

文化网格缩比实验的前端设计与实现

闫瑞瑞¹, 马建国², 周金治²

(1. 广州航海高等专科学校航海系 广州 510725; 2. 西南科技大学信息与控制工程学院 四川 绵阳 621010)

【摘要】提出了在缩比实验环境下文化网格前端的软件方案,使信息源提供的数据在自动加内容标引后,按给定的格式封装按给定的速率并行播出。该方案通过内容标引、并播复用、数据封装、播出控制四个主要功能部件实现。实验结果表明文化网格的原理可行性,并有较高的运行效率。

关键词 网格; 并播; 传输流; 统一内容定位
中图分类号 TN919; TP319 文献标识码 A

Design Decision for Front-End of Scaling Examination of Culture-Grid

YAN Rui-rui¹, MA Jian-guo², ZHOU Jin-zhi²

(1. Dep. of Navigation, Guangzhou Maritime College Guangzhou 510725;

2. Dep. of Control Theory and Control Eng. Southwest University of Science and Technology Mianyang Sichuan 621010)

Abstract This paper gives the software scheme of the front-end under the condition of experiment. The data of the front-end bring UCL (Uniform Content Locator) information automatically, and then is packed as the given format, parallel broadcasted. The design scheme is implemented by four key parts: Uniform Content Locator, parallel broadcast, data packing and controlling of broadcasting. The result of experiment indicates that culture grid is feasible and effective.

Key words grid; parallel broadcast; TS; uniform content locator

文化网格是一种覆盖全国的现代文化环境,是国家规模的信息共享平台,是一种重要的信息资源^[1]。文化网格融合数据广播技术和存储技术,以较低的成本向城乡家庭提供报纸、期刊、网站和课程信息,使“众多用户广泛共享众多的资源”,它是“多对多”的信息共享,而不是网络的“一对一”或“一对多”^[2]。

文化网格的运作分两步:(1)借助覆盖全国的卫星数据广播把文化资源的新鲜网页送到城乡家庭;(2)借助计算机对需求的判断,把其中有兴趣的内容(智能)下载存储。本地存储器长期积累广播网页,作为内容丰富的数据仓储供用户使用。为了使用户计算机能智能下载信息,必须在源端对所播出的节目加入有效的标引信息,为此引入了统一内容定位(Uniform Content Locator, UCL)技术。经长期存储积累形成的当地数据仓储可使用户摆脱广播网络的单向性限制和按节目时间表接收的限制,随时按需取用。本文以网站为例来讨论在缩比实验环境下文化网络系统前端的软件设计与实现。

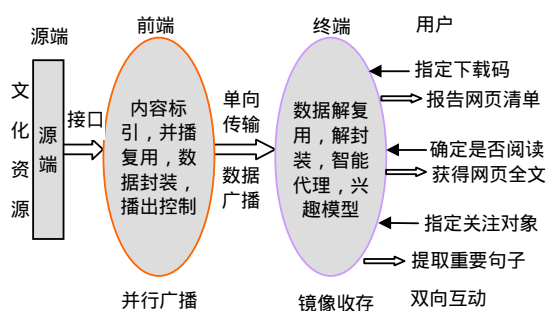


图1 文化网格系统结构

1 系统结构

图1表示文化网格的系统结构。广播“前端”与提供信息内容的“源端”在地理上是分开的,中间通过文件传输协议连接。信源服务器主动发起信息传播。前端整合源端提供的资源,控制信息的播放,通过大规模并播保证每个用户任何时刻都可以得到享用所有资源的实时服务。终端的并存技术是用户计算机全天在线,将符合用户兴趣模型的网页有选择地存入本地存储器,用

收稿日期: 2003-12-09

基金项目: 国家863计划资助项目(2002AA121063); 国家自然科学基金资助项目(60272014); 四川省应用基础研究计划资助项目(03JY029-017-2)

作者简介: 闫瑞瑞(1975-), 女, 硕士, 主要从事新一代信息网络、数据广播方面的研究。

时间来换取带宽。用户在阅读网页内容时,不再直接在广播网络上收取内容,而是用UCL代码,无带宽限制地从本地镜像仓储中读取^[2]。大规模并播是将一个8 MHz的模拟电视频道分为1 024个32 Kbps的子带宽,如果是实时数据流,可用若干32 Kbps的带宽传送。UCL作为资源与用户沟通的桥梁,使网页主动地为用户服务成为可能。万维网使用URL从空间上定位,告诉用户网页存于某服务器的某个目录下;而UCL从内容(语义)上定位,告诉用户该网页包含的是哪方面的内容^[3]。因此,大规模并播技术从空间上整合资源;大容量并存技术从时间上整合资源;而UCL技术从内容上整合资源^[4]。

2 前端软件设计

2.1 实验环境简介

本文实验的前端软件是针对缩比实验环境的,用5个镜像的网站代替前端千个网站到源端的文件传输,所用的镜像网站是www.sohu.com、www.sina.com.cn、www.163.net、www.263.net和www.163.com。前端软件对镜像网站的内容作自动标引、时间复用和并行播出。该软件在局域网环境下开发、调试,已成功移植到DVB-C环境下。在DVB-C下用户可通过有线接收卡接收文件。

2.2 软件方案要点

前端软件的主要功能包括自动标引、并播复用、数据封装和数据调度控制。软件设计的总体思想是,前端软件按时间复用规则组织源端节目(网站内容),对网页自动作统一内容标引;依照自定义协议规则,将要发送的网站文件数据封装成一系列固定格式的分组;将这些分组送入UDP/IP协议栈,以异步广播方式并行发送;前端软件对网站内容作数据调度、带宽分配、差错控制、速率控制和同步控制。前端软件结构如图2。接收端用户按兴趣从UDP/IP协议栈接收不同内容的数据,解复用多路节目,按各网站结构存储,并保存对应内容的标引信息。

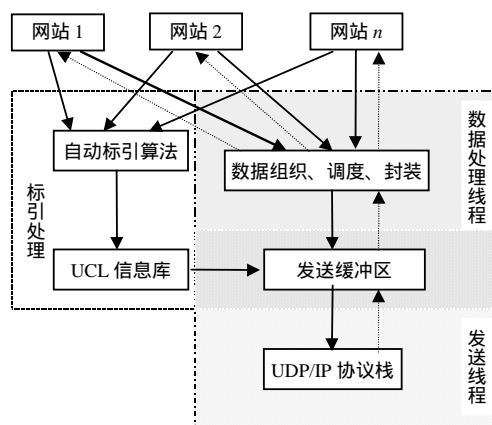


图2 软件结构图

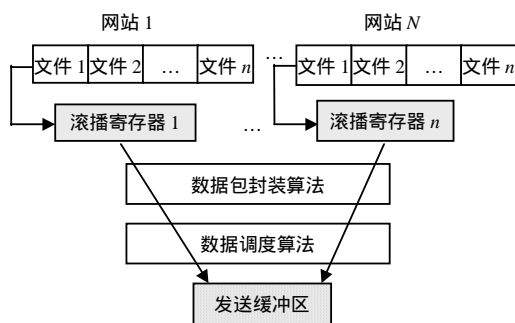


图3 数据的组织

1) 自动标引:不同网站的统一资源定位(Universal Resource Locator, URL)有不同的组织方式,但大体是按照栏目类别以树状结构组织,使提取UCL信息成为可能。以搜狐网和新浪网为例,从URL分析可以得到网站、栏目和文件名;网页文件标题从文件源代码获得;从网站输入到前端的时间得到UCL的时间代码;关键词由用户设置,匹配选择适合的网页。2) 并播复用(对源端网站内容的组织):在缩比实验环境下,源端网站数据的来源采用网站镜像。前端软件对网站数据的组织和传送采用时分复用,以固定的时间片来传送不同网站长度固定的数据包。数据包的结构按照协议定义,如表1所示。网站内部各文件采用滚播,程序中对各网站设立各自的滚播寄存器,按各网站设置的带宽不同,以不同的频度将对应的滚播寄存器的内容调入发送缓冲区。例如,网站1设置的带宽为 1×32 Kbps,网站2设置的带宽为 2×32 Kbps=64 Kbps,对应的滚播寄存器1和2的大小为一个数据包Data字段的长度,即174 Byte,经数据封装算法封装为完整的数据包,即188 Byte。数据调度算法读取数据包的间隔为固定的时间片 x ,根据网站1和2的不同带宽值,读取一次滚播寄存器1之后连续读取两次滚播寄存器2,如此反复所形成的传输流(Transport Stream, TS),如图3所示。3) 数

据封装:协议包结构是结合本文实验的单向大规模并播和内容定位的特点,并参考MPEG-2和DVB的TS包结构定义的,如图4所示。前8个字段与ISO/IEC13818传送流分组的4 Byte包头相同^[5,6]。TS包结构的后5个字段的说明如下。(1) Subject_code:8位字段,指示相同节目标识(Program Identification, PID)内不同栏目的编码值;(2) File_Num:8位字段,指示相同PID和相同Subject_code的数据文件序号;(3) File_Packets:32位字段,记录节目文件的总分组数;(4) Sending_Packet:32位字段,随着每一个具有相同PID、Subject_code和File_Num的传送流分组而增加;(5) Data:数据字节,不足174字节时,以0xFF填充。

以上字段中, PID和Subject_code为选择接收而设置,用户根据自己的兴趣设置网站和栏目,用户终端分析数据包中的PID和Subject_code字段,保存与用户设置码值一致的包,丢弃码值不一致的包。File_Packets及Sending_Packet是为可靠传输而设置的。由于数据广播是单向信道,不可能实行滑动窗口和检错重发机制,因此协议如何解决传输误码就显得尤为重要。本文的实验的解决办法是在数据包中设置File_Packets和Sending_Packet字段,对前端发出的数据包作序号标志,在终端按序号存放包,这种方式可以控制数据包的丢包、重包和包乱序。数据包的封装算法目前采用数据管道方式,将网站文件分成固定长度的分片(174×8 bit),按照自定义的协议包结构封装。4) 数据的调度控制:并行数据广播是以一个预先设定的固有速率 V_0 及其正整数倍数为量化单位,将拥有较高传输速率的整个信道,按照时间复用方式,划分为大量独立的子信道或子通道。一个或若干个子信道或子通道可供不同的信息内容提供者租用,为其提供全天候的、信息可通达所有用户的专用信息通道。固有速率 V_0 的典型速率可为32 Kbps、64 Kbps或128 Kbps。

前端软件的数据调度控制包括线程同步、精确速率控制以及差错控制。

(1) 线程同步:数据处理线程完成数据的组织、调度、封装和填充发送缓冲区,发送线程完成从发送缓冲区读取数据,调用接口函数发送数据。数据处理线程和发送线程共享“发送缓冲区”,可能发生缓冲区读写冲突,以及写到缓存的数据不能被即时取走或多次被取走等错误。因此必须要求严格的线程同步。实现线程同步的办法是采用两个CEvent同步对象 E_1, E_2 ,使两个线程连锁,如图5所示。程序初始化时置 E_2 有效, E_1 无效。数据处理线程执行到写缓存操作时调用WaitForSingleObject()函数检测事件 E_2 的状态, E_2 有效时执行写缓存,在写缓存结束后,将事件 E_1 设置成有信号状态, E_2 复位。在发送线程中调用WaitForSingleObject()函数,检测到事件 E_1 的状态为有效时开始读缓存操作,否则将发送线程挂起。发送线程读缓存结束后,将 E_2 设置成有信号状态, E_1 复位,为数据处理线程写缓存做准备。如此反复,完成一次写缓存紧接一次读缓存的操作,可避免读写冲突,以及两线程进度不一致的问题。(2) 精确速率控制:由于互联网本身的特点,本文的实验在互联网环境下时不能精确地控制速率。在DVB-C环境下,最根本的速率控制点在于控制硬件发送卡的发送速率。发送线程调用发送卡的接口函数,发送函数在返回之前使发送线程处于挂起状态,因此发送速率控制了发送线程的进度。通过对事件对象连锁的方法,转而控制数据处理线程的进度。(3) 差错控制:差错控制包括控制数据包的丢包、重包、包乱序和包错。如前所述,数据的丢包、重包和包乱序是通过在数据包中设置File_Packets和Sending_Packet字段来实现的。而因传输导致的包差错可通过CRC32循环冗余校验来判断。对于有差错的数据包,终端会作出标记,在下一轮的接收中弥补纠正。经过对文化网站的统计表明,多数文化实体每天生成的内容小于百兆,然而若某一网站带宽为32 Kbps,即每4 KB/s,14.4 MB/h,每天推送的字节量可达345 MB,因此在24 h内可以获得轮播多次的机会,足以正确接收,证明纠错方式是可行的。

Sync_byte (8 b)	transport_error_ indicator (1 b)	payload_unit_ start_indicator (1 b)	transport_ priority (1 b)	PID (13 b)	transport_ scrambling_ control (2 b)	adaption_field_ control (2 b)	continuity_ counter (4 b)
Subject_code (8)	File_Num (8 b)			File_Packets (16)			
File_Packets (16)				Sending_Packet (16)			
Sending_Packet (16)				Date(174×86)			

图4 协议字段结构

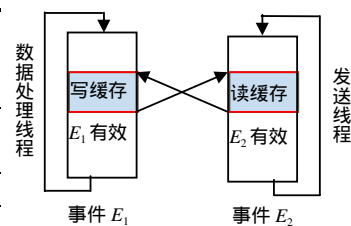


图5 线程同步

(下转第374页)

询发射波与回波,还是由之触发单片机中断再通过软件启停定时器均需要一定的时间(中断方式误差相对要小);(2)单片机的时间分辨率不是太高,如晶振频率为12 MHz,时间分辨率为1 μ s。而引入8254可大大提高测量精度,它用发射波与回波形成的GATE信号直接启停8254的定时器,同时8254具有更高的工作频率,时间的分辨率更高,系统采用2 MHz的工作频率。2) 8254的时标信号由2 MHz的多谐振荡器产生,其振荡频率的稳定性至关重要,电路为由晶振组成的多谐振荡器,振荡频率的稳定性好。3) 在软件上进行相应处理,8254三个定时器同时工作,当只有一次回波时,液位较高,系统仅读入 T_0 时间;当有两次回波时,系统读入 T_0 、 T_1 ,比较两者的差,超过一定范围时,舍去该组数据,否则保存 T_1 ;当有三次回波时,液位相对较低,系统读入 T_1 、 T_2 并比较,正常时,取 T_1 、 T_2 均值。每形成一次发射波,更新一次测量数据,以实现液位变化的动态显示。同时在由所测时间换算为液位时采用如下计算方法:将实测时间值乘以10 cm(标准试块高)再除以测标准试块时的时间值,若直接采用时间乘以速度的算法,运算过程的四舍五入将引起较大的误差,原因在于超声波传播很快(km/s以上),时间上小的舍入反应在距离上会有大的误差。

本文设计的系统已通过调试,运行正常,达到设计要求。

参 考 文 献

- [1] 邱瑞昌,姜久春. 泊车用超声波测距仪的研制[J]. 电子产品世界, 2001, (9): 11.
- [2] 胡汉广. 单片机原理及其接口技术[M]. 北京: 清华大学出版社, 1999.
- [3] 陈光东. 单片微型计算机原理与接口技术[M]. 武汉: 华中理工大学出版社, 1999.

编 辑 漆 蓉

(上接第348页)

3 实验结果

当改变信源为教育信息时,本系统可以有效地为远程教育服务,通过有线电视或卫星传输,能使人们在没有互联网的高山、荒原、海底享用信息,接受教育。本文的实验有良好的平台移植性,通过了局域网和DVB-C平台的验证。经测试,对26 MB的网站内容自动标引,生成的标引信息库为120 KB,占4.5%,平均用时67 s,且标引信息的生成不占发送时间,证明本文的实验效率较高。用户端利用标引信息中的PID和Subject_code字段在物理层控制选择接收,缓解了用户计算机的工作量。经过多次实验,用户端都能灵活控制并正确无误地接收信息。

参 考 文 献

- [1] 李幼平. 广播型文化网格[C]// 第二届中国数据广播论坛论文集, 成都: 2003.
- [2] 马建国,李在铭. 广播型网格[J]. 计算机科学, 2004, 31(8): 5-7, 21.
- [3] 马建国,邢玲,马建国,等. 数据广播中的UCL标引与传输机制[J]. 电子学报, 2004, 32(10): 1 621-1 624; 1 643.
- [4] Ma Jianguo, Liu Guihua, Xing Ling. An agent of data broadcasting based on UCL[C]// IEEE International Conference on Natural Language Processing and Knowledge Engineering, Beijing, 2003.
- [5] 闫瑞瑞,马建国. 数据广播中内容定位技术的软件实现[J]. 电视技术, 2003, (9): 42-44.
- [6] ETSI. ISO/IEC 13818-1: Information technology-generic coding of moving pictures and associated audio recommendation[S] 1994.

编 辑 熊思亮