

基于区域化的 P2P 流媒体直播系统模型

宋启昌^{1,2}, 胡君¹, 王栋¹

(1. 中国科学院长春光学精密机械与物理研究所, 长春 130033; 2. 中国科学院研究生院, 北京 100039)

摘要: 基于树状拓扑协议及扩展的模型和基于 Gossip 协议的模型在媒体数据分发方面存在缺陷。针对该问题, 提出一种基于区域化的 P2P 流媒体直播系统模型, 采用动态自适应的区域化自治机制, 在大规模的流媒体直播网络系统中降低系统管理控制负载的开销, 优化系统带宽和时延, 提高了流媒体数据资源的利用率。

关键词: P2P 流媒体; 树状拓扑协议; Gossip 协议; 区域化

Model of P2P Streaming Media Live System Based on Regionalization

SONG Qi-chang^{1,2}, HU Jun¹, WANG Dong¹

(1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033;

2. Graduate University of Chinese Academy of Sciences, Beijing 100039)

【Abstract】 Considering the shortcomings of the P2P streaming media technology, which based on the model of the tree topology and extensions and Gossip protocol in media data distribution, the model of P2P streaming media live system based on regionalization is proposed, and a dynamic self-governing regionalization mechanism is used. In a large-scale live streaming media network system, the new model reduces the probability of network congestion, and optimizes bandwidth and latency of the system, while improving the streaming data resource utilization.

【Key words】 P2P streaming media; tree topology protocol; Gossip protocol; regionalization

1 概述

随着 Internet 的快速发展, 网络中的流媒体业务在视频直播领域中占据了相当重要的角色。然而随着人们对视频直播的质量和时延要求的不断提高, 传统的 C/S 网络覆盖模式由于其本身体系结构的限制, 很难满足现有流媒体视频直播业务的要求。为了提高流媒体应用的 QoS, 人们提出了 P2P 流媒体技术^[1]。但随着网络流媒体数据的急剧增加, P2P 流媒体技术的应用也面临困境, 特别是在流媒体数据分发^[2]方面。目前 P2P 流媒体数据分发模型主要可分为 2 类: 基于树状拓扑协议及扩展的模型和基于 Gossip 协议的模型^[3]。本文通过对上述 2 类模型的分析, 针对它们在媒体数据分发方面的缺陷, 提出一种基于区域化的 P2P 流媒体分发模型。该模型将所有的节点划分为接入节点和普通节点, 根据源服务器的节目列表由接入节点对普通节点进行区域化自治管理。

2 基于树状拓扑协议及扩展的模型

在基于树状拓扑协议及扩展的模型中, 拓扑结构上的节点有明确定义的关系, 即“父节点-子节点”关系, 树的根节点是源节点。该模型又分为单组播树模型和多组播树模型。

2.1 单组播树模型

单组播树模型是一种常见的流媒体视频直播模型, 其网络拓扑如图 1 所示。

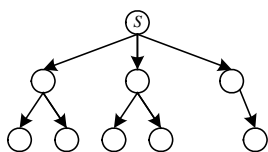


图 1 单组播树模型的拓扑结构

2.1.1 数据分发

在单组播树结构中, 数据是通过一棵组播树传输的。树的根节点是组播源, 每个非叶节点从自己唯一的父节点得到全部数据, 再复制转发给自己所有的子节点, 叶节点只从父节点得到数据, 不再复制转发。当组播树中的非叶节点退出时, 它的子节点将暂时得不到数据, 这时系统需要尽快重建连接, 保证所有节点都在组播树中。

2.1.2 典型单组播树模型及缺陷

在单组播树中, 节点离根节点越远, 数据的时延就越大, 因此, 树的深度应该尽可能小, 另外每个节点的有限输出带宽限制了节点的宽度, 理想的组播树是在深度和宽度之间能够有效地平衡。ZIGZAG 模型^[4]是典型的单组播树模型, 它定义了一整套完整的构建规则, 保证了树的深度维持在 $O(\log N)$, N 为单组播树中的节点数, 能够有效地构造一棵平衡的组播树。ZIGZAG 模型可能是最自然的方法, 不需要复杂的视频编码算法, 结构优化容易, 但 ZIGZAG 模型还是存在单组播树不可避免缺陷:

(1) 公平性差。所有的叶子节点都不需要转发数据, 而中间节点至少需要转发 2 份数据到它的子节点, 同时叶子节点的增长速度明显比中间节点快, 即基于单组播树的分发 P2P 网络的负载是不均衡的。

(2) 稳定性低。每个节点仅与它的父节点相连, 一旦父节

作者简介: 宋启昌(1983—), 男, 硕士研究生, 主研方向: P2P 流媒体, 网络优化算法; 胡君, 高级工程师; 王栋, 助理研究员、博士

收稿日期: 2009-06-14 **E-mail:** songqichang8301@163.com

点出现故障,那么这个节点和它所有的后续节点将无法得到数据,重新恢复需要一定的时间,对于实时流媒体来说,将造成难以估量的损失。

2.2 多组播树模型

面对单组播树的一些缺陷,人们又提出了多组播树模型,其网络拓扑如图 2 所示。

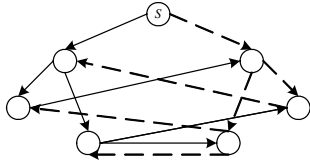


图 2 多组播树模型的拓扑结构

每个节点组播的完成依靠覆盖单播树,与单播树一样,中间节点需要转发数据到它的子节点,而叶子节点无须做这项工作,在图 2 中以实线箭头和虚线箭头区分。通过这种方式构建多个组播树的视频直播流媒体网络方案都建立在多表述编码(MDC)的基础上,采用 MDC 编码的媒体流,分成多个层分别传输于多棵树上,每个节点从它所在的多棵树上获取数据,然后再将各层数据整合,还原成可以播放的媒体数据。树的构造要求每个节点只在某棵树中为中间节点,而在其他的树中都为叶子节点。因此,多组播树的主要缺陷是这种设计很难优化,即使是在中心拓扑下也很困难,很可能使整个 P2P 网络处于异常复杂的环境中。

3 基于 Gossip 协议的模型

在基于 Gossip 协议的模型中,所有节点都处在一个无规则的网状结构中,节点随机地给系统中的部分节点发送消息,每个接收到消息的节点继续向其他节点发送消息,重复这个过程,直到消息被发送到系统中的所有节点,并通过 Gossip 协议维护系统中其他部分节点的视图,通过一定的调度算法在节点之间交换数据,每一个节点既是数据的接收者又是数据的提供者,其网络拓扑如图 3 所示。

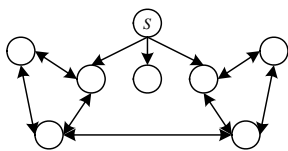


图 3 Gossip 协议模型拓扑结构

3.1 数据的分发

DONET 模型^[5]是典型的基于 Gossip 协议的模型。在 DONET 中节点的伙伴及伙伴之间数据的传输方向并不固定,伙伴之间根据各自缓存的数据情况进行数据交换,因此,节点和伙伴需要相互知道所缓存的数据的内容。视频数据被分割成相同大小的片断,用一个缓存映射 BM(Buffer Map)来表示节点中是否拥有某个片断的数据。节点和伙伴通过不断交换 BM 来了解相互间的缓存情况。

数据调度算法的描述为:节点首先获取只有一个提供的数据片断,然后再获取有 2 个提供者的片断,依此类推,当有多个提供者的片断时,在保证播放的 Deadline 之前获取该片断的情况下,选择带宽最大的提供者。算法被周期性地执行更新调度。调度完成后,同一个提供者的数据片断被表示成 BM 的形式送给相应的提供者,提供者通过 RTP 实时传输协议传输数据,这样就完成了媒体数据的分发。

3.2 DONET 的优缺点

相对于树型结构, DONET 模型的拓扑结构比较简单,

每个节点的数据来源并不依赖于某个特定的父节点,系统有更好的扩展性和可靠性,同时也提高了资源的利用率,优化了负载。但由于采用了随机调度和选择伙伴节点的洪泛机制传递消息,随着网络规模的扩大,流量急剧增加,导致网络拥塞,使带宽和时延的优化比较困难。

4 区域化 P2P 流媒体直播分发模型

单组播树、多组播树及网状结构视频直播流媒体网络系统在以下几方面的比较描述如表 1 所示。

表 1 3 种视频直播流媒体网络系统比较

类别	服务规模	控制开销	管理复杂度	上行带宽	可靠性	自适应性
单组播树	任意	较低	较低	较低	较差	较差
多组播树	大中型	较高	高	适中	很好	较好
网状结构	中小型	高	适中	高	好	好

除了上述差别以外,在其他方面也存在差别。在这 3 种 P2P 网络结构中只有单组播树结构易于优化,其余 2 种优化困难,但在网络的负载均衡方面,单组播树结构就明显不如后 2 种,后 2 种能达到较高的传输速率,更充分利用了各个节点的资源,但实现起来较为困难。

4.1 区域化流媒体直播模型的提出

流媒体直播系统的数据传输要求带宽高、时延小、可靠性高以及自适应能力强,上述的 2 类数据分发模型由于自身的缺陷,在流媒体直播网络规模很大时就很难同时满足这几方面的要求。因此,本文提出一种区域化流媒体直播模型。它根据视频直播的节目,将除源节点以外的其他所有节点按照节点加入算法划分成以节目为单位的虚拟式自治区域(VAS),在每个区域中有 2 类节点:接入节点(IN)和普通节点(PN)。接入节点在每个区域只有一个,接入节点是每个区域的获取媒体数据的接口,主要负责接收和分发数据流及普通节点的管理,相当于数据源节点,但它不具备源节点管理节点信息的能力。区域化模型的网路拓扑如图 4 所示。

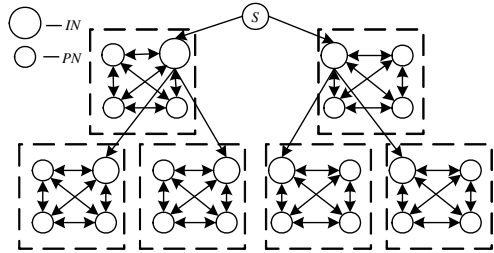


图 4 区域化模型的拓扑结构

4.2 新节点的加入

当有新节点 N 请求加入系统时, N 向源服务器提出加入申请,源服务器根据节点 N 请求的节目信息,查询接入节点管理信息表 IN_LIST,找出可以获取该节目的所有接入节点,如果 IN_LIST 中有相关的接入节点,源服务器根据区域性能评估模型对所有相关的接入节点进行性能评估,让节点 N 尽最大努力加入性能最优的虚拟区域,同时在源服务器中管理该区域普通节点的信息表 PN_LIST 中以及接入节点中普通节点管理信息表 IP_LIST 中添加节点 N 的信息。如果源服务器对所有的相关区域的性能评估都达不到要求或 IN_LIST 中没有相关的接入节点,则节点 N 被定义为接入节点,加入到系统中,同时更新 IN_LIST 表,添加节点 N 的信息。

区域性能评估的依据如下:

(1)接入节点的计算处理能力、内存情况和上传带宽大小;

(2)区域负载状况以及虚拟区域所允许的节点数是否达到上限;

(3)邻近原则,即查看某虚拟区域是否有邻近的节点存在。

只有满足前2个依据的要求,新普通节点才可以加入到该虚拟区域,而邻近原则是在满足要求的所有虚拟区域中选择最优加入区域的依据。

4.3 接入节点的管理

接入节点是每个虚拟区域的源节点,接入节点按照 ZIGZAG 模型构成一个单组播树。当一个节点被定义为接入节点时,它在源服务器中的接入节点管理信息表 IN_LIST 中加入节点信息,并周期性地与源服务器进行通信,把当前该虚拟区域内所有节点的状况反馈给源服务器,更新 IN_LIST 的参数信息以及该区域普通节点管理信息表 PN_LIST 的参数信息,以便有新的节点加入时进行区域性能评估。

当接入节点正常离开时,它向源服务器发出退去请求,源服务器根据该虚拟区域内的普通节点的性能指标选举出一个普通节点为接入节点,并更新 IN_LIST 和 PN_LIST,如果该区域内除了接入节点已没有其他普通节点,则采用 ZIGZAG 模型的节点正常离开处理机制进行处理。

当接入节点因网络故障或其他原因异常离开时,源服务器如果在规定的几个连续周期内没有收到该区域的接入节点的反馈信息,则认为该接入节点已经失效,然后源服务器按照接入节点正常离开的处理机制对该区域进行处理。

4.4 普通节点的管理

普通节点是每个虚拟区域的主体,普通节点按照 DONET 模型构成一个无结构的网状结构,由该虚拟区域的接入节点统一管理。

当普通节点加入到某个虚拟区域后,该普通节点向该虚拟区域的接入节点发出获取其他节点存在信息的请求,接入节点收到请求后返回该区域内普通节点的存在信息列表 PL_LIST,然后该普通节点按照 DONET 模型的节点管理机制和 PL_LIST 表进行处理,区域内每一个节点既是数据的接收者又是数据的提供者。

在虚拟区域内部,节点间的数据的表示和交换以及数据的调度算法,采用 DONET 模型的数据分发机制进行处理。

4.5 模型算法描述

区域化流媒体直播模型算法描述如下:

(1)客户端节点 N 从源服务器端 S 下载视频直播节目列表,然后选择自己需要的节目向 S 发送连接建立请求。

(2)源服务器节点 S 收到节点 N 的连接建立请求后,通过调用节目映射搜索函数 $SearchIN$ 从接入节点管理信息表 IN_LIST 中获得此节目所对应的接入节点对象数组 IN_ARRAY。

(3)如果 IN_ARRAY 为 NULL,转到步骤(4),否则,通过调用性能评估模型算法对数组 IN_ARRAY 中的接入节点进行评估,如果存在满足要求的接入节点 IN ,则转到步骤(5),否则,转到步骤(4)。

(4)请求节点 N 被定义为新接入节点 N_{IN} ,按照单组播树 ZIGZAG 模型的构建算法加入到系统中,即将节点 N_{IN} 信息添加到接入节点管理信息表 IN_LIST 中。转到步骤(9)。

(5)源服务器 S 根据步骤(3)得到接入节点 IN 信息后, S

向请求节点 N 返回该接入节点 IN 的信息,并同时分别向接入节点 IN 和请求节点 N 发送授权认证信息。此时 IN 节点是此虚拟区域 VAS 的源节点。

(6)请求节点收到源服务器 S 给它返回的接入节点 IN 的信息后,向接入节点 IN 发送连接建立请求,同时附加授权认证信息。

(7)接入节点 IN 收到请求节点 N 的连接请求时,在接入节点的普通节点管理信息表 IP_LIST 中添加节点 N 为此虚拟区域 VAS 的普通节点,并向源服务器发送节点 N 的加入信息,更新源服务器的 IN_LIST 表以及该接入节点对应的普通节点管理信息表 PN_LIST,同时接入节点还要向节点 N 返回该虚拟区域内的其他普通节点信息和节点 N 的代理节点。

(8)节点 N 接收到接入节点 IN 返回的数据后,按照 DONET 模型算法,获得初始的伙伴节点列表。

(9)如果节点 N 的加入是由步骤(4)跳转过来的,它采用 ZIGZAG 模型的数据分发处理机制获取流媒体数据,否则,采用 DONET 模型的数据分发机制,与其他伙伴节点之间进行流媒体数据的调度。

5 性能分析

为检测区域化流媒体模型的网络性能,在由8台PC机组成的机群上部署 PeerSIM 仿真软件,通过实现上述区域化流媒体模型构建算法来模拟实验,对原服务器控制负载和流媒体数据传输的绝对延迟进行分析。PC 机硬件配置为: Intel Core(TM)2 Duo CPU, 1 GB 内存,操作系统为 Windows XP。

5.1 源服务器控制负载分析

源服务器控制主要包括新节点的加入请求处理以及接入节点管理信息表 IN_LIST、普通节点的信息表 PN_LIST 和接入节点中普通节点管理信息表 IP_LIST 的维护等。在仿真实验中,设定虚拟区域内最大允许的节点数为30个,通过测试给出了在该 P2P 网络中有不同节点数时源服务器的控制负载和流媒体数据的传输负载的比值,如图5所示。

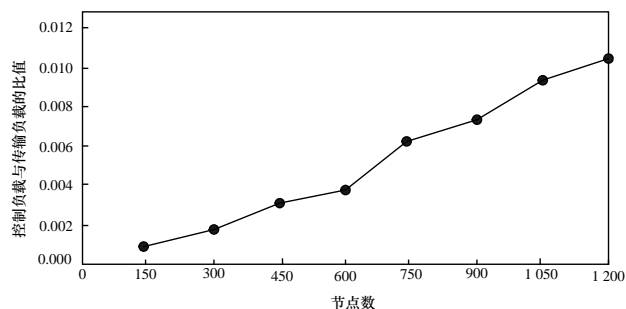


图5 源服务器控制负载与数据传输负载的比值

可以看出,在模型中源服务器的控制负载和流媒体数据传输负载的比值很小,在节点数为1200时比值小于0.012。

5.2 流媒体数据传输的绝对延迟分析

流媒体数据传输的绝对延迟是指流媒体数据在伙伴节点间交换数据时从一个节点到另一个节点的总传输延迟,反映了流媒体数据传输的实时性。在仿真实验中,设定虚拟区域内最大允许的节点数为50个,通过节点数的增加对该模型和 DONET 模型的数据传输延迟进行测试,测试结果见图6。可以看出,在节点数目为1200时区域化流媒体模型的数据传输的绝对延迟约为175 ms,延迟较小,而 DONET 模型中的数据传输绝对延迟平均要比区域化流媒体模型的绝对延迟高出20 ms左右。

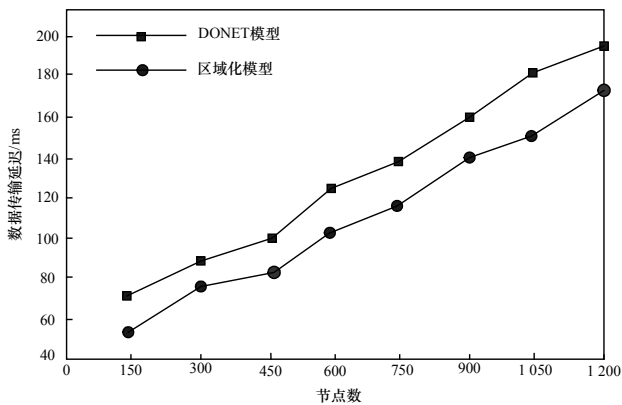


图6 区域化流媒体模型和DONET模型数据传输延迟对比

在此构建算法下的区域化流媒体直播模型充分发挥了ZIGZAG单组播树模型和DONET网状模型的优势。在视频直播网络规模越来越大的情况下,位于上层的接入节点利用ZIGZAG模型使整个系统易于优化,同时减小了控制的开销和管理的复杂度,接入节点中的叶子节点也需要转发数据,提高了数据资源的利用率,而位于下层的普通节点在小规模的虚拟区域内利用DONET模型能最大程度上避免因网络规模扩大产生网络拥塞而造成带宽和时延难以优化的问题。

但该模型同时也存在一些问题,加重了部分节点的负担,特别是接入节点的性能指标直接影响整个系统的发挥,本文通过区域性能评估模型,限制了虚拟区域内普通节点的数量,间接地对接入节点进行了优化,减少了接入节点对整个系统的不利影响。

(上接第111页)

不能确定基于预测模型与差分阈值的能量汇报融合融合机制比普通的汇报机制更优;在7 h~92 h之间,由于节点A的ADC降幅为581,而B的降幅为593,因此节点A利用基于预测模型与差分阈值融合机制比普通的汇报机制节省12个ADC值。为说明节省的12个ADC值的意义,先考虑实验中基于普通传输机制的节点B的单位能耗值的分布,如图5所示。

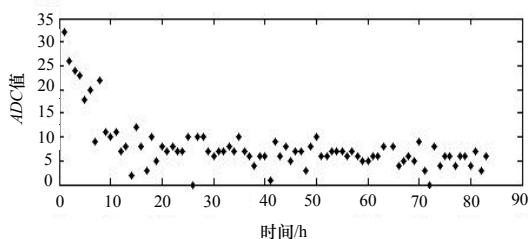


图5 节点单位时间的能耗量分布

从图5可以看出,节点的单位能耗值大都分布在3~10之间,所以,在普通情况下,节点A可以将其工作时间延长1 h~4 h,占其整个工作时间(92 h)的1%~4.3%。

在本实验中,感知节点的能耗主要包括4个模块:收发数据,存储数据,获取感知数据,微控制器耗能。虽然节点减少了耗能多的发送操作,却相应的增加了控制器的运算量,这也是节能效果不是很明显的原因之一。实验3并没有建立在一个复杂的网络上,节点没有进行能耗较大的接收、转发数据操作,这是节能效果不明显的又一原因;由于节点能量的有限性,能够只从能量汇报这一个操作上延长节点1%~4.3%的工作时间,其效果还是值得肯定的。

6 结束语

本文分析了目前主流的流媒体模型ZIGZAG和DONET的优缺点,提出区域化P2P流媒体直播系统模型及构建算法,并对模型的系统性能进行了仿真实验。实验结果表明,源服务器的控制开销占系统总开销的比重较小,数据传输的绝对延迟比DONET降低了20 ms左右。但该模型在提高接入节点性能方面还需做进一步研究。

参考文献

- [1] 马凌霄. 基于P2P网络的流媒体技术研究[D]. 杭州: 浙江大学, 2005.
- [2] Stegmaier B, Kuntschke R, Kemper A, et al. Adaptive Query Processing and Optimization in Streaming P2P Environments[M]. New York, USA: ACM Press, 2004.
- [3] 龚海刚, 刘明. P2P流媒体关键技术的研究进展[J]. 计算机研究与发展, 2005, 42(12): 2033-2040.
- [4] Tran D A, Hua K A, Do T T. ZIGZAG: An Efficient Peer to Peer Scheme for Media Streaming[C]//Proc. of IEEE INFOCOM'03. San Francisco, CA, USA: [s. n.], 2003.
- [5] Zhang Xinyan, Liu Jiangchuan, Li Bo, et al. CoolStreaming/DONet: A Data-driven Overlay Network for Peer-to-peer Live Media Streaming[C]//Proc. of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Miami, FL, USA: [s. n.], 2005.

编辑 顾姣健

4 结束语

本文通过基于预测模型的能量融合机制进行能量汇报,在节能与保持数据准确度上达到一种平衡,延长了节点的生命周期。下一步工作将在基于簇的复杂网络中进行实验,以比较本方法的节能效果。

参考文献

- [1] Krishnamachari B. Modelling Data-centric Routing in Wireless Sensor Networks[C]//Proceedings of the IEEE Int'l Conf. on Communication. New York, USA: IEEE Computer Society, 2002.
- [2] Rost S. Memento: A Health Monitoring System for Wireless Sensor Networks[C]//Proc. of the 3rd IEEE Communications Society Conf. on Sensor, Mesh and Ad Hoc Communications and Networks. Reston, VA, USA: [s. n.], 2006.
- [3] Deshpande A, Hellerstein J M. Approximate Data Collection in Sensor Networks Using Probabilistic Model[C]//Proceedings of the 22nd International Conference on Data Engineering. Washington D. C., USA: IEEE Computer Society, 2006.
- [4] Kulakov A, Davcev D, Trajkovski G. Application of Wavelet Neural-networks in Wireless Sensor Networks[C]//Proceedings of the 6th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. Washington D. C., USA: IEEE Computer Society, 2005.
- [5] 王振龙, 胡永宏. 应用时间序列分析[M]. 北京: 科学出版社, 2007.

编辑 陈文