

一种跨安全区的文件可靠性传输方法的研究

通过对电网在不同安全分区之间进行文件交互时存在的问题的研究，分析了目前在文件跨区传输的过程中在安全性、可靠性上存在的不足，并借鉴TCP可靠传输、断点续传及MD5加密等一系列成熟的技术，提出了一种跨安全区的文件可靠性传输方法，以此来保证信息在电网安全分区之间的可靠共享。

叶海峰 张炜 彭伟 杨劲松 田宏强 黄寅/安徽省电力公司



叶海峰/高级工程师

关键词/Keywords

- 跨区可靠传输 ·
- 心跳机制 ·
- 蚂蚁传输 ·
- 文件加密 ·

0 引言

在电网中，电力系统内按照安全级别要求不同，对电网进行了分区。不同安全区之间使用物理隔离装置进行分离，而分区之间的信息交互需要按照电网的文件格式要求通过特定的传输软件穿透隔离装置进行。但是，目前隔离装置及传输软件在交互内容的大小、格式及网络通道的可靠性等方面对这类信息的交互形成了极大的限制。

虽然隔离分区的存在在很大的程度上保证了电网的安全性，然而随着电网的快速发展，越来越多的信息需要在不同的安全分区之间形成共享，现有的信息交互方式已经不能满足需要，此时就有必要研究一种安全可靠的，在电网安全区之间进行文件传输的方法，以突破现有的隔离装置及传输工具限制，作为信息跨区交互的基础，为电网跨区业务的融合提供支撑。

1 跨区传输

1.1 跨区传输现状

目前，电力系统内，电网按照安全

级别从高到底被依次划分为I区、II区、III区和IV区，电网分区之间以正反向隔离装置进行分离。隔离装置均提供了文件传输通道，用于在不同分区之间进行信息交互。

图1所示的是跨区信息交互的架构，描述了信息文件从电力III区通过III区的文件传输服务器，穿透反向隔离装置提供的文件传输通道，到达I区的文件传输服务器，从而完成III区信息共享到I区的过程。隔离装置在文件传输的过程中，对文件内容进行了检查，保证文件的合法。

在信息量不大、实时性要求不高的情况下，这种交互方式已经能满足需

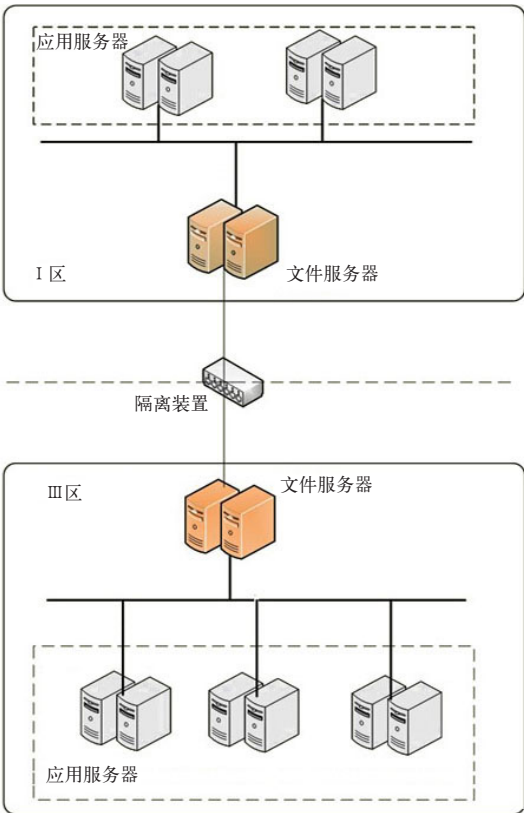


图1 跨区传输现状说明图

求。但是随着业务量扩大，需要交互的信息量增多，其种类也在增多，现有的跨区传输模式中存在的不足已经越来越突出，具体问题如下所述。

1.2 跨区传输问题

目前，在信息文件进行跨区传输的过程中，主要存在着可靠性和安全性两种比较突出的问题，具体体现在以下几个方面。

1) 传输通道的可靠性不能及时反馈。目前，由于物理环境、政策及软件使用情况等内外部不同因素的影响，正反向隔离装置偶尔会发生网络不通、拥堵等不同类型的问題，导致传输通道不畅。

跨区文件传输软件是作为一种后台信息共享的工具而存在的，是没有实时监督的，而需要跨区共享的信息则分类众多，信息的业务需求也不相同。有的信息是可以延期发送的，而有的则需要实时发送；有的信息是需要收取方给出回执的，而有的则不需要。不同类型的信息对传输通道的要求也不尽相同。所以，针对某些实时性要求高、需要进行送达验证的信息，传输通道的可用性极为重要。

然而，在当前的模式下，并不存在一种方法能够及时地将传输通道的状态及时通知到相关的系统、人员进行处理。这极大地影响了一些业务功能的使用，并可能会对电网的安全造成一定的影响。

2) 大文件传输导致隔离传输通道拥堵。随着业务的不断扩展，跨区传输内容的种类也不断增多。当有大文件需要穿透隔离装置时，一方面，隔离装置自身对所能通过的文件大小有着严格的限制；另一方面，由于隔离装置传输速度的限制，大文件的传输需要一定的时间，在其传输的过程中，也将会占有隔离装置的传输通道，导致其他待传输文件无法及时传送。

3) 文件传输的完整性有待验证。在文件传输的过程中，当前没有对传输完成的文件进行完整性的判断，以防止发生如网络丢包等情况导致文件内容的缺失。

4) 文件传输过程中存在被篡改的可能。在文件传输时，需要经过多个服务器、多个软件并多次落地，在此过程中，文件内容被篡改的可能性

极高，文件传输的安全性不能得到充分保证。

2 改进的跨区传输实现

针对目前在跨区文件传输过程中遇到的比较突出的上述几个问题，需要逐个进行研究，以提出一种合理可行的解决方法，具体实现方法包括下面几个部分。

2.1 心跳机制

针对传输通道的可靠性不能及时反馈的问题，可以借鉴TCP中心跳机制的做法。

TCP中的心跳机制是由客户端定时发送简单的消息到服务端，以保持客户端与服务器之间的连接。而在本文中所提出的心跳机制则如图2所示，是在两个分区的文件传输服务器上各自布置程序，程序启动后，一方（A侧）定时周期性地发送简单的、内定内容的文件，并传输到对端B侧，B接收到信息之后，生成回执并返回到A，若A在规定的时间内收到B发送的回执，表明正反向的文件传输通道均是正常的；若在3个心跳周期内没有收到对应的回执文件，则给出对应的网络问题反馈。

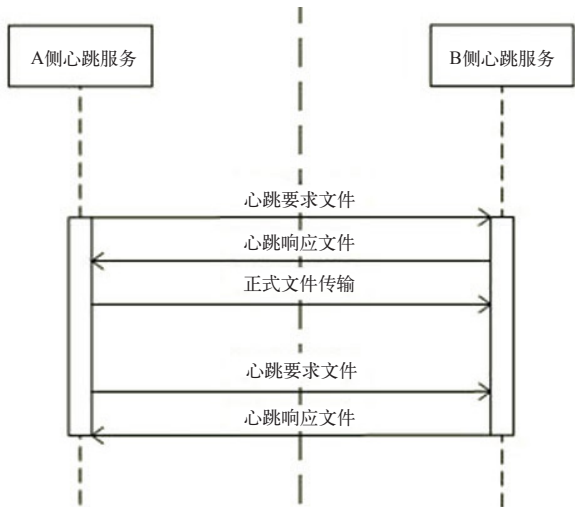


图2 心跳机制

2.2 断点续传及文件拆分

大文件在传输过程中，会对后续其他文件的传输造成一定的影响。究其原因，一方面是由于文件传输通道是单一的，导致文件较大时传输的时间比较长，而目前还不存在时钟机制来打断耗时过长的文件传输，从而无法保证其他业务文件

的及时传输；另一方面，则是因为隔离装置中文件传输通道的速度是有限制的，不能完全满足待传输文件大小的增长需要。

所以，参照其他成熟软件的做法，可以通过两种方式来解决在大文件传输中发生的相关问题。

1) 方式一：周期性中断传输。设定时钟机制，并借鉴HTTP中断点续传的做法，到时间周期则打断大文件的传输，并从待传输的文件队列中选择下一个传输目标。

HTTP断点续传的理解可以分为两部分：一部分是断点，一部分是续传。断点的由来是，在下载过程中，将一个下载文件分成了多个部分，同时进行多个部分的下载，当某个时间点任务被暂停时下载暂停的位置就是断点了。续传就是当一个未完成的下载任务再次开始时，会从上次的断点继续传送。

在HTTP的下载（或上传）过程中，如果由于网络故障、电量不足等原因导致下载中断，这时就需要使用到断点续传功能。下次启动时，可以从记录位置（已经下载的部分）开始，继续下载之前未下载的部分，避免重复部分的下载。断点续传的实质就是能记录上一次已下载完成的位置。

断点续传需要在下载过程中记录每条线程的下载进度，每次下载开始之前先读取数据库或者查看状态描述文件，查询是否有未完成的记录，有就继续下载，没有则创建新的文件传输状态记录。在每次向文件中写入数据之后，在数据库中更新下载进度，下载完成之后删除数据库中下载记录。

在文件传输时，可以设定时间周期，每个文件传输开始时启动时钟，如果到达规定时间文件还没有传输完成，则保留文件传输的位置并结束本周期的传输，开始传输队列中的其他文件，在下次轮到本文件的传输时，读取当前传输的位置，并继续后面内容的传输。

这种传输方式的好处是，保证了所有待传输文件公平占用传输通道，不存在文件传输的长等待。其缺点是需要额外的信息用来存储文件的传输状态，包括当前断点、本文件传输状况等。

2) 方式二：文件拆分传输。借鉴了压缩分包的做法，按照默认的最大文件大小限制，将大文件按照此大小进行拆分为多个小文件进行传输。

文件拆分传输的流程主要包括几个环节：提取信息—拆分—验证传输—验证接收—组装—检查。在文件拆分传输的过程中，还需要考虑到文件拆分后的组装的顺序、完整度等内容。

①提取信息。提取原文件的大小等属性信息，生成原文件对应的属性文件并优先传输到对侧B。

②文件拆分。按照目标大小，将文件进行拆分成为若干个小文件，并对小文件进行编号和排序，同时生成对应的小文件清单，包括文件顺序、组织内容等信息。

③验证传输。按照小文件的排序将其进行验证严格按照顺序传输，一方面保证了小文件的内容符合要求，另一方面则保证了小文件传输的严格的先后顺序。

④验证接收。按照小文件的排序先后进行验证按序接收；同时，在全部文件接收完成之后，按照拆分内容清单，对小文件进行检查。

⑤组装。在小文件接收完毕之后，将众多小文件，按照排序先后进行组装，以生成与原文件相同的单一文件。

⑥检查。对组装好的文件，按照原文件所对应的属性文件进行属性比对，如果确认无误，完成检查，否则要求重传。

经过文件的拆分传输，避免了隔离装置对所传输的文件大小的限制，同时防止了大文件传输对隔离通道的长期占用，保证了其他业务的可用性。这种传输方式的缺陷是需要对子文件传输的顺序进行严格的控制，以防止拆分后的子文件没有完全传输导致整个文件的传输失败。

本文中通过目前业界主要使用的两种对大文件的处理方式，建议主要使用断点续传的形式，来处理大文件在跨区传输过程中所造成的长期占用通道等问题。

2.3 MD5文件加密

文件在传输时，存在两个与传输安全相关的方面。

首先，在文件跨区传输的过程中，如何判断一个文件已经传输完整，如果文件已经传输完成，但是没有通过完整性的校验，则需要对文件进行重新传输。针对这个问题，参照互联网在这个方面的经验，可以使用MD5算法对文件的完整性进行检查。

其次，文件在跨区传输时，经历了多次落地，每次落地都是物理文件，在此过程中存在一定的风险，人工可以对文件进行修改，这时就需要一种防篡改的手段，用于保证文件在传输中的安全性。同样，经过调研成熟产品的经验，可以看到MD5算法能够满足要求。

MD5 是指信息－摘要算法，是一种密码的算法，它可以对任何文件产生一个唯一的MD5验证码，每个文件的MD5码就如同每个人的指纹一样，都是不同的，一旦这个文件在传输过程中，其内容被损坏或者被修改的话，那么这个文件的MD5码就会发生变化，通过对文件MD5的验证，可以得知获得的文件是否完整。

在应用MD5算法对文件进行处理分为如下几个步骤。

- 1) 打开文件，获取dos文件头。
- 2) 通过dos文件头的结构体内变量计算出PE文件头的位置。
- 3) 通过PE文件结构体变量计算出块表位置。
- 4) 通过块表第一个区块表计算出代码段的起始位置和长度。
- 5) 计算MD5值。
- 6) 将计算出的MD5值写入PE文件头之前的8个字节中。

在接收端收到相应的文件之后，同上述步骤对文件进行处理，并将MD5的算法进行比较，如果相同，则文件是完整的且未经过篡改，否则，文件需要重新进行传输。

另一方面，在使用MD5对需要传输的文件进

行加密后，文件内容以密文形式进行传输，保证内容在防篡改的同时，能够防止不被授权的查阅造成资料泄露。

3 结束语

安全区划分是电力系统的特色，对安全性的提升有目共睹。但是，随着电网和信息技术的发展，电网信息需要进行整合和共享，跨区通信的需求量也就越来越大。本文中主要从安全性、可靠性上对现有的跨区文件方式进行研究，通过利用业界成熟的心跳机制、断点续传及MD5加密等技术，提出了一种跨区安全可靠传输的方法，但是该方法的成效还需要与实际结合并不断进行完善和应用。

参考文献

[1] 陈增强, 郭嘉琳, 刘忠信, 等. 具有断点续传功能的文件传输系统的设计与关键技术[J]. 计算机工程, 2002, 第12期: 14-16.

[2] 鲁孙林. 基于Http协议的断点续传.

[3] 王坤, 林莉芸. 心跳机制在 DHCP协议中的应用[J]. 新乡学院学报, 2014, 31(2).

[4] 彭文波. MD5算法原理及应用[J]. 信息网络安全, 2004, 等5期: 44-46.

[5] 马赛牧, 王晶. 基于文件拆分和缓存预测的日志文件传输算法[J]. 电信工程技术和标准化, 2013, 第8期: 71-78.

作者简介

叶海峰（1978-），男，高级工程师，硕士，工作领域为电网运行。

(收稿日期: 2015-12-10) **EA**
(编辑: 杨得润)