

文章编号: 1005-8451 (2015) 08-0005-04

# 铁路互联网售票容灾网络互联技术的研究

刘相坤, 李 琪, 武振华

(中国铁道科学研究院 电子计算技术研究所, 北京 100081)

**摘 要:** 本文在铁路互联网售票系统的设计和维持的实践基础上, 对容灾网络进行总结和新的探索。分析了铁路互联网售票系统容灾网络的需求, 对主要网络互联技术发展作了介绍, 对数据中心间互联技术比较分析, 提出容灾的网络架构, 最后对铁路互联网售票数据中心间网络互联提出建议。

**关键词:** 铁路互联网售票系统; 系统容灾; 容灾网络

**中图分类号:** U293.22 : TP393 **文献标识码:** A

## Interconnection technologies of disaster recovery network for Railway Internet Ticketing System

LIU Xiangkun, LI Qi, WU Zhenhua

(Institute of Computing Technologies, China Academy of Railway Sciences, Beijing 100081, China)

**Abstract:** Based on system design and daily practice of Railway Internet Ticketing System, this article summarized and researched on the disaster recovery network, analyzed the requirements of disaster recovery network of Railway Internet Ticketing System, introduced the development of main interconnection technologies, compared the interconnection technologies between data centers, put forward a network architecture of disaster recovery, proposed suggestions for network interconnection among data centers in disaster recovery of Railway Internet Ticketing System.

**Key words:** Railway Internet Ticketing System; disaster recovery; disaster recovery network

随着企业生产对信息系统的依赖加大,灾难备份与恢复变得更加重要,需要通过风险管理、危机管理、应急管理、灾难恢复和业务连续管理,制定全面的连续计划和应急预案,确保组织的生存能力。容灾网络的实现是容灾恢复的前提和基础,网络的稳定性、安全性和时延在一定条件下可能决定容灾的成败,也可能会影响生产的效率。

容灾网络的架构和实现技术在不同的信息系统中不尽相同,实际网络条件及应用情况也有很大差异,本文结合多年对铁路客票发售与预定系统的设计、实施和维护的工作经验及体会,针对铁路互联网售票系统灾难备份与恢复的现状以及未来发展,对容灾网络的互联技术深入研究分析,提出铁路互联网售票系统容灾的网络实现方案。

### 1 铁路互联网售票系统容灾网络需求分析

#### 1.1 互联网售票同城两中心的互联需求

铁路互联网售票系统分客服外网、内网、客票网

专网 3 个网络层次,为实现同城容灾双活中心,存在客服内外网以及客票网的跨数据中心集群应用,集群的时延要求不超过 10 ms。

在客服内外网的互联网用户数据库、客票网的互联网售票数据库使用 VMware 虚拟化技术实现跨数据中心的应用集群时,VMware Metro VMotion 和 HA 技术都需要 2 层互联,并且往返时延(RTT)要求不超过 10 ms。未来如果规划 WEB/Cache/APP 服务器使用虚拟化,并在两个数据中心间形成 Metro HA,也需要 2 层互联。

由于 IP 地址空间受限,超大的应用集群系统会拆分成小的应用集群,以放置到多个不同的 IP 地址段,此时发生的跨数据中心应用互访逻辑,则需要提供数据中心间 3 层互访需求,且互访应用交易逻辑组件也对时延敏感。

#### 1.2 新型的云架构下容灾网络需求

新型的云架构下容灾网络不仅需要数据中心之间的 3 层互联,还需要具备 2 层互联的能力,因此不仅在数据中心内(云内)需要 2 层环境,在数据中

收稿日期: 2014-12-30

作者简介: 刘相坤, 副研究员; 李 琪, 副研究员。

心之间（多云融合）也需要构建2层延展能力，并需精心设计这种2层互联网络模型。

在构建数据中心间的2层互联时，要保持数据中心的灾备独立性，即经过互联的数据中心必须具备相当的系统运行独立性，不能因为主数据中心的任何故障或差错导致灾备数据中心被殃及，从而使整个灾备系统网络失效。在这种考虑下，建议采用L2 over L3技术隔绝故障域。即为了业务的灵活性，在实现数据中心间不同网段的路由需求时，把2层网络建立在可扩展的3层网络架构之上，提供一个既能实现3层、又支持2层网络跨数据中心扩展的通用网络互联平台。

## 2 数据中心间互联技术比较分析

### 2.1 交换机直连实现网络互联

数据中心间通过核心交换机直接互联，使用交换机的Trunk等方式实现2层扩展，并保证主中心和备中心能够使用同一个虚拟局域网（VLAN）和IP地址设置以确保分布在两个数据中心的业务集群（Cluster）、vMotion等应用的实现。为确保可靠，通常采用两条以上链路进行互联。

存在的问题：故障域隔离性不好。2层直接使用Trunk打通后，主中心和备中心将完全变为一个局域网，主中心所有VLAN的服务器产生的广播流量、未知单播流量都会泛洪到备中心，需要注意的是由于需要确保主中心和备中心之间连接的低时延需求（vMotion或Cluster需要低时延），不建议在连通链路上部署防火墙，同时也无法利用访问控制列表（ACL）隔绝广播。这就意味着如果主中心受到2层攻击，也会同时影响备中心的业务，同时如果主中心有大量的未知单播和广播流量，也会影响互联链路和备中心。

冗余性会带来网络复杂性，即每个数据中心采用一对交换机进行互联，在2层链路上就产生了天然环路，必须采用Spanning Tree等技术进行环路阻止，但Spanning Tree技术会阻断一条链路，使链路带宽不能完全利用，并且Spanning Tree技术无法实现域隔离，这使得一个数据中心的问题会波及其他。

### 2.2 采用VSS/IRF等双机虚拟化技术实现互联

主备中心的核心交换机组成多机集群，逻辑上两条链路可以捆绑为Etherchannel，链路带宽可以充分利用，同时可以避免使用Spanning-Tree技术增加网络复杂度。

存在的问题主要是故障域隔离性不好。由于还是采用传统2层技术互联，广播、未知单播等泛洪问题还未得到解决，故障域不能有效隔离。

冗余性带来网络复杂性，虽然利用双机虚拟化技术解决环路问题，有效地增加了互联带宽，但由于控制平面合一带来“脑裂”风险，影响数据中心业务的高可用性。

### 2.3 采用基于IETF标准的L2 over L3技术实现互联

数据中心间通过路由器采用3层互联，可以使用标准的3层路由协议（利用ISIS）避免环路，同时使用基于IETF标准的L2 over L3的技术使2层报文承载在3层网络上进行传输。由于使用了ISIS协议进行通告，没有环路问题，同时，由于控制平面分布方式，没有“脑裂”所带来的风险隐患，因此，同时使用L2 over L3技术完全不用启用Spanning Tree技术，这样数据中心的Spanning Tree域将会受到隔离。

在使用L2 over L3技术后，可以利用地址通告的方式（Mac地址由路由控制平面宣告）而不采用广播泛洪的方式，可以确保两个数据中心之间的广播、未知单播等泛洪流量被限制在单个数据中心内部，有效地隔绝故障域。同时网络IP层互联业务。

## 3 网络互联技术发展

### 3.1 网络架构演进与数据中心发展

网络架构演进与数据中心发展如表1所示。

### 3.2 数据中心互联与业务迁移的连接方式

针对不同业务需求和容灾需求，实现的互联方式也不同，常用的有存储区域网络（SAN）互联、2层互联和3层互联。

SAN互联：也称为后端存储网络互联。通常借助传输技术（波分复用、同步数字体系等）实现主中心和灾备中心间磁盘阵列的数据复制。

2层互联：也称为数据中心后端服务器网络互联。在不同的数据中心服务器网络接入层，构建一个跨数据中心的2层网络，以满足服务器集群或虚拟机

表1 网络架构演进与数据中心发展

	紧耦合3中心模式	松耦合3中心模式	多中心模式
组网特点	数据中心同时作为核心网络节点（紧耦合）。数据中心、分支、业务中心间需要全互联。	互联网络和骨干网与数据中心松耦合，分支机构就近接入，链路减少成本降低；实现网络接入的初步标准化。	数据中心、业务中心与网络节点完全松耦合。独立的数据中心与分支互联平面，数据中心互联与分支机构上联合二为一，完全标准化组网。
风险与可靠性	压力与风险分布在多个中心，业务与网络紧耦合。	业务系统与网络耦合度降低，灵活性增加，压力与风险降低，可实现平滑切换。	业务系统与网络系统的完全松耦合，规避风险对业务的大面积影响的可能。
成本与业务灵活性	链路成本较高、利用率不均衡，业务部署不灵活。	链路与维护成本得到降低、节省了分支上行链路，资源利用率得到提高。	成本最高，业务部署灵活度高、资源灵活配置。

动态迁移等场景对 2 层网络接入的需求。

3 层互联：也称为数据中心前端网络互联。不同数据中心的前端网络通过 IP 技术实现互联，分支客户端通过前段网络访问各数据中心。当主数据中心发生灾难时，通过前端网络快速收敛，使得客户端通过访问灾备中心以保障业务连续。

在铁路互联网售票双活系统中，存储互联采用 SAN 互联方式，集群和虚拟机迁移采用 2 层互联，业务互访采用 3 层互联方式。

**3.3 数据中心互联波分复用（DWDM）光传输技术**

建设同城数据中心 DWDM 平台满足当前的互联需求，并为将来业务的快速部署要求提供支持。

当前“两地 3 中心”的业务架构，在同城中心之间有巨大的互联带宽需求。双 / 多活应用、双活存储要求同城中心之间快速的数据同步，需要大带宽、低延迟、高可用的透明传输通道；同城中心之间的数据备份数据量巨大，也具有大带宽的需求。

自建 DWDM 平台有很多优势。虽然一次性投资较高，但是减少每年的光纤租金，综合成本低；基于可重构光分插复用器（ROADM）的 DWDM 平台提供类似裸光纤的透明传输功能，易于网络的维护及排错；自管理的 DWDM 平台，端口种类丰富，开通快速；如果网络设备直接使用裸光纤，虽然易用，但是提供的带宽有限，可靠性不足，复用方法复杂，升级及扩展困难。

DWDM 平台为将来提供了充足的扩充能力，为快速满足互联带宽需求打下基础。DWDM 提供可以

平滑升级的 400 G ~ 4 T 的带宽容量；DWDM 升级容易，对已有的业务没有影响。

数据中心 DWDM 系统组网需要注意存储应用要求、光纤距离、光纤性能、保护方式等方面问题。存储应用的业务数量和种类，端到端时延要求决定数据中心之间的光纤距离；实际距离并不等于光纤距离，光纤距离受到时延限制，目前国内比较流行的数据中心之间光纤距离在 80 km 以下；通常城内光缆质量较差，每公里衰减 0.25 dB ~ 0.4 dB 不等，距离较短时通常不需要考虑色散补偿；应设置双机保护，Y 型光纤保护，保护开关模块（PSM）光纤切换。

**4 铁路互联网售票2地3中心网络架构设计和互联建议**

**4.1 网络架构设计**

图 1 设计了铁路互联网售票 2 地 3 中心及铁路局之间的主要网络。同城的两个中心实现同步复制，同城两中心与异地灾备中心实现异步复制。同城两中心自建 DWDM 平台，实现客票网、客服内网、客服外网、SAN 网络的互联需求。铁路局通过上联路由器与 3 个中心的外联路由器互联。

**4.2 数据中心间网络互联建议**

不使用传统的交换机 2 层互联方式进行数据中心间的互联。这种互联模式会导致互联后的两个数据中心变成一个数据中心平面，在这样一个差错域、广播域、生成树域、首跳冗余性（FHRP）协议域内互联，会导致广播和差错的泛滥、故障的牵连，甚至导致整个灾备网络系统比单独系统还不稳定。

通过提供 3 层与 2 层互通并存的平台，即利用叠加模式实现 2 层互通。所谓叠加模型是指在已有 3 层协议上叠加 2 层协议而实现 2 层互联，而传统的 VLAN Trunking、紧耦合的 VSS/IRF 无法同时提供 3 层与 2 层互通。

路由器平台能够提供更佳的路由支持能力，由于数据中心的 2 层连通流量需要承载在 3 层流量中，所以路由的收敛速度也就决定了 2 层连通在故障环境下的收敛速度，这点对于 Cluster/vMotion 等时延敏感业务是至关重要的；通过提供纯粹的 L3 功能，没有 Spanning Tree 技术的制约，利用基于互联网工程



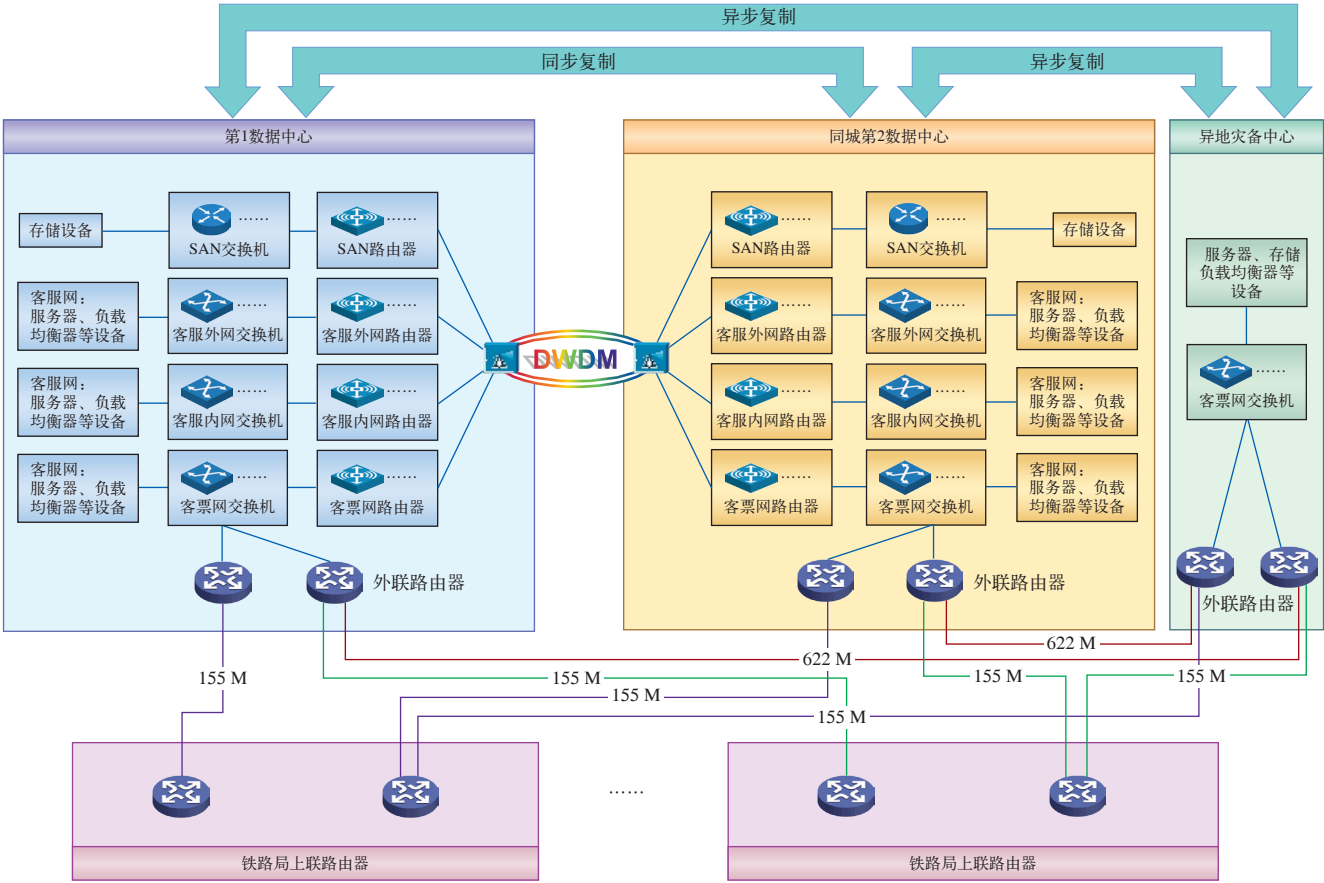


图1 铁路互联网售票2地3中心网络架构

任务组（IETF）标准的 L2 over L3 技术，承载 L2 流量的同时，利用路由控制平面技术解决地址学习泛洪、未知单播泛洪等多种传统局域网技术问题，解决数据中心流量路径优化的能力；同时通过提供比交换机平台更丰富的功能，可以实现更佳的 QoS 保障、更全面的流量管理识别能力，进一步提高数据中心业务调度的安全性和可用性。

所以建议采用 3 层方式实现 2 个数据中心的互联，并在 3 层基础上打通 2 层网络服务，可以使用标准的 L2 Over L3 技术，核心交换机与数据中心互联区路由器可单独建立 L2 和 L3 两条以上链路，各自承载 L3 和 L2 互联数据。同时建议 2 地 3 中心的网络逐步从紧耦合向松耦合发展。

## 5 结束语

信息系统容灾网络的技术在不断发展中，容灾需求的变化也对网络的要求越来越高，网络的设计应不断改进以适合自身容灾的需要。铁路互联网售

票系统业务以及传统售票业务逐渐趋向集中，容灾也越来越重要，容灾网络的设计需要综合考虑原有网络结构和新技术的发展，最终达到服务于业务容灾需求和安全的目标。

### 参考文献：

- [1] 李智慧. 大型网站技术架构核心原理原理分析 [M]. 北京：电子工业出版社，2013.
- [2] 刘 鑫. 高性能网站构建实战 [M]. 北京：人民邮电出版社，2013.
- [3] 康 楠. 数据中心系统工程及应用 [M]. 北京：人民邮电出版社，2013.
- [4] 郭 欣. 构建高性能 Web 站点 [M]. 北京：电子工业出版社，2014.
- [5] Jeffrey L.Whitten, Lonnie D.Bentley, Kevin C.Dittman. 系统分析与设计方法 [M]. 肖 刚，孙 慧. 北京：机械工业出版社，2003.

责任编辑 方 圆