

## 面向大规模分布式计算发布订阅系统核心技术\*

马建刚<sup>1,2+</sup>, 黄涛<sup>1</sup>, 汪锦岭<sup>1,2</sup>, 徐罡<sup>1</sup>, 叶丹<sup>1</sup>

<sup>1</sup>(中国科学院 软件研究所 软件工程技术研究开发中心,北京 100080)

<sup>2</sup>(中国科学院 研究生院,北京 100049)

### Underlying Techniques for Large-Scale Distributed Computing Oriented Publish/Subscribe System

MA Jian-Gang<sup>1,2+</sup>, HUANG Tao<sup>1</sup>, WANG Jin-Ling<sup>1,2</sup>, XU Gang<sup>1</sup>, YE Dan<sup>1</sup>

<sup>1</sup>(Technology Center of Software Engineering, Institute of Software, The Chinese Academy of Sciences, Beijing 100080, China)

<sup>2</sup>(Graduate School, The Chinese Academy of Sciences, Beijing 100049, China)

+ Corresponding author: Phn: +86-10-62630989, E-mail: mjpg@otcaix.iscas.ac.cn, <http://www.iscas.ac.cn>

**Ma JG, Huang T, Wang JL, Xu G, Ye D. Underlying techniques for large-scale distributed computing oriented publish/subscribe system. *Journal of Software*, 2006,17(1):134-147.** <http://www.jos.org.cn/1000-9825/17/134.htm>

**Abstract:** The publish/subscribe system is adapted to the dynamic large-scale distributed computing environment well and will be used widely due to its asynchronous, many-to-many and loosely-coupled communication properties. This paper analyzes the state-of-art of publish/subscribe systems. Many existing systems are classified according to the criteria, such as topology structure, event model and subscription model. Then the key techniques, such as matching algorithm, content-based routing algorithm, formal modeling, quality of service, are explained. Typical pub/sub prototypes and products are compared, and their shortcomings and limitations are discussed. The challenge of enhancing the system intelligence by introducing the event semantic and approximate matching, and the trends of supporting the mobile computing and P2P computing environment is forecasted.

**Key words:** Publish/Subscribe; event model; match algorithm; content-based routing; middleware; P2P

**摘要:** 发布/订阅系统技术具有异步、松散耦合和多对多通信的特点,适应了目前动态多变的大规模分布式计算环境的需求,有着广阔的应用前景.分析了国内外发布/订阅系统的研究现状,并从拓扑结构、事件模型和订阅模型等不同角度进行了系统的分类,然后分别就其关键问题从匹配算法、基于内容的路由算法、形式化建模和服务质量等方面进行了阐述,并对已有的典型系统进行了分析比较,指出了当前该领域研究存在的问题和不足.同时,分析了在支持语义和近似匹配来增强系统智能性所面临的挑战,展望了发布/订阅系统在支持移动计算、P2P等新型计算环境下的研究趋势.

**关键词:** 发布/订阅;事件模型;匹配算法;基于内容的路由;中间件;P2P

\* Supported by the National High-Tech Research and Development Plan of China under Grant Nos.2001AA113010, 2002AA413610, 2003AA413010, 2003AA115440 (国家高技术研究发展计划(863)); the National Grand Fundamental Research 973 Program of China under Grant No.2002CB312005 (国家重点基础研究发展规划(973))

Received 2005-06-23; Accepted 2005-07-18

中图法分类号: TP393 文献标识码: A

Internet 技术的广泛应用和移动计算、网格计算以及普适计算平台的快速发展,要求分布式系统能够满足大规模、分散控制和动态改变的要求.这就要求系统的各参与者之间,采用一种具有动态性和松散耦合特性的灵活通信范型和交互机制.发布/订阅(Pub/Sub)通信范型与传统的通信范型(消息传递、RPC/RMI 和共享空间)相比,具有异步、多点通信的特点,使通信的参与者在空间、时间和控制流上完全解耦<sup>[1]</sup>,能够很好地满足大型分布式系统松散通信的需求.

Pub/Sub 系统技术具有良好的应用前景,早期的基于主题的 Pub/Sub 系统已有很多成熟的产品广泛应用于银行、证券、制造业企业信息化等各个领域,较有影响的应用案例包括 NASDAQ 证券交易系统、费城股票交易所业务系统等.而面向大规模分布式计算的新型 Pub/Sub 系统,由于其关键技术尚未成熟,目前尚没有较有影响的应用案例,可以预见的应用场景包括:各类电子商务系统(网上拍卖系统、网上交易市场)、企业应用集成(EAI)、基于事件的供应链管理(ESCM)、电子新闻分发、在线网络游戏和大规模环境监测等.

目前,学术界和工业界从不同侧面对 Pub/Sub 系统技术进行了深入研究,工业界侧重于实用,关注 Pub/Sub 系统的服务质量(效率、安全性和可靠性等);而学术界则侧重于增强 Pub/Sub 系统的表达能力(如语义方面)和对新的应用环境(移动环境和 P2P 环境)的支持.针对现存的主要问题,有不同的研究工作和技术成果,例如匹配算法<sup>[2-10]</sup>、基于内容的路由算法<sup>[11-13]</sup>、安全策略<sup>[14-17]</sup>、基于 P2P 网络的 Pub/Sub 系统解决方案<sup>[18-23]</sup>、Pub/Sub 系统在普适计算和移动计算环境下的应用<sup>[24-30]</sup>、语义 Web 技术和本体论与 Pub/Sub 系统的结合<sup>[31-35]</sup>以及容错策略<sup>[36-40]</sup>和提高 Pub/Sub 系统可靠性的事务机制<sup>[31,41]</sup>的研究.同时,他们推出了各自的产品和原型系统,如 TIB/Rendezvous<sup>[42]</sup>(下面简称 TIB),Gryphon<sup>[2]</sup>,SIENA<sup>[11]</sup>,JEDI<sup>[12]</sup>和 Scribe<sup>[18]</sup>等.但从 Pub/Sub 系统的核心技术来看,一些关键问题仍有待解决,特别是新的计算环境下新的需求的产生,还有更为广阔的研究空间.

本文首先概述了发布/订阅系统的基本结构和特征;随后分析了其研究的主要问题和核心技术,分别从它的拓扑结构、模型、分类、匹配算法、路由算法、形式化建模和服务质量等方面进行阐述,并对已有的典型系统进行了比较;然后指出了当前该领域研究存在的问题和面临的挑战;最后,展望了未来的研究热点和发展方向.

## 1 发布/订阅系统概述

发布/订阅系统是一种使分布式系统中的各参与者,能以发布/订阅的方式进行交互的中间件系统.在发布/订阅系统中,信息的生产者和消费者之间所交互的信息被称为事件.如图 1 所示,生产者将事件发送给发布/订阅中间件;消费者则向发布/订阅中间件发出一个订阅条件,表示对系统中的哪些事件感兴趣,如果不再感兴趣,也可以取消订阅;而发布/订阅中间件则保证将生产者发布的事件及时、可靠地传送给所有对之感兴趣的消费者.信息的生产者称为发布者(publisher),信息的消费者称为订阅者(subscriber),发布者和订阅者统称为客户端.匹配算法(matcher)负责高效地找到与给定的事件相匹配的所有订阅条件;而路由算法(router)则负责选择适当的路径,将一个事件从发布者传送到订阅者.

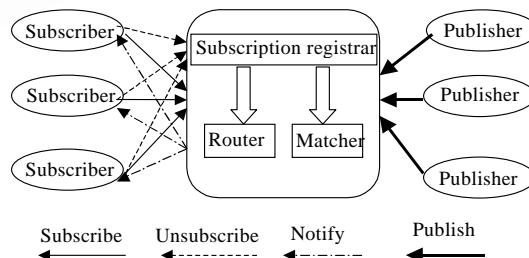


Fig.1 Concept model of Pub/Sub system

图 1 Pub/Sub 系统概念模型

一个典型的发布/订阅系统包括拓扑结构、事件模型、订阅模型、匹配算法、路由算法和提供服务质量保证的设施.其中:事件模型定义了事件的数据结构;订阅模型定义了系统能够支持的订阅条件,指明了订阅者如何表达对事件子集的兴趣;事件模型和订阅模型共同决定了系统的表达能力.匹配算法一般要结合相应的事件和订阅模型进行优化.拓扑结构决定了系统的可扩展性,路由算法一般要根据相应的事件代理网络的拓扑结构进行优化.Pub/Sub 系统的表达能力、效率、可扩展性和服务质量是其主要目标,但这些目标往往是相互矛盾的,需要根据具体需求权衡利弊,找到合理的均衡点.

## 2 发布/订阅系统的拓扑结构和分类

Pub/Sub 系统按照不同的功能特征,可以有不同的分类.其中通知模型可以刻画系统在事件发生时,通知订阅者的方式.它遵循推或者拉两种通信风格:在推的风格中,信息的提供者发起与接收者之间的通信,多用在移动和普通计算环境下;而在拉的风格中,信息的接收者来选择信息的提供者.Pub/Sub 系统按照通知模型可以分为 3 类:推的方式、拉的方式和两者结合的方式.典型系统如:TIB 采用了推的方式;而 JEDI 系统则支持推和拉两种方式.在文献[11]中提到了一个分类框架,把 Pub/Sub 系统从影响扩展性的拓扑结构的角分为集中型、层级客户端/服务器型和 P2P 型;从影响表达能力的订阅语言的角度分为基于渠道的、基于主题的、基于内容的和基于带模式内容的 4 种.文献[12]中则按照系统结构分为直接连接、广播、中心式、分布式和固定的 5 类;按照事件模型,分为基于元组的、基于记录的和基于对象的 3 种.但是这些分类比较模糊,也不能涵盖所有的 Pub/Sub 系统.第 2.1 节给出了一个按照系统拓扑结构的分类,第 2.2 节给出了按照事件模型和订阅模型的 Pub/Sub 系统分类.

### 2.1 发布/订阅系统的拓扑结构

为了提高系统的扩展性,大规模的 Pub/Sub 系统通常采用分布式的系统结构,其中分布着多个事件代理(event broker),每个事件代理为一定数量的本地客户端服务.这些事件代理按照一定的拓扑结构组织,负责事件在各节点之间的转发.Pub/Sub 系统的拓扑结构对其性能有着重要的影响.我们按照拓扑结构的稳定性把 Pub/Sub 系统在结构上分为静态模式和 P2P 模式.其中:静态模式又可分为星形、总线形、层级、环形、无环图和一般图;而构建于 P2P 网络之上的 Pub/Sub 系统的结构又可以分为纯 P2P 网络和混合的 P2P 网络两种.

在星形拓扑结构中,事件服务器采取中央服务器的结构,单一的事件服务器作为订阅者和发布者之间的代理,该拓扑结构可扩展性不好,实践中比较少用.在总线形结构中,采用一条公共总线来传递事件,可充分利用广播机制,典型的系统如 TIB.在层级结构中,事件代理之间构成了层级结构关系,事件代理之间、客户端和代理之间的通信采用同一种协议.事件代理会从所有的客户端接收发布的事件和订阅,并转发给关心该事件的子树.该结构中,层级越高的事件代理负载越重,容易产生性能瓶颈.SIENA, Gryphon 和 JEDI 等系统采用了层级拓扑结构.在环形拓扑中,事件代理的连接图是环形的,代理之间是平等的关系.代理之间的通信协议是双向的,但是代理之间的通信协议与客户端和代理之间的协议不同.在无环图结构中,其结构可以用无环无向图表示,信息是双向传送,能够确定任何两个代理节点只有一条路径.一般图结构进一步取消了无环性约束,两个代理间存在多条路径,冗余连接克服了单点失败,但却增加了路由算法选择最佳路径的复杂性.在 P2P 模式中,所有节点都是平等的,可以动态加入或退出.采用纯 P2P 网络的 Pub/Sub 系统是指 P2P 网络中的每个节点同时作为事件代理和客户端,如 Scribe 系统.而混合的 P2P 网络的 Pub/Sub 系统是指 P2P 网络中的每个节点只作为事件代理,每个代理连接若干个客户端,客户端不在 P2P 网络中,如 Hermes<sup>[19]</sup>.

### 2.2 基于事件模型和订阅模型的发布/订阅系统的分类

在发布/订阅系统发展过程中出现了各种各具特色的系统,其表达能力越来越强.下面,我们对发展过程中出现的各类系统进行回顾并分类,分析它们各自的特点和存在的问题.发布/订阅系统按照事件模型和订阅模型可以分为基于渠道、基于主题、基于内容、基于类型和混合型 5 类,其中,基于内容的系统又分为基于 Map 的和基于 XML 的 Pub/Sub 系统.

在基于渠道的系统中,订阅是通过指定渠道(channel)来实现的,订阅者接收所有发布到渠道中的事件,事件的接收与事件的内容无关.基于渠道的系统在功能上等同于在渠道和多播地址之间进行一对一映射的可靠多播.CORBA 事件服务规范<sup>[43]</sup>就是一个典型的基于渠道的系统,它定义了一个基本事件通知服务,发布者和订阅者可以通过若干个事件渠道进行交互和异步通信,不同的事件渠道可以像管道一样来组合.在这类系统中,订阅者不能指定它对渠道中的一部分事件感兴趣,表达能力很弱,但是实现简单.

基于主题的系统将所有的事件按照主题划分成组,每个事件只属于其中一个主题.订阅者在订阅信息时,指明对哪个主题感兴趣,将来一旦出现该主题下的事件,系统都会自动将其发送给订阅者,订阅者接收其订阅主题的所有事件.这种系统不去理解事件的内部结构,而将其看成是一个封闭的黑盒子,订阅者不能更细粒度地表达他对某主题下的一部分事件感兴趣.早期的 Pub/Sub 系统一般都是基于主题的系统,例如 TIB 的较早版本、IBM MQSeries 的较早版本、Scribe 和 Bayeux<sup>[20]</sup>等.后来,人们又在此基础上作了一些改进,将主题组织成一种层次结构.订阅者在订阅上层的主题时,即涵盖了下层的各主题,例如主题 B 可以定义为主题 A 的子主题,那么匹配 B 的事件会被所有订阅 A 和 B 的客户端接收.这样,对系统的表达能力作了一些提高,但是仍然改变不了主题概念的本质.这类系统的优点是实现简单、有效,但是表达能力较弱.

在基于内容的 Pub/Sub 系统中,订阅者在事件的内容上来指定约束条件,表达他们的兴趣.事件不局限于只属于特定的组,一个事件指向哪里的决策是由订阅者声明的查询和谓词决定的.基于内容的系统的优点是订阅的灵活性,订阅者在订阅前不需要了解一套定义订阅的主题名;带来的不利是给系统根据订阅条件来匹配大量事件,带来很大负担.订阅数目比基于主题的系统的主体的数目要大得多,因此匹配的效率和必须很高.目前的基于内容的 Pub/Sub 系统又可分为两类:一类是基于 Map 的;另一类是基于 XML 的.

在基于 Map 的 Pub/Sub 系统中,事件的内容为多个“属性=值”的集合(称由多个“属性=值”组成的集合为一个 Map).较有影响的原型系统包括 Gryphon,SIENA,JEDI 和 Rebeca<sup>[13]</sup>等.对于多数基于 Map 的系统而言,客户的订阅条件一般是由各个原子约束条件的“与”操作组成,每个原子约束条件中只涉及到其中一个属性,通常称为平面模式.在基于 Map 的系统中,每个事件的内容除了一个 Map 以外,也可以附加上一个系统不能理解的数据区.这样,Map 就成为“事件头”,后面的数据区就成为“事件体”,Map 中的数据用于对事件体进行描述,系统可以根据该描述信息进行相应的处理.

近年来,随着 XML 越来越多地作为不同系统之间数据交换的格式,人们开始提出了基于 XML 的 Pub/Sub 系统.在这种系统中,每个事件是一个 XML 文档,其表达能力比基于 Map 的系统有了很大的提高.客户的订阅条件一般是 XPath 表达式,其中既包括对 XML 文档结构的约束,又包括对某些元素和属性的约束,通常称为树模式.目前已有的原型系统包括 XFilter<sup>[5]</sup>,XTrie<sup>[6]</sup>,WebFilter<sup>[7]</sup>和 MatchMaker<sup>[8]</sup>等.虽然 XML 具有很强的表达能力,但是,订阅者必须预先知道被发布的 XML 文档所遵从的 XML Schema,才能根据该 Schema 定义出相应的订阅条件.

Eugster 等人将面向对象语言的类型模型与发布/订阅系统的事件模型结合起来,提出了一种“基于类型的发布/订阅系统”<sup>[44]</sup>.在这种系统中,事件被声明为属于特定类型的对象,事件对象中封装了属性和方法,事件对象根据类型进行分类;订阅对象中封装了订阅者的订阅条件,订阅者通过指定接收的对象类型来订阅.这种系统表达能力介于基于主题的和基于内容的系统之间,对类型的粗粒度的过滤类似基于主题的系统,在属性之上的细粒度的约束又类似于基于内容的系统,而在方法之上的约束又是面向对象技术的结果.但这类系统的表达式太复杂,很难抽取不同订阅条件之间的共性,从而很难设计一个高效的匹配算法,系统的效率较低.

有些 Pub/Sub 系统既具有基于主题的系统的特点,又具有基于 Map 的系统的特点,可以被看成是一种混合型系统.每个事件的内容包括一个 Map 形式的事件头和一个系统不能理解的事件体,同时,每个事件必须属于某个主题.客户在定义订阅条件时,必须先指定主题,再在其基础上,根据 Map 中的各项来定义详细的过滤条件.其典型系统包括 JMS 规范<sup>[41]</sup>和 CORBA 通知服务规范<sup>[45]</sup>,它们所规定的订阅语言具有很强的表达能力,类似于 SQL 语言.对于这种订阅模型,很难找到高效的匹配算法,当到达一个事件后,系统只能与各订阅条件逐个地进行匹配.因此,它们不具有较好的可扩展性.

### 3 发布/订阅系统的匹配算法

发布/订阅系统的匹配算法负责高效地找到与给定的事件相匹配的所有订阅条件,其设计目标主要包括:匹配的时间效率、匹配的空间效率和订阅维护的效率.其中,匹配的时间效率是最重要的目标.Pub/Sub 系统的匹配算法总是依赖于特定的数据模型:对于基于主题的 Pub/Sub 系统,因为给定主题下的每个事件都要被转发给该主题下的所有订阅者,所以不需要为其设计相应的匹配算法;对于基于 Map 和基于 XML 的 Pub/Sub 系统,已经提出了很多种高效的匹配算法;而对于其他具有复杂数据模型的 Pub/Sub 系统,尚未找到高效的匹配算法.大多数已有的 Pub/Sub 系统匹配算法都是基于这样的思想:对各个客户的订阅条件中的一些重复的成分,只需判断一次来优化,并充分利用订阅的索引结构,以缩短匹配时间.

#### 3.1 基于Map的发布/订阅系统匹配算法

基于 Map 的 Pub/Sub 系统匹配算法已经有了很多的研究成果<sup>[2-4]</sup>.Gough 等人在文献[2]中提出了一种基于搜索树的匹配算法,它将各订阅条件组织成一种树型结构,该树的深度为系统中的全部属性总数.当一个事件到达时,系统按某一路径从树根到达某叶子节点,就得出了所有匹配成功的订阅条件.该算法实际上是一种状态空间搜索方法.该算法的时间复杂度较低,速度很快,但其空间复杂度为指数级.同时,该算法的订阅维护的成本很高,每当客户增加订阅或取消订阅时,系统难以对该搜索树进行修改以反映订阅的变化,而必须要重建搜索树.因此,该算法不太适用于大规模的系统.

Aguilera 等人在文献[3]中提出了一种基于并行搜索树的算法.在这种树中,每个叶子节点表示一个订阅条件,每个非叶子节点表示对某个属性的一个操作,每个边表示操作的结果.当事件到达时,系统通过对该搜索树进行遍历,以得到所有匹配成功的订阅条件.因为多个订阅条件中的相同约束条件只需被判断一次,所以,其时间效率要优于将事件与各订阅逐个匹配的简单方法.该算法的空间复杂度与订阅数量呈线性关系,而且当客户增加或取消订阅时,系统可以通过对该搜索树进行局部修改来完成,因而订阅维护的成本较低.

Campailla 等人在文献[4]中提出一种基于二叉判定图(binary decision diagram)的算法.订阅条件中的每个原子约束条件可以被看成是一个布尔变量,从而每个订阅条件可以用一个二叉判定图表示.由于不同订阅条件中可能有相同的原子条件,则将各订阅的二叉判定图整合到一起,形成共享二叉判定图,其中的节点数为各订阅条件中的全部原子条件数量.当一个事件到达时,先求出各原子条件的值,然后再通过对二叉判定图的遍历,以求出所有匹配成功的订阅条件.该算法的优点在于它可以同时支持“与”、“或”两种操作,而其他的订阅匹配算法基本上都是只支持原子约束条件之间的“与”操作,对于那些存在“或”操作的订阅条件,需要将其转化成多个只包含“与”操作的订阅条件来处理.

#### 3.2 基于XML的发布/订阅系统的匹配算法

基于 XML 的 Pub/Sub 系统,出现了不少有影响的匹配算法<sup>[5,6,9,10]</sup>.它的客户订阅语言一般是 XPath,其匹配算法也是结合 XPath 语言的特点来优化的.文献[5]中提出一种称为 XFilter 的 XPath 树模式匹配方法,将每个树模式看成多个有限状态机(FSM)的集合,每个 FSM 负责其中的一个分枝.然后,所有这些 FSM 通过哈希表的方式来建立索引,以实现快速匹配.后来,又加以改进,提出一种 YFilter 方法<sup>[9]</sup>,将所有的树模式表示成一个统一的非决定性状态自动机(NFA),使不同树模式中的相同路径在匹配时只需被处理一次,以进一步提高匹配效率.文献[6]提出一种称为 XTrie 的索引结构,将每个树模式分解为多个“元素序列”的集合,对这些序列建立一种 XTrie 索引结构.在对 XML 文档进行匹配时,不同序列中的共同部分只需要处理一次,从而可以提高效率.文献[10]提出了一种基于层级下推自动机(HPDT)的方法来高效地处理 XPath 查询,并在原型系统 XSQL 上进行了验证.

### 4 发布/订阅系统的路由算法

Pub/Sub 系统路由算法就是要解决如何在事件代理网络中寻找一条恰当的路径,使事件低成本、高效和可靠地到达各相关的订阅者的问题.其追求的设计目标有:事件转发的网络效率、订阅维护的网络效率、容错能

力、负载均衡和每个节点的资源消耗.其中,事件转发的网络效率是最主要的设计目标.

常见的主要是基于内容的路由算法,它按照路由的准确性可以分为非精确型和精确型两大类:非精确型路由算法采用了洪泛法,会把事件转发给没有订阅的消费者,网络传输效率较低,但同时也降低了每个代理的负载;精确型路由算法只把事件发给有匹配的订阅条件的订阅者,旨在降低各事件代理之间的消息转发数量,但同时也增加了维护订阅路由表的成本和事件转发的复杂性等.目前研究得较多的是精确型路由算法,它主要包括简单路由算法和基于覆盖的路由算法:简单路由算法是一种基于过滤的算法,它通过在事件代理网络中,洪泛地发布新的和取消的订阅,使所有代理的路由表中都有每个活动的订阅.然而,由于所有的代理都必须具有所有活动订阅的知识,简单路由的路由表很大,而且订阅维护的成本很高,不适合大规模的应用系统.简单路由算法被 Gryphon 系统采用;基于覆盖的路由算法,不需要所有活动订阅的全局知识,它采用基于订阅覆盖测试的有选择性的转发策略,避免洪泛似的在代理网络中发送所有订阅,它被 SIENA 和 JEDI 系统采用.

Pub/Sub 系统路由算法的难点在于:当一个事件被发布时,并没有指明具体的接收者,每个事件代理收到一个事件后,它根据该事件的内容以及自己所知道的订阅条件,来决定下一步的转发.因为 Pub/Sub 系统中的每个消息的接收者都是不确定的,所以,已有的基于内容的路由算法都是以某种广播协议为基础的,采用一些优化措施,以避免不必要的消息转发.对于事件的传播,主要的优化措施是检测目标子网中是否有对本事件感兴趣的订阅者;对于订阅消息的传播,则是利用过滤条件之间的覆盖关系来优化.

Pub/Sub 系统的路由算法按照网络的拓扑结构的稳定性,分为基于静态网络的路由算法和基于 P2P 网络的路由算法两类:前者常常以固定不变的拓扑结构为基础,采用某种策略进行高效的消息转发;后者则利用 P2P 网络的路由协议的特性,来建立 Pub/Sub 系统优化的路由策略.

#### 4.1 基于静态网络的内容路由算法

在这类路由算法中,常常利用固定不变的拓扑结构的特性,采用优化的策略进行高效的消息转发.每个代理都汇合了其一部分邻居发来的订阅请求,作为自己的订阅请求向另一部分邻居转发.它将订阅消息中的信息分解为订阅者和过滤条件两部分:其中订阅者为发出订阅请求的节点;过滤条件为对事件的约束.当一个代理收到其邻居发来的订阅消息后,它将消息中的订阅者改为自己,而过滤条件保持不变,再转发给其他邻居.在大多数已有的研究<sup>[11-13]</sup>中,为了提高路由算法的效率,通常将事件代理网络的拓扑结构组织成层级结构或无环图结构,相应的路由算法基于这个给定的拓扑结构来选择合适的策略进行高效的消息转发.基于静态网络的内容路由算法一般可以分为生成树转发和逆向路径转发两大类,而利用广告技术作为优化基于内容的路由的附加机制.

在生成树转发算法中,整个系统中的所有事件代理被预先组织成一个层级结构.当某事件代理收到其客户端发来的订阅请求时,订阅消息由当前节点逐级地向根节点发送,并利用过滤条件的覆盖关系对消息转发进行优化.当某事件代理收到其客户发布的事件时,事件的传播采用“生成树转发”广播算法<sup>[46]</sup>,即该事件由当前节点逐级向根节点发送,途中每个节点将其转发到对其感兴趣的各个子树.典型系统 SIENA 和 JEDI 中都采用了这种路由算法.其缺点在于,在整个事件代理树中,由于每个事件都要被上传到根节点,因而根节点的负载较大,容易形成性能瓶颈.因为每个节点都保存了其子树中所有客户的订阅信息,所以越是上层的节点,其保存的订阅信息就越多,处理负担就越重,从而造成各节点负载的不均匀.而且,一个事件代理的失败会使所有与此相连的子网节点和其他节点断连,容易产生单点失败问题.

在采用基于逆向路径转发的路由算法中,订阅消息的传播采用“基于源转发”广播算法,事件消息的传播采用“逆向路径转发”广播算法<sup>[46]</sup>.每个事件代理都预先知道一棵以自己为根的生成树,当它收到其客户的订阅请求后,就按这棵生成树将订阅消息转发给其他各事件代理,并利用过滤条件的覆盖关系对消息转发进行优化.当每个事件代理收到其客户发布的事件时,就按照各订阅消息的逆向路径将该事件即转发到其他事件代理.固定网络上的基于内容的 Pub/Sub 系统大部分都采用这种路由算法.在一些主流的基于内容的 Pub/Sub 系统中(如 Rebeca, SIENA 的无环图结构等),为了简化路由策略,各事件代理被组织成一个无环的连通网络,因此,每个代理都有一个唯一的以其为根的生成树.这种协议的优点在于各事件代理的负载较为均衡.它的缺点在于,网络中的每个节点都同时处在多棵事件分发树之中,其中包含的路由信息难以被其他节点所替代,一旦某节点失效,整个

系统的路由重配工作将很困难。

广告(advertisement)是一种被发布者用来声明、暗示有发布事件意图的过滤器。每个事件代理通过维护广告路由表来转发新增和取消的广告。订阅路由表用来路由从发布者到订阅者的通知,而广告路由表用来路由从订阅者到发布者之间的订阅,订阅仅仅被转发到与接受的活动广告重合的邻居。SIENA 和 Hermes 等系统采取了利用广告的方法,对客户发布事件的能力加以限制,要求每个可能发布事件的客户必须首先发出广告消息,声明它将发布何种类型的事件;源代理采用“基于源转发”广播算法,将广告消息转发到其他各代理;当某客户发出订阅请求时,订阅消息则根据广告消息的逆向路径,到达各个可能发布相应事件的代理。这样可以减少订阅请求中的消息转发数量,但是却额外增加了广告消息的转发数量。因此,该方法只适用于系统中能够发布事件的客户不多、客户发布的事件满足确定的约束且不常变化的场景。

## 4.2 基于P2P网络的路由算法

P2P 网络由于其分散控制、容错和自组织等优点,受到学术界和工业界的广泛重视。许多人把 P2P 技术应用在构建 Pub/Sub 系统上,以提高对节点故障和拓扑结构变化的适应能力。P2P 网络可分为非结构化 P2P 和结构化 P2P 两种:在非结构化 P2P 网络(如 Freenet, Gnutella)中,各节点被随机地组织在一起,路由协议主要是基于洪泛法,因而效率较低;而在结构化 P2P 网络(如 Pastry, Tapestry, Chord 和 CAN)中,各节点被组织成一种具有特定结构的有向图,以使得任意两个点之间的路径长度一般不超过  $\log_k(N)$ ,其中  $k$  为预设的参数,  $N$  为网络中的最大节点数量。结构化 P2P 网络的特性使得在其上构建 Pub/Sub 系统更为方便、高效,因此,已有的构建于 P2P 网络上的 Pub/Sub 系统大都是构建于结构化 P2P 之上的,其中最典型的是 Scribe 系统。

很多构建在结构化 P2P 网络上的 Pub/Sub 系统所使用的路由算法都基于 Scribe 系统中提出的思想。Scribe 是建立在 Pastry 之上的基于主题的 Pub/Sub 系统。它为每个事件主题分配一个唯一的编号,在网络中节点的编号与该主题编号最接近的节点成为该主题的汇合点。各节点在订阅事件时,将订阅消息发送到相应主题的汇合点。在订阅消息转发途中,如果当前节点已经订阅了该主题,则订阅消息不必再向前发送。汇合点根据各订阅消息的逆向路径,构建一棵以自己为根的事件分发树。各节点在发布事件时,也将它发送到相应主题的汇合点,再由汇合点通过事件分发树转发到各订阅者。为了处理事件分发树中节点失效的情况,它要求树中的每个节点都定期地向其孩子节点发送心跳消息。如果某节点在一定时期内未收到其父节点的心跳消息,它就认为父节点已失效,然后它就以其他路径重新向汇合点发送订阅消息,以修复多播树。Scribe 只提供了弱的可靠性保证,在多播树的断裂和修复期间,汇合点所发出的事件可能无法到达某些订阅者。此外,它存在树根负载较大和单点失败的问题。

## 5 发布/订阅系统的形式化描述

目前,众多的发布/订阅系统之间一般都有一些细节上的差别,而且它们的语义往往是含糊的、非形式化的,这使得对这些系统进行比较变得很困难。如何以一种严格的、无歧义的方式来定义一个 Pub/Sub 系统,对于系统的理解和验证系统中各种算法的正确性是非常必要的。

SIENA 对 Pub/Sub 系统进行了一定程度的形式化描述,它使用一阶谓词演算和集合论,描述了订阅与事件之间的匹配关系、两个订阅之间的覆盖关系等。但其形式化描述是不完整的,没有形式化地证明其提出的路由算法的正确性,没有描述 Pub/Sub 系统的动态特征,更没有从整体上描述一个 Pub/Sub 系统。文献[13]基于线性时序逻辑的语法,给出了 Pub/Sub 系统的形式化语义规范,并给出了其文中提出的几种路由算法的正确性证明,且试图系统地描述一个 Pub/Sub 系统的动态特征。其缺点在于对 Pub/Sub 系统的动态特征的描述不够完整,如未描述取消订阅动作和通知动作之间的关系,也没有考虑事件代理之间的动态行为关系。文献[28]采用集合论和有限状态机,并引入订阅者代理和发布者代理,定义了分布式移动 Pub/Sub 系统的模型。但是,该模型还需要进一步扩展来支持发布者广告机制。文献[47]提到利用进程代数的理论,对 Pub/Sub 系统的主要特征进行了形式化描述。在文献[48]中提出了一种“参数化事件代数”的方法,以描述事件通知服务中“复合事件”的语义,它通过定义事件操作符来形成复合事件,并引入了事件实例选择和消费的参数,但是它没有对 Pub/Sub 系统的基本语义进

行描述.文献[49]提出了一种通用的 Pub/Sub 模型检测框架,它采用一个可重用的、可以捕获 Pub/Sub 运行时事件管理和分发策略的参数化的状态机模型,通过转换工具来产生特定 Pub/Sub 系统的模型并映射到框架中,用模型检测工具进行检测.

形式化描述对无歧义的理解、比较和验证 Pub/Sub 系统是很重要的,但是这方面的研究还比较少.在大部分系统中,至多只是给出了订阅语言的形式化语义,而忽略了事件服务本身的语义.

## 6 发布/订阅系统的事务和安全机制

在大规模分布式计算环境下,要应用 Pub/Sub 系统,通常要跨越异构的网络和系统平台,所以必须提供服务质量保证.通常通过事务和安全机制来提供服务质量保证.

在 Pub/Sub 系统中,事务用来把成组的消息合并成一个原子单元,要么发送或者接收这个消息序列,要么一个都不要做.IBM MQSeries 通过和数据库的紧密集成提供了大范围的事务机制.TIB 和 JMS 在单个会话的上下文环境中,能为组消息提供事务支持.DREAM<sup>[31]</sup>系统采用了中间件中介事务(MMT)的方式,把发布者、中介者和通知的接收者都集成到一个灵活的事务框架中,并给出了框架的原型实现.在处理 Pub/Sub 这样多对多的异步通信方式中,具有严格的 ACID 特性的传统事务已经不适用了,需要与 Pub/Sub 通信范型相结合,提出新的扩展事务模型.另外,把面向消息的事务和分布式对象事务集成起来,仍然是一个挑战.

安全性问题是 Pub/Sub 系统大规模应用需要解决的重要问题.Wang 等人给出了 Pub/Sub 系统的安全需求<sup>[14]</sup>.Miklos 等人假定存在一个可信任的代理网络,用 SIENA 的覆盖关系来约束允许的订阅和广告,但是它要求发布者和订阅者在发布和订阅时必须出示证书<sup>[15]</sup>.文献[16]把基于角色的访问控制(RBAC)和分布式通知服务结合了起来,并扩展了 Hermes,给出了原型系统.Feige 等人采用基于 PKI 的方法来建立发布和订阅事件的访问控制机制<sup>[17]</sup>.基于内容的 Pub/Sub 系统的安全问题,包括认证、机密性、完整性和审计性等方面,一些可以采用已有的方法(如数字签名来解决信息完整性),但还有不少问题有待解决.例如,基于内容的动态路由机制就给安全技术提出了挑战,呈现出一些相互矛盾的问题,路由要基于内容而又要保持内容的机密性,审计要基于订阅同时要满足不泄漏订阅的条件.另外,综合的 Pub/Sub 系统安全体系结构的研究也需要加强.

## 7 典型系统比较

表 1 从发布/订阅系统的分类、通知模型、拓扑结构、匹配算法、路由算法、服务质量和对移动客户端及事件复合的支持等方面,对国内外知名的 Pub/Sub 系统进行了分析和比较.从表中可以看出,大部分系统对移动客户端和事件复合都不支持.路由算法都是充分利用系统网络拓扑结构或者底层 P2P 网络协议进行优化.匹配算法根据事件模型和订阅模型,提出了各种各具特色的匹配算法.服务质量方面,工业界的研究成果支持的较好,学术界对系统服务质量的支持还有待加强.

## 8 发布/订阅系统的研究热点与发展趋势

随着新的计算环境和应用提出的一些新的需求,分布式网络环境下存在着信息的结构异构和语义异构问题,从而开始了对基于语义的 Pub/Sub 系统的探索.而 P2P 技术具有容错能力高与自组织等优点,使得将 Pub/Sub 与 P2P 技术结合起来研究,以提高 Pub/Sub 系统的可靠性.而一些研究者也借鉴主动数据库技术的复合事件处理的思想,展开了对支持事件复合的 Pub/Sub 系统的研究.对系统可靠性的要求,也使得容错路由算法的研究开展起来.随着移动计算技术的快速发展,要求 Pub/Sub 系统支持移动客户端,对此,许多人进行了尝试.但这些技术都还不太成熟,正处于探索阶段,有待深入.

### 8.1 基于语义的发布/订阅系统

现有的大部分 Pub/Sub 系统在表达能力方面尚存在不足,大多根据事件的结构信息来进行事件与订阅的匹配,而缺乏对事件本身语义的理解.因为发布者和订阅者都是匿名的,经常超越网络和系统边界,一般不能用同



一种语言或者语言格式.因此,在目前的 Pub/Sub 系统中增加语义是非常重要的,从语义上进行事件与订阅的匹配,以提高匹配的准确度,同时方便订阅者定义其订阅条件.近年来,一些研究者开始研究基于语义的 Pub/Sub 系统,从而使得系统具有一定的智能性.相关的工作包括:DREAM,S-TOPSS<sup>[32]</sup>,OPS<sup>[33]</sup>和 A-ToPSS<sup>[35]</sup>等.DREAM是德国 Darmstadt 理工大学的研究成果,提出了基于概念的方法,并把系统分为通知服务层和概念层两层,利用在不同抽象层次上的基本表示本体、设施本体和领域本体来建立共享的概念,利用 MIX 模型来表示语义事件,实现对语义异构的事件的支持;多伦多大学的 Burcea 等人在文献[34]中,指出了发布/订阅系统语义自省的基于内容匹配和路由算法的研究方向,并开发了 S-TOPSS 原型系统,利用 DAML+OIL 语言,建立各属性之间的关系,利用同义词、概念层次关系和映射函数进行匹配,以支持语义异构的事件.但是,S-TOPSS 没有解决基于语义的路由问题;国内的中国科学院软件研究所引入语义 Web 技术,提出了一种基于本体的 Pub/Sub 系统 OPS,它将事件表示为 RDF 图,把订阅表示为 RDF 图模式.该系统支持关系、树型和图状数据结构等格式的事件,表达能力很强,并提出了基于图的高效匹配算法.

Table 1 Comparison of typical Pub/Sub systems

表 1 典型的 Pub/Sub 系统的比较

| System              | Classification            | Notification model | Architecture            | Matching algorithm      | Routing algorithm              | Mobile client | Event composition       | Quality of service                             | Research organization                      |
|---------------------|---------------------------|--------------------|-------------------------|-------------------------|--------------------------------|---------------|-------------------------|--|--|
| CORBA event service | Channel based             | Push, Pull         | —                       | —                       | —                              | Not support   | Not support             | —  | OMG  |
| JMS                 | Hybrid                    | Push, Pull         | Centralized/distributed | Subject matching        | Flooding                       | Some support  | Not support             | Transaction, Priority, persistence             | SUN Microsystems                           |
| TIB                 | Hybrid                    | Push               | Bus                     | Subject matching        | Flooding                       | Not support   | Not support             | Transaction, Priority                          | TIBCO Company                              |
| Gryphon             | Map based                 | Push               | Hierarchical            | Parallel search tree    | Reverse path forwarding        | Not support   | Not support             | Exactly-once delivery                          | IBM T.J. Watson Research Center            |
| Scribe              | Subject based             | Push               | Pastry                  | Lookup table            | Multicast tree                 | Not support   | Not support             | Best-effort delivery                           | Microsoft Research Division                |
| JEDI                | Map based                 | Push, Pull         | Hierarchical            | Matching tree           | Spanning tree forwarding       | Support       | Partial (time sequence) | —  | Politecnico di Milano Technical University |
| Bayuex              | Subject based             | Push               | Tapstry                 | Lookup table            | Multicast tree                 | Not support   | Not support             | Fault-Resilient delivery                       | University of California at Berkeley       |
| SIENA               | Map based                 | Push               | Hierarchical            | Binary decision diagram | Spanning tree forwarding       | support       | Partial (time sequence) | Best-effort delivery                           | University of Colorado                     |
|                     |                           |                    | Acyclic graph           |                         | Reverse path forwarding        |               |                         |  |  |
| Hermes              | XML based                 | Push               | Pastry                  | —                       | Core-Based trees               | Not support   | Partially               | Fault tolerance security, congestion control   | University of Cambridge                    |
| Rebeca              | Map based                 | Push               | Acyclic graph           | Matching tree           | Reverse path forwarding        | Support       | Support (ECA)           | Transaction, security, self-repair             | Darmstadt University of Technology         |
|                     |                           |                    | Chord                   |                         |                                |               |                         |  |  |
| OPS                 | Content based (RDF graph) | Push               | Pastry                  | RDF graph matching      | Identifier range based routing | Not support   | Not support             | Exactly-Once delivery, concurrence consistence | Institute of Software, CAS                 |

目前, Pub/Sub 系统的事件/订阅匹配属于精确匹配,即一个事件要么满足一个订阅条件,要么不满足该订阅条件.而某些应用领域可能有“近似匹配”的要求,即一个事件可能部分地与某订阅条件相匹配,并可以用一个 0~1 之间的数值来表达近似匹配的程度.Liu 等人基于模糊集理论和概率论来处理订阅和发布中的不确定性,建立了支持不确定性表达的事件模型和订阅模型,并开发了支持近似匹配的 Pub/Sub 原型系统 A-ToPSS<sup>[35]</sup>.因为许多应用场景下没有特定订阅和发布的精确知识,用户发布或订阅采用的是不精确的词汇,所以在 Pub/Sub 系统中对近似匹配的支持是有前途的方向.

在基于语义的发布/订阅系统研究方面,语义匹配研究得较多,而基于语义的路由算法研究得较少.Pub/Sub系统如何与本体论的方法结合都需要进一步加以研究.

## 8.2 P2P网络上的发布/订阅系统

P2P网络的分散控制和容错性给 Pub/Sub 系统的研究带来了机遇,但其动态特性给 Pub/Sub 系统路由算法的设计也带来了许多新的挑战.陆续出现了 Hermes, Bayeux, P2P-ToPSS<sup>[21]</sup>和 Meghdoot<sup>[23]</sup>等在 P2P 网络上构建的 Pub/Sub 系统.Hermes 与 P2P-ToPSS 都是在 Scribe 的路由机制基础上加以扩展,以构建基于内容的 Pub/Sub 系统的.其思路是:首先将其全局事件空间分成很多个子空间,将每个子空间看成一个主题;然后再用 Scribe 的方法进行订阅和事件的转发.在 P2P-ToPSS 中,一个订阅可以跨越多个主题.这些系统只能提供与 Scribe 相同级别的可靠性保证.Bayeux 是一个建立在 Tapestry 之上的基于主题的 Pub/Sub 系统,它具有良好的效率和容错性.但是,事件分发树中的每个节点都要维护一个订阅者列表,其中记录了通过本节点可以到达哪些最终订阅者,该订阅者列表往往要占用较多的空间,从而增加了每个节点的负载.Terpstra 等人在文献[22]中提出了一种“many trees”方法,采用逆向路径转发的思想,在 Chord 上构造基于内容的 Pub/Sub 系统.它也只能提供弱的可靠性,不能保证事件能到达所有的订阅者.该方法要求每个节点不仅知道其输出边指向的各节点,而且应知道哪些节点的输出边指向自己,这将极大地增加网络自组织的成本.Meghdoot 是一个构建于 CAN 之上的基于内容的 Pub/Sub 系统,它创建了一个“订阅空间”,每个订阅条件和每个事件都是该空间中的一个点.将该空间进行分区,每个节点负责一个子空间,一个事件可能与若干个订阅子空间中的订阅条件匹配成功.当一个事件被发布时,系统将其转发到与之有关的各子空间的对应节点,然后进行匹配以找到所有匹配成功的订阅者.但是, Meghdoot 未能提出一个完整的路由方案.

P2P 网络节点的动态加入和退出,使得整个网络的拓扑结构处于不断的变化之中,从而使基于静态网络拓扑结构优化消息转发的思想难以应用在 P2P 网络中.同时在 P2P 网络中,由于两个节点之间的每次消息交互可能沿着不同的路径进行,从而使得保持消息的有序性也变成了很重要的问题.如果一个节点先后执行一次订阅操作和一次取消订阅操作,有可能取消订阅的消息会先于订阅消息到达目的节点,如果不采取其他措施,就会导致整个系统的状态出现不一致.因此,正如文献[38]中指出,虽然在 P2P 网络上构建 Pub/Sub 系统可以使 Pub/Sub 系统具有良好的可重配性(代理失败或加入新代理时,系统的重配工作较为方便),但这种系统能否提供高的可靠性,还是一个未知的问题.大部分已有的系统都只能提供尽力而为级别的可靠性保证,而不能提供恰好一次级别的可靠性保证.同时,已有系统大都没有考虑到:当消息的接收顺序与发送顺序不一致时,如何处理以保证系统状态的一致性.因此,在具有动态网络拓扑结构的 P2P 网络上提供可靠的传送语义,仍然是学术界有待解决的问题.

## 8.3 支持复合事件的发布/订阅系统

有的 Pub/Sub 系统不仅支持对单个事件的订阅,而且支持对“事件模式”或“复合事件”的订阅.即:当各发布者所发布的一系列事件所组成的事件序列满足某一条件时,才通知客户,这样可以改善系统的可扩展性和性能.这一思想是从主动数据库领域中的复合事件发展过来的, Pub/Sub 系统这方面相关的研究包括文献[19,31,48,50,51]等;Schwidorski 采用了 2g-precedence 模型来处理分布式事件的顺序和复合事件的检测<sup>[50]</sup>;文献[19]提出了一种通用的复合事件监测框架,它基于扩充了时序模型的有限状态机,提供了一种复合事件描述语言,并给出了在 JMS 系统上的验证; DREAM 系统使用了基于事件代数的事件复合监测机制和基于 ECA 规则的事件处理机制<sup>[31]</sup>;文献[51]中提出了一种事件关系的语言和形式化模型及其操作语义.目前,对 Pub/Sub 系统中的复合事件时序逻辑模型研究较少,支持复合事件的、便于用户订阅的订阅语言研究也需深入.由于对复合事件进行匹配所需要的系统开销很大,目前仍未找到高效的匹配算法.

## 8.4 容错的路由算法

在基于静态网络的内容路由算法中,事件代理网络被组织成层级结构或无环图结构,其中任何一个节点或边的失败,都可能会将整个网络分割成若干个不连通的子网,因而其容错能力较弱;而如果在网络中增加一些冗

余的边,固然可以提高网络的可靠性,但是每个节点就会有不止一棵以其为根的生成树,导致复杂的系统重配工作,从而会显著增加路由算法的复杂度.因此,这些路由算法缺乏对拓扑结构变化的适应能力,难以适用于大规模分布式计算环境.

近年来,人们在路由算法的容错能力,以及当故障发生时如何重配系统,以使其恢复到正常状态等方面作了较多研究<sup>[36-40]</sup>,以提高 Pub/Sub 系统的可靠性.这些研究大都基于代理网络的无环图结构.文献[36]着重于研究当无环图中某个边失败时,如何增加边和重配系统以使其恢复连通.文献[37]研究了当无环图中的边不可靠时,采用“感染算法(epidemic algorithm)”来保障事件传输的可靠性.文献[38]在有向无环图的基础上,将系统的拓扑结构保存在数据库中,以方便出现失败时的重配工作.文献[39,40]中将代理网络组织成有向无环图,并增加一些冗余边,以提高容错能力.但是这些关于容错能力的研究成果还不能令人满意.

## 8.5 支持移动计算的发布/订阅系统

随着无线网络技术的发展和移动终端的普及,移动计算正成为重要的研究领域.近年来,很多人开始研究将 Pub/Sub 系统扩展到移动环境中,以支持移动环境下的应用<sup>[24-30]</sup>.在移动环境下,常见的应用就是对移动客户端的支持,即 Pub/Sub 系统的各事件代理都位于固定的有线网络中,而其客户端可以是手机等移动计算设备.事件代理相当于移动客户端对 Pub/Sub 系统的接入点,移动客户端可以从某个事件代理处断开连接,经过一段时间以后,在新的事件代理处重新连接到 Pub/Sub 系统.在客户端漫游的过程中, Pub/Sub 系统必须适应客户端位置的变化,要支持持久化订阅,保证事件不丢失且不重复发送,要开发能够使客户端订阅过去事件的查询语言.

JEDI 是较早的支持移动客户端的 Pub/Sub 中间件,定义了 move-out 和 move-in 函数来支持断连和重连操作,允许移动客户端从一个事件代理迁移到另一个事件代理,但是它没有解决好同步问题.CEA 利用中介,在断连期间代表订阅者接收通知.中介作为订阅者代理,能够用来注册对订阅者位置信息的兴趣,当订阅者重连到系统时,中介会得到新的订阅者位置的通知,并把队列中的通知发送给订阅者<sup>[24]</sup>.然而这也存在严重的安全问题,恶意的一方会利用中介来跟踪订阅者,危害其位置隐私,在订阅者重连后发送假的通知.SIENA 的移动服务利用客户端代理和特定的客户端库,在订阅者断连和不同的访问点之间移交期间,代表订阅者管理订阅和通知,并实现了特定的同步机制以避免通知的丢失:客户端代理为断连的订阅者保存消息到特定的队列中,并在订阅者重连时,通过新旧代理之间的移交获取断连期间的消息<sup>[25]</sup>.文献[26,27]说明了扩展 REBECA 系统,以支持移动和位置依赖的应用,它支持物理移动性和逻辑移动性两种不同类型的移动性.Podar 与上述为断连的订阅者维护通知队列的方式不同,它通过通知的持久性来支持移动客户端,代理网络在通知有效期之前保存持久通知,在订阅者重连时发送有效通知<sup>[28]</sup>.Proto 是一个 JMS 兼容的中间件系统,通过持久化订阅和通知来支持 JMS 发布者和订阅者的移动性,实现了能运行在资源受限设备上的支持移动的 JMS API<sup>[29]</sup>.

引入移动客户端对已有的 Pub/Sub 技术带来了许多新的挑战,包括切换协议、移动过程中的同步问题、位置透明性问题、位置自省和上下文自省的事件服务等,都需要进一步研究.

## 9 结束语

面向大规模分布式计算的 Pub/Sub 系统技术是一个方兴未艾的研究方向,并有着广泛的应用前景.工业界和学术界都从不同侧面进行了深入研究,并从理论和实际应用的角度推出了各种原型系统和产品.本文分析了当前 Pub/Sub 系统的研究现状,对已有的系统按照不同标准进行了分类,并针对 Pub/Sub 系统的主要研究问题,从匹配算法、路由算法、事务和安全等核心技术角度进行了全面概述,既分析了存在的问题,又指明了研究的动机,并对解决同一问题的不同方法进行了比较.就 Pub/Sub 系统的核心支撑技术而言,尚存在很多有待解决的开放问题.在 Pub/Sub 系统设计中,对表达能力、可扩展性和服务质量相互矛盾的多方的权衡,刺激了具有较好表达能力和高效率的匹配算法、高效而可靠的路由算法的研究要进行下去,尤其是自稳定的容错路由算法. Pub/Sub 系统对移动客户端和复合事件的支持的研究也有待进一步加强.当前,也出现了增强 Pub/Sub 系统的智能性研究,包括引入模糊数学的理论,加强对近似匹配的支持;引入本体,加强对语义异构的事件的支持,我们

也正在从事这方面的研究工作.

## References:

- [1] Eugster PT, Felber PA, Guerraoui R, Kermarrec AM. The many faces of publish/subscribe. *ACM Computing Surveys*, 2003,35(2): 114–131.
- [2] Gough KJ, Smith G. Efficient recognition of events in distributed systems. In: *Proc. of the 18th Australasian Computer Science Conf. Adelaide: IEEE Computer Society*, 1995.
- [3] Aguilera MK, Strom RE, Sturman DC, Astley M, Chandra TD. Matching events in a content-based subscription system. In: *Proc. of the 18th ACM Symp. on Principles of Distributed Computing. Atlanta, 1999. 53–61.*
- [4] Campailla A, Chaki S, Clarke E, Jha S, Veith H. Efficient filtering in publish-subscribe systems using binary decision diagrams. In: *Proc. of the ICSE 2001. Toronto: IEEE Computer Society*, 2001. 443–452.
- [5] Altinel M, Franklin MJ. Efficient filtering of XML documents for selective dissemination of information. In: *Proc. of the 26th Int'l Conf. on Very Large Data Bases. Cairo: Morgan Kaufmann Publishers*, 2000. 53–64.
- [6] Chan CY, Felber P, Garofalakis M, Rastogi R. Efficient filtering of XML documents with XPath expressions. *The VLDB Journal*, 2002,11(4):354–379.
- [7] Pereira J, Fabret F, Llibat F, Jacobsen HA, Shasha D. WebFilter: A high throughput XML-based publish and subscribe system. In: *Proc. of the 27th Int'l Conf. on Very Large Data Bases. Roma: Morgan Kaufmann Publishers*, 2001. 721–724.
- [8] Lakshmanan LVS, Sailaja P. On efficient matching of streaming XML documents and queries. In: *Proc. of the 8th Int'l Conf. on Extending Database Technology: Advances in Database Technology. London: Springer-Verlag*, 2002. 142–160.
- [9] Diao Y, Altinel M, Franklin MJ, Zhang H, Fischer P. Path sharing and predicate evaluation for high-performance XML filtering. *ACM Trans. on Database Systems*, 2003,28(4):467–516.
- [10] Peng F, Chawathe SS. XPath queries on streaming data. In: *Proc. of the ACM SIGMOD Int'l Conf. on Management of Data. New York: ACM Press*, 2003. 431–442.
- [11] Carzaniga A, Rosenblum DS, Wolf AL. Design and evaluation of a wide-area event notification service. *ACM Trans. on Computer Systems*, 2001,19(3):332–383.
- [12] Cugola G, Nitto ED, Fuggetta A. The JEDI event-based infrastructure and its application to the development of the OPSS WFMS. *IEEE Trans. on Software Engineering*, 2001,27(9):827–850.
- [13] Mühl G. Large-Scale content-based publish/subscribe systems [Ph.D. Thesis]. Darmstadt University of Technology, 2002.
- [14] Wang C, Carzaniga A, Evans D, Wolf AL. Security issues and requirements for Internet-scale publish-subscribe systems. In: *Proc. of the 35th Hawaii Int'l Conf. on System Sciences. Washington: IEEE Computer Society*, 2002. 303–310.
- [15] Miklós Z. Towards an access control mechanism for wide-area publish/subscribe systems. In: *Proc. of the 22nd Int'l Conf. on Distributed Computing Systems, Workshops. Washington: IEEE Computer Society, IEEE Press*, 2002. 516–524.
- [16] Belokoztolszki A, Eyers DM, Pietzuch PR. Role-Based access control for publish/subscribe middleware architectures. In: Jacobsen HA, ed. *Proc. of the 2nd Int'l Workshop on Distributed Event-Based Systems. New York: ACM Press*, 2003.
- [17] Fiege L, Zeidler A, Buchmann A, Kilian-Kehr R, Mühl G. Security aspects in publish/subscribe systems. In: *Proc. of the 3rd Int'l Workshop on Distributed Event-Based Systems. Edinburgh: IEEE Computer Society*, 2004.
- [18] Rowstron A, Kermarrec AM, Castro M, Druschel P. SCRIBE: The design of a large-scale event notification infrastructure. In: *Proc. of the 3rd Int'l Workshop on Networked Group Communication. London: Springer-Verlag*, 2001. 30–43.
- [19] Pietzuch PR. Hermes: A scalable event-based middleware [Ph.D. Thesis]. University of Cambridge, 2004.
- [20] Zhuang SQ, Zhao BY, Joseph AD, Katz RH, Kubiawicz J. Bayeux: An architecture for scalable and fault-tolerant wide-area data dissemination. In: *Proc. of the 11th NOSSDAV. New York: ACM Press*, 2001. 11–20.
- [21] Tam D, Azimi R, Jacobsen HA. Building content-based publish/subscribe systems with distributed hash tables. In: *Proc. of the 1st Int'l Workshop On Databases, Information Systems and Peer-to-Peer Computing. Berlin: Springer-Verlag*, 2003. 138–152.
- [22] Terpstra WW, Behnel S, Fiege L, Zeidler A, Buchmann AP. A peer-to-peer approach to content-based publish/subscribe. In: Jacobsen HA, ed. *Proc. of the 2nd int'l workshop on Distributed Event-Based Systems. New York: ACM Press*, 2003. 1–8.
- [23] Gupta A, Sahin OD, Agrawal D, Abbadı AE. Meghdoot: Content-Based publish/subscribe over P2P networks. In: Jacobsen HA, ed. *Proc. of the 5th ACM/IFIP/USENIX Int'l Middleware Conf. LNCS 3231, Toroto: Springer-Verlag*, 2004. 254–273.

- [24] Bacon J, Moody K, Bates J, Hayton R, Ma C, McNeil A, Seidel O, Spiteri M. Generic support for distributed applications. *IEEE Computer*, 2000,33(3):68–76.
- [25] Caporuscio M, Carzaniga A, Wolf AL. Design and evaluation of a support service for mobile, wireless publish/subscribe applications. *IEEE Trans. on Software Engineering*, 2003,29(12):1059–1071.
- [26] Zeidler. A distributed publish/subscribe notification service for pervasive environments [Ph.D. Thesis]. Damastadt University of Technology, 2004.
- [27] Fiege L, Gärtner FC, Kasten O, Zeidler A. Supporting mobility in content-based publish/subscribe middleware. In: Endler M, Schmidt DC, eds. *Proc. of the 4th ACM/IFIP/USENIX Int'l Middleware Conf. Rio de Janeiro*: Springer-Verlag, 2003. 103–122.
- [28] Podnar I. Service architecture for content dissemination to mobile users [Ph.D. Thesis]. University of Zagreb, 2004.
- [29] Yoneki E, Bacon J. Pronto: MobileGateway with publish-subscribe paradigm over wireless network. Technical Report, UCAM-CL-TR-559. Cambridge: University of Cambridge, 2003.
- [30] Burcea I, Jacobsen HA, DeLara E, Muthusam V, Petrovic M. Disconnected operation in publish/subscribe middleware. In: *Proc. of the 5th IEEE Int'l Conf. on Mobile Data Management. Berkeley*: IEEE Computer Society, 2004. 39–51.
- [31] Buchmann A, Bornhövd C, Cilia M, Fiege L, Gärtner F, Liebig C, Meixner M, Mühl G. DREAM: Distributed reliable event-based application management. In: Levene M, Poulouvasilis A, eds. *Web Dynamics. Springer-Verlag*, 2004. 319–352.
- [32] Petrovic M, Burcea I, Jacobsen HA. S-ToPSS: Semantic Toronto publish/subscribe system. In: *Proc. of the 29th Int'l Conf. on Very Large Databases. Berlin*: Morgan Kaufmann Publishers, 2003. 1101–1104.
- [33] Wang JL, Jin BH, Li J. An ontology-based publish/subscribe system. In: Jacobsen HA, ed. *Proc. of the 5th ACM/IFIP/USENIX Int'l Middleware Conf. LNCS 3231, Toronto*: Springer-Verlag, 2004. 232–253.
- [34] Burcea I, Petrovic M, Jacobsen HA. I know what you mean: semantic issues in Internet-scale publish/subscribe systems. In: Cruz IF, Kashyap V, Decker S, Eckstein R, eds. *Proc. of the SWDB 2003. Berlin*: Morgan Kaufmann Publishers, 2003. 51–62.
- [35] Liu H, Jacobsen HA. A-ToPSS: A publish/subscribe system supporting imperfect information processing. In: *Proc. of the 30th Int'l Conf. on Very Large Databases. Toronto*: Morgan Kaufmann Publishers, 2004. 1281–1284.
- [36] Cugola G, Picco GP, Murphy AL. Towards dynamic reconfiguration of distributed publish-subscribe middleware. In: *Proc. of the 3rd Int'l Workshop on Software Engineering and Middleware. Orlanda*: Springer-Verlag, 2002. 187–202.
- [37] Costa P, Migliavacca M, Picco GP, Cugola G. Epidemic algorithms for reliable content-based publish-subscribe: An evaluation. In: *Proc. of the ICDCS 2004. Tokyo*: IEEE Computer Society, 2004. 552–561.
- [38] Bholra S. Topology changes in a reliable publish/subscribe system. Technical Report, RC23354. Yorktown Heights: IBM Thomas J. Watson Research Center, 2004.
- [39] Bholra S, Strom R, Bagchi S, Zhao Y, Auerbach J. Exactly-Once delivery in a content-based publish-subscribe system. In: *Proc. of the Int'l Conf. on Dependable Systems and Networks. Bethesda*: IEEE Computer Society, 2002. 7–16.
- [40] Chand R, Felber PA. A scalable protocol for content-based routing in overlay networks. In: *Proc. of the 2nd IEEE Int'l Symp. on Network Computing and Applications. Cambridge*: IEEE CS Press, 2003. 123–130.
- [41] Sun Microsystems Inc. JMS specification version 1.1, 2002. URL: <http://java.sun.com/products/jms>
- [42] TIBCO Corp. TIB/Rendezvous White Paper, 2000. URL: [http://www.tibco.com/software/enterprise\\_backbone/rendezvous.jsp](http://www.tibco.com/software/enterprise_backbone/rendezvous.jsp)
- [43] OMG. CORBA event service specification version 1.1, 2001. URL: <http://www.omg.org/corba>
- [44] Eugster PT, Guerraoui R, Sventek J. Type-Based publish/subscribe. Technical Report, DSC ID 200029, Lausanne: Swiss Federal Institute of Technology, 2000.
- [45] OMG. CORBA notification service specification version 1.0.1, 2002. URL: <http://www.omg.org/corba>
- [46] Dalal YK, Metcalfe R. Reverse path forwarding of broadcast packets. *Communications of the ACM*, 1978,21(12):1040–1048.
- [47] Baldoni R, Contenti M, Piergiorganni ST, Virgillito A. Modeling publish/subscribe communication systems: Towards a formal approach. In: *Proc. of the 8th IEEE Int'l Workshop on Object-Oriented Real-Time Dependable Systems. Guadalajara*: IEEE Computer Society, 2003. 304–311.
- [48] Hinze A. A-MEDIAS: Concept and design of an adaptive integrating event notification service [Ph.D. Thesis]. Freie Universität Berlin, 2003.
- [49] Garlan D, Khersonsky S, Kim JS. Model checking publish-subscribe systems. In: Ball T, Rajamani SK, eds. *Proc. of the 10th SPIN Workshop: Model Checking Software. Heidelberg*: Springer-Verlag, 2003. 166–180.
- [50] Schwiderski S. Monitoring the behaviour of distributed systems [Ph.D. Thesis]. University of Cambridge, 1996.

[51] Sanches C, Sankaranarayanan S, Sipma H, Zhang T, Dill D, Manna Z. Event correlation: Language and semantics. In: Proc. of the EMSOFT 2003. LNCS 2855, Philadelphia: Springer-Verlag, 2003. 323-339.



马建刚(1977—),男,河南郑州人,博士生,主要研究领域为分布式计算,中间件技术。



徐罡(1973—),男,博士,主要研究领域为分布式计算,企业应用集成。



黄涛(1965—),男,博士,研究员,博士生导师,CCF 高级会员,主要研究领域为软件工程,分布式计算。



叶丹(1971—),女,博士,副研究员,CCF 高级会员,主要研究领域为软件工程,分布式计算,虚拟企业。



汪锦岭(1974—),男,博士生,主要研究领域为分布式计算,中间件技术。



## 《软件学报》软件体系结构专刊 征文通知

专刊题目: 软件体系结构

特约编辑: 梅宏(北京大学)、吕建(南京大学)

### 一、征文范围

1. 软件体系结构的基本理论、方法和技术,如软件体系结构描述语言、软件体系结构形式化方法、软件体系结构模型、动态体系结构、软件体系结构风格等。
2. 基于软件体系结构的软件开发,如模型驱动体系结构、软件体系结构分析和评估方法、软件体系结构与服务质量关系、体系结构发现与恢复、基于体系结构的演化技术、体系结构转换、从需求到软件体系结构到系统实现的可追踪性问题等。
3. 软件体系结构与其他软件工程领域的交叉研究,如产品线体系结构、系统体系结构和软件体系结构的关系、领域特定的体系结构和参考体系结构、软件体系结构与面向对象技术、基于构件的软件开发、面向服务的体系结构等。
4. 软件体系结构与软件产业,如实例研究、工业界的最佳实践、软件架构师的角色和责任、互操作性和集成、软件架构师的教育、培训和认证、文化、经济和管理问题,软件体系结构对于提高中国软件产业生产力的作用等。

上述各项仅属举例性的,也非优选题目。论文内容并不仅限于此,所有与软件体系结构相关的内容均可投稿。

### 二、投稿要求

1. 投稿方式: 采用“软件学报在线投稿系统”(http://www.jos.org.cn)投稿。请在投稿时,在备注栏中注明“软件体系结构专刊投稿”字样。
2. 稿件格式: 参照《软件学报》论文格式(学报网站上提供了论文模版,可下载)。
3. 投稿文章未在正式出版物上发表过,也不在其他刊物或会议的审稿过程中,不存在一稿多投现象;保证投稿文章的合法性(无抄袭、剽窃、侵权等不良行为)。
4. 其他投稿须知请参阅《软件学报》投稿指南 <http://www.jos.org.cn/directory.htm>
5. 投稿作者需提交投稿声明;专刊投稿文章不收审理费。录用刊发文章将收取软件学报标准版面费。发表之后,将按软件学报标准支付稿酬,并赠送样刊及单行本。

### 三、重要时间

截稿日期: 2006年1月10日

录用通知发出时间: 2006年3月10日

录用修改稿提交日期: 2006年4月10日

出版日期: 2006年第6期