MDC 技术白皮书

Copyright © 2021 紫光恒越技术有限公司 版权所有，保留一切权利。
非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。
除紫光恒越技术有限公司的商标外，本手册中出现的其它公司的商标、产品标识及商品名称，由各自权利人拥有。本文中的内容为通用性技术信息，某些信息可能不适用于您所购买的产品。

目 录

1 概述 1
1.1 产生背景 1
1.2 技术优点 1
2 MDC技术实现 1
2.1 MDC的创建 1
2.2 MDC分配资源 1
2.3 MDC的启动 2
2.3.1 MDC-OS 2
2.3.2 MDC的数据同步 3
2.3.3 MDC的进程隔离 4
2.4 MDC二层转发运行机制 5
2.5 MDC三层转发运行机制 6
2.6 管理非缺省MDC 7
2.6.1 登录通道 7
2.6.2 配置方式 8
2.6.3 配置文件 8
2.6.4 日志输出 8
2.6.5 用户管理 9
2.7 停止MDC 10
2.8 删除MDC 10
3 IRF下的MDC功能 10
4 MDC的高可靠性和ISSU 12
4.1 MDC进程高可靠性 12
4.2 MDC的ISSU 13
5 典型组网应用 13
5.1 MDC典型组网应用 13
5.2 IRF+MDC综合组网应用 14

1 概述

1.1 产生背景

随着虚拟化服务器技术在数据中心的应用，用户对网络设备的虚拟化也提出要求。例如某公司新增一项业务，他们通过虚拟服务器搭建后台应用，同时需要对服务器连接的网络设备进行系列配置，重新购买、部署网络设备会增加预算费用。可以将现有设备划分出一个虚拟网络设备来承担新业务的接入和数据转发，划分出来的虚拟网络设备可以单独组网、单独配置，对已部署网络没有任何影响。划分出的虚拟网络设备具有物理网络设备的所有业务功能，也可以托管给业务部门单独管理。

将物理网络设备通过虚拟化技术划分成多台虚拟网络设备的技术称为 MDC（Multitenant Device Context，多租户设备环境）技术，这些虚拟网络设备称为 MDC。

1.2 技术优点

MDC 具有以下优点：

- 复用：同一物理设备上创建的所有 MDC 共用物理设备的物理接口、CPU、内存、磁盘资源，管理员可以根据需要为 MDC 分配资源，或者将一个 MDC 的资源收回分配给其它 MDC，使这些资源能得到最大利用。
- 隔离：每台 MDC 拥有自己专属的软硬件资源，独立运行、独立转发、独立提供业务。创建、启动、重启、删除一台 MDC，不会影响其它 MDC 的运行。
- 易管理：多台 MDC 集成在一台物理设备上，配置界面统一，可由一人维护，也可为每个 MDC 分派管理员，免去维护多台物理设备的烦恼，还可以节省机房空间。

2 MDC技术实现

2.1 MDC的创建

设备支持 MDC 功能后，整台物理设备就是一个 MDC，称为缺省 MDC。当用户登录物理设备时，实际登录的就是缺省 MDC。用户在物理设备上的配置实质就是对缺省 MDC 的配置。缺省 MDC 的名称为 Admin，编号为 1。缺省 MDC 不需要创建，不能删除。缺省 MDC 拥有对整台物理设备的所有权限，它可以使用和管理设备所有的硬件资源。缺省 MDC 下可以创建/删除非缺省 MDC。

与缺省 MDC 相对应的是非缺省 MDC，非缺省 MDC 下不可再创建/删除非缺省 MDC。

创建完成的非缺省 MDC 状态为 inactive，此状态下的 MDC 并没有启动。MDC 的创建相当于组建了一台具有硬件基础的虚拟设备。

2.2 MDC分配资源

对于系统的稀缺硬件资源，各 MDC 间可通过命令行分配协调，来达到资源最有效的利用。缺省 MDC 上可以给非缺省 MDC 分配的硬件资源包括接口、CPU 资源、磁盘空间和内存空间。MDC 创建后：

- 缺省情况下，设备上所有资源均属于缺省 MDC。非缺省 MDC 仅具有主控板的使用权限，不具有物理接口和业务板的使用权限。必须为非缺省 MDC 分配物理接口和业务板，否则非缺省 MDC 上将没有物理接口，不能转发业务报文。
- 可以根据需要在各 MDC 间进行业务板、接口、CPU、磁盘和内存空间的合理、灵活分配。
  - 业务板资源：一块业务板可以分配给多个 MDC 使用，这些 MDC 共享该业务板上的资源。
  - 接口资源：接口作为最关键的物理设备资源，可以在缺省 MDC 和非缺省 MDC 间分配。建议在 MDC 启动前将接口资源配置好，以减少资源分配带来的系统震荡。同时，为了能将新插入的业务板分配给非缺省 MDC，设备也支持在 MDC 运行状态下将接口分配给 MDC。
  - CPU 资源：所有的 MDC 共享设备的 CPU 资源。CPU 资源具有可抢占性。如果某个 MDC 分配了较少的 CPU 资源，当有突发的任务需要 CPU 处理，而其他 MDC 任务较少时，该 MDC 可抢占使用空闲的 CPU；当其他 MDC 同时也有大量任务需要 CPU 处理时，将按配置的比率在各 MDC 间进行 CPU 调度。以免某个 MDC 过多的占用 CPU，而导致其他 MDC 无法运行。
  - 内存资源：所有的 MDC 共享设备的内存空间。为了防止一个 MDC 过多的占用内存，而导致其他 MDC 无法正常运行业务，可以限制 MDC 对内存的使用。
  - 磁盘资源：创建 MDC 后，会在磁盘介质中创建对应的 MDC 目录，存放该 MDC 的配置文件、日志文件等掉电不丢失的数据。所有的 MDC 共享设备的磁盘空间。为了防止一个 MDC 过多的占用磁盘空间，而导致其他 MDC 无法保存配置文件、系统日志等，可以限制 MDC 对磁盘空间的使用。

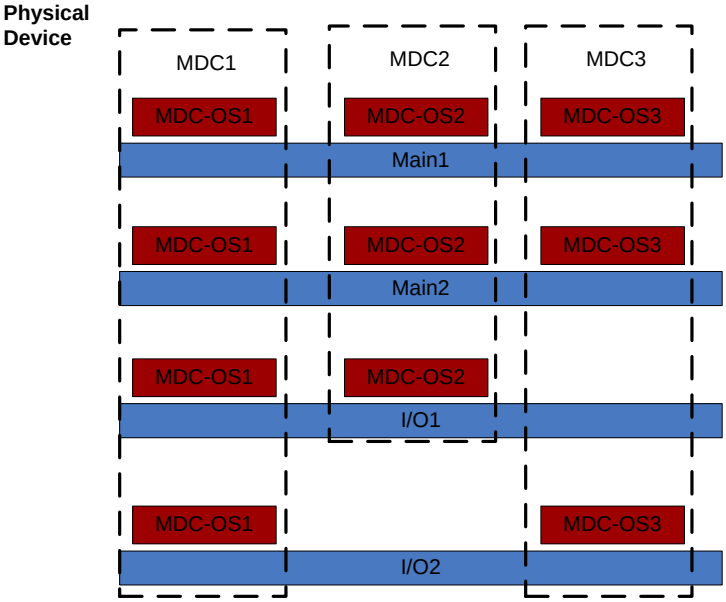
2.3 MDC的启动

MDC 的启动相当于将已经组成的虚拟设备上电启动，MDC 启动过程类似物理设备的启动。在 MDC 启动过程中，MDC 操作系统（MDC-OS）环境开始建立，MDC 内的进程开始初始化，并完成 MDC 的配置恢复。启动完成的 MDC 状态为 active，此状态下的 MDC 可以提供服务，可登录该 MDC 配置该 MDC 的运行参数。

2.3.1 MDC-OS

典型的分布式设备包括多块主控板和多块接口板。创建并启动MDC后，所有主控板、分配给该MDC的业务板上都创建并启动MDC。如 图 1 所示，每个单板上运行的MDC虚拟化环境称为MDC-OS，MDC-OS是该单板上支持虚拟化的所有进程和数据的总和。所有单板的MDC-OS组成一台分布式的MDC。

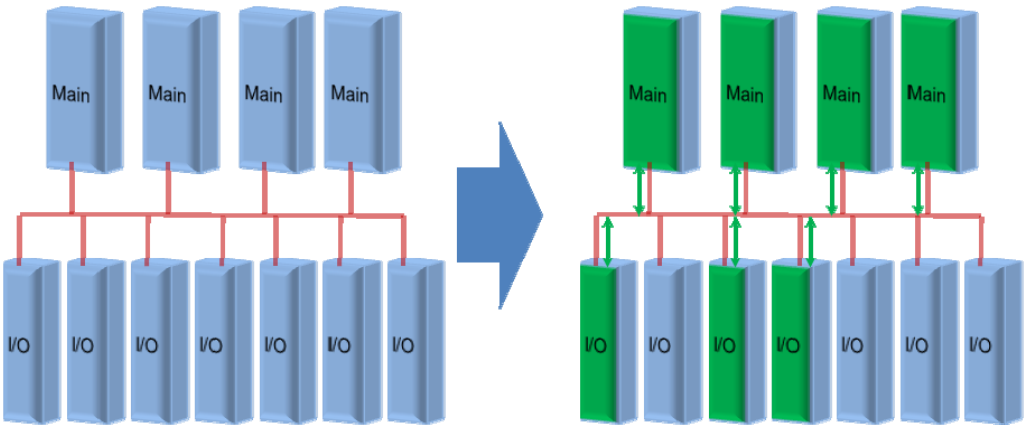
图1 分布式环境下的 MDC 布局



2.3.2 MDC的数据同步

用户登录该MDC后，只能对该MDC内的资源进行配置，即对存在MDC-OS的单板进行配置管理。进行数据同步时也只会在该MDC内的各MDC-OS单板间同步数据，与此MDC不相关的单板不会收到此MDC的任何信息。如 图 2 所示，数据只在MDC所属的单板间进行同步。可以将I/O单板只划分给某个MDC，由于该单板上的资源只归一个MDC所有，这样I/O单板能达到作为单独物理设备时的规格。

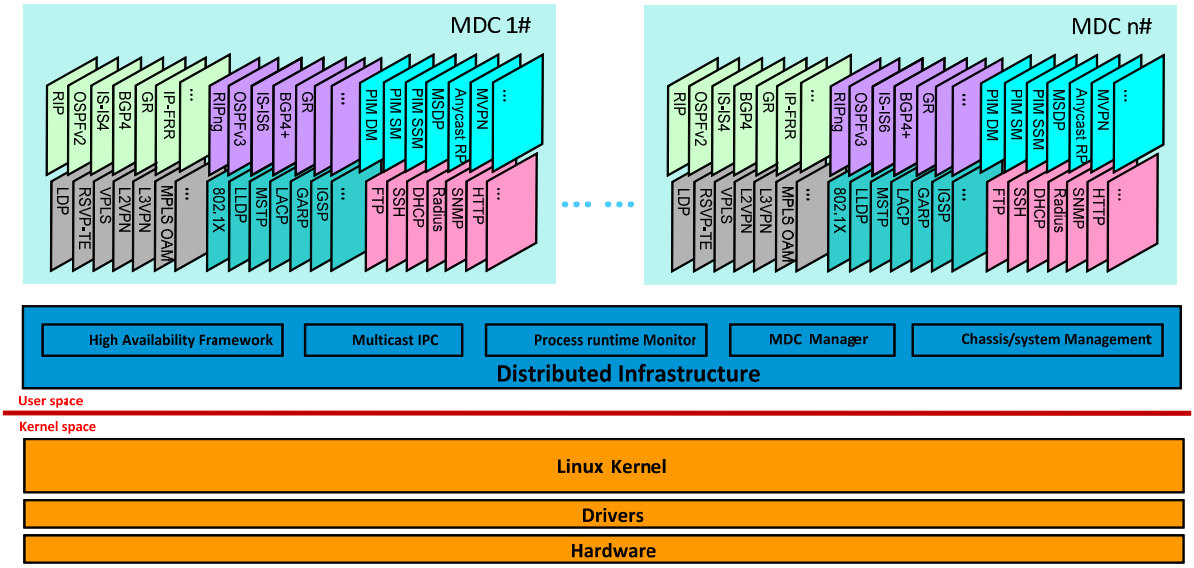
图2 MDC 的数据同步



2.3.3 MDC的进程隔离

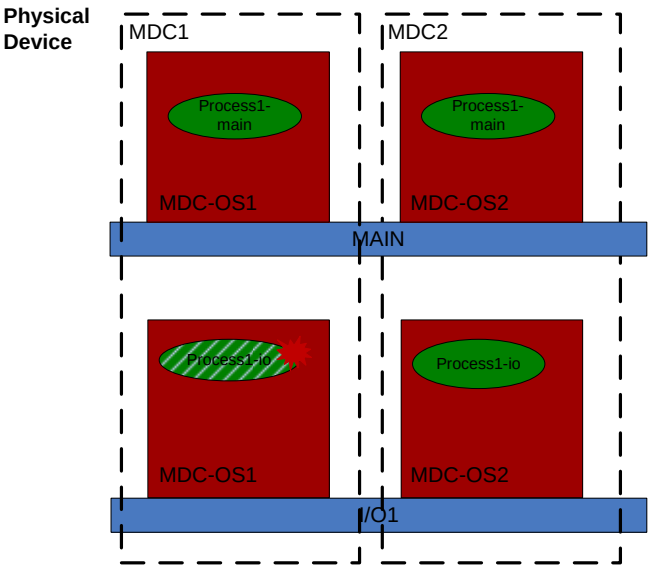
MDC将网络设备操作系统的数据平面、控制平面、管理平面进行了完全的虚拟化，各用户态进程在MDC内独立启动运行，各MDC虽然共用一个内核，但内核中MDC间数据也是分离的，从而实现每个MDC内的业务相互隔离，互不影响，给用户带来独立设备的体验。MDC的软件架构如 图 3 所示，Comware V7 的模块化实现，保证每个MDC可以分别运行自己的控制协议进程，各MDC的控制平面完全隔离，独立运行。

图3 Comware V7 模块化体系结构



Comware V7 对I/O单板的完全虚拟化的实现，让I/O单板上的所有MDC进程和数据隔离。如 图 4 所示，当I/O单板的process1-io发生异常进行进程重启等修复过程时，MDC-OS2 上的process1-io不会受到任何影响。在对各MDC进行CPU资源分配后，MDC1 的process1-io受到攻击、CPU负载比较高时，MDC-OS2 上的process1-io还是可以继续处理报文，不受影响。

图4 I/O 单板的进程虚拟化



2.4 MDC二层转发运行机制

MDC 对于二层转发的三要素（二层端口、二层转发表、VLAN）都进行了虚拟化，每个 MDC 都有独立的二层端口、独立的二层转发表和独立的 VLAN。

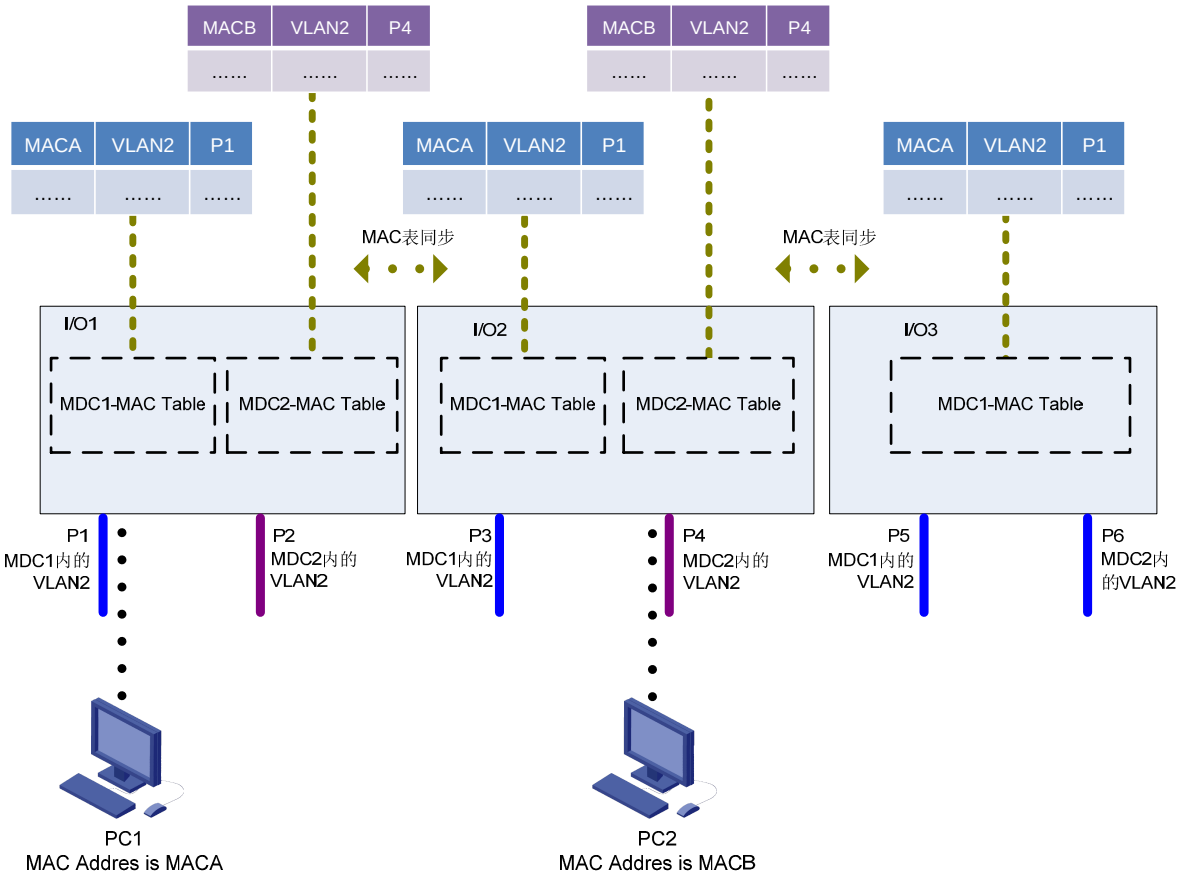
每个MDC可以创建相同编号的VLAN。如 图 5 所示：

- MDC1 拥有独享的二层端口 P1、P3、P5、P6；MDC2 拥有独享的二层端口 P2、P4。
- MDC1 和 MDC2 中都可以创建 VLAN2。
- PC1 的 MAC 地址为 MACA；PC2 的 MAC 地址为 MACB。

PC1 从 P1 端口接入设备，由于 P1 属于 MDC1，MDC1 中将生成一张 MAC 地址表，并添加表项 MACA-VLAN2-P1。该表项会在 MDC1 中同步，同步到 MDC1 内允许 VLAN2 报文通过的接口板 I/O1、I/O2 和 I/O3 上。

PC2 从 P4 端口接入设备，由于 P4 属于 MDC2，MDC2 中将生成一张 MAC 地址表，并添加表项 MACB-VLAN2-P4。该表项会在 MDC2 中同步，同步到 MDC2 内允许 VLAN2 报文通过的接口板 I/O1 和 I/O2 上。对于 I/O3 单板，由于无 MDC2 端口，则无需同步表项 MACB-VLAN2-P4。

图5 MDC 的二层转发



二层协议如 MSTP、LACP 等都支持虚拟化，每个 MDC 作为独立设备分配不同的桥 MAC 参与 MSTP 计算。如需将不同 MDC 中的接口聚合，需要配置跨设备聚合功能。

2.5 MDC三层转发运行机制

MDC相对VPN，做了更彻底地操作系统级别的隔离，安全性和操作维护的方便性都有更进一步的提高。各MDC独享VPN资源，如 图 6 所示：

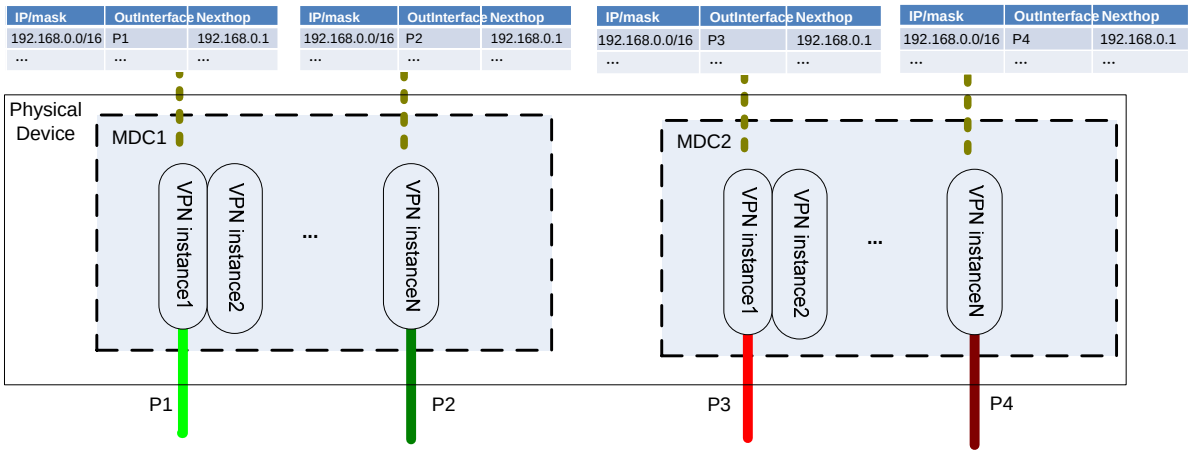
- 不同的 MDC 内可以有相同编号的 VPN 实例。
- 不同 VPN 内可以有相同的私网 IP。
- 每个 MDC 的每个 VPN 中都各自管理本 VPN 的三层转发表项（FIB）。

物理设备支持的 VPN 实例的最大数增加为：MDC 最大规格数 × VPN 最大规格数。

说明

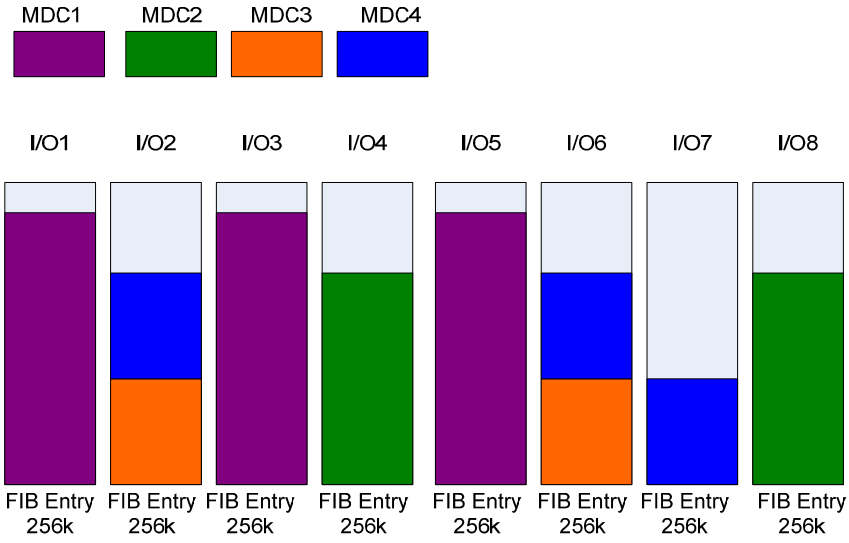
由于硬件资源限制，划分 MDC 后，物理设备支持的 VPN 实例的最大数可能达不到“MDC 最大规格数 × VPN 最大规格数”，请以设备实际情况为准。

图6 MDC 的三层转发



另外，通过合理部署MDC，可以在表项隔离的同时，每个MDC内也获得比较大的表项规格。如 图 7 所示，三层应用较多的MDC1 和MDC2，占用FIB表项较多，可将接口板仅分配给该MDC，其他MDC的FIB表项不会同步到此接口板，不会占用该接口板的FIB表项规格。如果MDC4 也需要较多的三层转发表项，通过查询发现I/O2、I/O6、I/O7 这几块接口板上的FIB表项规格剩余较多，可优先将MDC4 部署在这几块接口板上。

图7 MDC 的 FIB 规格



2.6 管理非缺省MDC

2.6.1 登录通道

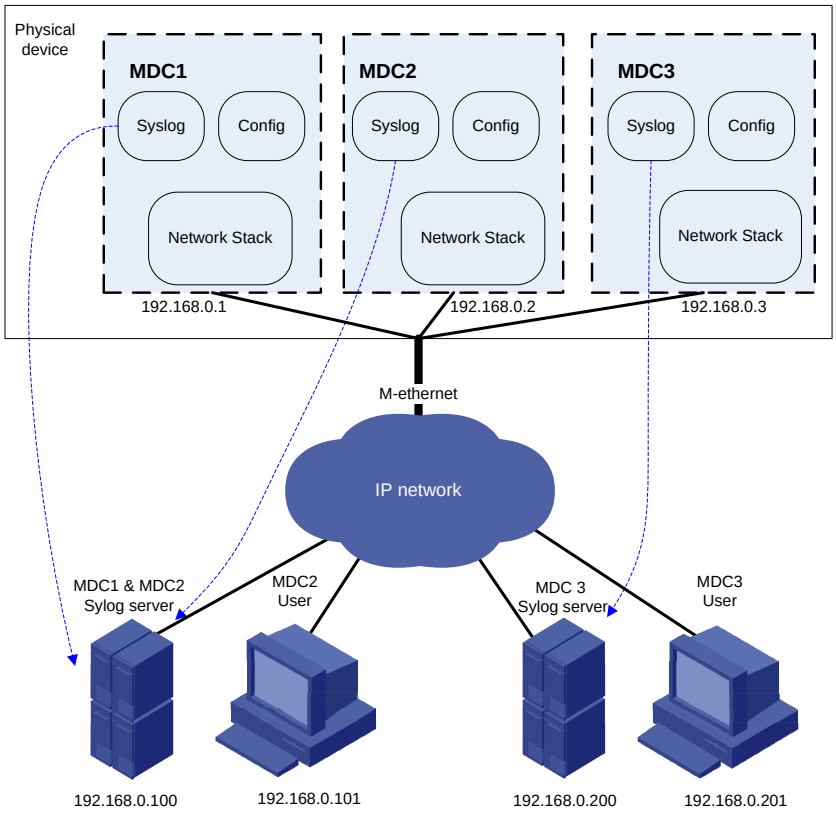
1. 内联接口

缺省 MDC 和非缺省 MDC 之间存在内联接口，其效果相当于 Console 口。如果用户当前已经登录了缺省 MDC，通过内联接口可登录非缺省 MDC，借用用户和缺省 MDC 之间的物理连接和会话连接来访问非缺省 MDC。此时，命令行视图将从缺省 MDC 的系统视图切换到指定 MDC 的用户视图。用户第一次登录 MDC 时，必须使用内联接口登录非缺省 MDC。

2. 虚拟以太网管理接口

MDC技术实现了对以太网管理接口的虚拟化。用户创建MDC时，系统会为每个非缺省MDC创建虚拟以太网管理接口，这些虚拟以太网管理接口共用物理以太网管理接口的通道资源。每个虚拟以太网管理接口都有独立的MAC。用户通过内联接口登录非缺省MDC后，可以为每个MDC的管理以太网接口配置不同的IP地址。用户通过和不同的IP地址建立连接来对各MDC进行管理。如 图 8 所示，MDC2 的用户通过和MDC2 的虚拟以太网管理接口建立连接，来对该MDC2 进行管理。

图8 通过虚拟以太网管理接口登录设备示意图



3. 物理接口

用户通过内联接口登录非缺省 MDC，在 MDC 上创建 VLAN 接口并配置 IP 地址，将 VLAN 接口和物理接口绑定，或者直接给三层以太网接口配置 IP 地址，完成路由配置后，可以通过非缺省 MDC 的物理接口使用 Telnet 或 SSH 等方式登录非缺省 MDC。

2.6.2 配置方式

如 图 8 所示，每个MDC都支持命令行、Web、SNMP、Netconf等配置方式。这些配置方式的使用方法和效果同物理设备。

2.6.3 配置文件

每个 MDC 有独立的配置文件，能独立保存本 MDC 的配置和进行配置恢复。MDC 内的用户可以对本 MDC 进行单独的重启操作，重启后 MDC 采用自己的配置文件进行配置恢复。

2.6.4 日志输出

如 图 8 所示，每个MDC在运行过程中，各自独立生成、输出Syslog。用户进行业务部署时，可以将多个MDC的Syslog信息输出到一台日志服务器中，也可以独立部署。

一些设备相关的系统事件，如重启整台设备等，在所有 MDC 内都会产生 Syslog 消息；一些单板相关的系统事件，如插入单板、拔出单板、重启单板等，在拥有这些单板使用权限的 MDC 内都会产生 Syslog 消息。

用户登录 MDC 后，只能看到当前 MDC 产生的 Syslog，无法看到其他 MDC 的 Syslog。管理员如需查看某个 MDC 的 Syslog，请登录到该 MDC 查看。

2.6.5 用户管理

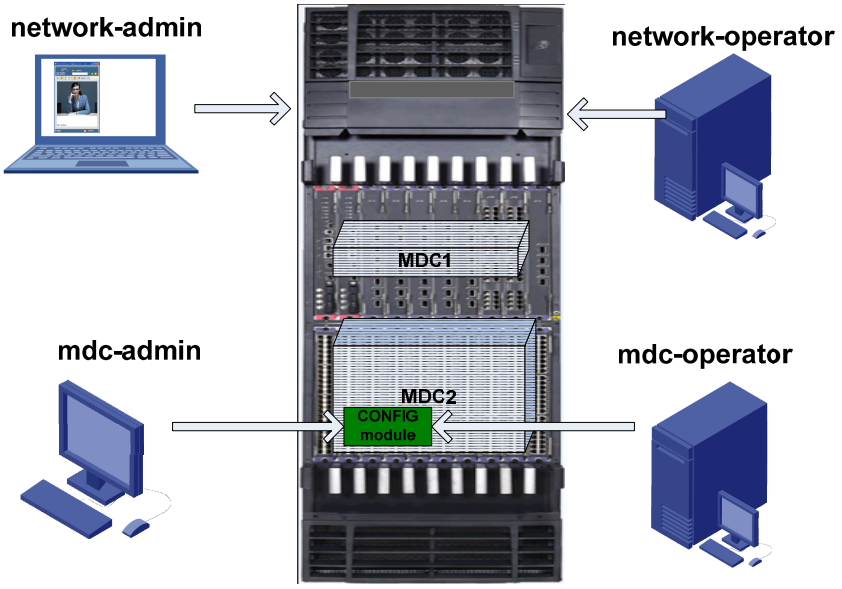
Comware 基于用户角色对用户进行权限管理。为用户赋予用户角色后，该用户登录设备后可执行指定范围内的操作。

如 图 9 所示，支持MDC的网络设备缺省支持 4 种用户角色，针对虚拟化有两级权限划分：

- 物理设备级
  - network-admin：可操作系统所有功能和资源，有创建、删除 MDC 等权限。
  - network-operator：可执行系统所有功能和资源的相关显示命令；可执行切换 MDC 命令，登录到 MDC 后可以执行 MDC 内的资源相关的显示命令。
- MDC 级
  - mdc-admin：可操作该 MDC 所有功能和资源，不能创建、删除 MDC，不能切换 MDC。
  - mdc-operator：可执行该 MDC 所有功能和资源的相关显示命令。

每个 MDC 都可以独立添加用户进行权限管理。MDC 的管理用户只能登录到对应的 MDC，对该 MDC 进行控制，而对其他 MDC 没有管理权限。当对一个 MDC 进行管理时，不会影响其他 MDC 的运行。由于管理用户只对一个 MDC 进行控制，也简化了管理环境，降低了维护难度。

图9 设备的用户角色划分



2.7 停止MDC

该操作只能对启动完成的 MDC 执行，和 MDC 的启动相反，执行停止操作后 MDC 状态为 inactive。停止 MDC 后将删除 MDC-OS 环境，该 MDC 内的所有进程将不再提供服务。但 MDC 的停止并没有释放分配的资源，停止后可以通过启动来重新恢复 MDC 的运行。MDC 的停止相当于将组成的虚拟设备下电。

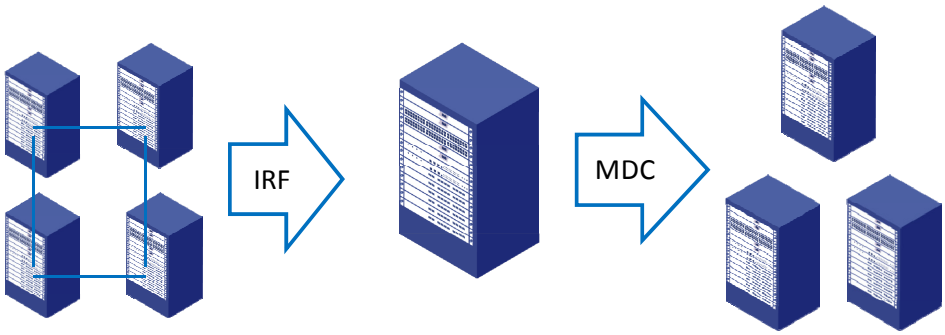
2.8 删除MDC

删除操作完成后 MDC 将不存在。删除操作可以在 MDC 状态为 active 和 inactive 时执行。对于 active 状态的 MDC，执行删除操作会先停止 MDC，再删除 MDC 删除。删除 MDC 后，该 MDC 所占用的资源将归还给缺省 MDC。

3 IRF下的MDC功能

IRF是一种N:1 虚拟化技术，使用它可以将多台物理设备虚拟成一台虚拟设备。MDC是一种 1:N的虚拟化技术。将这两种虚拟技术配合使用，可以达到多虚多的效果。即将多个设备通过IRF技术形成一个大的虚拟设备，再通过MDC技术将该虚拟设备进一步虚拟成多个MDC使用，如 图 10 所示。

图10 IRF+MDC 效果示意图



这两个技术一起使用，可以：

- 最大限度的整合设备资源，将多台设备按照组网需要虚拟成多台虚拟设备使用，提高设备的使用效率。
- 同时获得两种技术的优点。例如使得每个 MDC 同时获得 IRF 的主备主控板保护，相对于使用单个设备组网，提高了可用性。
- 提高系统的扩展性和使用的灵活性。在已经使用 MDC 的设备上，通过叠加 IRF 技术，可以在不改变网络部署的情况下扩展设备的端口数、带宽及处理能力。而在已经使用 IRF 技术的场合，直接在 IRF 的基础上使用 MDC 功能，部署新的网络，对现有网络不会产生任何影响

统一部署IRF和MDC后，IRF内所有的资源都可以在这些MDC间分配。不同物理设备的接口板可以分给同一MDC，进行报文转发时报文能通过IRF链路到达另一台物理设备上该MDC内的接口，等效于同一台物理设备上的转发。当跨成员设备的转发流量很大时，会给IRF链路带来压力。此时，请部署多条IRF链路来分担MDC的跨框转发流量，如 图 11 所示：

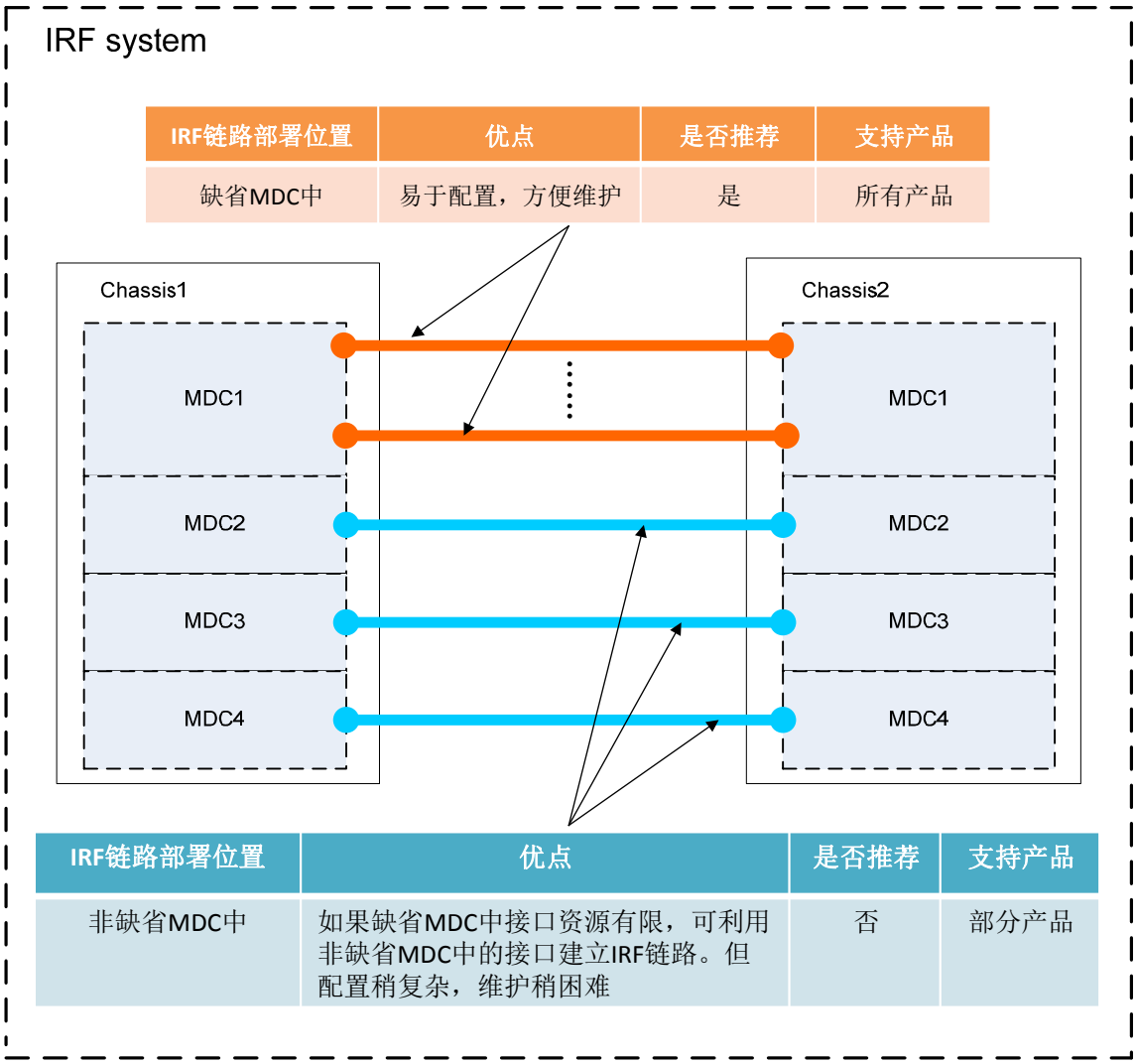
- 缺省 MDC 中的 IRF 链路用于转发 IRF 的控制协议报文以及所有 MDC 的跨框流量。

- 非缺省 MDC 中的 IRF 链路，仅转发本 MDC 的跨框流量。

说明

不同型号的设备由于硬件差异，对 IRF 链路的部署要求以及跨框流量的处理策略可能不同，请以设备的实际情况为准。

图11 IRF+MDC 组网环境 IRF 链路部署示意图



4 MDC的高可靠性和ISSU

4.1 MDC进程高可靠性

Comware V7 采用进程级 GR 技术来实现进程级的高可靠性。即每个进程在运行过程中备份状态和数据，进程重启时，与此进程相关的软硬件均不删除此进程的相关数据，待进程重新运行，恢复自身数据后，与相关软、硬件进行数据平滑，继续运行，保证运行状态一致。整个重启过程中业务不中断，系统其他部分功能不受影响，邻居设备不感知此过程。异常重启的服务只在重启期间不提供服务，重启后马上恢复服务。

在MDC中配置业务后，MDC系统会运行对应的进程：一个主进程（Active）和多个备进程（Standby），主备进程分别分布在不同的主控板节点上。各MDC内的进程隔离，使得MDC内的进程级主备倒换时，不影响其他MDC的进程布局。如 图 12 和 图 13 所示，MDC1 的Process1 主进程在node2 上，MDC2 的Process1 主进程在node1 上。MDC1 内的Process2 进程进行了主备倒换后，不影响MDC2 的进程布局。

图12 MDC 内的进程布局

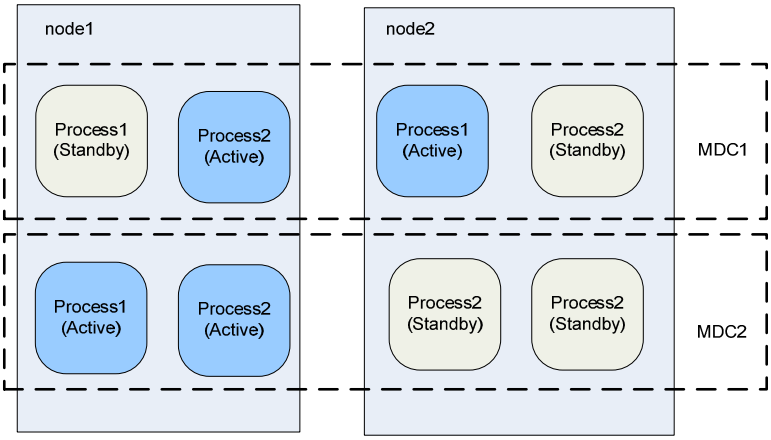
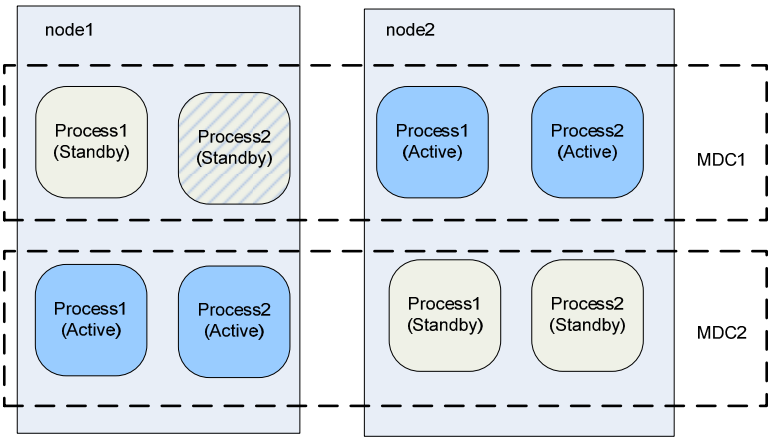


图13 进程级主备切换后的进程布局



4.2 MDC的ISSU

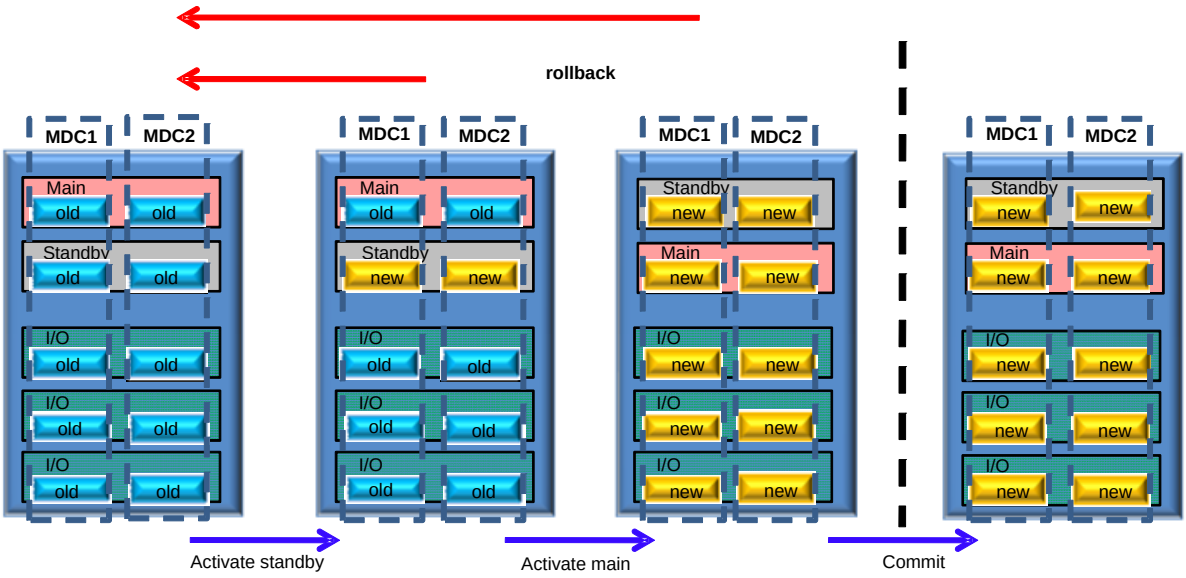
同一物理设备上的所有MDC共用一份启动文件，运行相同的软件版本。MDC的升级操作只在缺省MDC进行，升级成功后，各个MDC均完成升级。如 图 14 所示，MDC升级流程同物理设备的升级流程：

(1) 先对备用主控板进行升级，备用主控板升级完成之后，当选成新的主用主控板，替代原主用主控板工作。

(2) 原主用主控板重启升级，连带升级业务板。最终完成整台物理设备以及所有 MDC 的升级。

升级过程中如果出现异常也可以回滚到升级前的状态。

图14 MDC 的 ISSU 升级



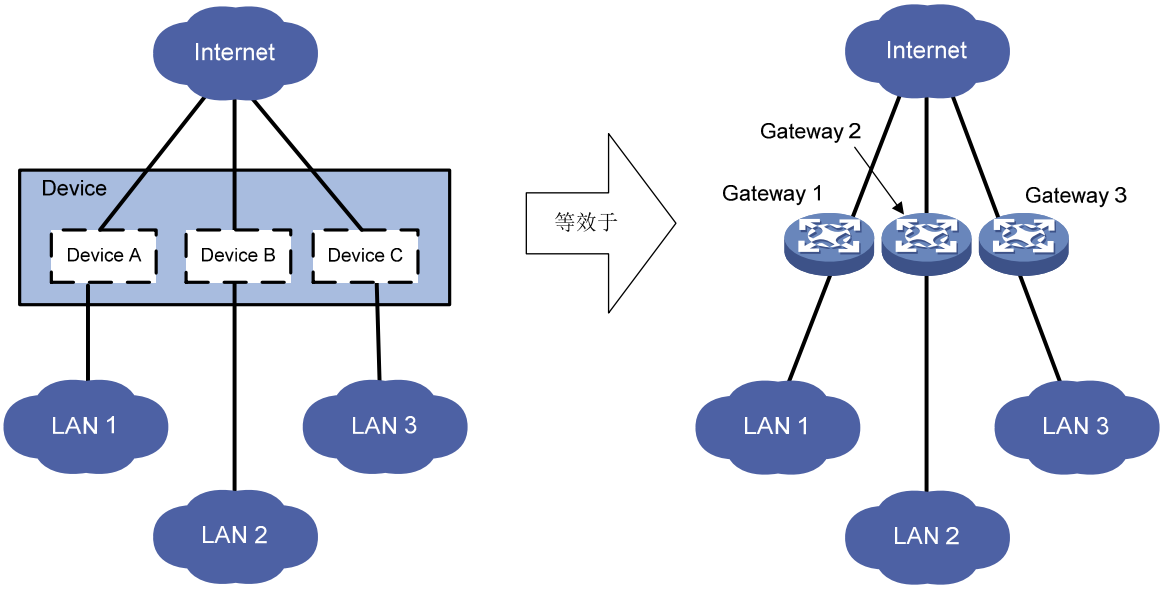
5 典型组网应用

5.1 MDC典型组网应用

MDC 的应用十分广泛，比如提供设备出租和业务托管、应用于实验教学等。

如 图 15 所示，LAN 1、LAN 2 和LAN 3 是三个不同公司的局域网，它们通过同一台物理设备Device连接到外网。通过虚拟化技术，能让一台设备当三台设备使用。具体做法是，在Device上创建三台MDC：Device A、Device B和Device C，分别负责LAN 1、LAN 2、LAN 3 的网络接入和控制。LAN 1、LAN 2、LAN 3 的网络管理员可以（也只能）分别登录到自己的接入设备进行配置、保存、重启等操作，不会影响其它网络的使用，其效果等同于LAN 1、LAN 2 和LAN 3 分别通过各自的网关Gateway 1、Gateway 2、Gateway 3 接入Internet。

图15 MDC 典型应用组网图



5.2 IRF+MDC综合组网应用

如 图 16 所示，某公司有两台设备已经组成了IRF。现要求创建两台独立的虚拟设备（MDC A和MDC B），分给该公司的A、B两个部门，用作接入公司网络的网关。通过MDC技术，可以使得各部门独享设备的表项资源，并且可以根据各部门的业务需求分配设备的内存、CPU资源。例如，A部门业务活跃，需要更多的CPU资源和内存资源；B部门人员和服务器较多，需要更多的接口资源。

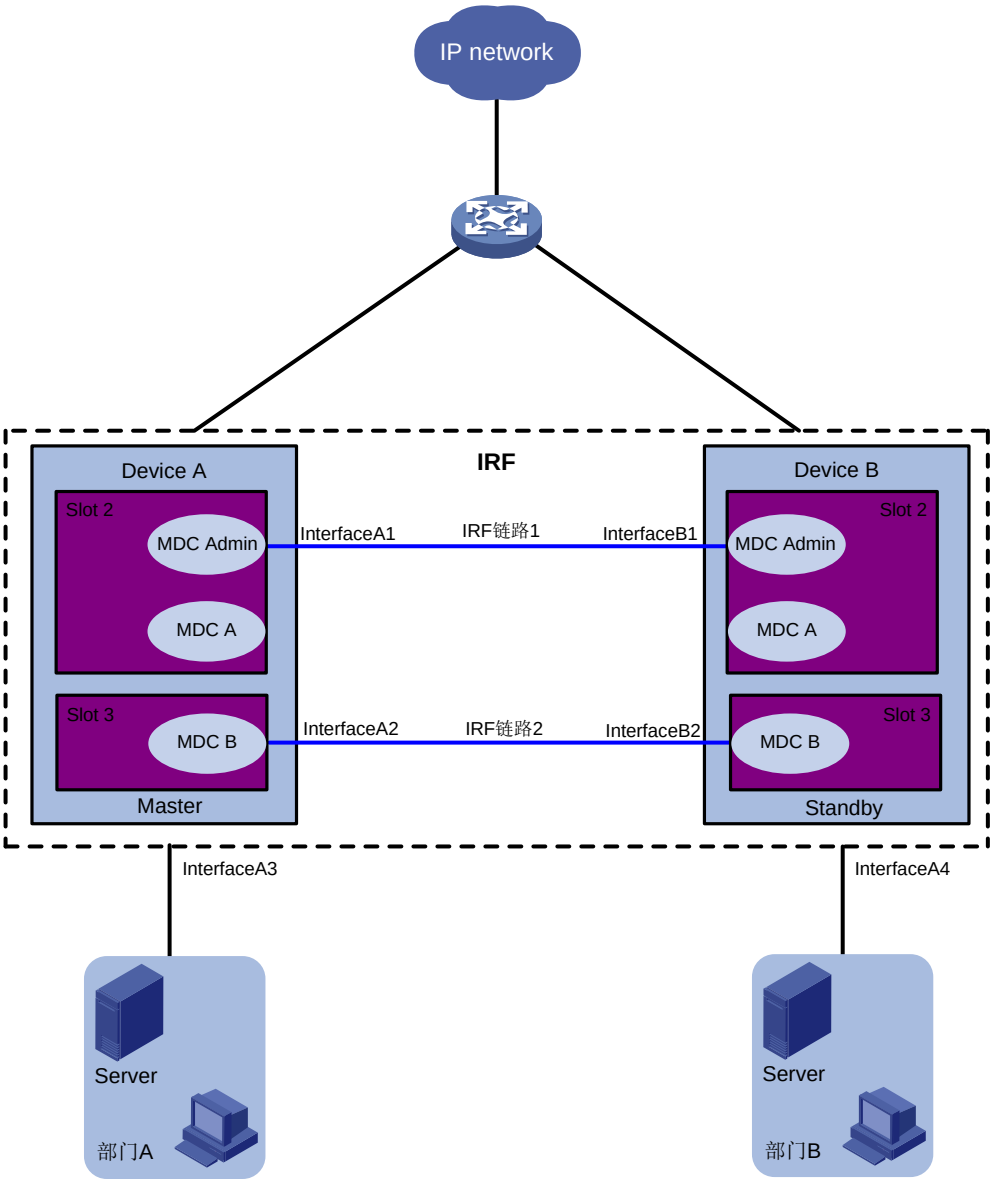



图16 IRF+MDC 综合应用组网图