

# P2P 异步传输点播系统设计与仿真

彭 昭 杨宗凯 陈京文 朱自珍

(华中科技大学电信系 湖北省智能互联网技术重点实验室 武汉 430074)

**摘 要** 互联网上大规模 Peer-to-Peer (P2P) 点播服务存在高效资源调度和不可预知用户行为等挑战。借鉴 BitTorrent (BT) 算法原理, 提出基于异步传输的 P2P 点播系统。该系统采用集中式节点定位以及非结构化拓扑, 通过异步传输机制提升文件资源共享度和利用率, 并引入视频数据块的优先级和稀有程度改进系统播放连续度。数据包及仿真实验表明本系统在播放连续度、启动时延以及服务器压力等性能指标上的良好性能, 充分验证了点播系统中异步传输机制的可行性。

**关键词** 对等网络, 异步传输, 点播

**中图分类号** TP393 **文献标识码** A

## Design and Simulation of Asynchronous Transfer System on P2P VoD

PENG Zhao YANG Zong-kai CHEN Jing-wen ZHU Zi-zhen

(Hubei Provincial Key Laboratory of Smart Internet Technology, Dept. of Electronics & Information Engineering, Huazhong University of Science & Technology, Wuhan 430074, China)

**Abstract** Providing video-on-demand (VoD) service over the internet remains some challenges in terms of efficient resource utilization and unpredictable user action. Inspired by the success of BitTorrent (BT) in the field of file sharing, we presented a Peer-to-Peer (P2P) VoD system based on asynchronous data transfer. This system is specified with centralized directory and unstructured overlay. By using asynchronous data transfer mode, it could improve sharing and utilization on resources. We also considered the priority and rareness of video data blocks to smooth video playback. Packet-level simulations show the benefits of system performance metrics on playback continuity, launch delay and server stress, and demonstrate the feasibility of the asynchronous transfer.

**Keywords** Peer-to-Peer network, Asynchronous transfer, Video-on-demand

互联网上的视频点播服务近年来受到广泛关注, 但由于流媒体传输的带宽占用高、服务时间长以及不可预知的动态行为<sup>[1]</sup>, 使得现有点播系统存在视频质量和服务器压力等瓶颈问题。通过共享网络中节点的计算、带宽和存储等资源, Peer-to-Peer (P2P) 模式彻底改变了视频点播的服务方式, 显著地减少服务器负载, 提高网络资源利用率。

一般来说, 在 P2P 模式下提供高质量的视频点播服务面临如下挑战: (1) 节点选择, 即如何在节点加入/中断恢复/VCR 操作场景中快速选择合适节点; (2) 媒体资源调度, 即针对不同网络拓扑设计合适的缓存策略, 高效调度视频数据块的需求和传输。

目前很多文献提出各种 P2P 点播系统及相关算法, 试图解决 P2P 模式中存在的上述问题。这些系统和算法根据节点定位方式分为: 集中式目录、分级拓扑、分布式 Hash 表和混合结构<sup>[2]</sup>。P2Cast<sup>[3]</sup> 和 DirectStream<sup>[4]</sup> 是基于树形拓扑结构的点播系统, 采用集中式目录提供节点定位, 分别通过补偿

和最近数据共享的方式进行资源调度传输。基于分级树形拓扑的 P2VoD<sup>[5]</sup> 系统则引入“代”和不定长缓存的概念解决系统抖动问题。借鉴 DHT 特点, 文献[6] 提出基于类 Pastry 的结构化系统方案, 提升系统扩展性, 减少抖动延迟。为有效结合树形拓扑的低延迟性和 Mesh 拓扑的可扩展性, 文献[7] 提出混合拓扑的 P2P 点播系统。

不同于上述算法, 文献[8, 9] 提出基于 BitTorrent (BT) 点播系统的研究思路和方法。BASS<sup>[8]</sup> 在原有 C/S 模式的流媒体服务器体系中添加节点之间的 BT 传输, 通过共享节点资源减少服务器压力, 但各节点不能优先下载后续视频数据。为进一步优化 BT 模式的播放连续度, BitoS<sup>[9]</sup> 提出高低优先级实现文件边下载边播放功能, 并通过修改节点选择算法支持视频连续播放。本文提出的算法与上述两种相似, 但仅借鉴 BT 异步传输机制, 并在此基础上根据不同应用场景提出新的节点选择算法和高低优先级缓存策略。

到稿日期: 2008-03-19 本文受国家自然科学基金(60773193), 国家发改委 CNGI 示范工程资助项目(下一代互联网示范工程 2006 年产业化及应用实验), 华中科技大学电信系基础研究基金(2008)资助。

彭 昭(1983-), 男, 博士研究生, 研究方向为 P2P 流媒体技术; 杨宗凯(1963-), 男, 博士生导师, 研究方向为现代信息技术; 陈京文(1969-), 男, 讲师, 博士, 研究方向为高速交换网络中的服务质量及传输性能分析、流量工程; 朱自珍(1982-), 男, 硕士研究生, 研究方向为 P2P 流媒体技术。

# 1 系统结构

系统由目录服务器  $S_D$ 、内容服务器  $S_C$  和普通节点集  $P$  组成。其中  $S_D$  负责维护全网节点信息,并为节点的请求、VCR 操作、中断恢复提供路由,这些信息包括节点地址、状态以及缓存映射; $S_C$  提供数据来源和调度; $P$  则构成非结构化拓扑,并相互提供数据共享和传输。

## 1.1 定义

假设视频源以恒定比特率(Constant Bit Rate,CBR)方式编码和传输,播放时长为  $T$ ,视频内容的最小存储和传输单元为数据块,每个数据块的播放时间为 1 个单位时间,所有数据块按播放先后顺序从 1-N 编号。Peer 节点  $i$  的 FIFO 缓存队列长度为  $L$ ,对应  $L$ -bit 的缓存映射 BitMap,其中比特位“1”和“0”分别表示相应视频数据是否存在。为了迅速定位节点, $S_D$  同时需要维护每个节点的最小数据块编号  $\min(i)$  和最大数据块编号  $\max(i)$  ( $0 < \min(i) < \max(i) < N$ )。算法其它重要参量定义如表 1 所示。

表 1 参量定义与说明

参量	描述
$N$	视频长度或视频数据块数目
$a_i$	节点 $i$ 的地址
$\min(S), \max(S)$	同一视频流中参与节点的最小和最大缓存数据块编号
$X_A$	备选资源节点集合
$l_i$	节点 $i$ 的现有伙伴连接数
$t_i$	节点 $i$ 的 BitMap 更新时间
$BM_i$	节点 $i$ 的缓存映射, BitMap

## 1.2 数据缓存

节点数据缓存用于节点选择和数据块调度。每个节点加入系统时,需根据数据缓存筛选能提供视频数据的资源节点;数据块请求时,需根据已连接节点伙伴的缓存映射 BitMap 分配请求。为此在节点缓存中定义两种数据块类型:已下载数据块和空缺数据块。Peer 数据块选择依据分段原则,将空缺数据块按照资源节点分为多段,每段对应一个资源节点的请求。具体来说,Peer 请求时获取伙伴节点 BitMap,并与自身 BitMap 比较决定请求数据块的调度。图 1 给出了某个时刻节点可能的数据块请求。由图可知,从 Peer A 获取编号为 3, 7, 8 的数据块,从 Peer B 获取编号为 9, 10, 11 的数据块。

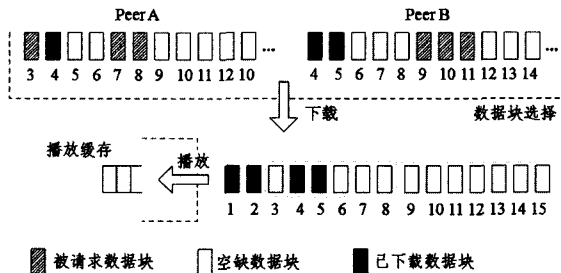


图 1 请求数据块的选择

此外缓存还按照高低优先级划分,表示离播放时刻的远近程度,保证异步传输机制下,视频播放的连续度。

# 2 协议与算法

P2P 异步传输点播系统的协议和算法涉及 3 个部分。第一部分是节点 BitMap 维护,即节点如何更新 BitMap 信息确

保精确定位资源节点和请求数据块。第 2 部分是节点的动态行为,包括加入、离开和中断恢复。第 3 部分涉及资源节点和视频数据块的选择原则。

## 2.1 BitMap 信息维护

本系统中,为了维护网络结构,保证节点间视频数据的高效准确传输,需在服务器和节点间维护 BitMap 信息。服务端,目录服务器  $S_D$  维护包括所有节点 BitMap 的列表信息  $\langle a_i, l_i, BM_i, t_i \rangle$ 。一般来说,每个节点会以周期  $T_i$  定时更新服务器上的 BitMap 信息。此外这种更新还会发生在节点初始加入、VCR 操作以及中断检测与恢复中。Peer 端,节点维护列表  $\langle a_i, l_i, BM_i \rangle$ ,包括所有连接节点 BitMap 信息。在数据传输和播放过程中,Peer 节点会从该列表选择替代节点取代现有伙伴节点中较差的一个,该操作以一定的疏通周期定时发生。在节点的初始加入时,需要更新获取资源节点的 BitMap。此外由于缓存窗口的滑动,BitMap 也会周期性更新以确保 BitMap 信息准确。

## 2.2 节点动态行为

节点的动态行为包括加入、VCR 操作和中断。

### A. 节点加入

当节点  $i$  加入一个视频会话时,会向目录服务器  $S_D$  发送请求消息  $\langle a_i, n_i \rangle$ ,其中  $a_i$  为地址, $n_i$  为请求数据块编号。 $S_D$  收到请求消息后,根据  $n_i$  对域内节点进行筛选,并返回可用节点列表  $X_A$ 。对于同一个节目源来说, $S_D$  维护该视频会话中节点缓存数据块编号的最小值  $\min(S)$  和最大值  $\max(S)$ 。根据请求数据块编号取值可分为以下两种场景:

(1)  $n_i < \min(S)$  或  $n_i > \max(S)$ ,即不存在能为节点  $i$  提供服务的资源节点。在这种场景中,节点  $i$  只能连接内容服务器  $S_C$  请求视频。如果请求失败,则表明目前服务是不可用的。

(2)  $\min(S) \leq n_i \leq \max(S)$ ,即存在能为节点  $i$  提供服务的资源节点。 $S_D$  筛选满足条件  $\min(j) \leq n_i \leq \max(j)$  的节点集  $\{P_j\}$ ,并随机选择部分节点列表  $X_A$  返回。

### B. VCR 操作

VCR 操作的动态行为可被看作一次新的节点请求,请求数据块编号为  $n_i$ ,且满足  $n_i \neq 1$ ,其余部分与请求操作类似。

### C. 中断

某个节点离开视频会话或崩溃,均会导致视频传输阻塞以及播放中断。为了减少此种情况产生的网络抖动问题,它的伙伴节点应该能够检测并恢复该中断。定义节点在一定时间内没有收到伙伴节点传送过来的任何数据,即可进入中断检测和中断恢复环节。向该伙伴节点发送检测消息,如果判断为中断节点,则选择新的替代节点并向服务器  $S_D$  汇报失败节点信息。

## 2.3 资源节点选择

节点  $i$  从  $S_D$  接收到可用资源节点列表  $X_A$  后,需确定如何选择合适的资源节点。具体的节点选择策略如下:

(1) 最近优先:选择包含高优先级数据块的资源节点。考虑视频播放时限,确保视频正常播放的数据块能够尽早获取。

(2) 稀有优先:选择包含稀有数据块的资源节点。考虑部分数据块的稀有特性,减少该类节点离开或中断造成的抖动影响。

(3) 地域优先:选择距请求节点较近的资源节点。考虑实际网络拓扑,减少不必要的骨干网络带宽消耗。

(4) 随机优先:随机选择可用资源节点,主要用于低优先级数据块的获取。

### 3 性能评估

为了验证算法的可行性和主要性能指标,本文采用基于数据包级的 NS-2 仿真<sup>[10,11]</sup>,对以下不同场景中的点播参数进行仿真分析。

(1) 播放连续度:视频停滞时间和次数与总节点数的关系。假设加入系统的总节点数为  $M$ ,每个节点的停滞时间为  $t_i$ ,停滞次数为  $m_i$ ,则播放连续度可由平均停滞时间  $t = \sum_M \frac{t_i}{M}$  和平均停滞次数  $m = \sum_M \frac{m_i}{M}$  衡量。

(2) 启动时延:节点发出加入请求到开始接收第一块视频数据的时间间隔。假设节点  $i$  发起请求的时刻为  $t_{is}$ ,接收视频的时刻为  $t_{ir}$ ,系统的总节点数为  $M$ ,则启动时延定义为  $\tau = \sum_M \frac{t_{ir} - t_{is}}{M}$ 。

(3) 服务器压力:从内容服务器  $S_C$  获取的数据块占总传输数据块的比例。假设从  $S_C$  获取的数据块数目为  $x_i$ ,视频长度为  $N$ ,则服务器压力定义为  $s = \sum_M \frac{x_i}{MN}$ 。

#### 3.1 仿真网络设定

仿真网络采用分层网络拓扑,包括核心路由层和边缘路由层。核心路由层由 6 个核心路由节点组成环形结构,路由节点间的链路带宽为 100Mbps。边缘路由层由 6 个域组成,每个域包含一个核心路由节点和多个普通节点,节点之间的链路带宽为 1.5Mbps。

视频长度为 1000 个数据块,以每秒 1 块的速率传输,且传输速率等于视频播放速率。节点列表  $X_A$  的更新周期为 120 秒,疏通周期为 30 秒。所有节点在 1000 秒内以随机的方式加入视频会话,衡量不同缓存大小  $\eta$  和连接数  $l$  对系统性能的影响,其中  $\eta$  定义为缓存大小与视频长度的比值。

#### 3.2 缓存大小

为了衡量不同缓存大小对播放连续度、启动时延和服务器压力的影响,仿真中设定节点连接数  $l=5$ 。图 2 给出了不同缓存大小和节点总数下系统性能统计曲线。

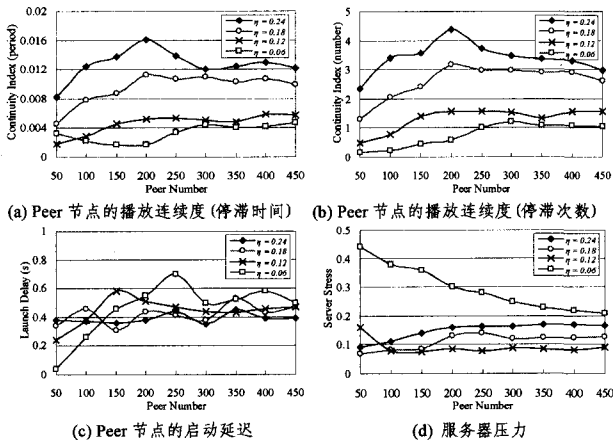


图 2 不同缓存大小的系统播放连续度、播放延迟与服务器压力对比 ( $l=5$ )

由图 2(a)和(b)可以看出,随着缓存大小的减少,系统的播放连续度有比较明显的改进。根据算法可知,缓存较大表明可供选择的资源节点较多,请求数据块区域较大,这将延长节点选择时间。图 2(c)表明系统的启动时延很小,且相对于视频长度而言,变化幅度不大,基本上集中在 0.4 到 0.6 秒的区间内。图 2(d)中, $\eta=0.12, 0.18$  和  $0.24$  时,随着  $\eta$  增加服务器压力有所增加,但取值较小。而当  $\eta=0.06$  时,初始服务器压力较大,且随着节点数的增多有下降趋势。其原因在于缓存较小导致可提供的资源有限,较少的节点数难以体现 P2P 共享特性,服务器提供资源相对较大,随着节点数的增多,peer 节点间共享的资源有所增加,从而减轻服务器压力。仿真结果充分验证了算法的可行性和可扩展性。

#### 3.3 连接数

该场景仿真主要分析节点连接数对点播参数的影响,设定缓存大小  $\eta=0.12$ 。图 3 给出了不同节点连接数和节点总数下系统性能的统计曲线。

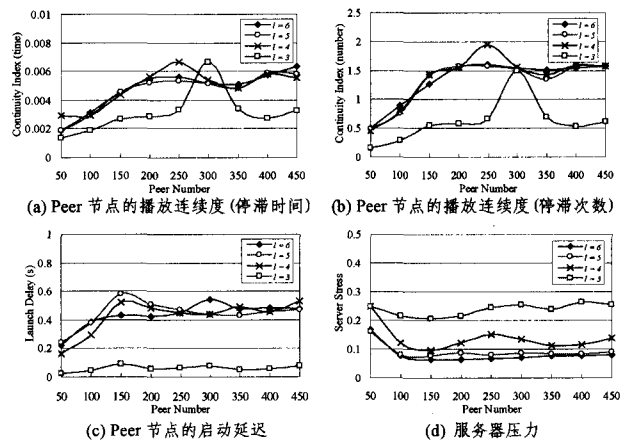


图 3 不同连接数的系统播放连续度、播放延迟与服务器压力对比 ( $\eta=0.12$ )

由图 3(a)和(b)可知在  $l=4, 5$  和  $6$  时,系统的播放连续度曲线几乎一致,且趋近于同一稳定值;而  $l=3$  时,系统连续度较好,但起伏较大。一般来说,资源节点更新需要一定的周期,在连接数较少时,不连续的数据传输可能导致较多的中断次数和较长的中断等待时间。在图 3(c)中,启动时延在  $l=4, 5$  和  $6$  时无太大变化,较为稳定;而在  $l=3$  时,启动时延较少,此种情况服务器相对来说会提供更多服务。图 3(d)明显反应这一状况,随着连接数的增大,服务器压力有所降低,但  $l \geq 4$  时这种变化较小,且处于较低的值。仿真结果充分显示连接数、节点数与点播参数的关系,且在  $l=4$  时可以达到比较稳定的结果。

**结束语** 本文参考 BT 算法提出基于异步传输机制的 P2P 点播算法。通过对缓存大小和节点连接数的仿真分析,验证该算法的可行性。仿真结果表明本算法具有很好的播放连续度、启动时延和扩展性。本算法的后续工作,将基于现有体系框架和算法研究高质量视频传输、骨干网络带宽优化、不同能力节点的自适应调度问题,进一步提升网络吞吐量和带宽利用率。

### 参考文献

[1] Yiu W, Jin X, Chan S. Distributed Storage to Support User In-

teractivity in Peer-to-Peer Video Streaming//Proc. of IEEE International Conference on Communication. June 2006

- [2] Yiu W-P, Xing J, Chan S-H G. Challenges and Approaches in Large-Scale P2P Media Streaming [J]. *IEEE Multimedia*, 2007, 14(2):50-59
- [3] Guo Y, Suh K, Kurose J, et al. P2Cast: peer-to-peer patching scheme for VoD service [A]//Proc. WWW' 03. Budapest, Hungary, May 2003
- [4] Guo Y, Suh K, Kurose J, et al. A Peer-to-Peer On-Demand Streaming Service and Its Performance Evaluation [A]//Proceedings of IEEE International Conference on Multimedia & Expo (ICME). Baltimore, MD, July 2003
- [5] Do T T, Hua K A, Tantaoui M. P2VoD: Providing Fault Tolerant Video-on-Demand Streaming in Peer-to-Peer Environment [A]// Proceedings of IEEE International Conference on Communications (ICC). Paris, France, June 2004
- [6] Yin Zuoning, Jin Zuoning. DHT Based Collaborative Multimedia Streaming and Caching Service [A]//TENCON 2005 IEEE Region 10. Nov. 2005:1-6
- [7] Zhou M, Liu J. A Hybrid Overlay Network for Video-on-Demand [A]//Proc. IEEE Int'l Conf. Comm. (ICC). IEEE Press, 2005: 1309-1311
- [8] Dana C, Li D, Harrison D, et al. Bass: Bittorrent assisted streaming system for video-on-demand//International Workshop on Multimedia Signal Processing(MMsP). IEEE Press, 2005
- [9] Vlavianos A, Iliofotou M, Faloutsos M. BiToS: Enhancing BitTorrent for Supporting Streaming Applications//25th IEEE International Conference on Computer Communications. INFOCOM 2006. April 2006:1-6
- [10] He Q, Riley G, Raj H, et al. Mapping Peer Behavior to Packet-level Details: A Framework for Packet-level Simulation of Peer-to-Peer Systems [C]// Proceeding of 11th IEEE/ACM Intl. Symposium on Modeling, Analysis and Simulation of Computer Telecommunications Systems. Florida, USA, 2003
- [11] Packet-level Peer-to-Peer Simulation Framework and GnutellaSim [EB/OL]. <http://www.cc.gatech.edu/computing/compass/gnutella/>. 2003

(上接第 34 页)

- [49] Miller B L. Noise, Sampling, and Efficient Genetic Algorithms [D]. Ph. D. Fissertation. Urbana-Champaign, Urbana, IL. Dept. Comput. Sci., Univ. Illinois, 1997. Available at: TR 97001
- [50] Das I. Robustness Optimization for Constrained Nonlinear Programming Problem [J]. *Eng. Opt.*, 2000, 32(5):585-618
- [51] Ong Y S, Nair P B, Lum K Y. Max-Min Surrogate-assisted Evolutionary Algorithm for Robust Aerodynamic Design [J]. *IEEE Transactions on Evolutionary Computation*, 2006, 10(4):392-404
- [52] Branke J. Efficient Evolutionary Algorithms for Searching Robust Solutions [C]// Parmee I C, ed. *Adaptive Computing in Design and Manufacture (ACDM 2000)*. Springer, 2000: 275-286
- [53] Branke J, Schmid C. Faster Convergence by Means of Fitness Estimation [J]. *Soft Computing*, 2005, 9(1):13-20
- [54] Parmee L C. The Maintenance of Search Diversity for Effective Design Space Decomposition Using Cluster-oriented Genetic Algorithms (Cogas) and Multi-Agent Strategies (Gaant) [C]//Proceedings of the Second International Conference of Adaptive Computing in Engineering and Control. 1996:128-138
- [55] Ray T. Constrained Robust Optimal Design Using a Multiobjective Evolutionary Algorithm [C]//Proc. Congr. Evol. Comput. 2002:419-424
- [56] Teich J. Pareto-front Exploration with Uncertain Objectives [C]//Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO-01). 2001:314-328
- [57] Hughes E J. Evolutionary Multi-objective Ranking with Uncertainty and Noise [C]//Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO-01). 2001:329-343
- [58] Deb K, Gupta H. Introducing Robustness in Multi-objective Optimization [T]. KanGAL Report No. 2004016
- [59] Deb K, Gupta H. Searching for Robust Pareto-optimal Solutions in Multi-Objective Optimization [C]// Proceedings of the Third International Conference on Evolutionary Multi-criterion Optimization (EMO-2005). Guanajuato, Mexico. Lecture Notes on Computer Science 3410, 2005:150-164
- [60] Barrico C, Antunes C H. Robustness Analysis in Evolutionary Multi-objective Optimization-with a Case Study in Electrical Distribution Networks [C]// II European-Latin-American Workshop on Engineering Systems (SELASI II). Porto, Portugal, 2006
- [61] Barrico C, Antunes C H. A New Approach to Robustness Analysis in Multi-objective Optimization [C]// Proceedings of the 7th International Conference on Multi-objective Programming and Goal Programming (MOPGP'06). Tours, France, June 2006
- [62] Barrico C, Antunes C H. Robustness Analysis in Multi-objective Optimization Using a Degree of Robustness Concept [C]//Proceedings of the 2006 IEEE Congress on Computational Intelligence (WCCI 2006). Vancouver, Canada, July 2006:6778-6783
- [63] Luo Biao, Zheng Jinhua. A New Methodology for Searching Robust Pareto Optimal Solutions with MOEAs [C]//Proceeding of the 2008 IEEE Congress on Evolutionary Computation (IEEE CEC 2008). HongKong, China, June, 008:580-586
- [64] Luo Biao, Zheng Jinhua. Efficient MOEAs with an Adaptive Sampling Technique in Searching Robust Optimal Solutions [C]//IEEE Proceeding of the 7th World Congress on Intelligent Control and Automation (WCICA7). Chongqing, China, June 2008:117-123
- [65] Pänke I, Branke J, Jin Y. Efficient search for robust solutions by means of evolutionary algorithms and fitness approximation [J]. *IEEE Transactions on Evolutionary Computation*, 2006, 10(4):405-420
- [66] Pänke I. Efficient search for robust solutions by means of evolutionary algorithms and fitness approximation [D]. Karlsruhe: Institute for Angewandte Informatics and Formale Beschreibungsverfahren University at Karlsruhe, 2004
- [67] Deb K, Gupta H. A constraint handling strategy for robust multi-criterion optimization [T]. KanGAL Report No. 2005001. 2005
- [68] Handa H. Fitness Function for Finding out Robust Solutions on Time-varying Functions [C]//Proceedings of Genetic and Evolutionary Computation Conference (GECCO'06). Seattle, Washington, USA, July 2006:1195-1200