

# 大规模分布式虚拟环境中一种基于事件模型的快速路由机制<sup>\*</sup>

陈继明<sup>1,2</sup> 徐 丹<sup>2</sup> 鞠时光<sup>1</sup> 潘金贵<sup>2</sup>

(江苏大学计算机科学与通信工程学院 镇江 212013)<sup>1</sup>

(南京大学计算机软件新技术国家重点实验室 南京 210093)<sup>2</sup>

**摘 要** 在大规模分布式虚拟环境中,大量的信息交互成为影响系统可扩展性的重要因素。事件模型作为新型的通信模型,以其具有异步通信、动态性等特点在分布式系统通信中的作用日益明显。本文借鉴了主动兴趣管理思想,对基于事件模型的事件路由和事件传递技术进行了研究,提出了一种利用 Filter Index 结构来实现事件快速匹配及转发的路由机制。实验结果表明,该机制是可行和有效的。

**关键词** 分布式虚拟环境,事件模型,主动兴趣管理,Filter Index

## A Fast Routing Mechanism Based on Event Model in Large-scale Distributed Virtual Environment

CHEN Ji-Ming<sup>1,2</sup> XU Dan<sup>2</sup> JU Shi-Guang<sup>1</sup> PAN Jin-Gui<sup>2</sup>

(School of Computer Science and Telecommunications Engineering, Jiangsu University, Zhenjiang 212013)<sup>1</sup>

(State Key Lab for Novel Software Technology, Nanjing University, Nanjing 210093)<sup>2</sup>

**Abstract** Recently, the increasing information communication in large-scale distributed virtual environment restricts the scalability of systems. As a new communication model, event model has plays an important role in distributed communication system because of its asynchronous and dynamic characteristics. In the paper, combined with active interest management, a research of event-model-based event routing and event forwarding is introduced. Moreover, a new routing mechanism, which implements fast event matching and delivery by using Filter Index structure, is presented, and also proved to be feasible and effective by experiment results.

**Keywords** Distributed virtual environment, Event model, Active interest management, Filter index

分布式虚拟环境<sup>[1]</sup> (Distributed Virtual Environment, DVE)提供了一个共享虚拟空间,可以通过计算机网络支持对象的仿真以及多个分布用户的交互和通讯。尽管近年来网络速度和主机处理能力在不断提高,但当虚拟环境中存在大量的用户和对象时,即使每个用户只需要少量带宽,成千上万的用户通讯需求累加起来也必将超过网络所能承受的限度,因此对系统的通信模型提出了更高要求,以降低整个虚拟环境的带宽需求和网络通讯量。目前,分布式虚拟环境设计中主要采用的组播<sup>[2,3]</sup>、对等单播(peer-to-peer unicast)<sup>[4]</sup>以及 Client/Server<sup>[5]</sup>等通信模型,大都属于一对一的同步通信模型,显然已经不能满足大规模分布式虚拟环境的通信需求。事件模型<sup>[6]</sup>作为一种新型的通信模型,具有多对多、异步通信、动态性和可匿名等特点,成为大规模分布式虚拟环境中一种可行的通信模型的解决方案。

大规模分布式虚拟环境中的状态是实时动态更新的,基于订购和发布的事件都在不断地更新和发送,需要事件模型能够实时动态地处理和传递这些事件,因此大规模分布式虚拟环境中事件模型的设计不仅需要能够实现分布式虚拟环境的主动兴趣管理技术,提供对称的发布/订购模式,同时需要可动态扩展和配置,满足实时动态需求。目前,事件模型一般采用扩展性较强的多层代理结构<sup>[7]</sup>,在此结构中各个节点的

订购信息决定了事件的传递方向,订购信息就相当于分布式系统中用来寻址的路由信息,而每个代理节点则像路由器一样负责传递和转发事件。本文通过分布式虚拟环境中的事件路由(Event Routing)和事件传递(Event Forwarding)的研究,提出了一种适合于大规模分布式虚拟环境的路由机制,并通过实验表明该机制是可行和有效的。

## 1 事件路由

在分布式虚拟环境中,每个用户有各自感兴趣的内容,它们之间没有相关性,也不按照地域分布,完全是分散和动态的。这就给路由交换带来了难题,不能像互联网上的路由结构一样按照固定的 IP 地址组织成树形结构。从事件提供者的角度考虑,在一个由单个事件提供者、多个事件消费者和多个中间代理节点组成的拓扑结构中,以事件提供者作为根、事件消费者为叶子节点、代理节点作为中间节点构成一个树形结构。以这样的树形结构为依托,事件路由从叶节点向根节点方向收集。树中的每个中间节点接受子节点的订购,将子节点的订购合并后向上发送给父节点,直到根节点为止。根据兴趣管理<sup>[8]</sup>的基本思想即每个对象都有一个感兴趣的范围,且该对象只与位于其兴趣范围内的其他对象进行通讯,而不相关的对象和数据则被过滤。图 1 (左)描述了多个事件消费

<sup>\*</sup> 本课题得到国家自然科学基金项目(60473113)、国家自然科学基金重点项目(60533080)、江苏省高校自然科学基金指导性计划项目(05KJD520051)资助。陈继明 博士研究生,主要研究领域为 XML、分布式虚拟环境;徐 丹 硕士研究生,主要研究领域为虚拟环境、Agent 技术;鞠时光 教授,博士生导师,主要研究领域为数据库、XML;潘金贵 教授,博士生导师,主要研究领域为多媒体信息处理、多媒体远程教育系统。

者对于一个事件提供者的订购,其中内容空间表示为一个大的正方形,深色的小方格表示所感兴趣的区域,订购区域从下往上传递直到事件提供者,中间的节点则将从下游节点接收到的订购合并后向上传递;图1(右)表示了事件提供者产生事件后的向下传递过程,每个代理节点将接收的事件与各个接口的订购区域比较,把事件发往订购匹配的接口。

在只有单个事件提供者的情况下,有着明确的订购传递方向。但是,在分布式虚拟环境中,每个节点同时是事件提供者和事件消费者,因此每个节点需要向各个方向传递订购信息。通过订购的交换,每个节点的每个接口都维护了此接口方向的订购且这样的订购交换是完全分布式的。路由策略的选定与系统采用的拓扑结构是相关的,不同的拓扑结构需要

不同的路由策略。我们采用了一种改进的 CBT (Core Based Tree)的拓扑结构<sup>[9]</sup>,通过静态配置每个中间节点维护了子节点和父节点的订购区域,每个节点需要把路由信息向上游以及向其他子节点方向扩散。节点之间的订购区域的交换和扩散则采用以下的路由策略,即每个节点除了维护一个路由表外,还维护了一个上游订购裁剪区域和下游订购区域,上游订购裁剪区域表示父节点接口的订购信息,下游订购区域表示所有子节点的订购区域的总和。在事件路由中采用基于 CBT 的双向共享组播树的主动兴趣管理技术,可以进一步减少使用的组播地址,并使组播树更加便于管理,甚至可以在不影响运行的情况下动态扩展,极大地简化了路由交换的策略。

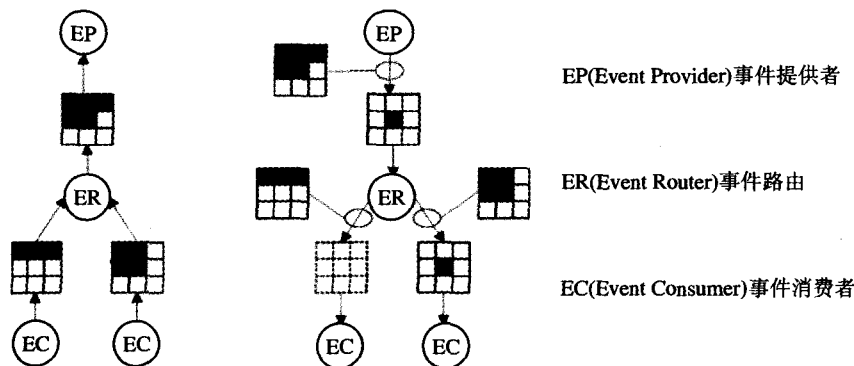


图1 单个事件提供者的订购传递和事件过滤

## 2 事件传递

事件传递决定了事件如何从事件提供者传递至每个事件订购者,以及在中间节点上收到的事件需要向哪些接口发送。事件传递建立在实现事件路由的基础上,事件传递需要根据节点的路由信息决定。在分布式事件模型的结构中,采用主动兴趣过滤<sup>[10]</sup>(Active Interest Filtering, AIF)技术来实现事件传递,即每个中间节点相当于提供了一个过滤器(Filter)和分发器(Dispatcher),对接收的事件根据路由表中各个接口的订购区域进行过滤和分发。其中每个代理节点有多个接口,每个接口表示该代理节点与某个客户端或者是其他代理节点建立了连接。在每个代理节点上,维护了一张路由表,其中记录了各个接口的订购区域。当事件服务器接收到来自一个接口的事件时,将事件中包含的发布区域与路由表中的所有订购区域进行匹配,决定事件的传递方向。

事件路由建立时每个代理节点都有一个局部的路由表,路由表是根据从各个接口收集的订购信息形成的订购与接口的映射关系。因为在每个节点之间已经建立了事件路由,即已经为事件的传递提供了寻址依据,所以只需在路由节点上实现事件的匹配和转发即可。在 CBT 结构中,每个节点将接收到的事件的发布区域分别与父节点和各个子节点的订购区域进行匹配,然后向匹配的接口发送事件的拷贝。由于在这种树形结构中没有回路,因此不会导致一个节点收到重复的事件,在同一条路径上同一事件不会发送多份拷贝,像应用层组播一样可以减少多余的通信量。

在基于主动兴趣管理的事件传递过程中,主机通过直接相连的路由器向系统发布事件  $E$ ,路由器则负责对事件  $E$  进行路由转发,并最终分发给合适的事件订购者。假设某个事件  $E_i$  通过  $VIF_k$  到达拥有  $m+1$  个虚拟接口的路由器  $ER_i$  ( $0 \leq i \leq n$ ),对于  $ER_i$  的某个虚拟接口  $VIF_j$  ( $0 \leq j \leq m$ ),如将该端口的订购区域为  $ES(VIF_j)$ ,将通过  $VIF_j$  转发事件的发布区域记作  $PR(E_i)$ 。对于是否通过其余的  $m$  个  $VIF$  转发,则根据  $E_i$  和  $VIF_j$  之间的匹配函数  $match: E_i \times VIF_j \rightarrow \{ture, false\}$  来进行计算,即当  $PR(E_i) \cap ES(VIF_j)$  的运算结果为  $ture$  则进行转发,否则不转发。当所有的虚拟接口都不需要进行转发,则该事件被丢弃。图2描述了事件  $E_i$  在路由节点上匹配和转发的过程。

图2描述了事件  $E_i$  在路由节点上匹配和转发的过程。

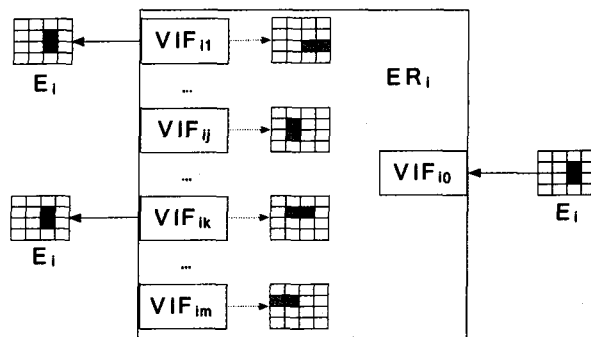


图2 事件  $E_i$  的匹配及转发过程

## 3 快速路由机制

事件传递的目标是实时性和准确性,要求用最短的时间将事件传给正确的接收者。由于在每个中间节点都需要根据路由信息对事件进行匹配然后传递,经过多个节点的累积,将带来很大的延迟,因此需要尽量的减少在每个中间节点的处理时间。我们注意到当事件  $E_i$  由虚拟接口  $VIF_{i0}$  抵达主路由器  $ER_i$  时,需将其发布区域分别与下游的虚拟接口  $VIF_{i1}$  到  $VIF_{im}$  所维护的订购区域分别进行逐个匹配,通过判断两

个区域是否相交来实现对  $E_i$  实现转发。但在大规模分布式虚拟环境中,下游路由器连接的主机或路由器数量一般很大时,如果采用逐个匹配转发的方法,路由器的处理时间将和主机的数量成正比,因此不利于系统的扩展。本文提出了一种 Filter Index 结构,此结构不仅提高了路由器节点的处理时间,同时可以降低系统的误送率,减少通讯时的冗余量,进一步提高了系统的可扩展性。

### 3.1 路由表结构描述

Filter Index 是以包含关系作为节点之间连接关系的一种图状结构(如图 3 所示),图中的节点至少包含关键字(key)以及指针(pointer),而图中节点之间的边则通过其所包含的关键字之间的关系(子集或超集)利用指针来进行描述。同时,为了方便对此结构进行检索,将没有任何子集的节点称为 root,没有任何超集的节点称为 top。对 Filter Index 的检索实际上是从 root 或 top 开始查找子集或超集的简单的递归过程。假设在图 3 中查找到  $V_2V_3V_4V_5$  所有的子集,首先从 boot 开始查询,找出其中属于  $V_2V_3V_4V_5$  的子集节点即  $V_2$  和  $V_4$ ,然后分别根据它们指向超集的指针找到各自的超集,判断它们是否为  $V_2V_3V_4V_5$  的子集,对属于其子集的那些节点递归地重复上述过程,直至抵达 top。值得注意的是在此过程中有些节点可能被访问了多次,为了进一步提高查找效率,需要对每个已经访问过的节点都进行标识。

在大规模分布式虚拟环境中,我们采用了 Filter Index 结构对路由器上所维护的所有事件的订购信息进行描述,结构中的每个节点的信息用四元组  $\langle \text{Key}, \text{VIFInfo}, \text{SuperLink}, \text{SubLink} \rangle$  来进行表示,其中 Key 表示关键字,即为事件的发布区域或接口的订购区域;VIFInfo 表示主动路由器虚拟接口的相关信息;SuperLink 表示指向该节点超集的指针;SubLink 表示指向节点子集的指针。考虑到在大规模分布式虚拟环境中包含了大量的场景虚拟空间,我们采用了基于区域(Locale)的方式对空间进行划分<sup>[1]</sup>,即将虚拟空间分为若干小的区域(region),这些区域可以是不均匀的或者手工分配的,事件的发布区域或接口的订购区域也采用 region 进行描述。同时,对于 VIFInfo 可详细描述为  $\langle \text{VIF id}, \text{Region ID}, \text{Flag}, \text{Status}, \text{TTL}, \text{Sequence Number}, \text{Acknowledge Number}, \text{Raw List} \rangle$ ,其中 VIF id 表示虚拟接口 VIF 的序号(区分具有相同 key 值的不同 VIF 端口);Region ID 表示虚拟接口对应的路由器或主机的地址;Flag 标记了虚拟接口是上游 VIF 还是下游 VIF;Status 表示虚拟接口的状态,这些状态包括虚拟接口刚刚建立尚未被确认时的状态(Pending)、已经确认过的可用状态(Active)和不可用的消亡状态(Closed);TTL(Time to Live)表示 VIF 当前可以存在的时间,以秒为单位;Sequence Number 和 Acknowledge Number 记录了订购消息和回答消息中的序列号;Raw List 用于记录本地网络中通过 Multicast 方式连接的主机的订购区域,用于主机更新或删除订购区域时,重新计算本地网络中其他主机的订购区域的并集。

### 3.2 路由传递策略

事件匹配是事件传递的主要工作,决定了哪些接口需要转发,哪些接口不需要转发,事件匹配时的效率直接影响路由器的性能。在大规模分布式虚拟环境系统中采用了基于 Filter Index 结构的路由传递策略,能够实现快速的事件匹配和传递,极大地减少了在每个中间节点(路由器)的处理时间。

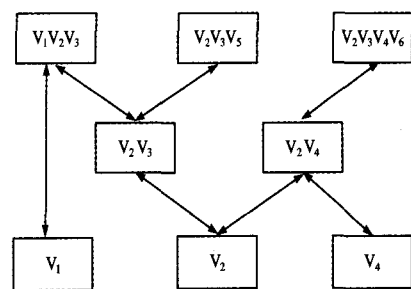


图 3 Filter Index 结构

具体策略描述如下:当事件  $E_i$  到达主动路由器后,须将其的发布区域与节点所包含端口的订购区域进行 match 运算,再根据其结果实现转发;事件的匹配转发过程将分为两步:第一步,“自下而上”即首先从 Filter Index 结构的 root 开始,一旦查找到与其发布区域相匹配的节点,即将该事件转发给此节点所对应端口,然后利用 SuperLink 指针找到其超集节点,同时将事件也转发给此节点所对应端口,直至抵达 top;第二步,“自上而下”即从 Filter Index 结构的 top 开始,查找未被访问过且相匹配的节点进行转发,然后利用 SuberLink 指针查找到其子集节点,同时也将事件也转发给此节点所对应端口,直至抵达已被访问过的节点。基于 Filter Index 结构的快速事件路由传递算法如下:

```

Procedure Fast_Router( $E_i$ , VI)
//root-to-top
1. for each node in root
2.   do
3.     match_sent(node) //事件  $E_i$  匹配及转发
4.     node = node->suplink
5.   until (node not in top) and (nodes is unmarked)
6. end for
//top-to-root
7. for each node in top
8.   while (nodes is unmarked)
9.     match_sent(node) //事件  $E_i$  匹配及转发
10.    node = node->sublink
11.  end while
12. end for

```

上述算法描述了基于 Filter Index 的事件路由传递的过程,子函数 match\_sent 实现了对事件  $E_i$  匹配及转发功能。对于事件的匹配,由于采用区域(region)的方式对空间进行划分,在匹配时只需将发布区域和订购区域进行交集运算来判断是否匹配。值得注意的是,在向虚拟接口 VIF 发送事件时,需对接口状态 Status 的判断,只有当接口处于 Active 状态时进行事件数据包的转发。在 Active 状态下,TTL(Time to Live)表示 VIF 当前可以存在的时间,以秒为单位。主动路由器周期性地减少 TTL,直到它的值减小到零为止。如果在 TTL 减小为零之前,该 VIF 对应的主动路由器或主机都没有发送任何消息给当前路由器,那么这个 VIF 就被设置为 Closed 状态;反之 TTL 值会被重新设置为初始值。当 VIF 处于 Closed 状态时,路由器将在不确定的时间内删除这个 VIF。如果在删除之前路由器从这个 VIF 收到了任何消息,则重新将状态标记为 Active 或者 Pending,并重新设置 TTL 值。

## 4 实验与评估

为了验证事件模型的设计及其优化结构,我们实现了一个基于事件模型的通讯结构的原型系统。此系统的主要功能是构造大规模分布式虚拟环境的底层通讯平台,在 CBT 的拓扑结构上,实现了事件路由和事件传递的基本功能,为应用层

(下转第 113 页)

Kalman 滤波系统的结果是比较稳定和平滑的,能够起到较好的数据融合效果,具有实用性和应用前景,如图 4。

**结束语** 本文通过不同的仿真实验结果和实际运行实验结果,验证了在使用北斗双星定位系统的基础上,使用联邦 Kalman 滤波进行组合导航的有效性,能够在不同的环境中提升组合导航的精度和稳定性,同时联邦 Kalman 滤波具有计算量小、性能稳定、收敛快的优点,有利于产品的实用化。

## 参考文献

- 1 杜尚丰,宫晓琳. GPS/SINS 综合定位系统仿真实验研究. 仪器仪表学报,2005(S1)
- 2 刘钧圣,朱文彪. 一种模糊自适应 INS/GPS 组合导航方法. 现代防御技术,2005(5)
- 3 杨坤明. 组合导航系统的卡尔曼滤波器的设计. 西华大学学报(自然科学版),2005(4)

- 4 华冰,刘建业,熊智. 联邦自适应滤波在 SINS/北斗/星敏感器组合导航系统中的应用. 东南大学学报(自然科学版),2004(S1)
- 5 Sasiadek J Z, Khe J. Sensor fusion based on fuzzy Kalman filter [J]. In: IEEE, The Second Workshop on Robot Motion and Control, October 2001. 18~20
- 6 Sasiadek J Z, Wang Q. Sensor Fusion Based on Fuzzy Kalman Filtering for Autonomous Robot Vehicle[C]. In: Proceedings of the 1999 IEEE Int Conf on Robotics and Automation
- 7 Sasiadek J Z, Wang Q, Zeremba M B. Fuzzy Adaptive Kalman Filtering for INS/GPS Data Fusion[C]. In: Proceedings of the 15th IEEE International Symposium on Intelligent Control, 2000
- 8 Lawrence P J Jr, Berarducci M P. Navigation sensor, filter, and failure mode simulation results using the distributed Kalman filter simulator (DKFSIM). In: Position Location and Navigation Symposium, April 1996. 697~710

(上接第 109 页)

仿真提供基础。同时,采用了基于 Filter Index 的快速事件匹配路由结构作为事件在路由器上进行寻址及转发的路由策略。实验环境为 CPU P4 1.5G、RAM 256M、System Win2K,在实验中一共插入 45000 条订购,每隔 5000 条订购进行取样,记录插入订购和匹配发送事件所需的时间。

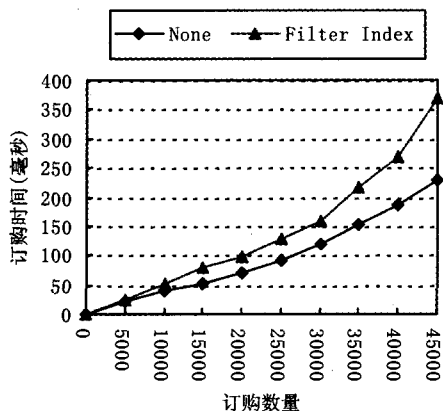


图 4 事件订购时间

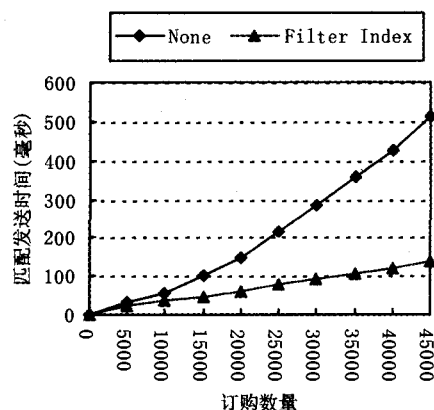


图 5 事件匹配发送时间

图 4 和图 5 分别是在原型系统上进行实验得到的订购时间和事件匹配发送时间的曲线。从图 4 中我们可以看出,在两组实验数据中,订购时间基本上都随着订购数量呈线性增长趋势,但是在订购数量增长到一定幅度后,采用 Filter Index 结构的订购时间大幅增长。分析原因,由于订购数量不断增大,导致 Filter Index 结构变得较为复杂,即在结构中插

入或删除节点的时间变长。在这方面,应该仍有改进的余地,如可根据事件订购的形式采用多组 Filter Index 结构以及 Cache 技术等。在图 5 中,事件匹配时间也基本上随着订购数量增长呈增长趋势,但比较两组实验数据可以明显看出采用 Filter Index 结构之后,事件的平均匹配传递时间得到了明显的改善,而且随着订购数量的不断增多,其效率的提高越显著。分析表明,实验结果达到了预期的效果,验证了本文提出大规模分布式虚拟环境中基于事件模型的路由机制是可行且有效的。

**结束语** 本文从事件路由和事件传递两个方面研究了分布式虚拟环境中基于事件模型的路由机制,提出了一种基于 Filter Index 结构的一种快速的事件匹配路由结构,并通过对依据本文设计方案实现的原型系统的实验和测试,验证了设计的可行性及正确性。今后的工作将从发掘事件之间的相关性,实现事件模板以及研究更丰富的订购语言等方面进行进一步的研究。

## 参考文献

- 1 潘志庚,姜晓红,张明敏,等. 分布式虚拟环境综述. 软件学报, 2000, 11(4): 461~467
- 2 Morse K L. An Adaptive Distributed Algorithm for Interest Management; [PhD Thesis]. University of California, 2000
- 3 Purbrick J, Greenhalgh C. Extending Locales: Awareness Management in MASSIVE-3. In: Proceedings of IEEE Virtual Reality, Feb. 2000
- 4 Abrams H A. Extensible Interest Management for Scalable Persistent Distributed Virtual Environments; [PhD thesis]. Naval Postgraduate School, 1999
- 5 Benford S, Greenhalgh C, Rodden T, et al. Collaborative Virtual Environments - to what extent is cyberspace a really space. Communications of the ACM, 2001, 44(7)
- 6 Starovic G, Tangney V C B. An Event Based Object Model for Distributed Programming. In: Proceedings of the International Conference on Object Oriented Information Systems, London, 1995
- 7 Yu H, Estrin D, Govindan R. A hierarchical proxy architecture for internet-scale event services. In: Proc. of WETICE99, Stanford, June 1999
- 8 Zabele S, Dorsch M, Ge Z, et al. SANDS: Specialized active networking for distributed simulation. In: DARPA Active Networks Conference and Exposition (DANCE), USA, May 2002
- 9 卢威,陈继明,徐晓阳,等. 分布式虚拟环境 AIMINET 的关键技术概述. 计算机科学,2006, 33(11)
- 10 Carzaniga A, Deng J, Wolf A L. Fast Forwarding for Content-Based Networking; [Technical report]. Dept. of Computer Science, Univ of Colorado, Nov. 2001
- 11 Purbrick J, Greenhalgh C. Extending Locales: Awareness Management in MASSIVE-3. In: Proceedings of IEEE Virtual Reality, Feb. 2000