基于 IRF2 和 LACP MAD 的气象网络设计研究

鲍磊磊[1] 唐红昇[2] 姜淑杨[1] 吴嘉伟[1] 李玉涛[2]

[1](南通市气象局 江苏 南通 226000)

[2](江苏省气象信息中心 江苏 南京 210000)

摘 要 气象内网传统架构基本采用 VRRP 和 MSTP 的组网模式。针对传统模式下,设备性能的利用率不高,网络管理工作繁杂,网络架构不能满足气象信息业务发展需求等问题,基于 IRF2 和 LACP MAD 对地市级气象内网进行优化设计和平滑升级。对升级后的虚拟化网络系统原理、工作过程、分裂冲突过程进行详细分析。通过模拟软件对升级后的网络系统进行分裂测试、抓包分析,结果表明该网络性能及可靠性均优于现有网络。该设计方案同样适用于省级气象网络核心和接入层的平滑升级。

关键词 VRRP MSTP 智能弹性架构 OSPF 链路汇聚控制协议 MAD

中图分类号 TP393 **文献标识码** A DOI:10.3969/j.issn.1000-386x.2019.01.025

DESIGN AND RESEARCH OF METEOROLOGICAL NETWORK BASED ON IRF2 AND LACP MAD

Bao Leilei[1] Tang Hongsheng[2] Jiang Shuyang[1] Wu Jiawei[1] Li Yutao[2]

[1](*Nantong Meteorological Bureau, Nantong 226000, Jiangsu, China*)

[2](*Jiangsu Meteorological Information Center, Nanjing 210000, Jiangsu, China*)

Abstract The traditional architecture of meteorological intranet basically adopts VRRP and MSTP networking mode. Under the traditional mode, the utilization rate of equipment performance is not high, the network management is complicated, and the network architecture cannot meet the needs of meteorological information business development. In order to solve these problems, based on IRF2 and LACP MAD, the optimization design and smooth upgrading of city level meteorological intranet were carried out. We analyzed in detail the principle, working process and conflict splitting process of the upgraded virtualized network system. Split test and packet capture analysis were performed on the upgraded network system through simulation software. The results show that the network performance and reliability are better than the existing network. The design scheme can apply to the smooth upgrading of the provincial meteorological network core and access layer.

Keywords VRRP MSTP IRF OSPF LACP MAD

0 引 言

气象信息网络的冗余性、稳定性和可靠性是保障气象数据通信的重要基础,是观测数据传输时效、天气视频会商、信息系统稳定运行的重要评价依据。目前气象系统主要是通过升级网络设备、提升带宽、合理规划网络结构来满足通信需求,地市级比较常见的双套备份网络方案还是采用 VRRP 技术将多台设备虚拟成一台备份组的方式[1-2]。

随着通信网络技术的发展,与时俱进地应用先进技术优化气象信息网络,从而达到系统稳定、管理简单的效果显得尤其重要。IRF 这种新兴的智能弹性架构可以将多台物理设备的软硬件资源虚拟化为一个逻辑意义上的 IRF 系统,极大地简化网络结构[3]。但是当 IRF 系统分裂后,网络中就会存在多台配置同样的独

收稿日期:2018-08-20。江苏省气象学会青年科研基金项目(KQ201809);南通市市级科技计划项目(YYZ16035)。**鲍磊磊**,工程师,主研领域:气象信息技术。**唐红昇**,高工。**姜淑杨**,助工。**吴嘉伟**,工程师。**李玉涛**,工程师。

立设备,因此需要一个冲突检测机制将分裂后冗余配置的设备从网络中分离出去[4]。本文通过模拟升级后的网络系统,采用 LACP MAD 的检测机制对 IRF 网络系统分裂后进行冲突检测,研究优化后的气象信息网络的稳定及可靠性。本文提出的平滑升级方案适合在气象信息领域推广和应用。

1 原系统概述

1.1 优化前网络结构

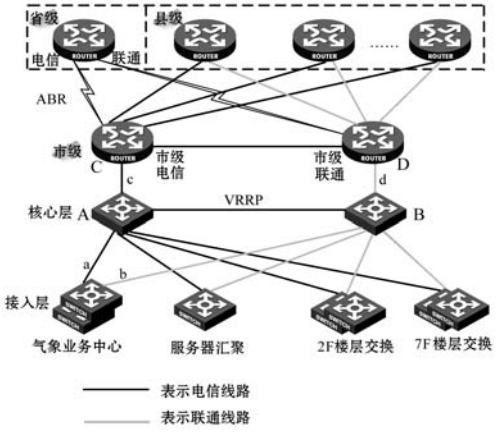

图 1 原有网络结构

如图 1 所示,顶层为省、市、县气象专线边界路由层,市级核心层由两台 H3C S10504 担当,分别命名为设备 A 和 B,A 和 B 启用 VRRP(虚拟路由冗余协议)[5-6]。设备 A、B 各自上联至市局电信和联通的 ABR 路由 C 和 D,C、D 之间通过心跳口相连,传递动态路由数据,形成主备模式。主备 ABR 路由器 C、D 与县局路由器组成三角型拓扑,可以实现市、县运营商线路双备份。

1.2 网络结构分析

现有网络结构中,核心和接入层采用了 VRRP 与 MSTP 配合的组网方式:在设备 A 和设备 B 之间增加物理连接可以为上下行链路提供冗余备份,但因此网络中多了回环链路,增加 MSTP 技术消除环路。这种组网方式分别为通信链路和网关设备提供冗余备份[7-9]:只要系统通信链路中上行链路 c、d 和下行链路 a、b 中各有一条线路可达,即正常通信;系统也不会因为链路 a 或链路 b 出现故障时,切换核心设备 A、B 的主备模式。

1.3 存在的问题

1)设备和链路资源浪费,备用核心设备及其相连的一半链路基本处于闲置状态,此外链路中的路径回环还需采用 MSTP 处理。

2)VRRP 配合 MSTP 可以实现线路的负载均衡,但是 VRRP 要做到负载均衡,必须在接入设备较多的情况下划分区域并且合理分配接入设备的地址,网络运维繁杂。

3)主备链路的切换速度慢,数据的转发和延迟高、现有网络环境实测切换速度为秒级,可靠性低[10]。

2 系统优化方案

2.1 方案设计

机房现有的两台 H3C S10500 系列交换机支持 IRF,在现有 VRRP 模式的基础上优化,不需要增加独立的物理设备,只需要增加板卡并优化软件系统[11]。考虑到数据传输速率和设备之间心跳的重要性,在主、备核心上各增加一块 40 Gbit/s 以太网光接口板卡,每个板卡各提供两个 40 Gbit/s 光接口模块用于实现两条 IRF 链路捆绑。本方案中分别将两台核心创建的 IRF 端口与新增板卡提供的 2 个 40 Gbit/s 物理端口绑定,以实现 IRF 链路的负载分担和备份[12]。具体规划如表 1 所示。优化后的网络拓扑如图 2 所示。

表 1 IRF 系统编号和端口规划

核心设备	成员编号	IRF 端口	绑定物理端口
A	1	IRF-Port2	FortyGigE 4/0/1 FortyGigE 4/0/2
B	2	IRF-Port1	FortyGigE 4/0/1 FortyGigE 4/0/2

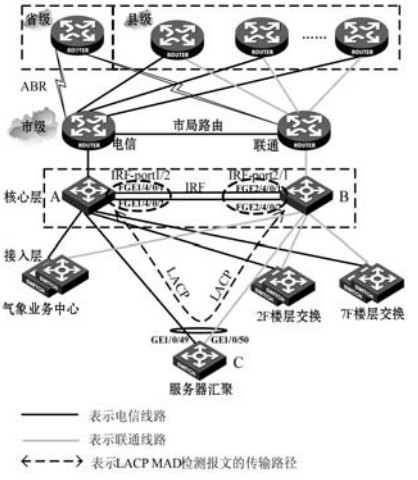

图 2 优化后的网络系统结构

IRF 运行在独立模式下,IRF 端口分为 IRF-Port2 和 IRF-Port1;在系统模式下,需要在 IRF 端口名称中加入设备自身的成员编号 n,标记为 IRF-Portn/2 和 IRF-Portn/1。同理,IRF 端口绑定的物理端口也是如此,即在原本的维数前面增加一维[13]。因此,表 1 中规划的 IRF 端口 IRF-Port1 和其绑定的物理接口 FGE4/0/1 原本分别采用一维格式和三维格式;一旦形

成 IRF 系统后,接口编号分别变为二维 IRF-Port2/1 和四维 FGE2/4/0/1,如图 2 所示。其中第一维表示核心设备的成员编号。

LACP MAD 部署链路还需要一台交换机作为 MAD 检测的仲裁设备,用于传递 LACP 报文。方案选用支持 LACP 协议的服务器汇聚交换机 C 分别和 A、B 部署聚合链路,同时不影响原先的数据转发。这样组网中既不存在任何链路资源的浪费,又不影响既有的网络拓扑。

2.2　IRF 系统工作流程

优化后的 IRF 系统将经历图 3 的四个阶段。

图 3　IRF 系统工作原理

具体过程如下:

1) 物理连接:将设备 A 和设备 B 的 FortyGigE 4/0/1 之间、FortyGigE 4/0/2 之间分别用 QSFP 线缆物理连接形成 IRF 系统。

2) 拓扑收敛:核心设备 A 和 B 会在自身的主用主控板上管理、记录收集到的拓扑信息,拓扑信息的收集是通过交互包含 IRF 端口、成员编号等信息的报文进行:设备 A、B 各自的主用主控板在启动时只有自身的拓扑信息,当 IRF-Port1 和 IRF-Port2 状态一旦 up ,A、B 便会根据收集到的拓扑信息更新自身的主用主控板。

3) 角色选举:即为确定 IRF 系统成员设备 A 和 B 为 Master 或 Slave 的过程。角色选举按照成员编号小的设备选举为主设备的原则,发生在 IRF 系统建立之初、A 和 B 其中有设备离开或者故障或者有新设备加入等拓扑发生变化的情况下。本设计通过设定设备 A、B 的成员编号,确定连接电信线路的设备 A 为 Master,连接联通线路的设备 B 为 Slave。IRF 系统形成后进入 IRF 管理与维护阶段。

4) IRF 的管理与维护:IRF 形成后网络中只存在一台同时拥有设备 A 和 B 上资源的虚拟设备,用主设备 A 命名并进行统一管理[14]。

2.3　LACP MAD 检测过程

IRF 系统对传统组网架构的突破和优化,简化了网络系统,从而减轻了网管的工作。但是当 IRF 链路故障导致单个 IRF 系统分裂成多个包含相同三层配置的独立 IRF 后,需要借用一种能够检测出多个 IRF 共存并且能够立即做出相应处理的协议机制,来降低网络中 IRF 系统分裂后冲突产生的影响[15]。本文选用 LACP MAD 检测协议来实现分裂检测,核心层和接入层设备 A、B、C 构成的 IRF 系统如图 4 所示。

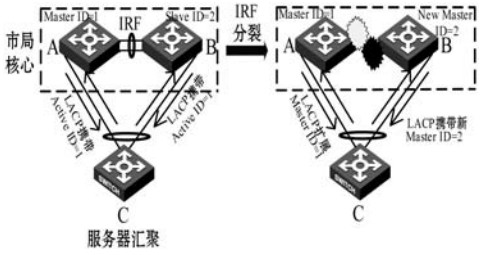

图 4　LACP MAD 检测过程

在 LACP 协议报文的扩展字段中携带了用于交互 IRF 的 ActiveID,该系统中,ActiveID 用 IRF 中主设备 A 的成员编号 1 来表示。系统正常运行时,A、B 传送的报文中的 ActiveID 值均为 1;当 IRF 系统分裂后,A、B 都变成 Master,即分裂为两个独立系统,分裂的系统产生不同的 ActiveID,和各自的成员编号一样,A = 1,B = 2。此时设备 A、B 之间通过交互 LACP 报文,就可以感知到系统中存在不同 ActiveID;系统检测到不同 ActiveID 后,设备 A 发现自己的 ActiveID 为成员编号 1,比设备 B 发现的自身 ActiveID = 2 要小,遵循给定规则:网络中保留 ActiveID 最小的设备,其他的设备全部隔离出来[16]。这样,ID 号小的设备 A 保持现状;ID 号大的设备 B 就会关闭自身所有业务端口(IRF 口除外)后,从网络中隔离出去,从而避免冲突,保证气象数据通过电信线路正常上传。

3　系统测试

3.1　测试模型

采用 H3C 的 HCL 仿真软件,模拟平滑升级后的设计方案建立如图 5 的网络模型,核心设备 coreA、coreB 配置 IRF 后上联本地路由器 MSR1,通过 OSPF 与远端路由 MSR2 通信,coreA、coreB 分别和下联接入层交换机 DevC 形成聚合链路。在 IRF 系统上启用 VLan194,通过 Trunk 模式和 DevC 连接,DecC 配置 Access 接口 GE0/3,接入 PC。

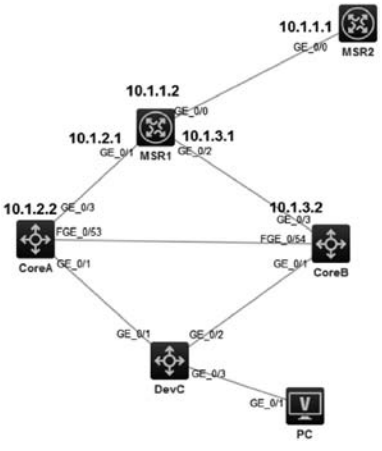

图 5　测试系统模型

3.2 功能测试

功能测试是验证本文提出的基于 IRF2 和 LACP MAD 的系统在 IRF 链路中断时，LACP MAD 检测方法可以检测到冲突发生，分裂后网络中只存在一台 IRF 设备正常工作，其他引起冲突的设备将关闭除保留端口外的全部业务端口后迅速从网络中分离。接入层设备能保障气象业务数据正常上传。

1）进入 IRF 系统，查看系统中 CoreA 和 CoreB 的工作模式。如图 6 所示，两台核心设备依据表 1 的规划形成了 IRF 系统，并使能了 LACP MAD 检测。

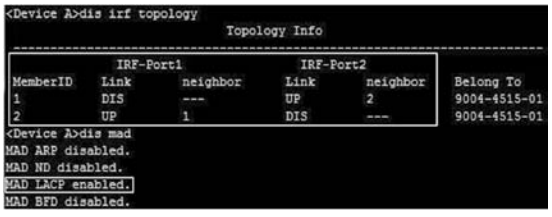

图 6 IRF 系统信息

2）断开 IRF 连线，验证 LACP MAD 检测到冲突从而将 MemberID 大的 CoreB 设备从系统中隔离出去。

如图 7 所示，IRF 系统分裂后，CoreB 的聚合端口 GE2/0/1 和上联口 GE2/0/3 均 down，CoreB 因为 MemberID 大，单板失效，聚合端口的 MAD 功能失效，被从系统中分离出去。此时分别登录 CoreA 和 CoreB，查看 IRF 信息，可以看出，系统分裂后，两台设备各自为 master。如图 8 所示。

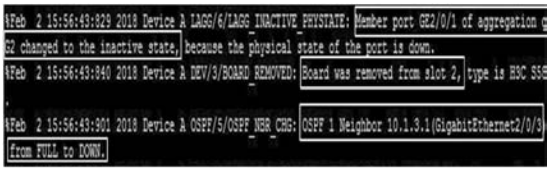

图 7 IRF 分裂过程

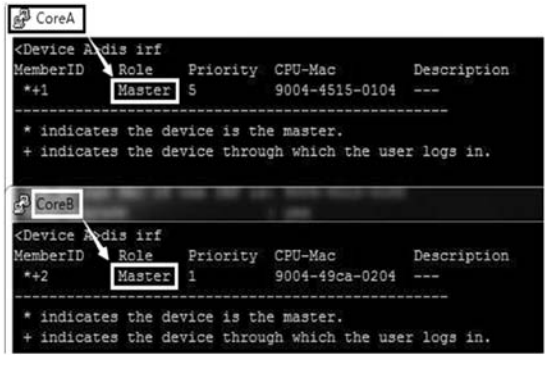

图 8 IRF 系统分裂后设备状态

3）整个过程使用命令 ping-c 4294967295 通过在 PC 上 ping 远端路由 MSR2 观察测试结果，从分裂到检测到冲突到最后分离 CoreB，期间只丢失一个 ms 级的数据包，系统实现了毫秒级的切换。

4）继续关闭 CoreA，同时在 CoreB 上使用 mad restore 命令将 B 恢复为 Active 状态。验证在 IRF 链路和高优先级设备同时故障的情况下，采取相应应急措施，可以将分裂出去的设备启用，实现通信。设备 B 的恢复过程如图 9 所示，聚合端口 GE2/0/1 和上联口 GE2/0/3 均启动，相应的 Vlan、网络协议相继启动。待上下行通信恢复正常后，启动 CoreA，并修复 IRF 链路，此时将重新形成 IRF 系统。

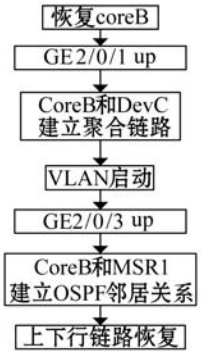

图 9 分离设备恢复过程

5）对 CoreA、CoreB 和 DevC 上的接口进行抓包，启用 Wiershark 抓包工具抓取 LACP 包进行协议分析。在 IRF 系统正常运行过程中，CoreA 和 CoreB 的 GE0/1 口分别和 DevC 的 GE0/1、GE0/2 口之间有 LACP 报文交互，截取 CoreA 的 GE0/1 口其中一个 LACP 报文，如图 10 所示。参考表 2 对其中的 Actor State 和 Partner State 字段的比特编码 0x3d（二进制值 00111101）进行分析：GE0/1 口在链路中 LACP_Activity，Aggregation，Synchronization，Collecting，Distributing 均为 1，表示端口链路可聚合，且被分配到聚合组中处于同步状态，收发包正常。

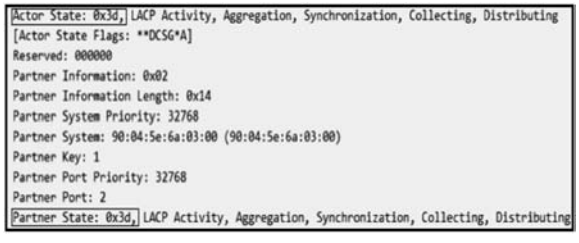

图 10 系统正常运行抓取的 LACP 包

表 2 比特编码对照表

位数	表示	值
0	LACP_Activity	1
1	LACP_Timeout	0
2	Aggregation	1
3	Synchronization	1
4	Collecting	1
5	Distributing	1
6	Expired	0
7	Defaulted	0

3.3　性能测试

性能测试需要借助 H3C 的通用测试平台,按照优化后的拓扑图 2 进行建模,主要测试系统中某台设备或者链路出现故障,系统自动切换时间。主要测试项目有:IRF 系统分裂,主、从设备故障,IRF 链路聚合、负载分担三种情况下的网络延迟。实验通过多次测试取平均值的方式,测试结果如表 3 所示。

表 3　性能测试结果　　ms

测试项目	切换时延
IRF 聚合链路 down	2.1
IRF 分裂	296
IRF 聚合链路流量	0
IRF 链路负载分担	15
Master 设备掉电	181
Slave 设备掉电	97

3.4　测试结果

实验不断地截取不同端口的报文进行分析:Actor State 字段的比特编码只有在 IRF 系统稳定后才是 0x3d。处于聚合组的两端端口稳定,IRF 系统形成,LACP 报文交互状态稳定,报文中携带的扩展字段相同,此时 MAD 检测不会生效。一旦 IRF 分裂为两个独立系统,LACP 报文交互的过程中检测到不同 Active-ID 存在后,系统即会在毫秒级的反应时间内将 Active-ID 不是最小的设备全部从网络中隔离出去,去除系统冲突产生的影响。此后,Wiershark 在 CoreB 的 GE0/1 和 DevC 的 GE0/2 口不再抓到 LACP 报文,从而验证了本系统的可靠性。整个分裂过程平均切换时延仅为 296 ms,而聚合链路交互 LACP 报文流量切换时延为 0,聚合链路负载分担平均切换时延也仅为 15 ms。主设备掉电,系统切换时延在 200 ms 之内,相比现有的 VRRP 拓扑链路实测秒级的切换时延,性能得到了很大的提升。对气象数据传输时效的影响基本可以忽略不计。

4　结　语

针对已有气象信息网络存在的问题,设计了新的网络结构,通过计算机测试得出优化后的网络具有以下优势:1)简化系统管理。尤其在中大型网络系统中的核心层,网管不再需要登录多台设备查看或更改配置,只需登录 IRF 系统,即可统一管理 IRF 内所有成员设备。2)高可靠性。体现在设备和链路的双重备份、业务和流量分担。相比 VRRP 的组网方式,IRF 组网中,Slave 设备既作为热备又处理业务。Master 设备一旦故障,系统会在毫秒级的范围内自动切换到新的 Master,切换时间短,对气象通信业务的影响可以忽略不计。3)强大的网络扩展能力。IRF 系统中所有成员设备的 CPU、端口等资源都可以为系统所用,提升系统处理、转发协议报文的能力。4)资源集约,双核心和接入层交换构成的 LACP 聚合链路既可以收发 LACP 报文,还可以转发气象业务数据,又不影响用户的网络层次模型,没有任何链路资源浪费[17]。该方案在部分省、市已实施完成,并投入业务应用,下一步将结合气象数据传输业务和数值预报业务进一步开展稳定性研究。

参　考　文　献

[1] 李进喜,戴维士. 气象信息网络运维保障典型个例分析[J]. 气象科技,2012,40(1):52-54.

[2] 马渝勇,方国强,向继涛,等. 省级气象信息网络系统的整体设计与实现[J]. 计算机应用研究,2012,29(4):1376-1377.

[3] 李长青,李红信,许凯. IRF 虚拟化技术在路由器高可靠性方面的研究与应用[J]. 数字技术与应用,2015(8):38.

[4] 朱梦莹. 通信设备虚拟堆叠系统的分裂检测与处理[D]. 南京:东南大学,2013:5-11.

[5] 许玮,王迎迎,秦运龙,等. 省级气象广域网网络优化的设计与应用[J]. 气象科技,2016,44(3):358-362.

[6] 燕东渭,杨艳,王垒,等. 面向业务保障的省级气象广域网络优化升级[J]. 气象科技,2015,43(2):211-215.

[7] 燕东渭,陈高峰,杨银见,等. 基于 OSPF 的陕西省气象宽带网络整合设计[J]. 气象科技,2012,40(4):585-590.

[8] 李军,李光,邸永强,等. 基于虚拟路由冗余协议和双向转发检测的基层气象通信网络设计和实现[J]. 气象科技,2017,45(2):281-284.

[9] 陈增吉. 基于 VRRP+MSTP 协议的可靠性网络设计[J]. 计算机应用与软件,2009,26(4):208-211.

[10] 虞谦. 江西服装学院校园网升级规划与设计[D]. 江西:江西财经大学,2017:39-45.

[11] 杭州华三通信技术有限公司. H3C S12500 系列交换机 IRF2.0 技术白皮书[Z]. 2012.

[12] 张玉芳,陈光礼,熊忠阳,等. 一种基于智能弹性架构的纵向异构方案[J]. 计算机工程,2014,40(9):96-101.

[13] 邱慧丽. 虚拟化技术在商业银行分行同城网中的应用研究[J]. 贵阳学院学报(自然科学版),2017,12(2):13-16.

[14] 王培英. 基于 LACP 多激活检测方法和处理机制的研究[D]. 成都:西南交通大学,2013:30-45.

[15] 王斌. 基于 IRF 的路由器堆叠设计与实现[D]. 重庆:重庆大学,2015:6-54.

[16] 邓少华. 基于 IRF2 技术的校园网架构研究[J]. 网络空间安全,2014,5(3):82-84.

[17] 秦丽娜. 基于虚拟化 IRF2 技术的网络可靠性分析[J]. 山西经济管理干部学院学报,2013,21(4):100-103.