

卫星广播系统 CMACast 省级接收站缺收数据 补调服务的设计与实现

韩书丽 谭小华 李湘 王鹏

(国家气象信息中心, 北京 100081)

摘要 中国气象局卫星广播系统(CMACast)在广播播发过程中存在接收小站文件错误接收或丢失的情况。建立数据下载中心,基于地面宽带网提供对广播缺收文件的补调下载,可以有效地解决上述问题。数据补调服务客户端完成了数据补调流程的建立,并设计了两级补调的机制以及灵活的可配置方式,在程序实现上采用多级并行处理、基于优先级的动态调度等方式来满足省级补调应用的需求。补调服务客户端已完成在省级通信系统的部署,在日常气象业务中发挥了实际效益。详细介绍了数据补调服务客户端的设计机制与实现方法,并对部署应用情况进行了分析。

关键词 卫星广播 气象资料 缺收 补调

引言

中国气象局卫星广播系统(CMACast)是国内和国际通信系统的重要组成部分,是气象观测资料和产品分发的主要通道^[1]。2012年6月1日,CMACast正式投入业务运行,其数据广播业务覆盖全国各级气象部门、气象部门外用户和国外用户,与地面广域网络系统CMANet协同实现天地一体化的数据传输。目前,CMACast已拥有接收站2562个,日播发数据超过150GB,日播发文件数超过130万个,业务运行以来,发挥了重大作用。

同时,由于受到卫星信道传输误码、接收干扰、雨衰、馈线电缆传输损耗等因素影响,接收小站文件接收错误或丢失情况时有发生,影响了资料接收的完整性。表1、表2分别给出了山西省气象局和北京市气象局(以下称山西局和北京局)CMACast广播接收的统计结果。根据统计结果,在统计时段,两个地方的CMACast接收错误率大部分大于1/10000,最差的时候达到104.5/10000。气象资料收集的完整性对气象业务至关重要,尤其是关键资料(如数值预报产品、自动站资料等)的缺收,将给预报预测和气象服务带来直接影响。

为了提高实时气象资料分发服务质量^[2],有效

表 1 山西局 CMACast 接收文件数及错误率

日期	完整接收	不完整接收	接收错误率/%
20120821	483074	5101	10.45
20120822	485272	1006	2.07
20120823	510050	91	0.18
20120824	496667	83	0.17
20120825	498672	96	0.19
20120826	523187	112	0.21
20120827	525903	1384	2.62
20120828	525840	632	1.20
20120829	510392	133	0.26
20120830	494149	105	0.21

表 2 北京局 CMACast 接收文件数及错误率

日期	完整接收	不完整接收	接收错误率/%
20121021	370304	263	0.71
20121022	356847	161	0.45
20121023	355941	41	0.12
20121024	361093	7	0.02
20121025	358166	137	0.38
20121026	373017	11	0.03
20121027	365440	69	0.19
20121028	374891	194	0.52
20121029	364794	193	0.53
20121030	364142	148	0.41

保障 CMACast 接收站数据收集的完整性,国家气象信息中心组织建设实时气象数据下载平台,通过地面宽带网提供对 CMACast 广播缺收文件的补调支撑。在国家级,实时气象数据下载平台实现对实时气象数据的组织管理,并负责响应补调请求,提供基于 FTP 方式的数据下载。在省级,需要开发一套数据补调服务客户端,实现对缺收数据文件的补调,是整个工作的重点所在。

本文详细介绍了省级数据补调服务客户端的设计、实现和应用效果,依次对设计思路、具体实现及应用效果进行详细描述,并对工作进行了总结,阐述了下一阶段的工作方向。

1 方案设计

1.1 总体思路

对现有通信系统架构及实际业务需求进行综合分析,CMACast 省级接收站缺收数据补调服务客户端(以下简称数据补调服务客户端)的总体目标为:

①完整性:保证广播缺收文件清单的完整性,提高文件补调的成功率,最大程度降低由于广播资料缺收带来的影响。②时效性:设计合理的实现框架,提高程序处理效率,保证补调时效满足省级资料应用需求。③可配置性:支持灵活的可配置方式,满足不同省级用户的应用需求。

1.2 机制设计

根据上述总体目标,结合实际的业务场景和功能需求,对程序机制进行如下设计。

1.2.1 两级补调机制

采用两级补调机制保证数据收集的完整性。

(1)缺收数据文件补调机制。缺收数据文件的补调是整个应用的主体框架,在程序设计中应包含两项基本功能:缺收文件清单的生成及缺收文件的补调,其机制原理如图 1 所示。

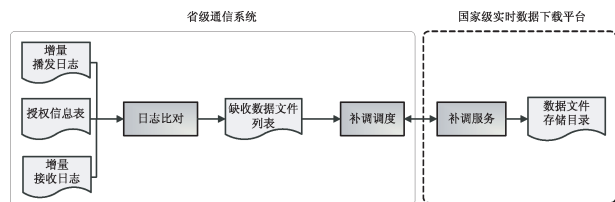


图 1 缺收数据文件补调机制

缺收文件清单的生成依赖于对 CMACast 的日志信息进行分析。这些日志信息包括:主站播发日

志、小站接收日志及授权信息表。其中,主站播发日志由主站播发服务器定时生成,日志中记录了增量周期内主站成功播发的数据文件信息,包括播发时间、文件名、文件大小等;接收日志文件由接收小站服务器定时生成,日志中记录了增量周期内小站接收的数据文件信息,包括播发时间、文件名、文件大小等;授权信息表由接收小站应用程序生成,该文件给出了接收小站通道目录的授权状态,当通道目录的配置更新时,授权信息表也会更新。通过对 3 类日志信息的比对,可得出接收小站接收数据文件的情况。

对于 CMACast 广播播发的数据文件,如果其通道目录为授权打开状态,说明该数据文件为应当接收的文件;而如果接收日志中没有此文件的信息,说明该数据文件播发丢失,应该进行补调。获取到缺收数据文件列表后,省级应用经过补调调度从国家级下载平台下载缺收数据文件,并存放到指定目录。

(2)增量播发日志补调机制。根据前文所述,增量播发日志由 CMACast 主站按照一定的时间间隔 ΔT 产生(ΔT :单位为 min,可配置,取值为 60 的约数),日志文件生成后和数据文件一起下发,在广播播发过程中也有可能出现丢失;播发日志信息是日志比对的基础,播发日志的丢失将直接影响到缺收文件清单的生成,进而影响到补调的完整性。因此,需要设计一级补调来保证播发日志的完整性。图 2 给出了增量播发日志补调的流程图。

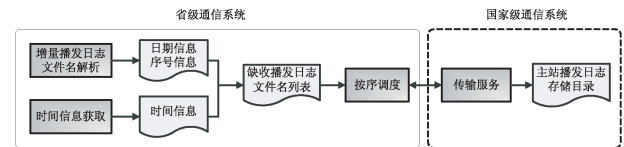


图 2 增量播发日志补调机制

增量播发日志补调的基本原理为:增量播发日志按照格式 `sendYYYYMMDDxxxx.log` 进行命名,其中,YYYYMMDD 代表文件播发日期;xxxx 代表日志序号,4 位编码,取值范围为 $[0001, 0002, \dots, \frac{1440}{\Delta T}]$ 。以 ΔT 取值为 1 min 为例,每天的第 1 个增量日志 `sendYYYYMMDD0001.log` 于当日 00:01:00 生成并广播,日志中记录所有 00:00:00—00:00:59 完成播发的文件信息;第 2 个增量日志

sendYYYYMMDD0002.log 于当日 00:02:00 生成;依此类推,系统定时生成增量播发日志,文件名中依序进行序号编码;最后一个增量日志 sendYYYYMMDD1440.log 于次日 00:00:00 生成。在此规则下,每一个播发日志文件对应一个时间预期,文件名和时间具有强相关性,先后到达的日志文件序号应满足连续性。对播发日志文件名进行解析,通过检测序号的连续性以及与时间预期的匹配可得到缺收播发日志文件名列表。当存在缺收播发日志时,按照序号依次从国家级国内通信系统补调。

1.2.2 补调清单可配置机制

通过日志比对得到的缺收数据文件列表给出的是缺收文件的全集,补调清单可配置机制,提供对补调清单进行过滤的功能。用户可根据本省实际需求及通信系统带宽条件,通过配置实现对数据文件的特殊筛选。设计配置文件的结构如表 3 所示。

表 3 配置文件结构

配置项名称	意义
SWITCH	定义开关状态,“ON”代表开启,“OFF”代表关闭
MINFSIZE,MAXFSIZE	定义文件大小范围,可不定义或者取缺省值
NAME	广播通道名称
FNAME	定义文件名模板

例如,广播通道 SATE_FY2D_L2L3 下播发的是 FY2D 的 2、3 级产品,如果用户仅需补调 FY2D 的兰勃托投影图像产品,则可以在配置文件中进行如下配置:

```
[CHANNEL]
SWITCH=ON
NAME=SATE_FY2D_L2L3
FNAME=Z_SATE_C_BAWX*_P_FY2D*R01*AWX
```

1.3 总体流程

数据补调服务客户端的实现依托于省级通信系统,需要与 3 个外部系统进行交互:国家级通信系统、CMACast 广播系统及国家级实时资料下载平台,其总体流程如图 3 所示。

省级 CMACast 接收小站服务系统将主站播发日志、小站接收日志及授权信息表存放在与省级通

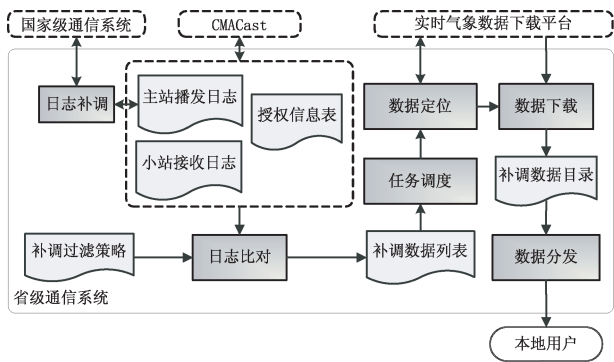


图 3 CMACast 补调总体流程

信系统共享的集群文件系统指定目录中。省级补调服务客户端对存在缺收的主站播发日志从国家级通信系统进行补调,同时,经过日志比对及补调策略过滤生成补调数据列表。

获取到补调数据列表后,客户端按照通道优先级进行任务调度。由于实时资料下载平台目录结构对用户透明,省级应用需要首先发起查询请求,获取缺收文件在下载中心的存储位置,然后进行基于 FTP 方式的下载。成功下载的文件按约定的接口目录进行存储,经省级通信系统按预定分发配置分发给本地用户。

2 具体实现

2.1 Libcurl 传输库

传输库主要用来实现不同网络主机之间的通信,在本系统中,客户端和服务端之间主要存在两方面的通信,分别是基于 HTTP 协议的查询获取补调地址和基于 FTP 协议的文件传输。为了保证通信传输的效率及可靠性,传输库应满足下列要求:

- (1)支持 HTTP 协议、FTP 协议。
- (2)支持用户名和密码认证,保护实时资料下载平台服务器数据安全,方便进行用户权限及用户下载行为管理。
- (3)支持下载文件断点续传,对于下载失败的文件可从断点处继续下载,提高总体下载效率,尤其是对于大文件的下载。
- (4)线程安全,能够满足并发应用的开发。
- (5)接口简单,能有效提高开发效率。

Libcurl 是一个免费的开源传输库,提供基于 C 的应用程序接口,支持文件的上传与下载,是一款综合性的传输工具。因为它能够很好地满足对传输库

的要求,所以在本应用中采用 Libcurl 作为传输库。

Libcurl 中利用回调函数来实现传输功能,对于不同的应用,需定义不同的回调函数。对于基于 HTTP 的应用,在回调函数中实现将数据写进内存,由于在一次会话中,根据数据量的大小,可能会多次调用回调函数,因此在程序中必须做好数据传递指针的处理,否则将出现接收数据的丢失;对于基于 FTP 的应用,在回调函数中实现将数据直接落盘保存成文件。

2.2 并发技术

并发技术指的是对任务进行并行化处理,使得不同任务从逻辑意义上可以在同一时间同时执行,提高任务完成效率。并发包括多进程并发和多线程并发两种^[3]。在数据补调服务客户端编码实现过程中,综合应用了这两种技术,实现了多级的并行处理。

如前文所述,数据补调服务客户端包含两项基本功能:CMACast 日志处理及缺收文件的补调。程序实现时,进行任务拆分,将这两个功能放在两个不同的进程中实现。其中,CMACast 日志处理进程负责生成缺收文件补调清单,而另一个进程负责缺收文件的补调,两个进程互相独立,构成异步的“生产者—消费者”架构。一方面,日志处理进程作为生产者实时生成补调清单并将信息入库,另一方面,补调进程作为消费者实时检索待补调文件信息进行处理。两个进程间采用 MySQL 数据库表作为缓冲区,通过设置状态位来对已完成任务和待完成任务进行识别。

在缺收文件补调进程中进一步细化并发处理的粒度,采用多线程下载技术^[4]。在单线程的下载中,可以直接应用 Libcurl 应用流程框架。而在多线程的应用中,由于线程间共享数据空间,为了避免多个线程共同修改共享区数据导致访问冲突,需要通过互斥锁、信号量同步等方式来保证线程安全。获得资源的线程执行加锁操作,任务执行完后进行解锁,在此期间其他线程如果要访问共享资源需挂起等待。

在线程同步时,要保证程序逻辑的正确性,以避免死锁的产生,编写程序时应注意的问题包括:

(1)保证加锁和解锁操作成对出现,尤其是当程序逻辑较复杂,涉及到 break、continue、goto 等语句跳转时;此外,应避免同一个线程多次执行加锁操作。

(2)尽量减少加锁和解锁之间的代码,简化程序

逻辑,将和共享资源无关的操作放在互斥锁语句外执行。

(3)避免使用复合锁,如果必须用,应注意解锁顺序,保证解锁顺序和加锁顺序相反。

2.3 基于优先级的动态调度

CMACast 卫星广播系统按照通道划分优先级,广播播发时遵循优先级分发机制,优先分发时效要求高的资料。因此,在进行补调任务调度管理时,也应按照资料优先级进行排队,保证优先级高的数据先进行补调。

但是,单队列是先进先出(FIFO, First Input First Output)的模型,不能满足基于优先级的调度。为了解决这个问题,在程序中设计多个队列,每个队列对应一个广播优先级。在进行任务调度时,针对入队操作,根据资料的优先级在相应队列的队尾入队;针对出队操作,优先从级别高的队列的头部取任务。通过这种方式,实现基于优先级的动态调度,调度示例如图4所示。

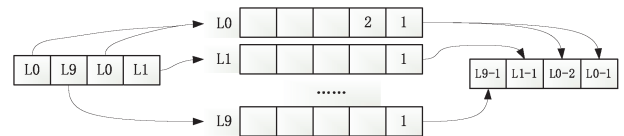


图4 基于优先级的动态调度示例

2.4 容错处理

考虑到进程重启、连通性故障等因素,在程序中考虑下列容错处理情况:

(1)由于网络故障导致文件下载失败时,按照一定的策略在监控界面进行告警,处理进程定位在未成功进行补调的任务节点,待故障恢复后,程序自动从该节点进行处理,保证缺失文件下载的完整性。

(2)新增基于内存的重传任务队列,当出现由于下载平台数据收集出现延迟导致文件下载失败或由于线程异常崩溃退出导致任务未成功完成处理时,将该任务添加至重传任务队列,并按照配置的重传时间间隔、重传次数进行任务的重新执行,提高缺收文件补调的成功率。

(3)当进程重启时,重传任务队列中的信息丢失,为了能够重新获取未完成处理的文件列表并处理,在数据库表中增加了一类状态位的标识,用来表示正在处理任务。进程重新启动后,优先将进程停止前未完成处理的任务进行处理。

(4)按照预定策略经过重补调后仍未成功下载的文件,在省级通信系统监视界面中显示“补调失败”状态,提示业务人员通过其他方式(如,手动补调)获取缺收数据文件。

3 应用情况

目前,补调应用客户端已完成全部功能开发,并部署到省级通信系统实时运行。为了对补调应用性能进行评估,选取山西局和北京局连续 10 天的运行数据,从完整性和时效性两个方面进行分析。

3.1 完整性

图 5 和图 6 分别给出了山西局和北京局 CMACast 补调的总体情况,图中深、浅色垂直矩形分别代表日应当补调文件数、日成功补调文件数,圆点代表日补调成功率,即补调成功率=成功补调文件数/应当补调文件数。

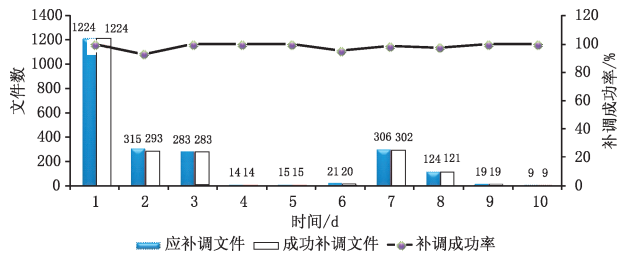


图 5 山西局 CMACast 补调总体情况

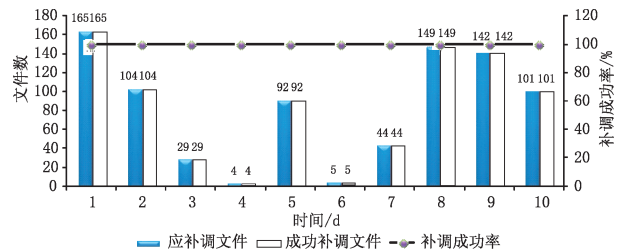


图 6 北京局 CMACast 补调总体情况

根据图示,统计时段内,山西局日补调成功率接近 100%,最差的时候为 93%;北京局日补调成功率为 100%。经分析,山西局补调成功率低于 100% 的原因:新增广播资料在节目表中没有正确进行配置,属于业务执行过程的问题。

3.2 时效性

将文件从国家级主站播发到成功完成补调的时间间隔定义为文件的补调时效,将某时间段内所有文件的补调时效进行平均得到该时段的平均补调时效。

图 7、图 8 给出的是山西局和北京局的日平均补调时效情况。根据图示,在补调时段内,山西局的日平均补调时效在 15 min 之内,北京局的日平均补调时效在 8 min 之内。

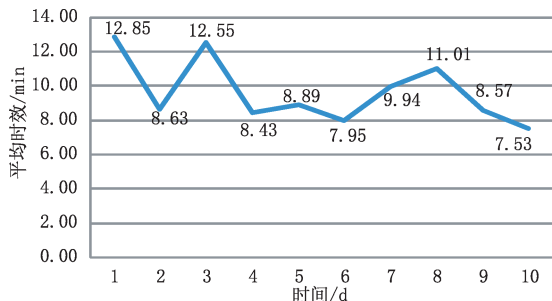


图 7 山西局 CMACast 补调总体时效

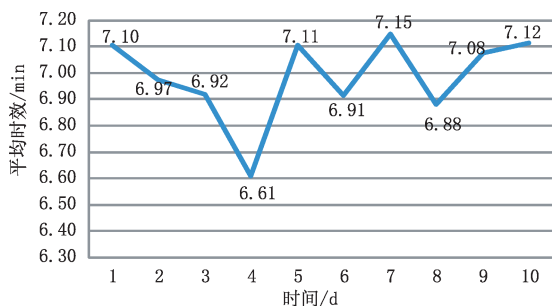


图 8 北京局 CMACast 补调总体时效

CMACast 播发的关键业务资料的时效分析结果如图 9、图 10 所示,其中实线线代表的是不同通道下资料连续 10 天的平均补调时效,虚线线代表的是不同通道下资料的平均文件大小。根据两条曲线的趋势对比我们可以看出,在山西局,不同通道下资料的补调时效差异比较大,大部分资料在 10 min 内能完成补调,而在 NWP_MCTR_002 通道下播发的资料补调时效能达到 30 min,资料的平均补调时效和文件大小基本上呈正比关系;而在北京局,所有关

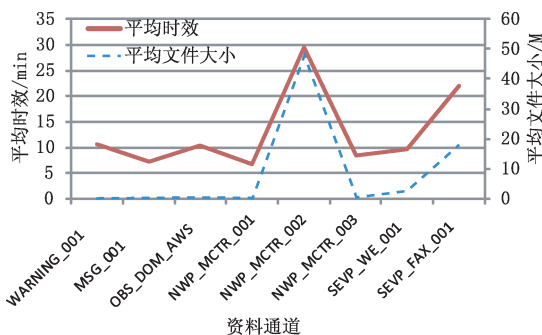


图 9 山西局关键资料补调时效

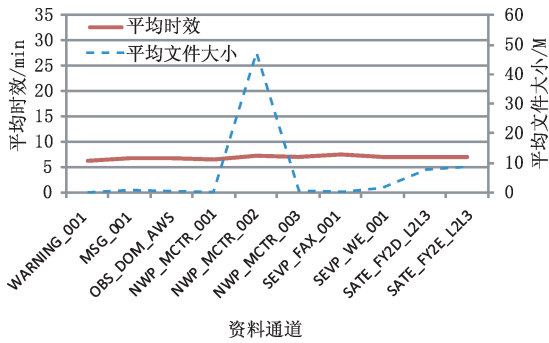


图10 北京局关键资料补调时效

键业务资料在6、7 min之内就能完成补调,资料的平均补调时效和文件大小基本上无关。产生上述差异的具体原因:山西局采用MPLS VPN线路,省级接入带宽为8 Mb/s(注:气象宽带网提速后,接入带宽为16 Mb/s),不同大小的文件传输用时差别较大,和文件大小有关;北京局采用光纤连接,网络传输速率很高,补调时效受文件大小因素影响很小。

4 总结

CMACast省级接收小站补调功能在全国省级

通信系统的部署,为CMACast卫星广播系统的业务化提供了有力保障,在日常气象业务中发挥了实际效益。本文详细介绍了数据补调服务客户端的设计方案和技术路线,为其他通信分发补调服务提供了借鉴思路。由于CMACast广播数据缺收情况具有不确定性,当出现极端情况,如极端天气现象或者是接收小站出现失锁时,将导致大数据量级的数据缺收。因此,在大负载情况下,如何更好地提高补调性能,是一个值得深入研究的问题。在未来的工作中,仍需要不断深入业务需求,持续优化补调性能,使其能够更好地发挥气象服务保障作用。

参考文献

- [1] 王春芳,李湘,陈永涛,等. 中国气象局卫星广播系统(CMACast)设计[J]. 应用气象学报,2012,23(1):113-120.
- [2] 王甫棣,郑波,胡英楣. 国务院气象数据服务系统的升级改造[J]. 气象科技,2012,40(5):745-749.
- [3] 刘学勇,陈建伟. 精通Linux C编程[M]. 北京:清华大学出版社,2009.
- [4] 李兴宝,侯方. 基于C/S模式的气象资料自动下载系统[J]. 气象科技,2010,38(5):609-612.

Design and Implementation of Meteorological Data Downloading Service for CMACast Provincial Receiving Stations

Han Shuli Tan Xiaohua Li Xiang Wang Peng

(National Meteorological Information Center, Beijing 100081)

Abstract: Affected by satellite channel transmission errors, interference signals and rain attenuation, the files broadcasted by CMACast might be lost or received mistakenly at the receiving stations, resulting in an incomplete collection of broadcasting data, which further impacts the forecasting and other meteorological service. To build a data center and provide downloading service through the ground wideband network is an effective solution to cover the shortage of the satellite network. The data downloading client implements data downloading process to better meet the demands of the provincial-level downloading service, in which a two-stage downloading mechanism is designed and a flexible configurable mode is adopted. The client adopts such technologies as multi-level parallel processing and priority-based dynamic scheduling in the program. The downloading client has been deployed on the provincial telecommunication systems and received practical benefits in daily meteorological service. The mechanism and implementation of the downloading service are introduced in detail.

Key words: satellite broadcast, meteorological data, receiving failure, download