

基于索引排序的快速动态区域匹配算法

姚益平¹⁾ 张颖星¹⁾ 陈 欣²⁾ 季利春¹⁾

¹⁾(国防科学技术大学计算机学院 长沙 410073)

²⁾(北京邮电大学计算机学院 北京 100876)

摘 要 数据分发管理实现基于值的过滤,可进一步减少大规模仿真中盟员接收冗余数据的可能性和网络中的数据流量.数据分发管理实现的关键是区域匹配算法的实现,高效、精确的区域匹配一直是数据分发管理追求的目标.现有的区域匹配算法如直接匹配法、网格法、基于分类的算法等都不很理想,或者过滤效果不佳,或者耗时较大,难以适应大规模分布式仿真的需要.论文针对在区域比较多的大规模分布式仿真系统中,区域大都需要随着仿真的推进而频繁地被动态修改的实际特点,提出了一个基于索引排序的快速动态区域匹配算法——IOBM算法,该算法将范围的上下界分别各用一个节点表示,使用指针数组来索引每一维上的范围节点,数组元素的下标值表示对应的节点值,利用区域范围更新前后的信息,将匹配限定在移动区间之内,通过对移动区间之内的链表进行直接操作来完成匹配工作,因而大大减少了匹配计算的时间,实现了高效、精确的动态匹配.该算法尤其适合区域比较多的大规模分布式仿真的需要.

关键词 高层体系结构(HLA);运行支持环境(RTI);数据分发管理(DDM);区域匹配;动态

中图法分类号 TP391

DOI号: 10.3724/SP.J.1016.2009.01375

Dynamic Region Matching Algorithm Based on Index-Order

YAO Yi-Ping¹⁾ ZHANG Ying-Xing¹⁾ CHEN Xin²⁾ JI Li-Chun¹⁾

¹⁾(College of Computer, National University of Defense Technology, Changsha 410073)

²⁾(School of Computer Science and Technology, Beijing University of Posts and Communications, Beijing 100876)

Abstract The HLA Data Distribution Management (DDM) service provides an abstract, application-driven data filtering capability. It can reduce the transmission and reception of irrelevant data. The key problem of its implementation is region matching algorithm. The current algorithms such as directly matching, grid-based, and sort-based approach are all not so perfect. They are either time consuming or inaccuracy in filtering, and are difficult to support large-scale simulation. Aiming at the characteristic of frequently region changing in large-scale simulation systems, a more effective matching algorithm based on intersecting information of region moving is proposed. It represents the upper bound and lower bound of a range with two nodes, and uses two indexed ordered tables to store the publication and subscription nodes of a dimension respectively. It uses range intersecting information during a range moving, and limits the matching computing only in the area of moving. Thereby it can greatly decrease the candidate ranges that need to do matching computing, thus can decrease the matching complexity greatly, and reach

收稿日期:2007-08-04;最终修改稿收到日期:2009-06-13. 本课题得到国家自然科学基金(60773019)、国家教育部博士点基金(200899980004)和国家“八六三”高技术研究发展计划项目基金(2006AA01Z330)资助. 姚益平,男,1963年生,博士,教授,博士生导师,主要研究领域为并行与分布仿真、虚拟现实等. Email: ypyao@nudt.edu.cn. 张颖星,男,1981年生,博士,主要研究方向为并行离散事件仿真、基于Agent仿真等. 陈欣,男,1978年生,博士研究生,主要研究方向为计算机仿真和虚拟现实等. 季利春,男,1978年生,博士,主要研究方向为分布交互仿真、并行离散事件仿真等.

precise matching with high performance. This algorithm is extremely fit for the need of large-scale distributed simulation that has a large number of regions.

Keywords high level architecture; runtime infrastructure; data distribution management; region matching; dynamic

1 引言

大规模分布式交互仿真如 J9901 (Joint Operations 1999—1)、AO00 (Attack Operations 2000)、MC02 (Millennium Challenge 2002) 等的实体数往往达到几万甚至几十万^[1-2], 保证这类仿真效率的重要途径是数据分发管理 DDM (Data Distribution Management) 的使用^[3]. DDM 是高层体系结构 HLA^[4] (High Level Architecture) 的重要管理服务之一, 它是在声明管理实现基于类的数据过滤后, 实现基于值的过滤, 其目的是进一步减少大规模仿真中盟员接收冗余数据的可能性和网络中的数据流量^[5]. 在 HLA 1516 接口规范^[4]中, 数据分发以区域为基础, 公布盟员通过“带区域登记对象实例”或“为更新关联区域”等服务来声明自己的更新区域, 订购盟员通过“带区域订购对象类属性”等服务来声明自己感兴趣区域. DDM 实现的关键是更新区域和订购区域的匹配(区域交叠)判断, 由于在联盟执行过程中, 区域大都随着仿真的推进而频繁地被动态修改, 而每次区域变化都需要进行匹配判断, 因而动态区域匹配算法的设计决定着 DDM 的效率和性能, 也直接影响到仿真系统的性能.

2 几种典型的区域匹配算法

现有的典型区域匹配算法概括起来主要有直接匹配法、网格法和基于分类的算法等几种. 其中直接匹配法(brute force)查询当前所有的订购(或公布)区域. 该算法最简单, 但计算开销也最大, 假设共有 N 个 d 维区域需要进行匹配, 则所需的查询时间为 $O(d \times N)$.

网格法(grid-based)将 d 维全域预先划分为规则的 d 维网格^[6], 每一个网格逻辑上都对应一个多目广播组(或数据传输通道), 对区域边界进行计算即可确定区域覆盖的网格, 进而确定数据传输通道. 网格法仅需 $O(d \times K)$ 的时间去计算动态 d 维区域相交(K 是被覆盖的网格的数量), 但网格法会产生

虚假连接和冗余连接^[3]. 就整体而言, 网格法由于过滤效率和组播组存在的矛盾而难以适用于规模较大的仿真需要^[7].

基于分类的匹配算法(sort-based)将所有区域中相同维上的范围在该维度上进行投影, 若两个或多个范围在投影上有重叠, 则表明这些范围相交. 对所有维度进行上述处理, 当且仅当两个区域的所有公共维度上的范围相交时, 这两个区域才相交. 基于分类的匹配算法的计算复杂度为 $O(d \times (M+N) \times \log(M+N))$, 其中 M 、 N 分别表示更新区域和订购区域的数目, d 为维数. 基于分类的算法一次可完成对所有区域的交叠判断, 虽然这种判断可通过采用位操作或位向量^[8-9]而提高匹配效率, 但由于每次匹配都得对所有范围节点进行排序和相交检测, 因而它实际上更适合一次性的静态匹配, 对于只需对个别区域进行动态匹配的情况(个别区域动态修改), 其匹配效率相对较低^[10-11]. 为了提高动态区域匹配的效率和文献[12]对文献[8]提出的基于分类的匹配算法进行了进一步的改进, 但其计算复杂度仍然为 $O(\max RS / DUB \times N) + O(\max LBS / DUB \times N)$, 即与区域的数量 N 成正比.

由以上介绍可知, 目前实现的 DDM 区域匹配算法都不很理想, 或者过滤效果不佳, 或者耗时太大, 难以适用于大规模分布式仿真.

3 基于索引排序的快速动态区域匹配算法

3.1 算法思想

为解决上述问题, 在这里提出一种建立在索引有序表基础上的算法. 该算法基于如下的原理: 当区域移动时, 其它区域与该区域交叠情况的变化必定与该区域刚才的移动有关, 也就是说: 区域在动态更新时不用与所有区域进行匹配, 即无需考虑在区域移动区间之外的区域, 因为与这些区域的交叠情况是不可能变化的, 如图 1 所示.

图 1 是二维空间中区域移动交叠变化情况, 区域 2 发生移动时, 在 X 维度上途中经过区域 1 和区

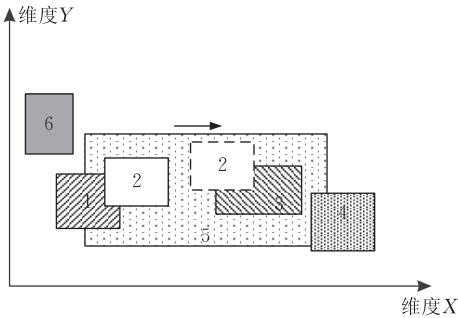


图 1 区域移动相交变化示意图

域 3 的边界(与它们的交叠情况发生了变化),而没有经过区域 4、5、6 的边界(与它们的交叠情况没有变化);而在 Y 维度上没有经过任何区域. 这样我们利用更新前后的区域确定区域移动时经过的区域即可获得区域交叠变化的情况. 在这里我们假定区域的移动是小范围的,在实际应用中绝大多数区域变化也符合这种情况^[13].

根据区域交叠的定义,要判断两区域是否交叠,必须首先判断这两个区域公共维度上的范围是否交叠,当且仅当这两个区域的所有公共维度上的范围交叠时,这两个区域才交叠. 因此,区域匹配算法的关键实际上是范围交叠的判断. 为此,我们分两种情况进行考虑:

第 1 种情况(见图 2):当范围 1 增大移动时,我们发现,若范围 1 的开节点(空心圆表示)遇到范围 3 的闭节点(实心圆表示),则表示范围 1 与范围 3 从不相交变成了相交;若范围 1 的闭节点遇到范围 2 的开节点,则表示范围 1 与范围 2 从相交变成了不相交.

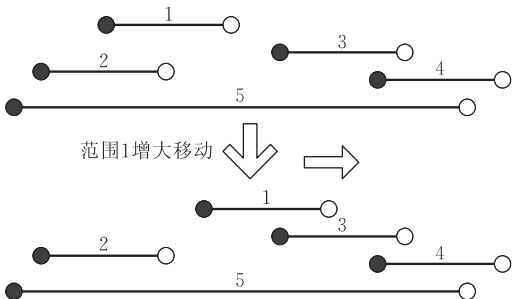


图 2 范围 1 增大移动后交叠变化情况示意图

第 2 种情况(见图 3):当范围 1 减小移动时,我们发现,若范围 1 的闭节点遇到范围 2 的开节点,则表示范围 1 与范围 2 从不相交变成了相交;若范围 1 的开节点遇到范围 3 的闭节点,则表示范围 1 与范围 3 从相交变成了不相交.

上述范围 1 在移动过程中,没有经过范围 4、5 的边界,因此,与它们交叠变化情况不用考虑.

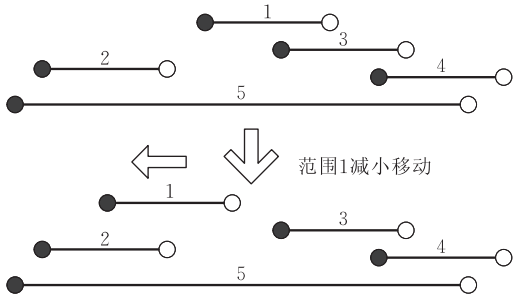


图 3 范围 1 减小移动后交叠变化情况示意图

3.2 区域数据的存储

区域数据的存储方式决定了查询区域数据的时间. 这里我们使用索引有序表来存放区域的每一维数据,即全域空间的每一维对应两组索引有序表,一组索引有序表用来存放所有公布区域在该维上的范围,另一组索引有序表用来存放所有订购区域在该维上的范围. 实现时区域每一维上的范围 $[a,b)$ 使用两个节点表示:第一个节点称为闭节点,其节点值为 a ,第二个节点称为开节点,其节点值为 b . 节点的数据结构定义如下:

```
struct Dim {
    int RangeID;           //该节点所属的范围标识
    struct Dim *Next;      //节点指针
}
```

为了便于范围移动和加快查找速度,我们将两个索引有序表作为一组分别存储范围的闭节点和开节点. 每个索引有序表通过一个指针数组进行索引,我们将指针数组的大小设为维度的上界,指针数组的下标从 0 开始,由于维度的下界为 0^[4],因此,我们可以使用数组元素的下标值来表示对应的节点值,即指针数组元素所指向的节点其节点值与该数组元素的下标值相等. 如一个含 6 个范围 $[0,1)$ 、 $[0,4)$ 、 $[1,4)$ 、 $[3,6)$ 、 $[5,7)$ 、 $[6,8)$ 的有序表表示如图 4 所示.

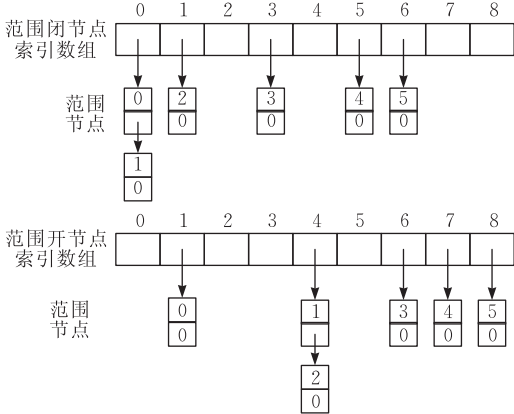


图 4 索引有序表示意图

若范围 2 从 $[0,4)$ 移动到 $[2,6)$,则相应的索引有序表的变化如图 5 所示.

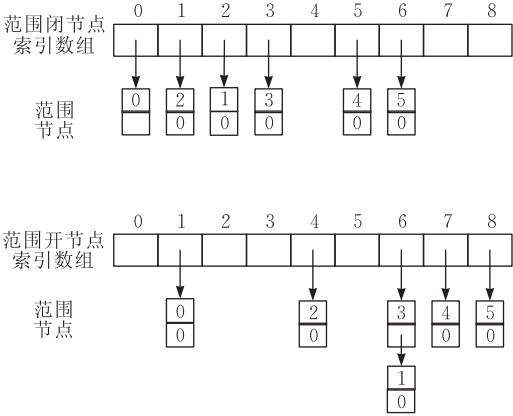


图 5 范围 2 从 $[0,4)$ 移动到 $[2,6)$ 后节点变化示意图

若一个从 $[0,2)$ 移动到 $[2,4)$ 的范围(设为 R)需要与图 4 中的范围进行匹配(注意:匹配之前范围 $[0,2)$ 与图 4 中的范围交叠情况需继承下来),则我们首先找到范围闭节点索引数组中 $A[2+1] \sim A[4]$ 指向的所有范围节点:这里有一个即范围 4 的闭节点,表明范围 R 与范围 4 从不相交变成了相交,然后取范围开节点索引数组中 $A[0] \sim A[2-1]$ 指向的所有开节点,这里有一个即范围 0 的开节点,表明范围 R 与范围 0 从相交变成了不相交.

从图 4 中可以看出,对于不在移动范围之内的节点是不需要对它进行匹配的.

对于新创建的范围:(1)要创建两个新节点,并在对应的指针数组中的相应位置链入这两个新节点;(2)为了统一处理,我们假设其移动之前的范围为 $[0,0)$.

3.3 区域匹配算法

基于以上原理,我们给出基于索引排序的动态区域匹配算法——IOBM(Intersecting-Information-Based Matching Algorithm)算法如下:

- 1. For(修改区域 RG 中的每一个范围 R) Do
{
 - ① 设范围 R 修改之前的区间为 $[a,c)$ (对于新创建范围, $a=c=0$),修改之后的区间(新范围)为 $[b,d)$,该范围所属维上的公布范围索引表为 Ub,Uk ,订购范围索引表为 Sb,Sk ;
 - ② 若范围 R 为公布范围,则令 $Modb=Ub,Modk=Uk,Matb=Sb,Matk=Sk$;若范围 R 为订购范围,则令 $Modb=Sb,Modk=Sk,Matb=Ub,Matk=Uk$;原有与 $[a,c)$ 交叠的范围的列表为 $R_{OverlapSet}$ (对于新创建范围,该列表为空);
 - ③ 若 R 为新创建范围,则为 R 分别创建一个闭节点

和开节点,将闭节点链入 $Modb[b]$ 所指链表中,将开节点链入 $Modk[d]$ 所指链表中;否则,将 $Modb[a]$ 所指链表中 R 闭节点移入 $Modb[b]$ 所指链表中(注意,只移 R 闭节点,若 R 闭节点后链有节点,则应仍将它们留在 $Modb[a]$ 所指链表中;以下同),将 $Modk[c]$ 所指链表中 R 开节点移入 $Modk[d]$ 所指链表中;

- ④ if ($c < d$) then 将 $Matb[c] \sim Matb[d-1]$ 所指链表中的所有闭节点对应的范围加入到 $R_{OverlapSet}$ 中;
- ⑤ if ($a < b$) then 将 $Matk[a+1] \sim Matk[b]$ 所指链表中的所有开节点对应的范围从 $R_{OverlapSet}$ 中删除;
- ⑥ if ($a > b$) then 将 $Matk[b+1] \sim Matk[a]$ 所指链表中的所有开节点对应的范围加入到 $R_{OverlapSet}$ 中;
- ⑦ if ($c > d$) then 将 $Matb[d] \sim Matb[c-1]$ 所指链表中的所有闭节点对应的范围从 $R_{OverlapSet}$ 中删除;
- ⑧ 相交列表 $R_{OverlapSet}$ 中即为与 R 有交叠的所有范围.

}

2. 对 RG 中所有范围的相交列表进行综合处理,即可得到与修改区域 RG 相交的所有区域.

不难看出,上述算法的匹配计算量 $\leq O(d \times K)$,其中 d 为区域的维数, K 为区域 RG 中的范围在移动过程中所遇到的节点数的平均值.由此可知,IOBM 算法计算量与要匹配的区域数量无关,而只与修改区域中的范围在移动过程中所遇到的节点数有关.因此它非常适合区域较多的大规模分布式仿真的需要.

4 实验结果

为了验证上述算法,我们设计测试程序如下:该程序根据用户输入的维度上界 m 及范围数 n ,随机生成该维上的 n 个公布范围和 n 个订购范围;然后对每一种算法,分别进行如下操作:对 n 个公布范围中的每一个范围,依次进行多次不同的修改,对每一个范围的每一次修改,采用匹配算法求出与修改后的公布范围相交的订购范围集合;记下每一种算法进行上述操作所花费的总时间,将总时间除以公布范围修改的总次数,即可得每一次公布范围修改时,为求与之相交的订购范围集合匹配算法所需的时间.测试在 IBM R40 笔记本电脑上进行,配置为 Intel P4 2.0GHz,256MB 内存,Windows XP 操作系统.为了获得准确的测试结果,我们对不同的 m,n 值进行了大量的测试,现给出有代表性的测试结果如图 6~图 8 所示.

图 6 显示了所有匹配算法在范围数固定为 500 时的耗时.由于其它算法结果耗时较多,若在一个图

中给出,则很难看出 IOBM 算法的时间变化趋势,所以图 7、图 8 只给出了与网格数为 10 的网格法的对比测试结果.由测试结构可以看出 IOBM 算法随着维度上界的增大,其一次匹配所需时间相应减少(甚至优于网格数为 10 的网格法),这也表明 IOBM 算法的性能与范围在移动过程中所遇到的范围数有关,因为维度上界越大,则范围分布越广,公布范围在移动过程中所可能遇到的订购范围数越少.其它匹配算法(包括直接匹配法、网格法、基于分类的算法等)的匹配时间则基本上与维度上界无关,而主要

与范围数相关.

图 9~图 11 显示了维度上界固定时,算法运行时间随范围数变化的情况.从测试结果可以看出,IOBM 算法性能大大优于基于分类的算法、直接匹配法及网格数为 100 的网格法,而与网格数为 10 的网格法性能相当(在某些情况下甚至更优),但网格数为 10 的网格法数据过滤效果很差,并且存在虚假连接和冗余连接,而 IOBM 算法实现了精确匹配.基于分类的算法匹配时间迅速增加的原因在于它每次都对所有公布区域和订购区域进行匹配计算.

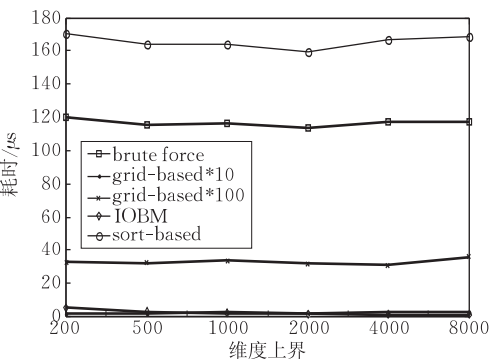


图 6 范围数=500,一次匹配计算所需时间随维度上界变化测试结果

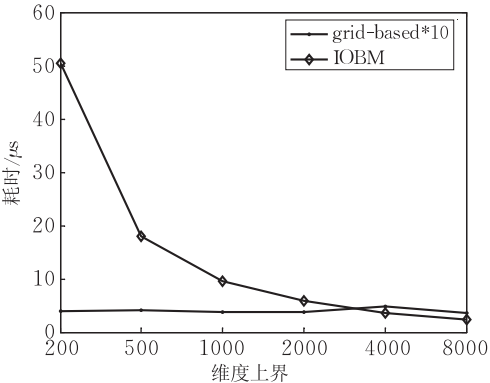


图 7 范围数=2000,一次匹配计算所需时间随维度上界变化测试结果

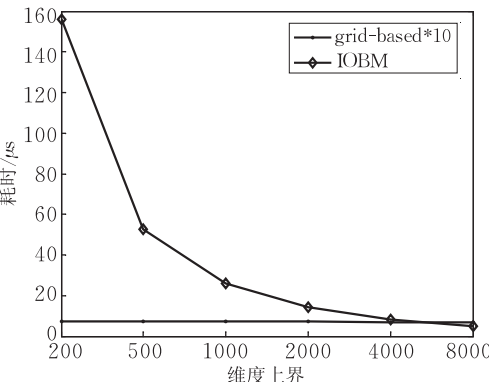


图 8 范围数=3500,一次匹配计算所需时间随维度上界变化测试结果

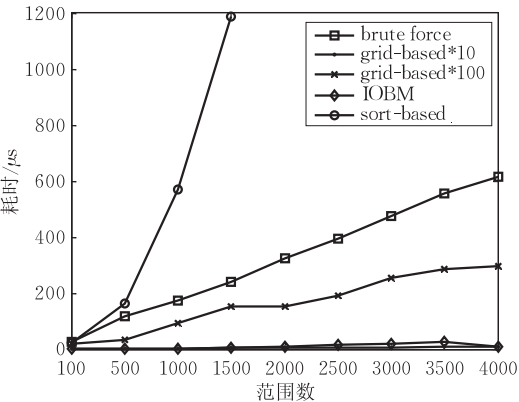


图 9 维度上界=1000时,一次匹配计算所需时间随范围数变化测试结果

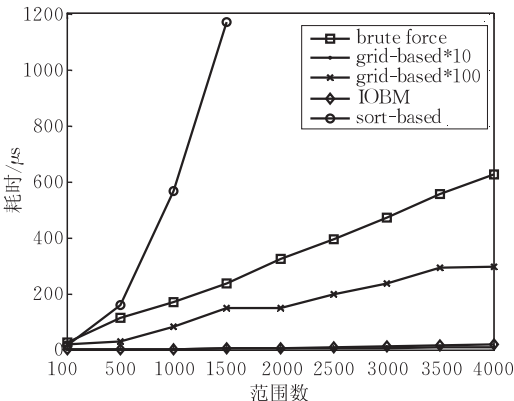


图 10 维度上界=2000时,一次匹配计算所需时间随范围数变化测试结果

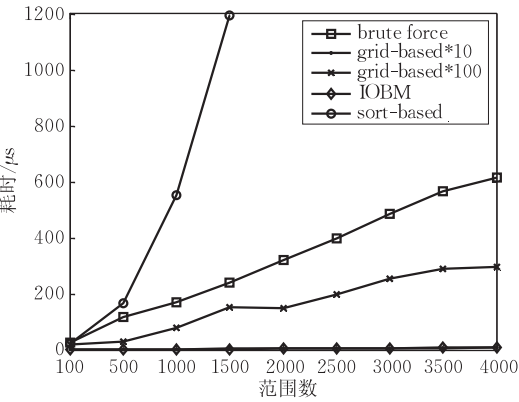


图 11 维度上界=4000时,一次匹配计算所需时间随范围数变化测试结果

5 结束语

论文通过分析几种区域匹配算法的优缺点,提出了一种基于索引排序的快速动态区域匹配算法.该算法的创新在于:使用指针数组来索引每一维上的范围节点,数组元素的下标值表示对应的节点值,利用区域范围更新前后的信息,将匹配限定在移动区间之内,通过对移动区间之内的链表进行直接操作来完成匹配工作,因而大大减少了匹配计算的时间.测试表明,该算法不仅实现简单,而且匹配效率非常高,并且是精确匹配,因而可满足大规模分布式仿真动态 DDM 的需要.

参 考 文 献

- [1] Torpey M, Wilbert D, Helfinstine B, Civinskas W. Experiences and lessons learned using RTI-NG in a large-scale, platform-level federation//Proceedings of the Simulation Interoperability Workshop. Orlando, Florida, USA, 2001
- [2] USJFCOM. Millennium Challenge 2002 Model and Simulation Federation, August 2002
- [3] Helfinstine B, Wilbert D, Torpey M, Civinskas W. Experiences with data distribution management in large-scale federations//Proceedings of the Simulation Interoperability Workshop. Orlando, Florida, USA, 2001
- [4] IEEE Standard for Modeling and Simulation (M&S) High Level Architecture(HLA) — Federate Interface Specification (IEEE Std 1516.1-2000). USA: The Institute of Electrical and Electronics Engineers, Inc., 2001
- [5] Yao Yi-Ping, Lu Xi-Cheng, Wang Huai-Min. Design and implementation of hierarchical RTI server. Chinese Journal of Computers, 2003, 26(6): 716-721(in Chinese)
(姚益平, 卢锡城, 王怀民. 层次式 RTI 服务器的设计与实现. 计算机学报, 2003, 26(6): 716-721)

- [6] Hyett Mark, Wuerfel Roger. Implementation of the data distribution management services in the RTI-NG//Proceedings of the Simulation Interoperability Workshop. Orlando, Florida, USA, 2002
- [7] Tan Gary, Ayani Rassul, Zhang Yu-Song, Moradi Farshad. Grid-based data management in distributed simulation//Proceedings of the 33rd Annual Simulation Symposium. Washington, USA, 2000
- [8] Racz C, Tan G, Yu J. A sort-based DDM matching algorithm for HLA. ACM Transactions on Modeling and Computer Simulation, 2005, 15(1): 14-38
- [9] Huang Hong-Bing, Ye Chao-Qun, Jin Shi-Yao. An improved sort-based algorithm for region matching. Computer Engineering & Science, 2006, 28(9): 80-82(in Chinese)
(黄红兵, 叶超群, 金士尧. 一种改进的基于排序的区域匹配算法. 计算机工程与科学, 2006, 28(9): 80-82)
- [10] Racz Come, Yu Jun, Tan Gary, Tay Seng Chuan. Adaptive data distribution management for HLA RTI//Proceedings of the European Simulation Interoperability Workshop. London, UK, 2002
- [11] Yu Jun, Racz Come, Tan Gary. Evaluation of sort-based matching algorithm for the DDM//Proceedings of the 16th Workshop on Parallel and Distributed Simulation. Washington, USA, 2002
- [12] Pan Ke, Turner Stephen John, Cai Wentong, Li Zengxiang. An efficient sort-based DDM matching algorithm for HLA applications with a large spatial environment//Proceedings of the 21st International Workshop on Principles of Advanced and Distributed Simulation (PADS'07). San Diego, California, USA, 2007: 70-82
- [13] Petty M. Geometric and algorithmic results regarding HLA data distribution management matching//Proceedings of the Simulation Interoperability Workshop. San Diego, California, USA, 2000
- [14] Sheng Wei, Zhong Zhou, Wei Wu. An interest filtering mechanism based on LoI//Proceedings of the ICAT 2006. Hangzhou, China, 2006: 334-343



YAO Yi-Ping, born in 1963, Ph.D., professor, Ph.D. supervisor. His major research interests include parallel and distributed simulation and virtual reality.

research interests focus on parallel discrete event simulation and agent-based simulation.

CHEN Xin, born in 1978, Ph.D. candidate. His current research interests focus on computer simulation and virtual reality.

JI Li-Chun, born in 1978, Ph.D. candidate. His current research interests focus on distributed simulation and parallel discrete event simulation.

ZHANG Ying-Xing, born in 1981, Ph.D.. His current

Background

The High Level Architecture (HLA) is an architecture for reuse and interoperation of distributed simulations, which has been standardized as IEEE1516 in September 2000. It provides several Data Distribution Management (DDM) services to filter out irrelevant data. These services rely on the computation of intersection between update and subscription regions, which is called matching. Currently, there are several typical DDM matching algorithms, however, these algorithms are still time consuming and not good enough for scenarios when the regions have a large number of dimensions and need to be modified along with the simulation advancement dynamically. The paper introduces a algorithm using range intersecting information during a range moving, and

limits the matching computing only in the moving area. This algorithm also use an index ordered table data structure in order to improve its query efficiency.

This work is supported by the National Natural Science Foundation of China under grant No.60773019, the Ph. D. Programs Foundation of Ministry of Education of China under grant No. 200899980004 and the National High Technology Research and Development Program of China under grant No. 2006AA01Z330. The aim of the research is to achieve precise matching with high performance in dynamic scenarios. Until now, the research team has published more than 20 papers about the technology of parallel and distributed simulation.