

基于贝叶斯网络的通信网告警相关性和故障诊断模型

邓 歆 孟洛明

(北京邮电大学网络与交换国家重点实验室 北京 100876)

摘 要: 该文采用贝叶斯网络建立告警相关性和故障诊断模型。首先介绍了基于贝叶斯网络推理的基本概念。提出了通信网功能分层结构的思想,建立不同网络层次间的故障传播模型。详细讨论了从故障传播模型中构造贝叶斯网络,以及分布式告警相关性模型的实现框架。最后结合 SDH over DWDM 系统,具体分析了基于贝叶斯网络的告警相关性分析过程及实验结果。实验证明利用贝叶斯网络能够准确定位通信网根故障点。

关键词: 故障管理;告警相关性;故障传播模型;贝叶斯网络

中图分类号: TN915.07

文献标识码: A

文章编号: 1009-5896(2007)05-1182-05

Bayesian Networks Based Alarm Correlation and Fault Diagnosis in Communication Networks

Deng Xin Meng Luo-ming

(State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: This paper proposes the alarm correlation and fault identification based on Bayesian networks in communication networks. At first, the basic concepts of Bayesian networks are introduced. Then the paper presents an approach for modeling large communication networks that are divided into their constituting sub-networks. And the causal relation is used to model the functional relationship among the sub-networks. The paper discusses how to construct Bayesian networks from the causal relation and presents a distributed alarm correlation framework based on CORBA. Finally, the realization and results of alarm correlation and fault identification is discussed in SDH over DWDM systems. The experimentation has proved that using Bayesian network based alarm correlation is benefit to detect and localize the root faults in communication networks.

Key words: Fault management; Alarm correlation; Fault propagation model; Bayesian networks

1 引言

通信网的特点是规模大、结构复杂、异构性、高度分布。管理如此复杂的网络变得越来越困难,尤其在故障管理方面,迫切需要新的方法分析网络中的告警相关性,辅助网络管理员查明产生故障的原因。

在网络管理领域,故障被定义为产生功能异常的原因。故障是产生告警的原因。告警是由在特定事件发生时被管对象发出的事件通知,用于传递告警信息。告警相关性分析是指对告警进行合并和转化,将多个告警合并成一条具有更多信息量的告警,确定能反应故障根本原因的告警,准确诊断故障^[1]。告警相关性可以用于网络故障定位和告警过滤。

故障诊断是故障管理中的一个重要部分。尽管大型通信网出现故障是无法避免的,但是快速检测和识别网络故障的原因能够提高系统的稳健性和可靠性,提高通信服务质量。通信网故障诊断过程大致可以分为 3 个阶段。(1)第 1 阶段是故障检测(fault detection)。故障检测是指在线指示网络中出现故障的设备。通常,通信网设备失效时,以告警的形式提供故障指示信息。因此,故障检测阶段完成告警信息的采集

和告警信息的压缩。(2)第 2 阶段是故障定位(fault localization)。故障定位是指分析网络设备上报的告警信息的相关性分析,定位出可能的故障原因。当网络中出现故障时,会引发一系列告警,但并不是所有告警都表明故障原因,所以需要网络中发生的告警事件进行相关性分析,确定产生故障的根本原因。(3)第 3 阶段是故障识别(fault identification)。故障识别是指从一些可能故障中识别出真正的故障,并隔离故障。在这个阶段常常采用测试方法。

本文重点讨论故障识别方法。大多数故障识别方法都采用的确定性的故障模型^[2]。在确定性的故障模型中,当链路 a 失效引起链路 b 失效,那么认为链路 a 与 b 间存在依赖关系。通信系统实体间不确定性的依赖关系用概率来表示。在不确定性模型中,告警相关性分析就是对所观察到的告警寻找最有可能的解释。有一种解决不确定性的方法是贝叶斯网络(Bayesian networks)^[3, 4]用于故障诊断。通常假设:(1)给定一个故障 a ,由故障 a 引起的故障 b 和 c 的情况是相互独立的;(2)假设故障 a 和 b 出现都引起故障 c ,那么故障 a 或故障 b 引起故障 c 的情况是相互独立的;(3)故障间彼此独立。通过在因果关系图中引入多失效模式(multiple failure modes),解决

性能问题间的因果关系。在多失效模式下采用概率矩阵来表示对象间的故障传播模型。

2 基于贝叶斯网络的概率推理

2.1 贝叶斯网络简介

贝叶斯网络是一个有向图 $B = (V, L, P)$, 其中每个节点都标注了定量概率信息, 每个节点在给定父节点下都有一个条件概率分布。详细描述如下^[5]:

(1) 随机变量集组成网络节点 V 。变量可以是离散的或者是连续的。

(2) 一对连接节点对的有向边的集合 L 。如果存在从节点 X 指向节点 Y 的有向边, 则称 X 是 Y 的一个父节点。

(3) 每个节点 X_i 都有一个概率分布 $P(X_i | \text{Parents}(X_i))$, 量化其父节点对该节点的影响。

(4) 贝叶斯网络是一个有向无环图(DAG)。

贝叶斯网络的拓扑结构(即节点和边的集合)用一种精确简洁的方式描述了在域中成立的条件独立关系。在一个正确构造的网络中, 有向边表示 X 对 Y 有直接影响。在确定贝叶斯网络的拓扑结构之后, 还需要为每个节点指定其对父节点的条件概率。拓扑结构与条件概率表(CPT)结合起来, 可以隐含地指定所有变量的全联合概率分布。

2.2 贝叶斯网络的语义

贝叶斯网络指定了全联合概率分布, 全联合概率分布中的每个条目都可以表示为贝叶斯网络条件概率表(CPT)中适当元素的乘积, 公式表示为

$$\begin{aligned}
 P(x_1, \dots, x_n) &= \prod_{i=1}^n P(x_i | \text{parents}(X_i)) \\
 &= \prod_{i=1}^n P(x_i | x_{i-1}, \dots, x_1) \quad (1)
 \end{aligned}$$

贝叶斯网络提供了一种表示域中的条件独立关系的简洁方式。贝叶斯网络中的条件独立关系可以描述为: 给定父节点, 一个节点与它的祖先节点之间是条件独立的。即给定一个节点的马尔可夫覆盖, 这个节点和网络中的所有其它节点都是条件独立的。其中马尔可夫覆盖(Markov blanket)由一个节点 X 的父节点 U , 子节点 Y 以及子节点的父节点 Z 组成(如图 1 所示)。因此, 构造贝叶斯网络拓扑结构的正确次序是: 首先添加“根节点”, 然后加入受根节点影响的变量, 依次类推, 直到“叶子节点”, 即对其它变量没有直接因果影响的节点。

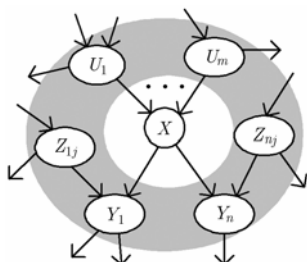


图 1 贝叶斯网络的条件独立关系

2.3 贝叶斯网络的概率推理

采用 X 表示查询变量, E 表示证据变量 E_1, E_2, \dots, E_m, e 表示一个观察到的特定事件, Y 表示非证据变量集(也称为隐变量) Y_1, Y_2, \dots, Y_n 。贝叶斯网络的概率推理的基本任务是: 在给定一组证据变量 E 的观察值 e , 计算一组查询变量 X 的后验概率分布 $P(X | e)$ 。可以通过贝叶斯公式计算:

$$P(X | e) = \frac{P(X, e)}{P(e)} = \alpha P(X, e) = \alpha \sum_y P(X, e, y) \quad (2)$$

α 是一个归一化常数, 保证其所包含的概率相加等于 1。其中的求和针对所有可能的 y , 即非证据变量 Y 的值的的所有可能组合。

贝叶斯网络推理大致分为以下 4 种推理模式:

(1) 诊断推理 从影响结果(证据)推理出原因, 也称为后向推理。通信网故障诊断就属于这种推理。例如: 给定一组证据, 求解引起这些证据的最可能的解释原因。

(2) 因果推理 从原因推理出影响结果, 也成为前向推理。例如: 给定原因, 求解最有可能产生的影响。

(3) 交互因果推理 多个原因共同产生一个影响结果。给定影响结果和其中一个原因, 推理出另一个原因的概率。

(4) 混合推理 以上 3 种推理模式的混合。

目前有很多概率推理方法, 大致分为: 精确推理和近似推理两种。精确推理方法包括: 枚举法, 变量消元法, 团算法等。在单连通网络中, 精确推理的时间复杂度与网络规模呈线性关系。近似推理包括: 似然加权, MCMC 算法等。随机近似技术能够提供对网络的真实后验概率的合理估计, 能够处理比精确算法规模大得多的网络。

3 基于贝叶斯网络的网间告警相关性模型

3.1 网间故障传播模型

通信网按照功能可以分为: 基础媒质层、传输层、交换层、接入层、用户层。大量研究分析表明: 应该针对不同网络的特点, 选择适应各自网络特性的告警相关性方法。对于复杂网络的问题, 还必须将多种方法组合在一起共同解决。

(1) 底层通信网(如 SDH, DWDM)的目标是隔离影响网络资源可用率的故障, 如光缆断、设备接口故障等。底层通信网最常采用确定性模型, 假设故障与症状间的依赖关系和因果关系必须百分之百准确, 如基于规则的告警相关性模型。

(2) 高层通信网(如 IP 网)拓扑结构经常发生变化, 告警包含大量不确定性的内容。高层通信网应该采用不确定性模型。

(3) 由于下层子网和上层子网间的依赖关系频繁改变, 确定性模型无法和实际配置情况保持一致, 应该采用不确定性模型分析网间的依赖关系。

本文采用故障传播模型, 描述不同层网络间的功能依赖关系(如图 2 所示)。故障传播模型是一个有向无环图 $G =$

$(V, E)^{[6]}$ 。其中 V 由一组非空的节点组成，对应不同的子网。 E 由一组有向边 (v_i, v_j) 组成，表示在子网 v_j 依赖于子网 v_i 所提供的服务。且 $v_i \neq v_j$ 。如果 (v_i, v_j) 属于 E ，则 (v_j, v_i) 不属于 E 。模型分 1 到 $N+1$ 层(Layer)，表示不同层网络。

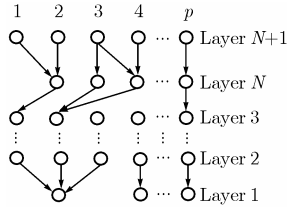


图 2 子网间的故障传播模型

目前有很多通信网建模的方法。采用面向对象的技术，ITU-T M.3100, TMF 814.2 定义网络信息模型(management information models)。网络信息模型中的被管对象间采用 TMN 接口交换管理信息。通信网体系结构模型(architectural models)用于描述通信网内部结构和功能，如 ITU-T G.774 描述 SDH 管理信息模型。还有其它一些方法，如 MODEL 提供一种规范的对象描述语言。

3.2 从故障传播模型中构造贝叶斯网络

从故障传播模型中构造贝叶斯网络分为 3 步^[4]：

(1) 对应故障传播模型中的每个节点，定义贝叶斯网络的一组变量 V_i ，其中变量的值域等于 {fault, normal}。当 $V_i = \text{fault}$ 时，表示对应的故障传播模型中的子网出现故障，无法向上层子网提供正确的服务。

(2) 对应故障传播模型中的每一条有向边 $X \rightarrow Y$ ，定义贝叶斯网络的一条有向边 $Y \rightarrow X$ ，表示节点 Y 是节点 X 的父节点， X 与 Y 之间存在功能依赖关系。

(3) 初始化定义“根节点”的先验概率和有向边的条件概率。

对应图 2 中的网间故障传播模型，构造的贝叶斯网络如图 3 所示。其中，“根节点”表示底层子网，“叶子节点”表示高层子网。在贝叶斯网络中，“根节点”的先验概率和有向边的条件概率的初始值可以根据历史数据统计得出，或是根据通信专家的经验进行估计。在此基础上，利用贝叶斯网络参数学习 EM 算法，验证、更新贝叶斯网络结构。

定义 X 表示根故障源； E 表示证据变量，即可观察(或计算)的子网内部失效情况， e 表示一个观察到的特定事件； Y 表示隐变量，即不可观察的未知子网内部失效情况。基于贝叶斯网络的网间告警相关性任务是：给定一组证据变量 E 的

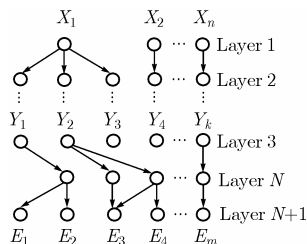


图 3 贝叶斯网络

观测值 e (即已知子网内部失效情况)，计算一组查询变量 X 的后验概率分布 $P(X|e)$ ，即诊断出根故障源。

3.3 告警相关性模型的实现框架

通信网规模和结构复杂度的增加对网络管理的稳定性和效率提出了更高的要求。因为采用集中网管模式的故障诊断系统执行效率低，系统健壮性差。很多研究指出，在大型网络中的告警相关性模型应该采用分布式结构^[7]。本文提出一种基于 CORBA 技术的分布式告警相关性模型(如图 4 所示)，CORBA 为各个模块间提供消息传递服务。

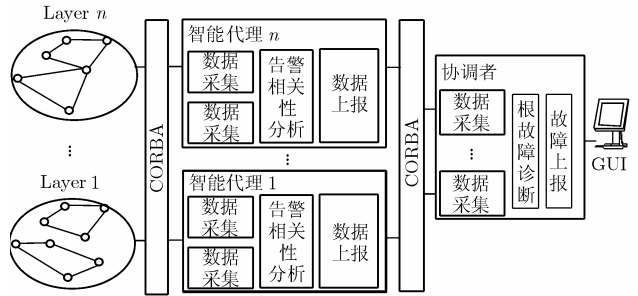


图 4 分布式告警相关性模型

被管网络被分成几个不同的管理域(domains)，管理域是一个抽象的概念，可以是子网、网元或是网元的功能单元等。每个管理域包含一个智能代理，负责域内的告警相关性分析。智能代理也可以称为域内故障诊断代理(Domain Diagnostic Agent, DDA)，负责监控和分析域内的网络状态，提供故障检测和故障定位功能。智能代理包括以下 3 个功能模块：

(1) 数据采集模块 采集被管网络的告警信息和配置信息，并将数据格式转化为适于数据库存储的统一格式。告警消息包括多个字段，如告警发生时间、告警清除时间、告警发生的位置信息、告警级别、告警内容描述等等。配置信息包含网元层和网络连接层的具体信息。网元层描述的是被管网络中物理或逻辑上的网元，包括子网、局站、机框、机槽、机盘、端口等等。网络连接层则描述的是网元之间的连接关系。

(2) 告警相关性分析模块 根据子网自身的特点选择相应的告警相关性分析方法，压缩告警，上报分析结果(相关告警组和可能故障)。提供通信专家在线维护和配置告警相关性规则的功能。

(3) 数据上报模块 向上层协调者上报可能的故障。

此外，必须有一个协调者(coordinator)共享网间的信息，协调解决网间的根故障识别。协调者主要包括以下 3 个功能模块：

(1) 数据采集模块 负责采集不同管理域上报的可能的故障集合。

(2) 根故障诊断模块 根据网间实体的功能依赖关系，建立故障传播模型。采用基于概率的方法识别根故障。

(3) 故障上报模块 向用户上报故障，并提供用户图形界

面。

通信网的告警相关性分析的过程如下:

步骤 1 数据采集模块采集告警和网络配置信息;

步骤 2 告警和配置格式标准化, 并将信息按时间窗口和所属子网分类, 存储在数据库中;

步骤 3 根据通信子网的特点选择不同的告警相关性模型, 并行执行告警相关性分析;

步骤 4 根据网间实体的功能依赖关系, 建立故障传播模型。根据网间的故障传播模型, 将子网状态(相关告警组和可能故障)传递给上层子网;

步骤 5 构造贝叶斯网络对整个通信网进行根故障诊断;

步骤 6 图形化显示根故障, 具体的故障内容包括: 故障系统代号, 故障区间, 故障源, 故障描述, 故障发生时间等。并为用户提供与故障相关的告警信息查询功能。

4 实验模拟

通信网故障大致可以分为: 线路故障和设备故障。因为线路故障对通信网服务质量的影响最大, 本文只讨论底层网络出现的线路故障对高层网络服务的影响。即当上层 SDH 系统出现线路故障时, 计算底层光缆段出现线路故障的可能性。

实验环境中是采用朗讯 ADM16 SDH 设备和 OLS80G DWDM 设备组建网络, SDH 系统采用 MS-SPRING 保护方式。SDH 与 DWDM 系统的承载关系如图 5 所示。

(1)物理传输层由 4 个光缆段组成{1#, 2#, 3#, 4#}, 光缆上承载一个 DWDM 系统。

(2)DWDM 系统由 a ~ d 4 个 OADM 站组成链状结构, 组成 3 个光传送段{a-b, b-c, c-d}。

(3)4 个 SDH 系统{R1, R2, R3, R4}承载在 DWDM 系统之上。其中 R1 承载在 b-c 段上, R2 承载在 b-c 和 c-d 段上, R3 承载在 b-c 和 c-d 段上, R4 承载在 a-b 和 b-c 段上。

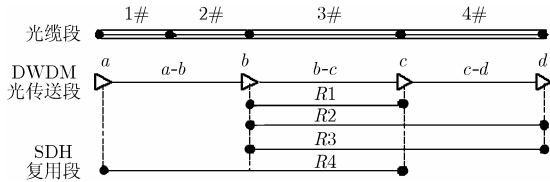


图 5 SDH over DWDM 实验模型

定义 查询变量 $X = \{X_1, X_2, X_3, X_4\}$ 表示光缆段信息, 隐变量 $Y = \{Y_1, Y_2, Y_3\}$ 表示 DWDM 光传送段信息, 证据变量 $E = \{E_1, E_2, E_3, E_4\}$ 表示 SDH 复用段信息。其中变量的值域等于 {fault, normal}。当变量状态等于 fault 时, 表示线路故障, 无法为上层提供可靠的传输服务。

根据 SDH over DWDM 故障传播模型, 构造贝叶斯网络(如图 6 所示)。贝叶斯网络的条件概率表根据历史故障数据统计获得(参见表 1-表 3)。

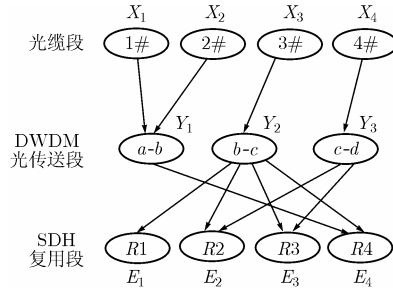


图 6 贝叶斯网络

表 1 光缆段故障的先验概率

	X_1	X_2	X_3	X_4
故障	0.3	0.1	0.2	0.1

表 2(a) DWDM 光传送段 Y_1 故障的条件概率表

X_1	X_2	Y_1
		故障
故障	故障	0.95
故障	正常	0.95
正常	故障	0.95
正常	正常	0.1

表 2(b) DWDM 光传送段 Y_2 故障的条件概率表

X_3	Y_2
	故障
故障	0.95
正常	0.1

表 2(c) DWDM 光传送段 Y_3 故障的条件概率表

X_4	Y_3
	故障
故障	0.90
正常	0.1

案例 1 在某个时间段内, R2 和 R3 同时上报告警, 告警指示 R2 和 R3 复用段线路故障。

$$\begin{aligned} P(X_3 | e_3, e_4) &= \alpha P(X_3, e_3, e_4) \\ &= \alpha P(X_3)P(Y_2 | X_3)P(e_3 | Y_2)P(e_4 | Y_2) \\ &= (0.94, 0.06) \end{aligned}$$

$$\begin{aligned} P(X_4 | e_3, e_4) &= \alpha P(X_4, e_3, e_4) \\ &= \alpha P(X_4)P(Y_3 | X_4)P(e_3 | Y_3)P(e_4 | Y_3) \\ &= (0.88, 0.12) \end{aligned}$$

分析结果: 引起“R2 和 R3 同时出现告警”的根故障是: 3# 光缆段出现线路故障。

案例 2 在某个时间段内, R1, R2 和 R4 同时上报告警, 告警指示 R1, R2 和 R4 复用段线路故障。

$$\begin{aligned} P(X_1 | e_1, e_2, e_4) &= \alpha P(X_1)P(Y_1 | X_1)P(e_4 | Y_1) = (0.74, 0.26) \\ P(X_2 | e_1, e_2, e_4) &= \alpha P(X_2)P(Y_1 | X_2)P(e_4 | Y_1) = (0.42, 0.58) \\ P(X_3 | e_1, e_2, e_4) &= \alpha P(X_3)P(Y_2 | X_3)P(e_1 | Y_2)P(e_2 | Y_2) \\ &\quad \cdot P(e_4 | Y_2) = (0.99, 0.01) \end{aligned}$$

$$P(X_4 | e_1, e_2, e_4) = \alpha P(X_4)P(Y_3 | X_4)P(e_2 | Y_3) = (0.75, 0.25)$$

分析结果: 引起“R1, R2 和 R4 同时出现告警”的根故障是: 3# 光缆段出现线路故障。

定义故障诊断准确率 HR, 用于衡量系统的执行效率。根据实验室多次测量结果统计平均, 根故障诊断率 $HR \geq 92.5\%$

$$HR = \frac{\text{诊断出的根故障数}}{\text{实际故障总数}}$$

表 3(a) SDH 环 1 故障的条件概率表

Y_2	E_1
	故障
故障	0.9
正常	0.2

表 3(b) SDH 环 2 故障的条件概率表

Y_2	Y_3	E_2
		故障
故障	故障	0.9
故障	正常	0.3
正常	故障	0.3
正常	正常	0.1

表 3(c) SDH 环 3 故障的条件概率表

Y_2	Y_3	E_3
		故障
故障	故障	0.9
故障	正常	0.3
正常	故障	0.3
正常	正常	0.2

表 3(d) SDH 环 4 故障的条件概率表

Y_2	Y_1	E_4
		故障
故障	故障	0.9
故障	正常	0.2
正常	故障	0.2
正常	正常	0.1

5 结束语

贝叶斯网络是目前不确定知识表达和推理领域最有效的理论模型之一, 适用于不确定性和概率性的知识表达和推理, 特别适用于有条件地依赖多种控制因素的决策。贝叶斯网络是一种基于网络结构的有向图解描述, 具有多源信息一致表达与信息融合能力, 能进行双向并行推理, 并能综合先验信息和样本信息, 使推理结果更为准确可信。因此, 贝叶斯网络在故障诊断领域中的应用具有重要意义。

贝叶斯网络的粗细粒度取决与系统设计者, 取决于故障诊断需要的精确程度。选择贝叶斯网络建立告警相关性模型的重要原因有以下几点: (1)从贝叶斯网络结构中识别与问题

不相关的条件, 简化告警相关性处理复杂度; (2)从故障传播模型中推导出贝叶斯网络结构; (3)从历史数据或者专家经验中估计初始的先验概率和条件概率; (4)利用 EM 算法进行参数学习, 验证、更新贝叶斯网络结构; (5)稳健性; (6)算法支持。

本文提出了一种基于贝叶斯网络的告警相关性分析和故障诊断模型。本文详细讨论了从故障传播模型中构造贝叶斯网络, 以及分布式告警相关性模型的实现框架。最后结合 SDH over DWDM 系统, 利用贝叶斯网络准确定位根故障点, 具有很强的可操作性。今后将重点研究基于贝叶斯网络的通信网故障诊断的参数学习。

参考文献

- [1] Gardner R D and Harle D A. Methods and systems for alarm correlation. Global Telecommunications Conference, 1996. GLOBECOM '96. 'Communications: The Key to Global Prosperity, London, UK, 18-22 Nov., 1996, vol.1: 136-140.
- [2] Bouloutas A T, Calo S, and Finkel A. Alarm correlation and fault identification in communication networks. *IEEE Transactions on Communications*, 1994, 42(2/3/4): 523-533.
- [3] Ekaette E U and Far B H. A framework for distributed fault management using intelligent software agents. IEEE CCECE 2003, Canadian Conference on Electrical and Computer Engineering, Canada, 4-7 May, 2003, vol.2: 797-800.
- [4] Steinder M and Sethi A S. End-to-end service failure diagnosis using belief networks. Network Operations and Management Symposium (NOMS), Florence, Italy, 2002: 375-390.
- [5] Russell S and Norving P. Artificial Intelligence: A Modern Approach (Second Edition). USA, Prentice-Hall, 2003: 540-546.
- [6] Choi Jaesung, Choi Myungwhan, and Lee Sang-Hyuk. An alarm correlation and fault identification scheme based on OSI managed object classes. ICC '99. IEEE International Conference on Communications, Vancouver, BC, 6-10 June, 1999, vol.3: 1547-1551.
- [7] Li H, Yang S, and Baras J S. On system designs for distributed, extensible framework for network monitoring and control. Tech. Rep. CSHCN TR 2001-12, Center for Satellite and Hybrid Communication Networks, University of Maryland, 2001.

邓 歆: 女, 1977 年生, 博士生, 研究方向为通信软件与网络管理。

孟洛明: 男, 1955 年生, 教授, 研究方向为通信软件与网络管理。