

基于 JMS 的消息过滤改进算法

姚 刚, 邓江沙

(长沙理工大学 计算机与通信工程学院, 湖南 长沙 410076)

摘 要:发布/订阅模式是 JMS 中一种多对多的消息传递模式,它允许客户使用消息选择器在主题中注册感兴趣的内容, JMS 提供者负责对消息的过滤和传递。文中基于 JMS 规范,介绍了一种改进的消息选择器处理算法,通过分析传统方法在性能上的不足,描述了这种消息过滤机制。

关键词:发布/订阅;JMS;消息选择器;消息过滤

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2006)07-0091-03

An Improved Message Filtering Algorithm in JMS

YAO Gang, DENG Jiang-sha

(College of Computer and Communication, Changsha University of Science and Technology, Changsha 410076, China)

Abstract: The publisher/subscriber model is a many-to-many message delivery model in JMS. It allows a client using message selectors to register in the topic it's interested in. The JMS provider is responsible for the message filtering and delivery. This paper introduces an improved message selector processing algorithm based on JMS, and describes the message filtering mechanism through analyzing the performance insufficiency in traditional method.

Key words: publisher/subscriber; Java message service; message selector; message filtering

0 引言

消息中间件现在被广泛用于实现分布式系统中可靠的、高效的、实时的跨平台数据传输,它具有高效可靠的消息传递机制,用来屏蔽掉各种平台及协议之间的特性,实现在不同平台之间通信。消息中间件的发布/订阅消息传递模式为消息的提供者 and 消费者提供了一种异步的、松耦合的通信方式。JMS(Java message service)是 Sun 提出的旨在统一各种消息中间件系统接口的规范,它给出了标准的基于消息中间件的主要访问方法,支持各种同步和异步通信机制。JMS 定义了一组接口和相应语义的规范,提供了诸如持久、验证和事务的消息服务,它最主要的目的是允许 Java 应用程序访问现有的消息中间件。现在 JMS 已成为 J2EE 平台的一个组成部分,许多消息中间件厂商通过实现 JMS 接口,为企业提供一个完整的消息传递功能^[1~3]。

文中主要分析了 JMS 中发布/订阅模式下的消息处理机制,并介绍了一种改进的消息过滤算法,它的主要思想就是在众多的信息中提取出它们相同的部分,减少对重复信息的处理^[4,5],从而提高 JMS 提供者对消息的处理效率和性能。

1 JMS 中的发布/订阅消息传递模式

JMS 定义了两种消息传递模式:点对点和发布/订阅模式。在点对点模式中,应用程序围绕着消息队列、发送者和接收者的概念而创建,每一个消息都包含具体队列的地址,接受客户从为放置发送给他们的消息而建立的队列中提取消息。发布/订阅模式下,发布者把消息提交给一个主题,接收者向主题订阅感兴趣的内容。主题是由 JMS 提供者提供并管理的逻辑地址,客户用来指定它的消息的目标和它消费的消息地来源的对象,发布者和订阅者一般是匿名的,可以动态地发布或订阅内容,系统负责把发布者发送给一个主题的消息分发给该主题的多个订阅者。发布/订阅模式特别适用于将一个消息发送多个用户的情形,它比点到点的模式更加节省资源,不需要建立多个连接,系统的结构也更加清晰、简单。

2 JMS 中消息选择器的处理

消息是消息系统重要的一个组成部分,在 JMS 中消息由消息头、属性和消息体三个部分组成。消息头包含了客户端和提供者用于标识和路由由消息使用的值,所有的消息头都支持相同的头字段;每一个消息都包含了一个内置的功能来支持应用程序定义的属性值,属性为支持应用程序定义的消息过滤提供了一个有效的机制;消息体是消息的具体内容,在 JMS 中定义了 stream、map 等类型的消息体。

收稿日期:2005-10-11

作者简介:姚 刚(1978-),男,湖北公安人,硕士研究生,研究方向为计算机网络应用技术;邓江沙,硕士生导师,副教授,研究方向为网络应用技术、面向对象技术。

在 JMS 中可以使用消息选择器来指定消费者感兴趣的₁消息,消息选择器允许 JMS 消息消费者从一个特定的主题接收消息时更加具有选择性,它是一个包含条件表达式的 String,以消息中的头字段和属性值作为标准,客户通过头字段引用和属性引用指定它感兴趣的₂消息。消息选择器的语法是以 SQL92 条件表达式语法的一个子集为基础的,可以包含字面量、标识符、表达式等。在创建消息消费者时,可以指定一个消息选择器作为其参数,然后消息消费者只接收其消息头和属性与选择器相匹配的那些消息。如果消息选择器的值是一个空串,那么表明消息消费者没有消息选择器。

下面这个例子中显示对消息设置了 username 属性:

```
TextMessage message = session.createTextMessage();
message.setText(text);
message.setStringProperty("username", username);
publisher.publish(message);
```

这时消费者就可以使用这个属性对消息进行过滤,消息选择器再创建消费者声明。

```
TopicSubscriber subscriber =
session.createSubscriber(topic, "username = 'Tom'", false);
```

消费者通知 JMS 服务器只将 username 属性为 Tom 的消息传递给它。

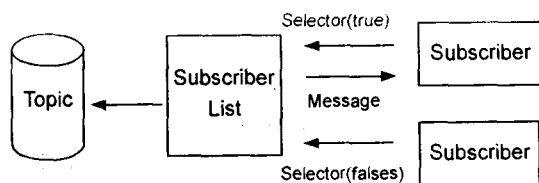


图 1 消息选择器的处理

当一个消息的订阅者将向一个主题提交请求时,它会指定一个消息选择器,表明自己所感兴趣的₃消息内容, JMS 服务器则对每条消息进行处理,浏览所有订阅者的消息选择器,检查它们是否与消息的相关属性匹配,当结果为真时,就将这条消息传递给订阅者。

3 改进的消息过滤算法

由于消息选择器把过滤消息的任务分配给 JMS 提供者,而不是给应用程序, JMS 服务器需要对每个订阅者的消息选择器进行检查,这样处理消息所花费的代价就与订阅者的数量、每个消息选择器处理所耗的时间密切相关。而在实际的消息系统中,针对一个主题会有大量的消息发布和订阅,逐个地匹配消息选择器需要占用大量的系统资源,从而影响了整个系统的性能。

基于对提高消息系统的处理能力和运行效率的考虑,对消息选择器的处理算法进行改进。在实际消息系统中,针对一个主题的消息订阅者其实有很多相同的请求内容,也就是它们的消息选择器中有很多共同的部分,而 JMS 服务器对消息选择器的处理是单独的、分离的,没有考虑之前已经做过的工作和其它选择器的内容,这样就做了很

多重复的匹配工作。下面是 3 个消息选择器的条件表达式:

- 1) Author = "Tom" and Publisher = "O'Reilly"
- 2) Author = "Jack" or Price < 100
- 3) Author = "Jack" and Publisher = "O'Reilly" and Type = "Computer"

这 3 个条件表达式中,就有一些相同的部分,如 Publisher = "O'Reilly", Author = "Jack", 消息选择器的求值次序在同一优先级内是由左到右的,假如正在处理的这条消息在匹配选择器 1 时, Publisher = "O'Reilly" 的结果为 false,由此可判断消息选择器 3 也不成立,因为它也包含有 Publisher = "O'Reilly" 这部分信息。而 JMS 服务器在对选择器 3 进行匹配时,会按从左至右的顺序依次进行匹配,遇到 Publisher = "O'Reilly" 这部分时,会重新匹配一次,浪费了之前匹配过的已知的处理结果。

本算法的基本思想就是找出消息选择器中相同的信息,尽量减少对重复单元的处理。

首先将 Selector 的表达式分解,以便于提取它们公共的元素,并且新增一个 cache unit 结构保存这些公共的单元。一个新的选择器添加到一个主题的订阅列表中时,例如:

Author = "Tom" and Publisher = "O'Reilly"

将它分解成 2 个单元:

Unit 1: Author = "Tom"

Unit 2: Publisher = "O'Reilly"

检查这些单元与其它的消息选择器是否有相同的部分,若有,则将这个单元添加到 cache unit 结构中,当所有的选择器添加完成时,就生成了一个公共的消息选择器,它包含了这个主题订阅列表中所有消息选择器中相同的信息单元。 JMS 服务器先对这个消息选择器进行匹配,并将每个单元的匹配结果存储到 cache unit 结构中(如图 2 所示):

Expression	Result
Author = "Tom"	True
Publisher = "O'Reilly"	False
Price < 100	True
⋮	⋮

图 2 Cache unit 结构

JMS 服务器对主题列表中的每个消息选择器依次进行处理,在匹配每个选择器时,首先查找在 cache unit 中是否存在对应的单元,若有,则直接取出已经匹配的结果,若没有,就提交到服务器进行匹配处理。

4 算法的性能评估

改进的消息过滤算法通过提取主题列表中消息选择器中相同的单元,节省了 JMS 提供者匹配重复信息的代价,但同时₄对消息选择表达式的分解、对消息选择器列表

中公共单元的提取又增加了系统资源的耗费。

这样,比较JMS服务器匹配消息选择器的代价和查找选择器公共信息所费的代价就显得十分重要。

消息选择器中条件表达式的类型和复杂程度是影响JMS服务器所耗代价最重要的一个因素。如 Author = "Tom", 处理这个选择器所耗的代价就很低,而如 Title LIKE "% network %", JMS服务器在处理这种类型的消息选择器时所花费的时间就要多一些。而像 Abstract LIKE "% JMS %" and Publisher = "O'Reilly", 表达式的复杂度比前两个都要高,在JMS对它进行匹配时,由于Abstract中的数据内容一般比Title要多,JMS服务器进行处理时,对关键字查找所花费的时间也就加大了整个消息处理过程的代价。

通过比较传统的消息系统与采用改进过滤算法的消息系统,可以发现,在消息订阅者数量不多、消息选择器复杂度不高的情况下,两种消息过滤机制的差异性不大;当消息订阅者数量众多,并且加大了消息选择器的复杂度的情况下,基于改进过滤算法的消息系统具有很明显的优势,它能明显地提高JMS服务器处理消息的速度,改善整个消息系统的效率和性能,如图3所示。

5 小结

改进的消息过滤算法在匹配消息选择器时,采用类似缓存的机制,避免了重复信息的处理,提高了服务器处理的速度,具有更高的适用性,可以大大地改善消息系统的性能。改进的消息过滤算法还不能针对消息的具体内容

进行处理,而基于内容的消息订阅是消息系统中的一个重要研究方向,这也是日后需要进一步开展的工作。

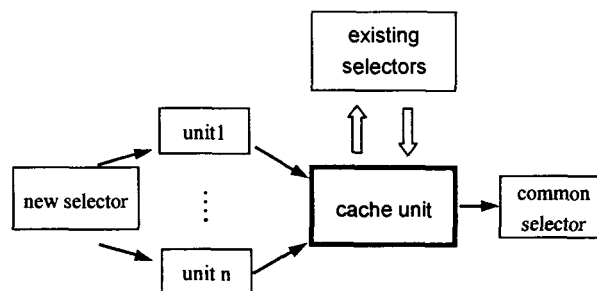


图3 改进的消息选择器处理模型

参考文献:

- [1] Sun Microsystems. Java Message Service Specification [EB/OL]. <http://java.sun.com/products/jms/docs.html>, 2002.
- [2] Hapner M, Burrige R. Java 消息服务 API 参考指南 [M]. 康博译. 北京:清华大学出版, 2002.
- [3] Monson-Haefel R, Chappell D A. Java Message Service [M]. Sebastopol: O'Reilly, 2001.
- [4] Aguilera M K, Strom R E, Sturman D C, et al. Matching events in a content-based subscription system [A]. In Eighteenth ACM Symposium on Principles of Distributed Computing [C]. Georgia, Atlanta: 1999. 53-61.
- [5] Gorton I, Almquist J, Cramer N, et al. An Efficient, Scalable Content-Based Messaging System [A]. In Proc The 7th IEEE International Enterprise Distributed Object Computing Conference [C]. Brisbane: [s. n.], 2003. 278-285.

(上接第90页)

* 可靠的消息传送,保证消息发送到目的地,有并且仅有一次。

* 消息存储在数据库中和文件系统中。

* 提供错误处理机制、日志备份、系统监控、系统恢复等系统维护功能。

信息交换平台实现了富日物流和主要客户之间的平等电子交易,提高了双方合作的效率、准确率和实时率,给双方企业带来了不少好的效益。首先,由于企业之间采用电子票据进行商务往来,减少了纸张和传真的费用,以及相应的管理费用。其次,富日物流可以及时地对各种电子票据,如订单、发票、合同等进行统计,并作出可行的决策和计划。最后,企业之间商务往来的电子化,提高了企业的生产力,增强了企业的竞争力,并使人为错误降到相对较低的水平。

4 结论

富日物流开发信息交换平台给客户使用,客户只需进行简单的配置即可使用,如需与应用交互,则在企业应用中增加相应的数据发送和接收功能即可,或者提供开放源码的API,让对方使用API开发信息交换平台和应用系

统。这种方式实现了真正意义上的平等交易,是未来电子商务发展的方向。但是在研究的过程中,也遇到了很多问题:首先,没有政府和相关权威行业协会的支持,构建注册表具有一定的难度。其次,EDI不可能被全面取代,研究如何把XML, ebXML和EDI(还包括现有的其它标准如UDDI等)相互集成,取长补短,这将是下一步研究的目标。

参考文献:

- [1] Webber D. The Benefits of ebXML for e-Business [EB/OL]. <http://www.idealliance.org/proceedings/xml04/papers/44/webber.pdf>, 2004-12.
- [2] Builder.com. ebXML: 一种商业协作发展方向 [EB/OL]. <http://zdnet.com.cn/developer/tech/story/>, 2004-04.
- [3] van der Eijk P. The ebXML Messaging Service [EB/OL]. <http://Webservices.xml.com./pub>, 2003.
- [4] OASIS ebXML Messaging Services Technical Committee. Message Service Specification Version 2.0 [EB/OL]. <http://www.ebxml.org/specs/ebMS2.pdf>, 2002.
- [5] ebXML Technical Architecture project Team. ebXML Technical Architecture Specification v1.0.4 [EB/OL]. <http://www.ebxml.org/specs/ebTA.pdf>, 2001.