了数量庞大的监测中心和探针, 形成了一个分布式任务调用/执行环境。在这种情况下, 测量任务的调度问题是一个分布式调度问题^[3,4]。

通常, 测量任务的执行有一定的实时性要求, 即其所在的探针(有的测量任务要运行在一组探针上)须保证其执行性能。例如, 实时流测量要求探针能提供实时(及时)处理网络分组的 CPU 和存储资源。因此, 能同时在某个(组)探针上执行的测量任务是受限的。

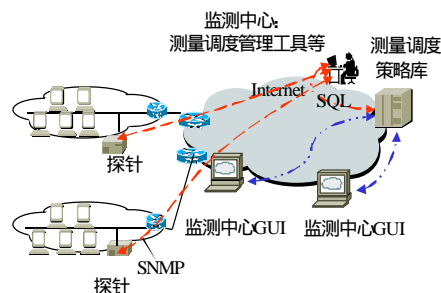


图 1 大型网络测量系统中的测量调度

测量任务能否执行还受到探针拓扑的影响。例如, 对某一路径单向时延的测量, 需要在该路径的两端都有探针, 而测量 RTT 仅需一端有探针即可(两个方向上的选路不一定相同, 因此 $RTT/2 \neq$ 单向时延)。

一个测量任务可能在 A(组)探针或 B(组)探针上执行的结果是相同的, 即针对一个测量目标, 测量调度可能会有多种选择。然而, 任何系统的资源都是有限的, 人们总是希望用最小的代价做尽可能最多的事情。研究测量调度问题也是希望通过对测量任务合理地调度, 高效地使用网络测量系统资源。

综上所述, 测量调度问题是基于测量需求(测量目标), 并根据系统当前资源和拓扑状况, 合理地规划组件配置、资源分配和测量任务调用/执行, 以达到高效使用系统资源的目的。

收稿日期: 2002-03-18

修回日期: 2002-04-15

基金项目: 国家 863 高技术研究发展计划基金 (2001AA112090)

作者简介: 蒋序平(1975-), 男, 湖北武汉人, 博士生, 研究方向为网络管理和分布式计算; 陈 鸣(1956-), 男, 教授, 博导, 研究方向是计算机网络和分布式计算。

许多受限的调度问题属于 NP 完全问题^[3, 5]。 NP 完全问题在较大规模的情况下, 找不到一个现实的求解。一个解决的方法是将大问题分解为多个独立的较小规模的问题^[5]。测量调度作为一个受限的分布式调度问题, 也是一个 NP 完全问题。本文通过建立一个形式化的测量调度模型, 分析了将大规模的测量调度分解为多个较小调度问题的原理。此外, 还对测量调度的拓扑约束和资源约束, 以及测量任务排队等问题进行了研究。该模型为网络测量系统的用户或管理员较好地规划系统组件的配置、资源分配和测量任务分布奠定了基础。

本文第二节建立了测量调度模型。第三节分析了大规模调度问题分解、拓扑约束和资源约束, 以及测量任务排队; 第四节给出了不考虑优化资源利用率的情况下, 进行测量调度的步骤。第五节是总结和进一步的工作。

2 测量调度模型

定义 1: 资源耗费函数 $C(t, p)$ 表示测量任务 t 在探针 p 上执行所耗费的资源量。资源量可以是性能瓶颈资源(例如, CPU 或内存)的数量, 或是一个资源矢量(每个分量表示一类资源的数量)。

定义 2: 探针资源函数 $RES(p)$ 表示探针拥有的资源量。

定义 3: 测量任务分布图 $L<V_p+V_t, E_t>$ 为一个无环有向图。其中结点集 V_p 是系统中探针的集合; 结点集 V_t 是测量任务集合; 边集 E_t 是测量任务在探针上执行关系的集合, 从 V_t 中结点 t 指向 V_p 中结点 p 表示测量任务 t 在探针 p 上执行; V_p 中结点 p 的权值是 $RES(p)$; E_t 中的边 $\langle t, p \rangle$ 的权值是 $C(t, p)$; 实际系统中, 在任何一个探针上执行的所有测量任务耗费的资源量不会超过该探针拥有的资源量, 因此测量任务分布图必须满足表达式:

$$\forall p \in V_p \quad \sum_{\langle t, p \rangle \in E_t} C(t, p) \leq RES(p)$$

定义 4: 给定测量任务分布图 $L<V_p+V_t, E_t>$, 测量任务集 $V_t \subseteq V_t$, 称 $L<V_p+V_t, E_t>$ 为 $L<V_p+V_t, E_t>$ 关于 V_t 的分布子图, 当且仅当:

$$V_p' = \{p | \langle t, p \rangle \in E_t, t \in V_t'\}$$

$$V_t' = V_t'$$

$$E_t' = \{\langle t, p \rangle | \langle t, p \rangle \in E_t, t \in V_t'\}$$

定义 5: 给定测量任务分布图 $L<V_p+V_t, E_t>$, 测量任务 $t \in V_t$, $L<V_p+\{t\}, E_t>$ 是 $L<V_p+V_t, E_t>$ 关于 $\{t\}$ 的分布子图, 则称 $L<V_p+\{t\}, E_t>$ 为测量任务 t 的执行拓扑图。

定义 6: \Leftrightarrow 表示测量任务的执行相关关系。 $t_m \Leftrightarrow t_n$, 当且仅当, 测量任务分布图 $L<V_p+V_t, E_t>$, $\{t_m, t_n\} \subseteq V_t$ 且 t_m 和 t_n 是弱连通^[6]的。关系 \Leftrightarrow 是自反、对称、和传递的, 因此是一个等价关系。

定义 7: 任务相关图 $T<V_p, E>$ 是一个无向图。其中结

点集 V_t 表示测量任务集合; 边集 E 是测量任务间 \Leftrightarrow 关系的集合, V_t 中结点 t_m 连接 t_n , 记为 (t_m, t_n) , 表示 t_m 与 t_n 有执行相关关系。

定理 1: 给定测量任务分布图 $L<V_p+V_t, E_t>$, V_t 上关系 \Leftrightarrow 对应的划分^[6] $A = \{V_1, V_2, V_3, \dots, V_n\}$, $LD = \{L_i < V_{pi} + V_{ti}, E_{ti} > | L_i \text{ 是 } L<V_p+V_t, E_t> \text{ 关于 } V_i \text{ 的分布子图}\}$, 则所有分布子图 $L_i < V_{pi} + V_{ti}, E_{ti} > \in LD$ 都满足表达式:

$$\forall p \in V_{pi} \quad \sum_{\langle t, p \rangle \in E_{ti}} C(t, p) \leq RES(p)$$

即 LD 中每个分布子图都是一个测量任务分布图, 如图 2(d)所示。证明略。

定义 8: 资源分配函数 $ALLOC(m, p)$ 表示监测中心 m 被授权使用探针 p 上资源的数量(资源量)。

定义 9: 资源分配拓扑图 $R<V_m+V_p, E_m>$ 为一个无环有向图。其中结点集 V_m 是系统中监测中心的集合; 结点集 V_p 是探针的集合; 边集 E_m 是监测中心对探针资源分配关系的集合, 从 V_m 中结点 m 指向 V_p 中结点 p 表示监测中心 m 可使用探针 p 的一定资源量; E_m 中边 $\langle m, p \rangle$ 的权值是 $ALLOC(m, p)$; V_p 中结点 p 的权值是 $RES(p)$ 。在实际系统中, 探针被分配的资源量要小于其拥有的资源量, 因此资源分配拓扑图必须满足表达式:

$$\forall p \in V_p \quad \sum_{\langle m, p \rangle \in E_m} ALLOC(m, p) \leq RES(p)$$

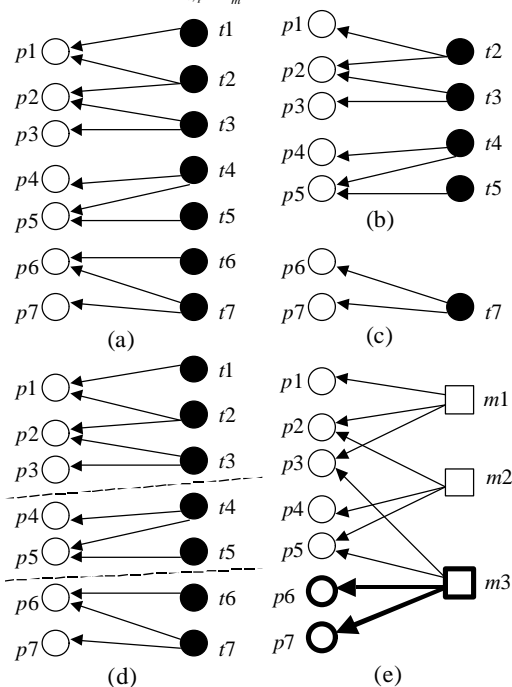


图 2 测量调度模型示例(权值略)

定义 10: 测量调度可形式化表示为一个八元组 $(M, P, T, C, RES, ALLOC, L, R)$ 。 M 是监测中心集合; P 是探针集合; T 是测量任务集合; C 是一个资源耗费函数; RES 是一个探针资源函数; $ALLOC$ 是一个资源分配函数; L 是测量任

务分布图 $L \langle P+T, E \rangle$; R 是资源分配拓扑图 $R \langle M+P, E_m \rangle$ 。

如图 2 所示: (a)是一个测量任务分布图, (b)是(a)关于 $\{t_2, t_3, t_4, t_5\}$ 的分布子图, (c)是测量任务 t_7 的执行拓扑图, (d)是(a)关于 \Leftrightarrow 对应的划分中各测量任务子集的分布子图(虚线隔开), (e)是一个资源分配拓扑图。

3 分析

3.1 大规模测量调度的分解

测量调度问题是一个 NP 完全问题。 NP 完全问题在较小的规模时, 才可能找到一个现实求解^[4]。测量任务分布图(定义 3)本质上表示了网络测量系统中测量任务的分布情况。在此基础上, 利用 \Leftrightarrow 关系在测量任务集合上对应的划分, 可将大的测量调度分解成多个独立的较小规模的调度问题(定理 1)。 \Leftrightarrow 关系的现实意义是: 凡具有 \Leftrightarrow 关系的测量任务在资源使用上存在一定的(直接或间接)相关性, 它们在执行时可能会相互影响。定理 1 中测量任务集 V_i 上 \Leftrightarrow 对应的划分 A 表示了 n 个在执行时完全独立的测量任务集合, 也就是说, 调度不同集合的测量任务不会相互影响。大的测量任务分布图被分解成 LD 中的 n 个分布子图, LD 表示了 n 个独立的较小规模的测量调度。

3.2 拓扑约束

测量任务能否执行受到探针拓扑的影响。测量任务需要某种特定的探针拓扑才能执行, 可用该测量任务的执行拓扑图(定义 5)来表示。资源分配拓扑图(定义 9)反映了系统组件的拓扑。由此, 可以得到以下结论: 要能调度执行某个测量任务, 资源分配拓扑图必有一个子图与该测量任务的执行拓扑图同构^[6]。如图 2(e)所示资源分配拓扑图的一个子图(加粗部分)和图 2(c)所示任务 t_7 的执行拓扑图同构, 这表明如果没有资源使用上的限制, m_3 能够调用 t_7 在探针 p_6 和 p_7 上执行。

3.3 资源约束

测量调度除了拓扑约束以外, 还有资源使用上的限制: 由测量中心调用, 在任何一个探针上执行的所有测量任务所消耗的资源量之和不能超过探针分配给该监测中心的资源量。例如, 图 2 中 m_3 是唯一能够调用 t_6 和 t_7 的监测中心(拓扑约束), 因此要同时调度执行 t_6 和 t_7 , 必须满足如下表达式:

$$ALLOC(m_3, p_6) \geq C(t_7, p_6) + C(t_6, p_6) \wedge$$

$$ALLOC(m_3, p_7) \geq C(t_7, p_7)$$

这一条件是测量调度的资源约束(资源使用上的限制)。当然, 如果资源分配拓扑图有另一个子图与 t_7 或 t_6 的执行拓扑图同构, 那么即使分配给 m_3 的资源不够用了, t_7 或 t_6 仍可能由其他的监测中心发起调用。

3.4 测量任务排队

在实际系统中, 满足拓扑约束但不满足资源约束的测

量任务可能会竞争分配给监测中心的资源。在这种情况下, 对资源使用的竞争与保证测量任务的执行性能没有矛盾。后者是指测量任务一旦被调用必须在执行结束之前拥有保证其性能的资源。资源使用竞争的解决方法是设置测量任务的优先级(例如, 按测量任务类型设置), 并基于优先级在监测中心处对测量任务进行排队。

4 不考虑优化的测量调度

基于测量调度模型, 网络测量系统的用户或管理员如果不考虑对资源利用率的优化, 可以通过以下步骤进行测量调度的规划:

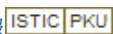
- (1) 根据系统组件的情况, 确定探针集合 P 和监测中心集合 M ;
- (2) 根据测量需求, 确定测量任务集合 T ;
- (3) 确定探针资源函数 $RES(p)$ 和资源耗费函数 $C(t, p)$;
- (4) 不考虑拓扑和资源约束, 确定(另)一个可能的测量任务分布图 L ;
- (5) 基于 L , 依据拓扑约束(3.2 节)的条件, 确定一个可能的资源分配拓扑图 R 和资源分配函数 $ALLOC(m, p)$, 如果做不到则(4);
- (6) 判断 L 是否满足资源约束(3.3 节)的条件, 如果满足, 这就是一个可行的测量调度方案, 否则(4)。

5 结论

本文建立了一个形式化的测量调度模型, 为测量调度问题的分析和解决奠定了基础。我们最终的目标是实现一种可编程的测量调度设施^[2], 帮助用户根据其需求制订高效合理的测量调度方案(组件配置、资源分配、测量任务分布等等)。下一步的研究工作是: 测量调度优化算法, 以及测量任务动态调度的概念、原理和实现技术。其关键技术包括: ①拓扑和资源约束判定算法; ②测量调度的优化评估; ③测量任务的优先级管理; ④合理的资源管理机制和相关的组件授权机制; ⑤基于策略^[7]的可编程调度设施, 作为动态测量调度的基础。

参考文献:

- [1] 陈 鸣. 网络测量方法和系统[Z]. 北京: 国家高技术研究发展计划(863 计划)课题申请书, 2001.
- [2] 沈振宇. 网络监测系统中监测中心的设计与实现[D]. 南京: 解放军理工大学硕士论文, 2002.
- [3] Cheriton D. Distributed Systems: Course Notes [M]. Stanford: Computer Science Department, Stanford University, 1999, 191-209.
- [4] 陈国良. 并行计算—结构、算法、编程[M]. 北京: 高等教育出版社, 1999, 10: 177-178.
- [5] Sipser M. 计算理论导引[M]. 北京: 机械工业出版社, 2000.
- [6] 王元元, 李尚奋. 离散数学[M]. 北京: 科学出版社, 1994.
- [7] Stone G, Lundy B, Xie G. Network Policy Languages: A Survey and a New Approach[J]. IEEE Network, 2001, 15(1): 10-21.

作者: 蒋序平, 陈鸣
作者单位: 解放军理工大学指挥自动化学院计算机教研室, 南京, 210007
刊名: 系统仿真学报 
英文刊名: JOURNAL OF SYSTEM SIMULATION
年, 卷(期): 2002, 14(11)
被引用次数: 1次

参考文献(7条)

1. 陈鸣 网络测量方法和系统 2001
2. 沈振宇 网络监测系统中监测中心的设计与实现 2002
3. Cheriton D Distributed Systems: Course Notes 1999
4. 陈国良 并行计算-结构、算法、编程 1999
5. Sipser M 计算理论导引 2000
6. 王元元;李尚奇 离散数学 1994
7. Stone G;Lundy B;Xie G Network Policy Languages: A Survey and a New Approach 2001(01)

本文读者也读过(10条)

1. 邢长友, 李万林, 陈鸣, XING Chang-you, LI Wan-lin, CHEN Ming 太比特路由器多元超立方体交换结构时延仿真[期刊论文]-解放军理工大学学报(自然科学版) 2005, 6(3)
2. 张国敏, 陈鸣, 王亮, 周晓, ZHANG Guo-min, CHEN Ming, WANG Liang, ZHOU Xiao 支持大规模分布式网络测量的组件自配置机制[期刊论文]-吉林大学学报(信息科学版) 2007, 25(1)
3. 郝继红, 陈鸣, 赵洪华, 张睿 NWS预测子系统的应用[期刊论文]-解放军理工大学学报(自然科学版) 2004, 5(6)
4. 李拥新, 陈鸣, 蒋序平, 宋丽华 一种用于网络管理的基于逻辑的Policy定义语言[期刊论文]-计算机研究与发展 2002, 39(11)
5. 仇小锋, 陈鸣, 蒋序平 基于策略系统的SYN Flooding攻击防御机制[期刊论文]-电信科学 2004, 20(1)
6. 赵金, 陈鸣 IP网络链路带宽的测量[期刊论文]-电信科学 2002, 18(5)
7. 何中枢, 陈鸣 网络即插即用技术及其一种基于目录服务的方案[期刊论文]-解放军理工大学学报(自然科学版) 2001, 2(3)
8. 张睿, 陈鸣 基于IP的主动测量协议IPMP[期刊论文]-解放军理工大学学报(自然科学版) 2003, 4(2)
9. 宋丽华, 陈鸣, 仇小锋 网络流量特征对排队性能影响的仿真分析与比较[期刊论文]-系统仿真学报 2005, 17(1)
10. 宋丽华, 王海涛, 陈鸣, SONG Li-hua, WANG Hai-tao, CHEN Ming 面向高速对等网应用的拥塞控制机制[期刊论文]-上海交通大学学报 2006, 40(3)

引证文献(1条)

1. 蒋序平, 陈鸣 BMNM: 一种构建网络测量系统体系结构的新方法[期刊论文]-电信科学 2003(10)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_xtfzxb200211022.aspx