OOB 远程控制系统文档

-------20130201，哥儿

本文的大体结构与思路：

所有最详尽的细节和最完整的全貌，全部都在里面。

从零开始，一步步搭建起来一个完整的 OOB 远程控制网络。

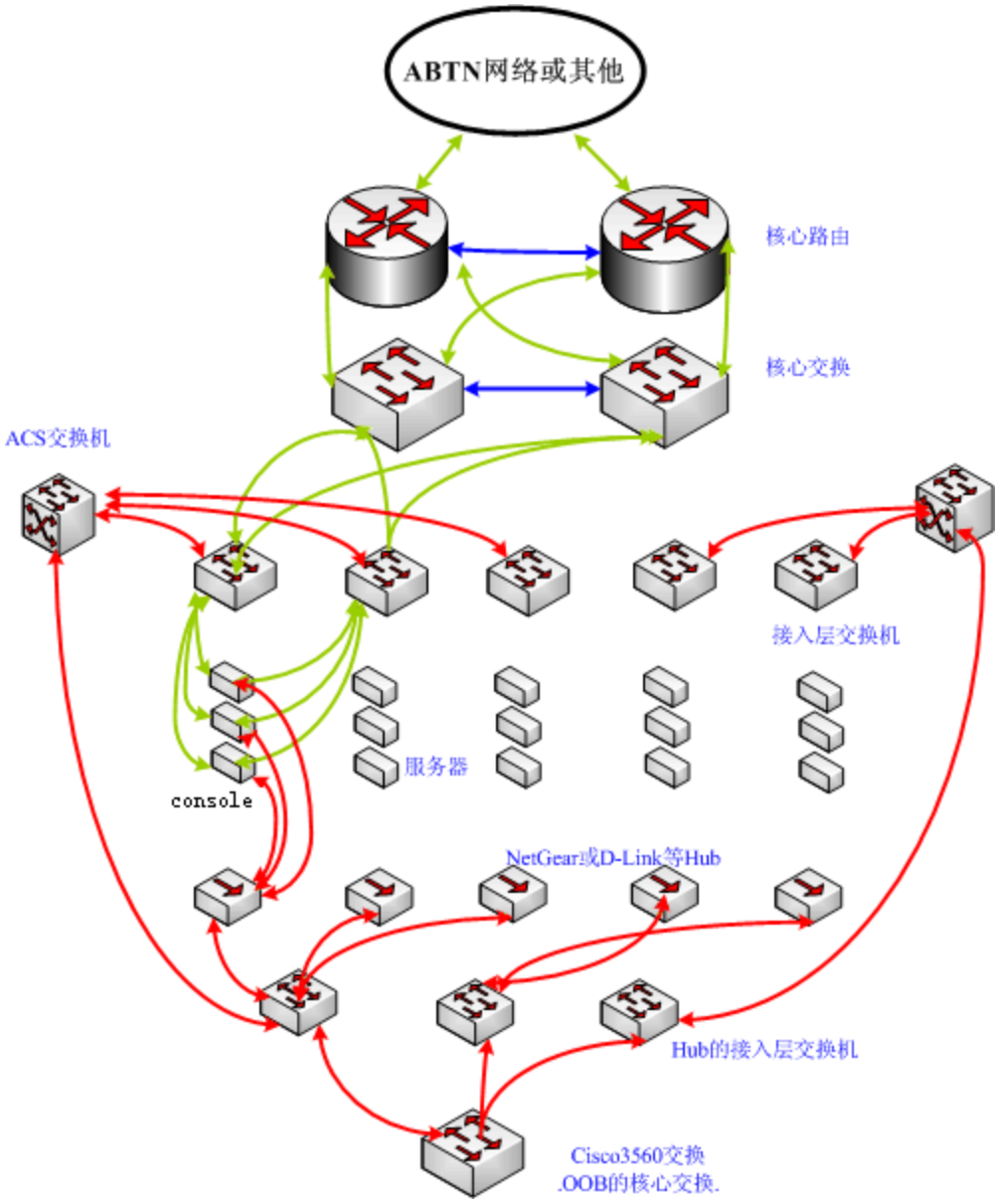
包含：硬件，系统，程序，网络，自动化等等。

根据清晰的索引提示，随意翻看，即刻使用。

##

##########

附上一幅图，OOB 的网络拓扑图，好初步有个直观的了解。
至于本图的解释，以及中间是如何构造起来的，本文后续陆续会讲到。



红色线为 OOB 网络的连线，绿色线为生产网络的连线。

####################################正文#######################################

####1：含义：

OOB 即：Out of Band Management System，带外管理系统。带外即带宽之外，生产网之外，
即独立于生产网之外的一种服务器远程控制管理系统。
由于这个系统走的是与生产网无关的另一套网络，所以可以在生产网出故障无法访问时，
作为一种应急通道来远程修复处理生产网，以达到最快速地挽救损失的目的。
最早期，服务器出故障了，得手工接上显示器键盘，手动按电源按钮。
但是当服务器规模比较大，故障比较多的时候，接显示器键盘就远远达不到要求了。
此时 OOB 派上用场了，可以远程管理控制服务器，包括电源的控制。

####2：用途：

应急故障处理，远程控制服务器：
当服务器死机时，不用去机房现场接显示器键盘，无需体会鞭长莫及的感觉。
而直接远程控制服务器（或交换机）的电源的开和关和重启，
以及观察设备状态。极大地节省时间，提高效率。
远程观察服务器状态，将运维人员解放出来，代替人工必须在机房现场：
可以远程查看目前服务器的状态，硬件如 CPU,内存，磁盘，温度等，以及 OS 的状态。
远程操作，犹如在服务器前接了显示器键盘一样直观。降低人力成本。
由于独立于生产网之外，所以可以在生产网的交换机或服务器出故障无法连连接时，
恰好发挥这套系统的作用，通过带外通道连接过去处理比如生产用的核心交换机之类的。
（当然我们不希望这种故障情况真的发生。）

####3：实现原理：

基于 IPMI：Intelligent Platform Management Interface. ，即智能平台管理接口，
一种为管理访问外围设备制定的工业标准.
BMC
-Baseboard Management Controller
-基板管理控制器.IPMI 的核心部件。
-独立于服务器的硬件及 OS 之外。
BMC 硬件卡及服务器 BIOS 设置
-BMC 硬件卡的设置
-服务器的 BIOS 设置
ipmitool，ipmiconsole（freeipmi 包含）
控制远程服务器的命令行工具。

基于 iLO：Intergrated Light-Out，惠普独家开发的远程管理系统。
惠普独有，区别于一般的 IPMI 标准。当然现在的 iLO2 和 iLO3 也支持 IPMI。
工具：hponcfg 和 lo100cfg（ILO 工具），conrep（修改 BIOS）。
HP 的 SSTK 工具包里有大量的这些需要用到的实用工具。堪称 HP 设备的瑞士军刀。

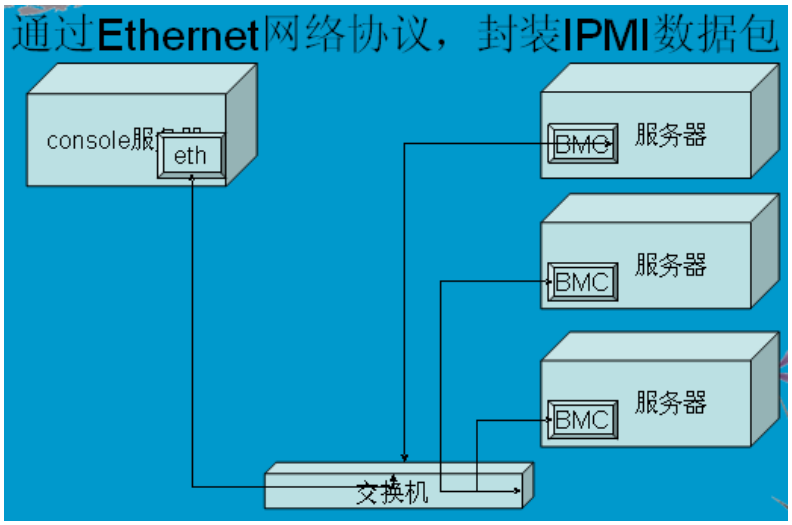
基于 RSA 的：Remote Supervisor Adapter。IBM 专有的远程控制卡。

基于硬件的 Console 串口交换机及 BootBox 电源管理的整套系统。
（目前只有 CM2 清江路以前雅虎环境有少量，但是即将清退。）

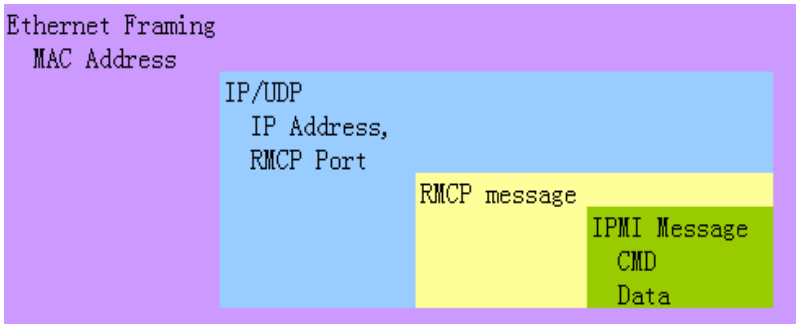
管理端与被管理端服务器之间的链路图：

管理端我们统称为 console 服务器，因为从这台服务器上我们可以以类似 console 控制台的方式控制别的机器。

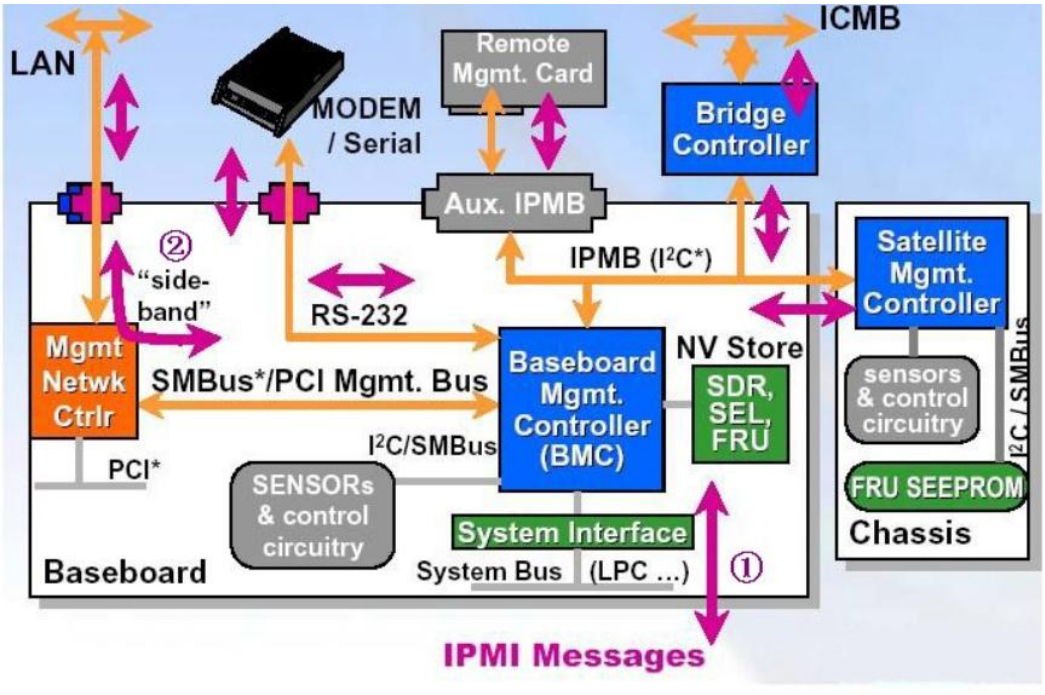
通过 Ethernet 网络协议，封装 IPMI（或 iLO）数据包：



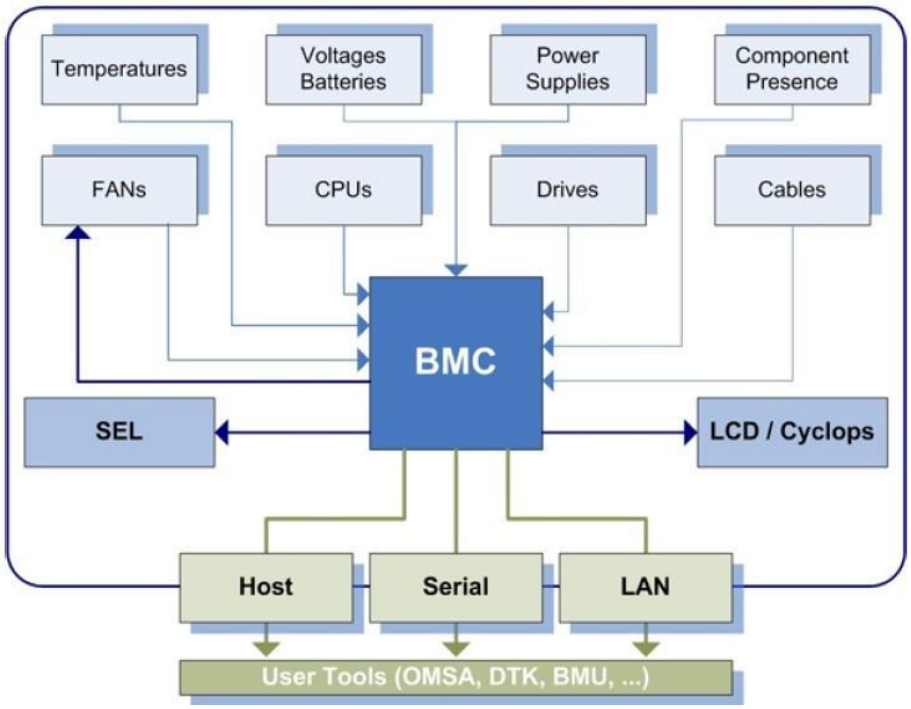
数据包帧结构：



DRAC 硬件卡的芯片逻辑图：



简化逻辑图：



####4：规划：

##4.1 系统规划：

三淘的：CNZ,CM3,4,5,6,8,9,10

阿里云的：余杭，北土城，三墩，兴义（分 xyi,xy2），东冠，

B2B 的：主要是兴义的（xyi）

这几个(20~30 个)大机房，各放一台 console 控制机，控制本机房内所有机器。

Console 服务器的配置，及其他被控制的服务器的配置。后文会陆续讲解。

##4.2 网络规划：

采用标准的网络拓扑，级联方式：

机柜内所有服务器的远程控制卡

---->每个机柜 DLINK

---->每排 Cisco2960

---->每个机房的 OOB 核心 Cisco3750

---->所有机房的最终核心 Cisco3750

本文的结尾会有一幅所有机房的 OOB 全部连接起来后的整体级联拓扑图。

网段：目前是这样规划的（以三淘的为例。 阿里云和 B2B 等在本文附带的表格里）：

CNZ 的 OOB 网段： 10.10.0.0/16

CM3 的 OOB 网段： 10.13.0.0/16

CM4 的 OOB 网段： 10.14.0.0/16

CM5 的 OOB 网段： 10.15.0.0/16

CM6 的 OOB 网段： 10.16.0.0/16

（CM7 的 OOB 网段： 10.17.0.0/16，在建机房）

CM8 的 OOB 网段： 10.18.0.0/16

由于 CM3,4,6 这类机房有很多个 room，机器数量过多，为避免维护困难及广播风暴，所以每个 room 的单独划分一个 VLAN。

##4.3：权限规划及安全控制：

所有机房的 OOB 全部连接起来后，NETOPS 拥有拨进这个网络的 VPN 权限及交换机控制权限；

所有 SITEOPS 及监控值班拥有控制所有机房的服务器的控制权；

PE 及 DBA 拥有控制所有 DB 服务器的权限。

但是以上所有人员都需通过统一的 WEB 入口进入，Console 机上没有个人账号，避免安全问题。

####5：实施方案

##5.1：BIOS 及 BMC 硬件设置

###5.1.1 各种机型：HP，DELL，HUAWEI，IBM

方案 A：新到机器，没有 OS 的情况下的首次设置，接显示器：

方案 B：有 OS，在 OS 内调用专用工具设置。

方案 A：

##DELL 的：

带 DRAC 远程控制卡的 DELL PowerEdge R610 和 R710 和 R900

加电--✳F11--✳Ctrl+E

IPMI Over LAN：On

LAN Parameters--✳ 记下 MAC Address。此即 DRAC 的 MAC。

IPv4 Settings：

IPv4 Address Source：DHCP

LAN User Configuration--✳Username：taobao

Password：(密码，文档里就不不写了)

Password：(密码，文档里就不不写了) 然后一定要按 Enter 回车保存。

保存退出。

重起后，会自动进入 BISO：

Integrated Devices

--✳ Embedded NIC1 and NIC2：Enabled

--✳ Embedded Gb NIC1：Enable with PXE。（开启 NIC1 的 PXE 功能）

--✳ Embedded Gb NIC2：Enable。

Serial Communication--✳ Serial CommunicatioN--✳On with console Red ……：COM2

--✳External Serial Connector：Remote Access Device.

--✳Failsafe Baud Rate： R610 和 R710 和 R900 此处是 115200

--✳Remote Terminal Type：VT100/VT220

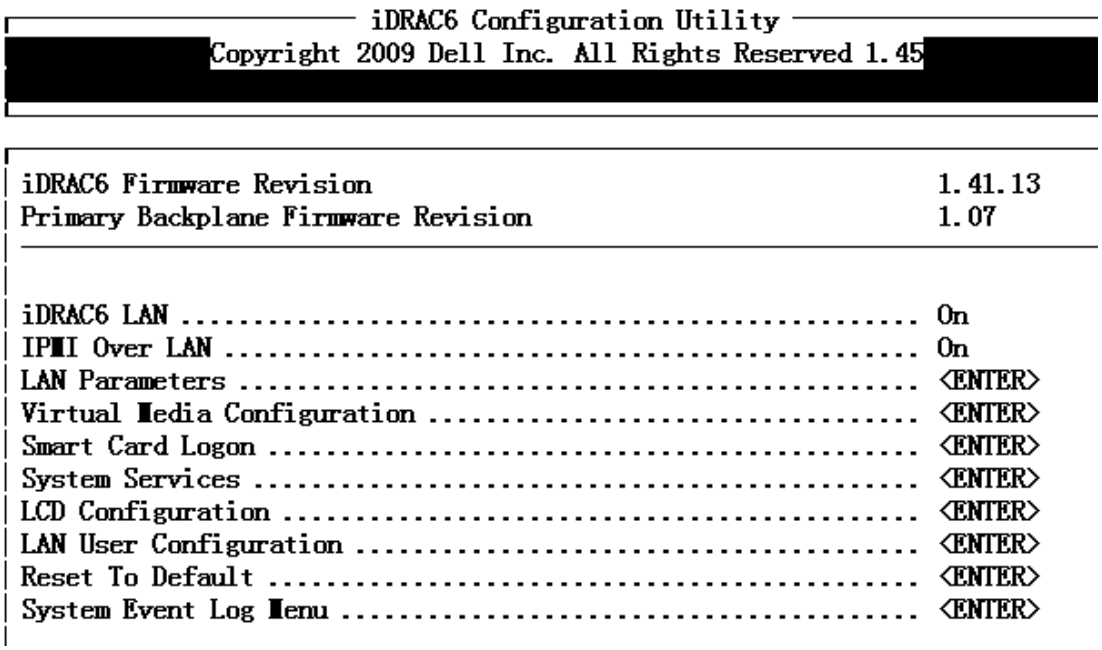
--✳Redirection After Boot：Enable.

BootSequence： C 盘要先启动，NIC1 排在最后。以免重启后不停地刻系统。

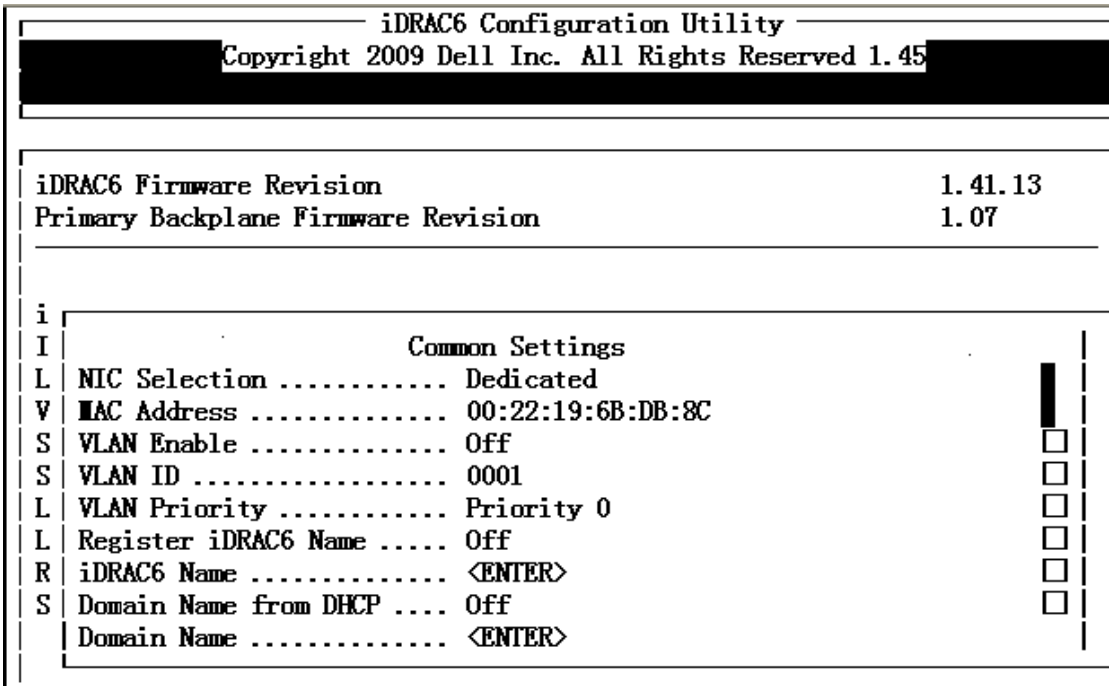
ESC 保存退出。

如截图所示：（DELL 其他型号，HUAWEI 及 HP 及 IBM 的原理类似，不再截图）：

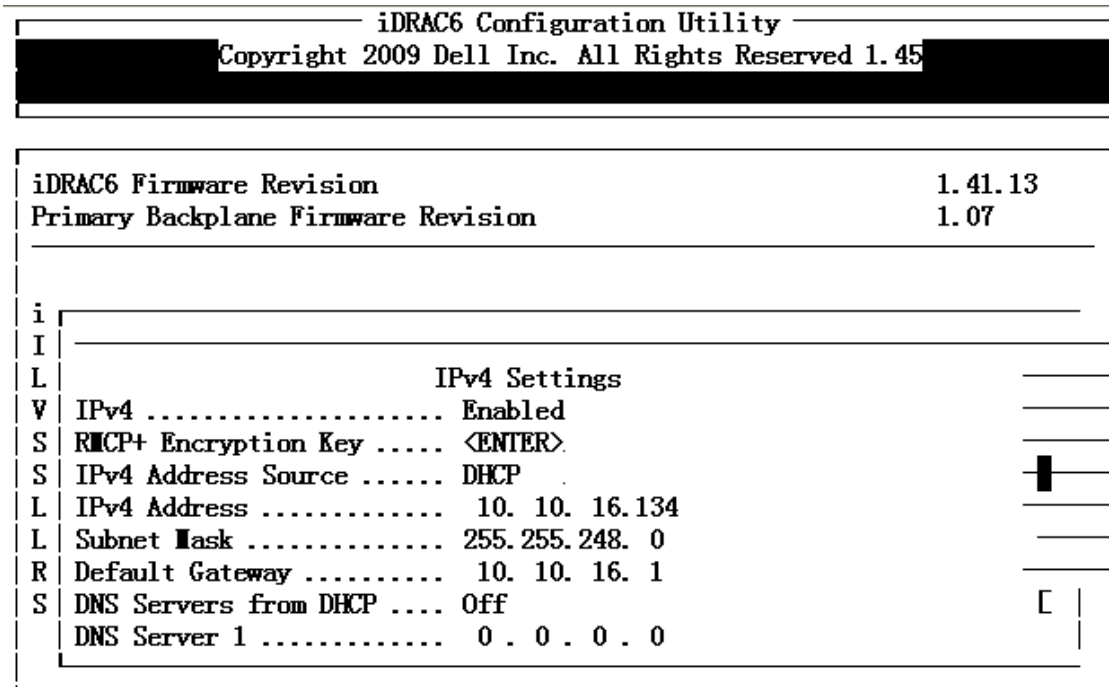
第一步：IPMIoverLAN 打开 ON（默认是 Off）：



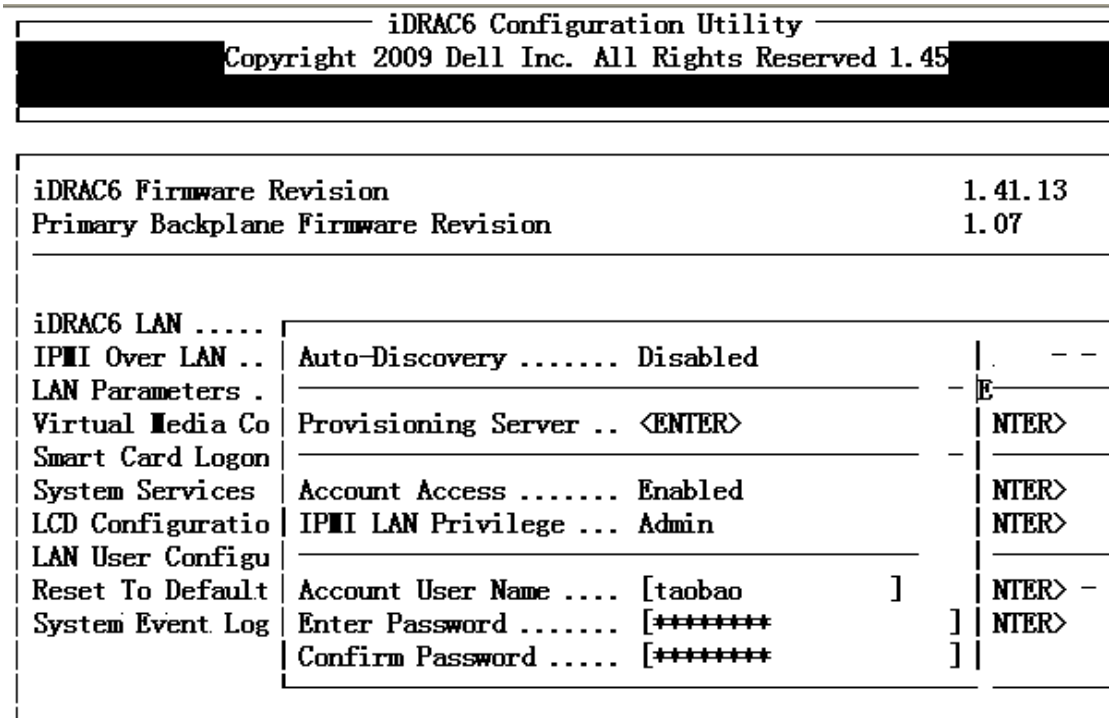
第二步：有独立 DRAC 卡的，Nic Selection 模式必须为 Dedicated；
无 DRAC 卡的，与 NIC1 共享的，此处应为 Share。
记录 DRAC 的 MAC：



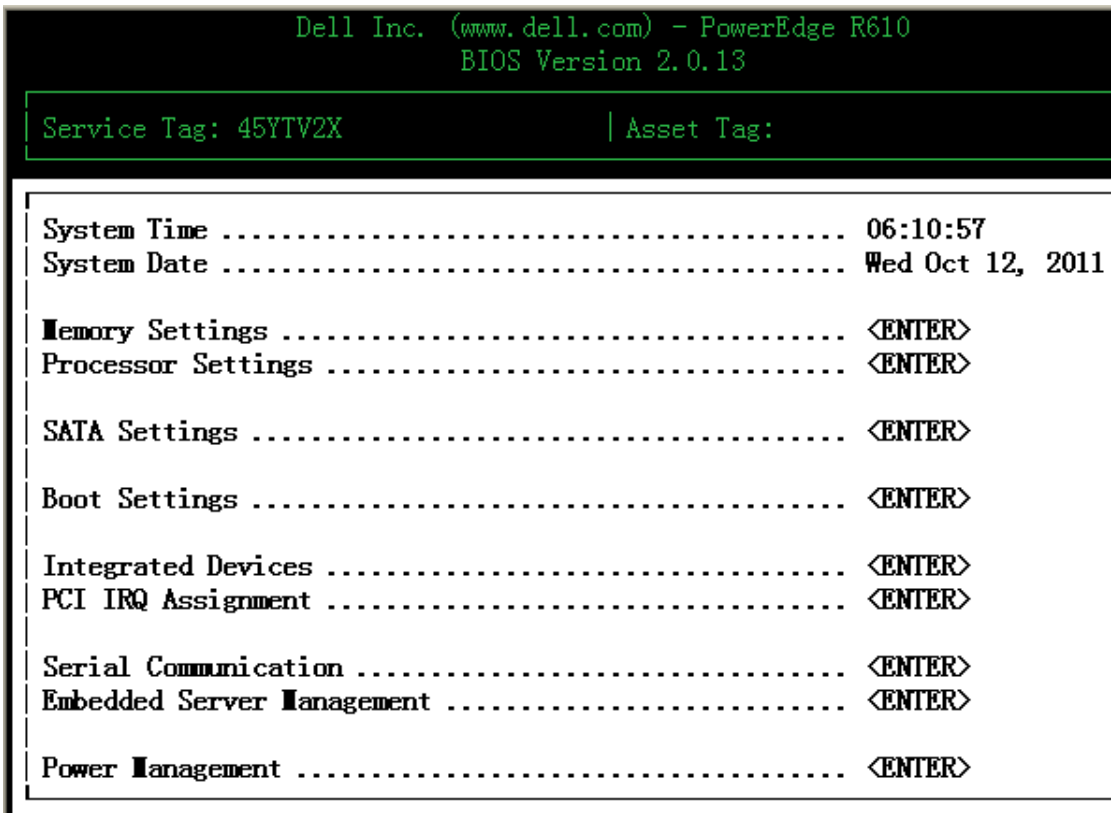
第三步：DRAC 卡的 IPv4 的 DHCP 功能打开（默认是 static）：



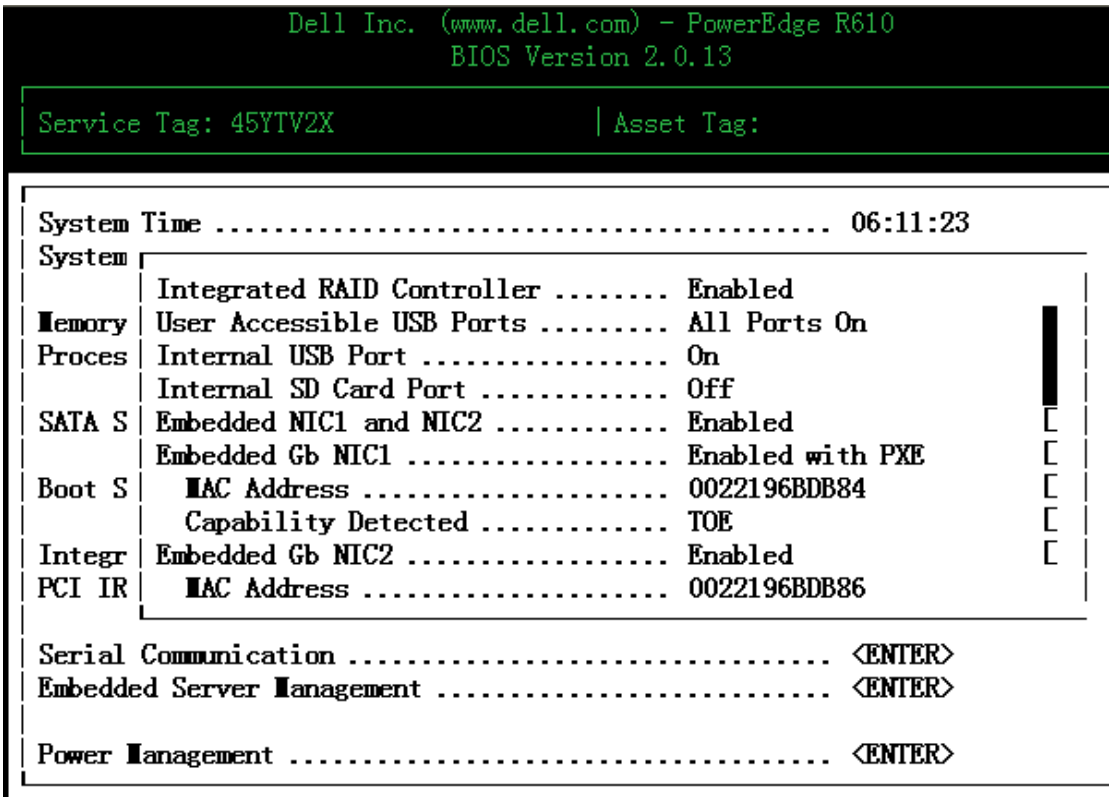
第四步：设置 DRAC 卡的管理用户名和密码：(taobao / 密码...)



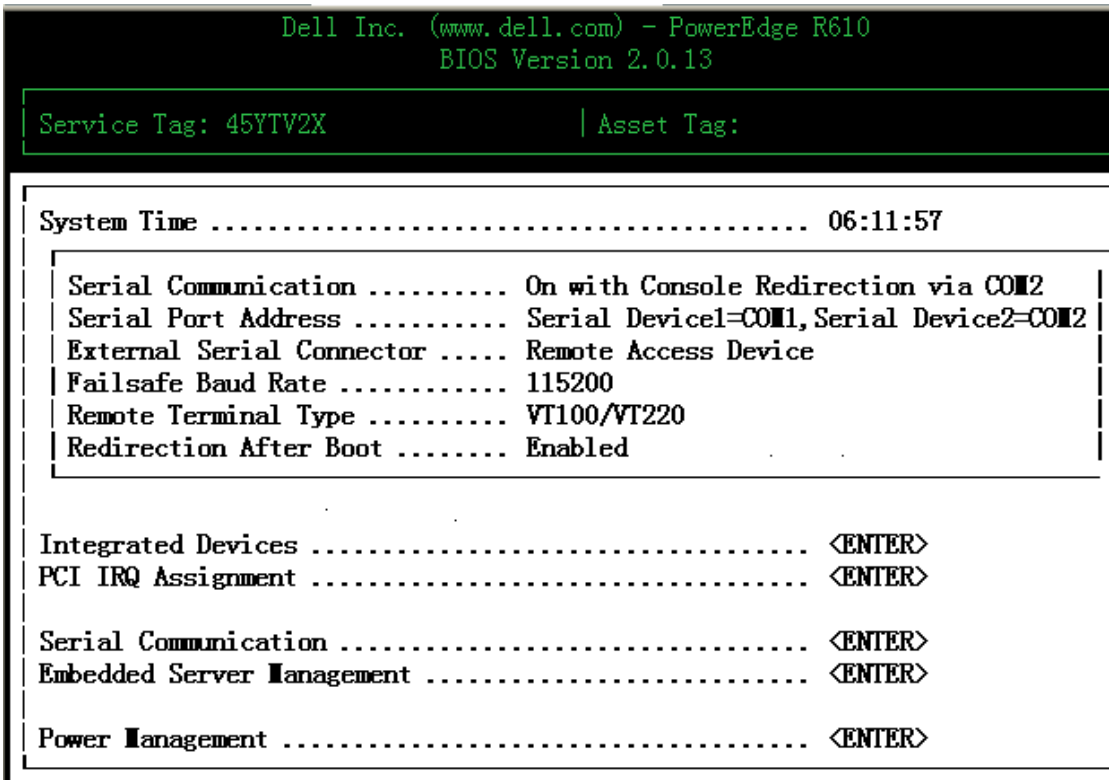
第五步：其他设置：



第六步：设置 NIC1 为 PXE 启动，以方便自动刻录 OS （这一步不属于 DRAC 的范畴）：



第七步：设置 DRAC 卡的波特率之类的参数：



然后按 F10 保存退出。

其他型号的文字描述（不再配以截图）：

不带 DRAC 卡的 DELL PowerEdge PE1950/PE2950：

加电--✳F2--✳Ctrl+E

IPMI Over LAN：On

IPMI LAN Channel--✳IP Address Source：DHCP

记下 MAC Address。此即 DRAC 的 MAC。

User Configuration--✳Username：taobao

Password：(文档里就不不写了)

Password：(文档里就不不写了) 然后一定要按 Enter 回车保存。

保存退出。

重起后，会自动进入 BISO：

Integrated Devices--✳ Embedded Gb NIC1：Disable。（即禁用 NIC1）

--✳ Embedded Gb NIC2：Enable with PXE

Serial Communication--✳ Serial CommunicatioN--✳On with ……：COM2

--✳External Serial Connector：COM2

--✳Failsafe Baud Rate：9600

--✳Remote Terminal Type：VT100/VT220

ESC 保存退出。

带 DRAC 远程控制卡的 DELL PowerEdge PE1950/2950/2970：

加电--✳F2--✳Ctrl+E

IPMI Over LAN：On

IPMI LAN Channel--✳IP Address Source：DHCP

记下 MAC Address。此即 DRAC 的 MAC。

User Configuration--✳Username：taobao

Password：(文档里就不不写了)

Password：(文档里就不不写了) 然后一定要按 Enter 回车保存。

保存退出。

重起后，会自动进入 BISO：

Integrated Devices--✳ Embedded Gb NIC1：Enable with PXE。（开启 NIC1 的 PXE 功能）

--✳ Embedded Gb NIC2：Enable。

Serial Communication--✳ Serial CommunicatioN--✳On with console Red ……：COM2

--✳External Serial Connector：Remote Access Device.

--✳Failsafe Baud Rate：57600 (R710 和 R900 此处是 115200)

--✳Remote Terminal Type：VT100/VT220

--✳Redirection After Boot：Enable.

BootSequence： C 盘要先启动，NIC1 排在最后。以免重启后不停地刻系统。

ESC 保存退出。

##HP 的：

HP ProLiant DL360 G5/G7，DL380 G5/G7, DL580 G5

加电--✳F9

BIOS Serial Console & EMS--✳ BIOS Serial Console Port：COM1。然后一定要按 F10 保存。

--✳ BIOS Serial Console Baud Rate：9600

System Options--✳Virtual Serial Port：COM1。然后一定要按 F10 保存。

(DL580 的如果查看 MAC，需要按 Tab 键才能显示。)

2009 年元旦后到货的 DL580 需要调整 Stand Boot Order 里的网卡启动顺序(看情况)。

保存退出。

加电--✳F8

Network--✳NIC and TCP/IP

记下 MAC Address。此即 ILO 的的 MAC。

--✳DNS/DHCP--✳DHCP Enable：On。（确保 On）

User--✳Add--✳User name：taobao

Password：(文档里就不不写了)

Password：(文档里就不不写了) 然后一定要按 F10 保存。

HP ProLiant DL140

加电--✳F10

Advanced--✳8042E……：Disable

记下下面两个网卡的 MAC ADDRESS：

NIC1 MAC Address

Dedicatecd NIC MAC Address。 此即 ILO 的的 MAC。

Console Redirection --✳Console Redirection：Enable

--✳Console type：VT100

--✳Flow Control：None

Continue C.R. after POST：Enable

IPMI--✳LAN Setting--✳IP Address Assignment：DHCP

并确保：

--✳BMC ……：Enable

--✳BMC ……：Enable

--✳BMC ……：Enable

--✳BMC ……：Enable

--✳BMC ……：Enable

保存退出。

HP ProLiant DL160

加电--✳F10

MAIN--✳

记下下面网卡的 MAC ADDRESS：

Dedicatecd NIC MAC Address。 此即 ILO 的的 MAC。

Advanced--✳Remote Access Configuration

--✳Remote Access：Enable

--✳EMS support(SPCR)：Disable

--✳Serial Port Mode：确保是[09600 8,n,1]

--✳Flow Control：Hardware

--✳Redirection After BIOS POST：Always

--✳Terminal Type：VT100

按 F10 保存退出。

之后激活 ILO100。

之后断电，重起。

HP ProLiant DL180 G5

加电--✳F10

Main

记下 NIC 的 MAC（只有一个 NIC）

Advanced--✳Console Redirection--->BIOS Serial Console：Enable 。

Advanced--✳Remote Access Configuration

--✳Remote Access：Enable

--✳EMS support(SPCR)：Disable

--✳Serial Port Mode：确保是[09600 8,n,1]

--✳Flow Control：Hardware

--✳Control Type：VT100

--✳Continue C.R. after POST：Always

按 F10 保存退出。

之后激活 ILO100。

之后断电，重起。

##5.2：OS 的设置

OS 内的系统参数设置

OS 内使用专用工具以替代接显示器设置 DRAC 卡

##5.2.1 OS 内的系统参数设置：

GRUB：

/boot/grub/grub.conf 内的 kernel 启动项里加上 console=ttyS1,115200：

ttyS0 代表 BIOS 里所指的 COM1，ttyS1 代表 COM2，以此类推。

如：

title Red Hat Enterprise Linux Server (2.6.18-164.el5)

root (hd0,0)

kernel /vmlinuz-2.6.18-164.el5 ro root=/dev/sys/root console=ttyS1,115200

initrd /initrd-2.6.18-164.el5.img

INITTAB：

/etc/inittab 里 c0 那一行修改为： 如：

Run gettys in standard runlevels

c0:2345:respawn:/sbin/agetty ttyS1 115200 vt100-nav

1:2345:respawn:/sbin/mingetty tty1

2:2345:respawn:/sbin/mingetty tty2

3:2345:respawn:/sbin/mingetty tty3

4:2345:respawn:/sbin/mingetty tty4

5:2345:respawn:/sbin/mingetty tty5

6:2345:respawn:/sbin/mingetty tty6

注意：重定向端口，波特率：

COM1 和 COM2 是指重定向后的端口，HP 的一般是 COM1，其他 DELL,HUAWEI 一般是 COM2。

115200 是指 DRAC 和 ILO 卡的波特率，不同的机型不一样，以实际调试为准：

DELL 1950,2950 是 9600。

DELL R610，R710，R900，R910，R510 可以是 57600，也可以是 115200。

HP 的全部是 9600。

HUAWEI 的可以是 9600 也可以是 57600 也可以是 115200。

##5.2.2 OS 内使用专用工具以替代接显示器设置 DRAC 卡，及添加 DRAC 卡的账号：

##5.2.3 OS 内刷 HP 机器的 ILO 的 License：

关键的就是连进 ILO 后执行：set /map1 license=247RHZPJ8S6375C4FA7F9361E 这种命令。

如 HP 170e 的 ILO100：

```
#!/bin/bash -

FILE=$1
hosts=`cat $FILE`
for host in $hosts
do
expect 2>&1 <<EOF
set timeout 10
spawn -noecho ssh -t -o ConnectTimeout=3   taobao@$host
expect {
        "yes/no" { send "yes\r";exp_continue }
        "assword:" { send "OOB 的密码\r";}
}
expect "/.->"
```

```
send "cd map1\r\n"
expect "/./map1/->"
send "set license=35QWZV2ZTS6MJDQ6WYVRPQ5J7\r\n"
expect "/./map1/->"
send "exit\r\n"
EOF
done
```

HP ILO2 和 ILO3 的则可直接 ssh 过去执行 set。（ILO100 因为无法退出所以需要 expect）：
（比如用 pgm 命令批量处理）

```
iLO1
pmg -f oobips -A -l taobao "set /map1 license=247RHZPJ8S6375C4FA7F9361E"
iLO2
pmg -f oobips -A -l taobao "set /map1 license=35GNTDG5RB6LHDMHPTKM2WRQH"
iLO3
pmg -f oobips -A -l taobao "set /map1 license=35VH4X6LN3GQQW5KDBKPJ6ZLH"
```

IBM 的：

由于在设置 IBM 的 RSA 卡时曾经引起服务器死机，并且重现的几率很高，
并且这些 IBM 服务器基本都快过保，也没多少售后服务，
所以决定不再对 IBM 做处理，就只使用默认的账号和密码。
IBM 的用户名和密码是 USERID / PASSW0RD ， 保持不变。

5.3 交换机部署：

本文说的基本都是 OOB 专用的。 生产用的 Cisco 不在讨论范围

5.3.1 Cisco/H3C （级联所用，目前 OOB 没有用到 H3C 交换机做上联）：

DLINK（只是 HUB，没有交换功能）：每个机柜一个；
Cisco2960：每排一个，为本排的 DLINK 做上联；
Cisco3750：每个机房一个，为本机房的所有 Cisco2960 做上联；
核心 Cisco3750：只有一个，为所有机房的那个 Cisco3750 做核心汇聚。

5.3.2 ACS：

ACS 即为 Avocent Console Switch，专门为 Cisco 或 H3C 等这种交换机准备的
Console 串口控制交换机。即交换机的 console 交换机。

5.4 交换机设置【需要由 NETOPS 兄弟执行】：

5.4.1 OOB 所用的 Cisco2960 交换机初始化设置：

用串口线连接交换机的 Console 口至本人笔记本，打开超级终端，设置超级终端属性：

端口速度：9600
数据流控制：硬件
数据位：8
奇偶校验：无
停止位：1

即可进入交换机的控制界面。然后复制粘贴如下命令。完整命令如下：
以 CM4 的一台 Cisco2960 交换机 OSW-T5F-06-1.cm4 为例（本排机柜所有 DLINK 的上联）：

```
!
no service pad
```

```
service timestamps debug datetime msec localtime show-timezone
service timestamps log datetime msec localtime show-timezone
service password-encryption
!
hostname
!
enable secret xxxxx
!
aaa new-model
aaa authentication login default group tacacs+ local none
aaa authorization config-commands
aaa authorization exec default group tacacs+ local none
aaa authorization commands 1 default group tacacs+ local none
aaa authorization commands 15 default group tacacs+ local none
!
aaa session-id common
clock timezone gmt 8
vtp mode transparent
ip subnet-zero
!
no ip domain-lookup
!
!
!
no file verify auto
spanning-tree mode pvst
spanning-tree portfast bpduguard default
spanning-tree extend system-id
!
vlan internal allocation policy ascending
!
vlan 32,102,105,115
!
interface GigabitEthernet0/1
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/2
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/3
 switchport access vlan 32
 switchport mode access
```

```
 spanning-tree portfast
!
interface GigabitEthernet0/4
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/5
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/6
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/7
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/8
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/9
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/10
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/11
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/12
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
```

```
!
interface GigabitEthernet0/13
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/14
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/15
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/16
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/17
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/18
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/19
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/20
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/21
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
```

```
interface GigabitEthernet0/22
 switchport access vlan 32
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/23
 switchport access vlan 102
 switchport mode access
 spanning-tree portfast
!
interface GigabitEthernet0/24
 switchport trunk allowed vlan 32,102,105,115
 switchport mode trunk
!
interface Vlan1
 no ip address
 no ip route-cache
 shutdown
!
interface Vlan115
 ip address 10.14.115.1 255.255.255.0
 no ip route-cache
!
ip default-gateway 10.14.115.247
no ip http server
logging facility local5
logging 10.14.102.102
access-list 99 permit 10.0.0.0 0.255.255.255
access-list 99 permit 172.19.0.0 0.0.255.255
access-list 99 permit 172.23.0.0 0.0.255.255
access-list 99 permit 172.24.0.0 0.0.255.255
snmp-server community xxxxx RO 99
snmp-server community xxxxx RW 99
tacacs-server host 10.14.102.102
tacacs-server directed-request
tacacs-server key xxxxx
radius-server source-ports 1645-1646
!
control-plane
!
banner exec C
**************************************************
*            ATTENTION!!! Config files live on:         *
*                 netops1.cmX      DO NOT                        *
*              MAKE MODIFICATIONS HERE WITHOUT             *
*                    UPDATING netops1.cmX. !                 *
```

```
*              Please do a chkin after configs            *
*        /tftp/bin/chkin.sh -m "changes" device     *
*              THANK YOU FOR YOUR COOPERATION.            *
**************************************************
!
line con 0
line vty 0 4
 password xxxxx 此处省略密码
line vty 5 15
!
```

如此，即可完成新上架的 OOB 所用的 Cisco 交换机的初始化。

##5.4.2 TRUNK：由 netops 兄弟帮忙设置交换机的 Trunk 模式。

因为机房太多，机器太多，所以要划分很多个 VLAN，而这些 VLAN 之间可能需要互相访问，所以交换机上需要做 VLAN。

##5.5 Console 控制机设置

##5.5.1 DHCP 服务配置：

以三淘的 console-oob.cm4 这台 console 机的配置为例：

这台 console 为 CM4 的所有服务器提供 DHCP server 服务器。

（特殊情况说明：

比如由于 CM4 大节点有 T5，T9，T11，T18 等几个小机房房间，每个机房都做了个大 VLAN，以避免服务器太多造成交换机出现广播风暴。）

DHCP 服务配置：

```
[root@console-oob.cm4 ~]# cat /etc/dhcpd.conf
#
# DHCP Server Configuration file.
#     see /usr/share/doc/dhcp*/dhcpd.conf.sample
#
ddns-update-style none;
ignore client-updates;
authoritative;

subnet 10.14.0.0 netmask 255.255.224.0 {
        option routers 10.14.31.247;
        option broadcast-address 10.14.31.255;
        option subnet-mask 255.255.224.0;
        option domain-name "oob.tbsite.net";
        option domain-name-servers 10.14.4.243;
        #option time-offset -18000;
        option ntp-servers 10.14.4.243;
        max-lease-time 86400;
        default-lease-time 86400;
        range 10.14.4.10 10.14.4.240;
        range 10.14.5.10 10.14.5.240;
```

```
        ##.....省略.....
        range 10.14.30.10 10.14.30.240;
        range 10.14.31.10 10.14.31.240;
}

subnet 10.14.32.0 netmask 255.255.224.0 {
        option routers 10.14.63.247;
        option broadcast-address 10.14.63.255;
        option subnet-mask 255.255.224.0;
        option domain-name "oob.tbsite.net";
        option domain-name-servers 10.14.32.243;
        #option time-offset -18000;
        option ntp-servers 10.14.32.243;
        max-lease-time 86400;
        default-lease-time 86400;
        range 10.14.32.10 10.14.32.240;
      #range 10.14.33.10 10.14.33.240; # 3-4 C2960 预留
        range 10.14.34.10 10.14.34.240;
        range 10.14.35.10 10.14.35.240;
        ##.....省略.....
        range 10.14.62.10 10.14.62.240;
        range 10.14.63.10 10.14.63.240;
}

subnet 10.14.64.0 netmask 255.255.224.0 {
        option routers 10.14.95.247;
        option broadcast-address 10.14.95.255;
        option subnet-mask 255.255.224.0;
        option domain-name "oob.tbsite.net";
        option domain-name-servers 10.14.64.243;
        #option time-offset -18000;
        option ntp-servers 10.14.64.243;
        max-lease-time 86400;
        default-lease-time 86400;
        range 10.14.64.10 10.14.64.240;
      #range 10.14.65.10 10.14.65.240; # 3-2 C2960 预留
        range 10.14.66.10 10.14.66.240;
        range 10.14.67.10 10.14.67.240;
        ##.....省略.....
        range 10.14.94.10 10.14.94.240;
        range 10.14.95.10 10.14.95.240;
}

subnet 10.14.128.0 netmask 255.255.224.0 {
        option routers 10.14.159.247;
        option broadcast-address 10.14.159.255;
```

```
        option subnet-mask 255.255.224.0;
        option domain-name "oob.tbsite.net";
        option domain-name-servers 10.14.128.243;
        #option time-offset -18000;
        option ntp-servers 10.14.128.243;
        max-lease-time 86400;
        default-lease-time 86400;
        range 10.14.128.10 10.14.128.240;
        range 10.14.129.10 10.14.129.240;
        ##.....省略.....
        range 10.14.158.10 10.14.158.240;
        range 10.14.159.10 10.14.159.240;
}
[root@console-oob.cm4 ~]#
```

保存，重启 DHCPD 服务： service dhcpd restart 即可。

##5.5.2 TRUNK 及 VLAN 设置：

交换机端口做 Trunk：

接 OOB 网络的 NIC2，所接的交换机的端口，要做 Trunk。

如 C3750：

Cisco3750# interface GigabitEthernet0/15 switchport mode trunk

（假如机器的 NIC2 的网线接到了交换机的 Gi0/15 端口了）

##5.5.3 Console 服务器的 OS 设置：

服务器网卡配置：

接 OOB 网络的 NIC2：

```
[root@console-oob.cm4 network-scripts]# cat ifcfg-eth1
# Broadcom Corporation NetXtreme II BCM5708 Gigabit Ethernet
DEVICE=eth1
HWADDR=00:1E:0B:E9:87:76
BOOTPROTO=none
ONBOOT=yes
USERCTL=no
[root@console-oob.cm4 network-scripts]# cat ifcfg-vlan0031
DEVICE=vlan0031
PHYSDEV=eth1
BOOTPROTO=none
IPADDR=10.14.4.243
NETMASK=255.255.224.0
ONBOOT=yes
USERCTL=no
[root@console-oob.cm4 network-scripts]# cat ifcfg-vlan0032
DEVICE=vlan0032
PHYSDEV=eth1
BOOTPROTO=none
```

IPADDR=10.14.32.243
NETMASK=255.255.224.0
ONBOOT=yes
USERCTL=no
[root@console-oob.cm4 network-scripts]# cat ifcfg-vlan0064
DEVICE=vlan0064
PHYSDEV=eth1
BOOTPROTO=none
IPADDR=10.14.64.243
NETMASK=255.255.224.0
ONBOOT=yes
USERCTL=no
[root@console-oob.cm4 network-scripts]# cat ifcfg-vlan0128
DEVICE=vlan0128
PHYSDEV=eth1
BOOTPROTO=none
IPADDR=10.14.128.243
NETMASK=255.255.224.0
ONBOOT=yes
USERCTL=no
[root@console-oob.cm4 network-scripts]#

打开网卡的 VLAN 支持：
[root@console-oob.cm4 network-scripts]# cat ../network
NETWORKING=yes
NETWORKING_IPV6=no
HOSTNAME=console-oob.cm4
GATEWAY=172.24.102.247
VLAN=yes
VLAN_NAME_TYPE=VLAN_PLUS_VID
[root@console-oob.cm4 network-scripts]#

Vconfig：
sudo vconfig add vlan0031
sudo vconfig add vlan0032
sudo vconfig add vlan0064
sudo vconfig add vlan0128

重启网络生效：
sudo service network restart

##5.5.4 ARP 表空间设置：

ARP 表的空间大小设置：（DHCP 的客户端太多，会导致 ARP 缓存空间不够用）：

这个是因为，大规模 fping 时会因为表空间太小而出现的“**No buffer space available**”错误。

修改/etc/sysctl.conf 里的 ARP 缓存设置为：

net.ipv4.neigh.default.gc_thresh3 = 32768
net.ipv4.neigh.default.gc_thresh2 = 16384
net.ipv4.neigh.default.gc_thresh1 = 8192
net.ipv4.route.gc_thresh = 262144
sudo sysctl -p　　即可生效。

##5.5.5 Console 控制机的权限规划安全控制

Console 控制机删除了所有人的个人账号，不允许任何个人账号的登陆，以避免误操作等意外，导致生产机的电源被误关机 or 重启。

Console 控制机只允许 WEB 接口的 API 来调用。

Console 控制机的维护人员登陆，需要 用户名+密码+RSA 动态令牌密码 实现严格控制。

####6： 设置 OOB 以 DHCP 的方式获取 IP，及在 OOB 硬件卡里添加用户名和密码：

##6.1 BMC 的 IP 地址获取方式更改为 DHCP：

由于机器量太多，一般根据 OOB 的 MAC 地址来通过 DHCP 动态分配 IP 地址。

##6.1.1 DELL 及 HUAWEI 及 HP 的 DL180G5 和 DL170e 的 IPMI：

IPMI 协议的 DRAC（DELL 及 HUAWEI）及 HP 180G5 和 DL170e（DL2000 的片）：

设置为 DHCP:

/usr/bin/ipmitool lan set 1 ipsrc static; \
/usr/bin/ipmitool lan set 1 ipaddr 1.1.1.1; \
/usr/bin/ipmitool lan set 1 ipsrc dhcp; sleep 4; \
/usr/bin/ipmitool mc reset cold

20 秒后可查看状态：

/usr/bin/ipmitool lan print 1

注意：最后的那个 1 是 channel 的代码。绝大部分机器都是 1，不行就试试 2。

以上设置可能有时会失败，多执行几次试试。再不行就断电解决。

HUAWEI T8000 刀框的设置：

先 ssh 登陆到刀框控制面板：再：

设置静态 IP(华为 T8000 无法设置为 DHCP)：

如：smmset -l smm -t eth0 -d floatip -v *10.13.22.24　255.255.224.0　10.13.31.255*

查看状态：smmget -t eth0 -d floatip

确认 SD 卡文件内容是否也修改了：

cd /flash/smmdir/usr/local/smm
cat SMM.CFG |grep　*10.13.*

##6.1.2 HP 的 ILO2，ILO3，ILO100：

hponcfg　-f　Mod_Network_Settings.xml 即可。Xml 文件内容如下：

```
----------------------------------------------------------------------
<RIBCL VERSION="2.0">
  <LOGIN USER_LOGIN="adminname" PASSWORD="password">
```

```
   <RIB_INFO MODE="write">
     <MOD_NETWORK_SETTINGS>
       <ENABLE_NIC value="Yes"/>
       <REG_DDNS_SERVER value="Yes"/>
       <PING_GATEWAY value="No"/>
       <DHCP_DOMAIN_NAME value="Yes"/>
       <SPEED_AUTOSELECT value="No"/>
       <NIC_SPEED value="100"/>
       <FULL_DUPLEX value="Yes"/>
       <DHCP_ENABLE value="Yes"/>          激活 DHCP
       <IP_ADDRESS value="0.0.0.0"/>
       <SUBNET_MASK value="0.0.0.0"/>
       <GATEWAY_IP_ADDRESS value="0.0.0.0"/>
       <DNS_NAME value=""/>
       <DOMAIN_NAME value=""/>
       <DHCP_GATEWAY value="Yes"/>
       <DHCP_DNS_SERVER value="Yes"/>
       <DHCP_WINS_SERVER value="Yes"/>
       <DHCP_STATIC_ROUTE value="Yes"/>
       <REG_WINS_SERVER value="Yes"/>
       <PRIM_DNS_SERVER value="0.0.0.0"/>
       <SEC_DNS_SERVER value="0.0.0.0"/>
       <TER_DNS_SERVER value="0.0.0.0"/>
       <PRIM_WINS_SERVER value="0.0.0.0"/>
       <SEC_WINS_SERVER value="0.0.0.0"/>
       <STATIC_ROUTE_1 DEST="0.0.0.0" GATEWAY="0.0.0.0"/>
       <STATIC_ROUTE_2 DEST="0.0.0.0" GATEWAY="0.0.0.0"/>
       <STATIC_ROUTE_3 DEST="0.0.0.0" GATEWAY="0.0.0.0"/>
     </MOD_NETWORK_SETTINGS>
   </RIB_INFO>
   </LOGIN>
</RIBCL>
```

查看结果：

cat Get_Network.xml

```
<RIBCL VERSION="2.0">
<LOGIN USER_LOGIN="adminname" PASSWORD="password">
      <RIB_INFO MODE="read">
           <GET_NETWORK_SETTINGS/>
      </RIB_INFO>
</LOGIN>
</RIBCL>
```

然后 hponcfg -f Get_Network.xml

如果要实现类似 ipmitool mc reset cold 这种硬重置的动作，则：

```
cat > /tmp/RESET_RIB.xml << HPILORESET
    <RIBCL VERSION="2.0">
        <LOGIN USER_LOGIN="adminname" PASSWORD="password">
            <RIB_INFO MODE="write">
                <RESET_RIB/>
            </RIB_INFO>
        </LOGIN>
    </RIBCL>
HPILORESET
```

然后 hponcfg -f /tmp/RESET_RIB.xml 即可硬重置 ILO 卡。

HP DL460C G1 和 G6 也是 ILO2，也可以通过刀框设置刀片的 DHCP：

ssh admin@刀框 IP

show network 查看 OA 的 MAC 和 IP 信息

set ipconfig dhcp 可让 OA 通过 DHCP 得到新 IP

show ebipa 查看每片的 IP

show server info all 可看见每片的 MAC，SN，IP 等信息，非常全面

disable ebipa server all 为每片刀片以 DHCP 的方式分配 IP

如果失败，则：

OA-189239asdf> restart OA

执行 restart 重启 OA。（不会重启 OS）

附上一些 HP 460C 的 OA 用的到的自动化脚本：

（因为 HP 460C 类型的机器很多，操作 OA 时需要交互过程，且等待时间较长，所以自动化）

```
#!/usr/bin/perl -w

use Expect;
my $OAcmd = "show network";                    ---->查看 OA 的 MAC 和 IP 等信息
#my $OAcmd = "set ipconfig dhcp";              ---->设置 OA 以 DHCP 方式获得 IP
#my $OAcmd = "disable ebipa server all\nyes"; ---->设置所有刀片以 DHCP 方式获得 IP
#my $OAcmd = "add user taobao 密码";           ---->给 OA 添加新用户和密码
#my $OAcmd = "set user access taobao administrator";        ---->设置 OA 用户为管理员权限
#my $OAcmd = "assign oa taobao";               ---->确认 OA 用户有效

my $timeout = 10;
my $OAuser = "OA 的用户名";
my $OApass = "OA 的密码";

sub HP460Cadduser {
        my $ip = $ARGV[0];
        my $exp = new Expect;
        $exp->raw_pty(1);

        $exp = Expect->spawn('ssh', ($OAuser . "@" . $ip))
```

```
                or die "Cannot spawn ssh: $!\n";

        $exp->expect($timeout,        ---->因为 OA 交互过程等待时间较长，需要添加超时
                [ qr/password\:\ /i, sub { my $exp = shift;
                        print $OApass . "\n";
                        $exp->send("$OApass\n");
                        exp_continue; }],
                [ qr/continue connecting (yes\/no)?/i, sub { my $exp = shift;
                        $exp->send("yes\n"),
                        exp_continue; }],
                [ qr/\>/i, sub { my $exp = shift;
                        $exp->send("$OAcmd\nexit\n");
                        exp_continue; }],
        );
        $exp->hard_close();
    }
    &HP460Cadduser;
-----------------------------
```

##6.1.3 IBM 的 RSA（rebootrsa 可能会导致死机，50%概率）及 IBM 的刀框的控制面板 AMM 设置：

IBM 的 RSA 设置：

先下载这个脚本 http://yum-src.tbsite.net/yum-src/ops/sbin/local_hw_oob_init.sh 初始化，包括安装必要的工具，如 asu。

重置 RSA 卡：【!!! 这一步极有可能导致系统死机!!! 概率是 50%!】

（x86_64 下是 asu64， i386 的 32 下是 asu。 以下以 x86_64 为例）

/opt/ibm/toolscenter/asu/asu64 rebootrsa

激活 RSA 的 DHCP 模式：

/opt/ibm/toolscenter/asu/asu64 set RSA_DHCP1 Enabled

查看是否获得到了 IP:

/opt/ibm/toolscenter/asu/asu64 show |grep RSA_DHCPAssignedHostIP1

IBM 的刀框的 OOB 控制面板 AMM 设置：

型号：IBM System x 系列

先 ssh 登陆到刀框面板： ssh USERID@IBM 刀框的 IP

切换到 AMM 面板： env -T mm[1]

更改刀框 OOB 网卡为 DHCP 方式获得 IP： ifconfig -eth0 -c dhcp

查看 OOB 网卡状态信息： ifconfig -eth0

重启 OOB 面板： reset

此时即可获得 dhcp 的新 IP。

##6.2　BMC 的用户名密码统一化：新增或修改 用户名和密码：

##6.2.1　IPMI 类的（DELL，HUAWEI，HP 的 DL180G5 和 DL170e）：

```
yum install -y OpenIPMI OpenIPMI-tools; \
/sbin/service ipmi start; \
/usr/bin/ipmitool user set name 6 taobao; \
/usr/bin/ipmitool user set password 6 密码; \
/usr/bin/ipmitool user enable 6; \
/usr/bin/ipmitool user priv 6 4 1; \
/usr/bin/ipmitool user enable 6; \
/usr/bin/ipmitool sol payload enable 1 6; \
/usr/bin/ipmitool channel setaccess 1 6 callin=on ipmi=on link=on privilege=4; \
/usr/bin/ipmitool user list 1|grep taobao
```

即可。

解释：6 是用户的 ID，4 是管理员最高级别，1 是 channel，部分机型上是 2。

##6/.2.2　ILO2 和 ILO3 类的（HP）：

```
yum install -y hponcfg
rm -f hp-ilo-8.5.0-1.rhel5.x86_64.rpm
wget http://yum.corp.taobao.com/tmp/hp-ilo-8.5.0-1.rhel5.x86_64.rpm
rpm -ivh hp-ilo-8.5.0-1.rhel5.x86_64.rpm
/etc/init.d/hp-ilo start
/etc/init.d/hp-ilo restart
```

##sudo /sbin/hponcfg -f 添加或修改的 xml 文件：

##Add_User.xml 添加用户：

```
cat > /tmp/oob_add_user.xml << HPILOADD
<RIBCL VERSION="2.0">
   <LOGIN USER_LOGIN="adminname" PASSWORD="password">
   <USER_INFO MODE="write">
      <ADD_USER
         USER_NAME="taobao"
         USER_LOGIN="taobao"
         PASSWORD="密码">
         <ADMIN_PRIV value ="Y"/>
         <REMOTE_CONS_PRIV value ="Y"/>
         <RESET_SERVER_PRIV value ="Y"/>
         <VIRTUAL_MEDIA_PRIV value ="Y"/>
         <CONFIG_ILO_PRIV value="Yes"/>
      </ADD_USER>
```

如若同时添加多个账号，可在此处继续添加<ADD_USER　　…内容…　</ADD_USER>标签对。

```
   </USER_INFO>
   </LOGIN>
</RIBCL>
HPILOADD
```

/sbin/hponcfg -f /tmp/oob_add_user.xml　　即可导入设置。

##为避免已经存在了同名用户，得修改密码：
##Mod_User.xml 修改密码：

```
cat > /tmp/oob_mod_user.xml << HPILOMOD
<RIBCL VERSION="2.0">
   <LOGIN USER_LOGIN="adminname" PASSWORD="password">
   <USER_INFO MODE="write">
      <MOD_USER USER_LOGIN="taobao">
         <USER_NAME value="taobao"/>
         <PASSWORD value="密码"/>
         <ADMIN_PRIV value="Yes"/>
         <REMOTE_CONS_PRIV value="Yes"/>
         <RESET_SERVER_PRIV value="Yes"/>
         <VIRTUAL_MEDIA_PRIV value="Yes"/>
         <CONFIG_ILO_PRIV value="Yes"/>
      </MOD_USER>
如若同时修改多个账号，可在此处继续添加<MOD_USER 等 …内容… </MOD_USER>标签对。
   </USER_INFO>
   </LOGIN>
</RIBCL>
HPILOMOD
```

然后 /sbin/hponcfg -f /tmp/oob_mod_user.xml 即可修改设置。

##读取配置，验证是否添加或修改成功：

```
/sbin/hponcfg -w /tmp/oob_tmp.xml
cat /tmp/oob_tmp.xml|grep taobao
```

##6.2.3 RSA 的（IBM）：
BMC 的用户名密码统一化：
（请参考本文最开始处的[OS 内使用专用工具以替代接显示器设置 DRAC 卡]这一节）

####7： 监控系统：

##7.1 OOB 的监控机器列表 (截止到 20130201)：

网络环境跟各个机房的 console-oob.cm?完全一致。

机房地点	机房代号	OOB 监控机. 备机	OOB 监控机. 主要
三淘：			
联通滨盛路	cnz	tool1.ops.cnz.alimama.com	tool2.ops.cnz.alimama.com
电信清江路	cm2	dbconsole-oob.cm2	
电信兴议村	cm3	console-oob.cm3	console-oob2.cm3
电信聚园路	cm4	console-oob.cm4	console-oob2.cm4
电信聚园路	cm4.T7.sqa	console-oob.sqa.cm4	
联通青岛	cm5	console-oob.cm5	console-oob2.cm5 console-oob3.cm5(将清退)
电信东冠	cm6	console-oob.cm6	console-oob2.cm6
	cm7		

联通滨江	cm8	console-oob.cm8	console-oob2.cm8
联通青岛崂山	cm9	console-oob.cm9	console-oob2.cm9
联通萧山谷易	cm10		

阿里云:

兴议 i	xyi	oob.xyi.aliyun.com	oob2.xyi.aliyun.com
三墩	sd	oob.sd.aliyun.com	oob2.sd.aliyun.com
余杭	yh	oob.yh.aliyun.com	oob2.yh.aliyun.com
东冠	dg	oob.dg.aliyun.com	oob2.dg.aliyun.com
BTC	btc	oob.btc.aliyun.com	oob2.btc.aliyun.com
兴义 2	xy2	oob.xy2.aliyun.com	oob2.xy2.aliyun.com
兴义 2 的 17# (聚石塔)	xy2	oob.a17.xy2.aliyun.com	

B2B:

兴义 2 期 11#和 12# (国际站) 兴义 1 期 7# (中文站)	xyi	oob.xyi.en.alidc.net	
兴议 2 期 14#接 cm3	cm3	console-oob.cm3	console-oob2.cm3

特殊：兴义

兴义所有 room	直达 cm3,xyi,xy2 等 20 几个 ROOM	oob.xyallroom.cn.alidc.net 172.30.127.221	

这些监控机的 NIC2 都接到了 OOB 用的 Cisco2960 交换机，并且交换机的这个端口做了 TRUNK。
交换机设置请参考：[交换机设置]一节，与 console 服务器所需要的完全一致。
OS 及网卡设置请参考：[Console 控制机设置]一节，与 console 服务器完全一致。

##7.2 监控方式及规则：

Ping 的方式：

这几个监控机上，都安装了 nagios，通过插件 check_ping 做简单的 ping 检测。

每天 ping 一次，每次 5 个包，全部 OK 则视为正常，否则邮件报警。

（这个都可以手动调整，在监控中心的脚本里。）

用户名密码检测：

通过 ipmitool 或 ssh 等工具检测能否通过用户名密码正常的连进这些机器的 OOB。

##7.3 OOB 的 MAC 和 IP 地址信息的准确性和及时性的保证

所有 OOB 的 DHCPD 服务器上，root 账号的计划任务里，每 5 分钟执行一次更新动作，

把 dhcpd.log 里的日志过滤出来，通过 API 更新到集团基础服务器信息数据库里。

其中 MAC 和 IP 地址的重复，自动处理+人工处理，保证唯一性。

####8 【最最常用的】OOB 的统一 WEB 入口或各 Console 机入口及具体操作，及故障处理：

其实以上所有的动作，都是为了 console 脚本做准备的。

##8.0 OOB 的统一 WEB 入口

http://oob.alibaba-inc.com/

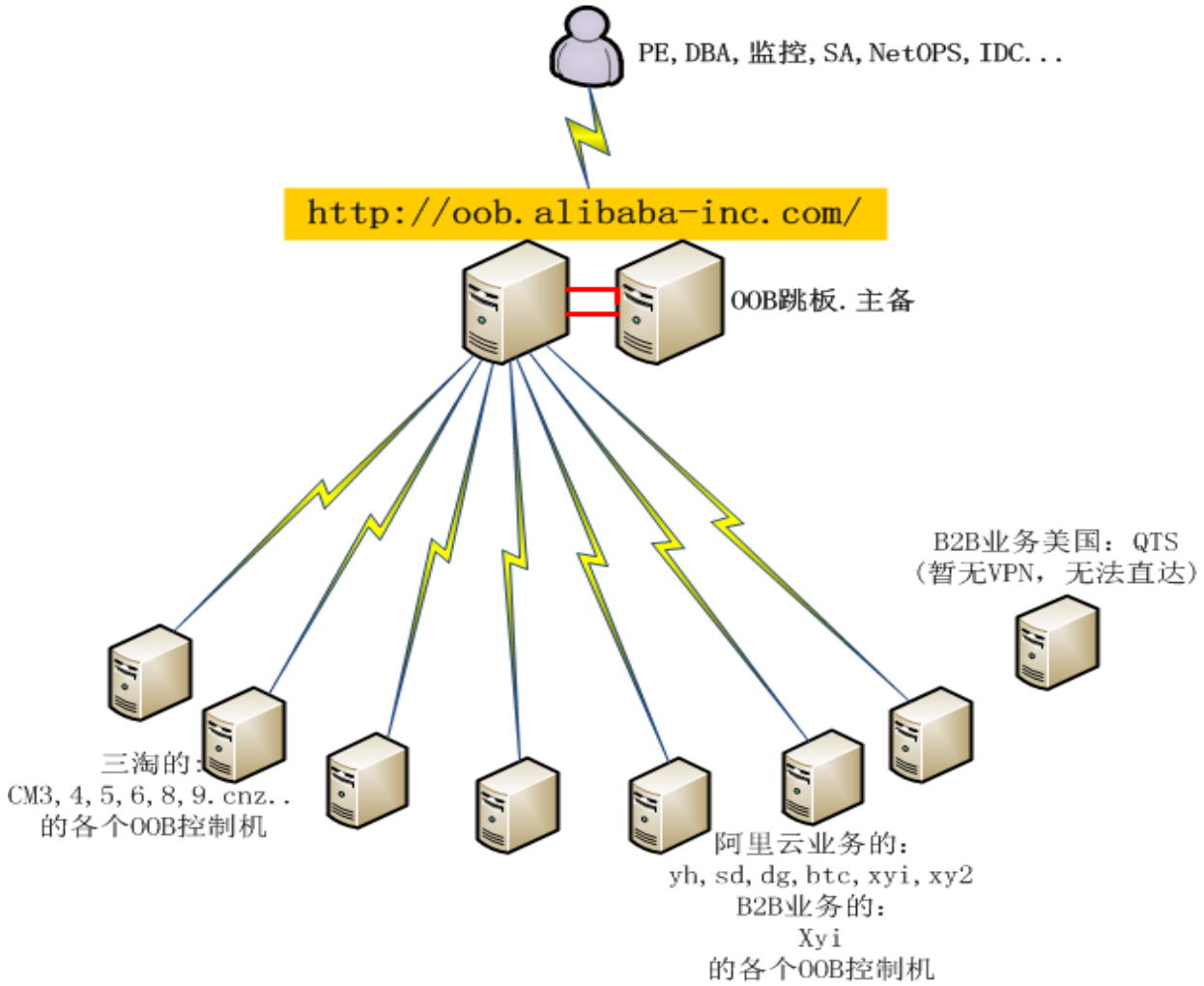
以域账号密码登陆这里，会自动根据用户进行权限识别和控制，权限限定在最小范围内。

所有 PE，DBA，监控，SA 及 IDCOPS 等都从这个入口进入，从这里管理控制服务器的 OOB。

这个入口简化了很多命令行等复杂操作，方便快捷。

本文所有的硬件，系统，网络等，都是为这个统一 WEB 入口提供支撑，供其调用。

逻辑图：



##8.1 Console 机入口和相关工具脚本：console 脚本，ipmitool 和 ssh：
Console 脚本，ipmitool 和 ssh 命令的真实入口：

Console 机的入口： 即/usr/local/bin/console 脚本所在的服务器： （全是 LINUX 平台的）

机房地点	机房代号	服务器 console 机	服务器 console 机的备机
三淘:			
联通滨盛路	**cnz**	tool1.ops.cnz.alimama.com	tool2.ops.cnz.alimama.com
电信清江路	**cm2**	dbconsole-oob.cm2	
电信兴议村	**cm3**	console-oob.cm3	console-oob2.cm3
电信聚园路	**cm4**	console-oob.cm4	console-oob2.cm4
电信聚园路	**cm4.T7.sqa**	console-oob.sqa.cm4	
联通青岛	**cm5**	console-oob.cm5	console-oob2.cm5 console-oob3.cm5(将清退)
电信东冠	**cm6**	console-oob.cm6	console-oob2.cm6
	cm7		
联通滨江	**cm8**	console-oob.cm8	console-oob2.cm8
联通青岛崂山	**cm9**	console-oob.cm9	console-oob2.cm9
联通萧山谷易	**cm10**		
阿里云:			
兴议 i	xyi	oob.xyi.aliyun.com	oob2.xyi.aliyun.com
三墩	sd	oob.sd.aliyun.com	oob2.sd.aliyun.com
余杭	yh	oob.yh.aliyun.com	oob2.yh.aliyun.com
东冠	dg	oob.dg.aliyun.com	oob2.dg.aliyun.com
BTC	btc	oob.btc.aliyun.com	oob2.btc.aliyun.com
兴义 2	xy2	oob.xy2.aliyun.com	oob2.xy2.aliyun.com
兴义 2 的 17# (聚石塔)	xy2	oob.a17.xy2.aliyun.com	
B2B:			
兴义 2 期 11#和 12# (国际站) 兴义 1 期 7# (中文站)	xyi	oob.xyi.en.alidc.net	
兴议 2 期 14#接 cm3	**cm3**	console-oob.cm3	console-oob2.cm3
特殊：兴义			
兴义所有 room	直达 cm3,xyi,xy2 等 20 几个 ROOM	oob.xyallroom.cn.alidc.net 172.30.127.221	

(WINDOWS 的 OOB 控制机，目前只有三淘有，未来统一补足：)

机房	服务器 console 机.(OS 为 windows)
cnz	oobweb.ops.cnz.alimama.com
cm2	110.75.125.18 (windows,要拨 VPN)
cm3	win-oob.cm3
cm4	win-oob.cm4
cm4.T7.SQA	win-oob.sqa.cm4
cm5	win-oob.cm5 (win-oob2.cm5 备用)
cm6	win-oob.cm6
cm7	
cm8	win-oob.cm8
cm9	
cm10	

其他机房暂未搭建 windows。

Console 机的登陆方法：

在公司：

可以用 Putty 或 SecureCRT 工具以 ssh 方式，依次登陆：

先登陆 login1.cm4.tbsite.net ＊ 再 ssh 到以上那些 console 机。

公司以外：

先拨生产 VPN， vpn.alibaba-inc.com ，再用 Putty 或 SecureCRT 工具以 ssh 方式，依次登陆：

先登陆 login1.cm4.tbsite.net ＊ 再 ssh 到以上那些 console 机。

特例：Windows 机：

OS 为 windows 的 console 机，无论在哪里，都要先拨生产 VPN，

再用远程终端客户端 mstsc 登陆过去。

（一般只有 SiteOPS 兄弟用的着 ）

OOB 工具：

console 脚本：

路径：/usr/local/bin/console

作用：登录远程机器的串口，对远程机器进行操作。

用法：

[geer@console-oob.cm3 ~]# console

USAGE:

/usr/local/bin/console [[--force] options] hostname_OR_oob_ip

[options]: Be Careful !!!

--force : only used before poweron/off/reset/cycle ! For purpose of not confirm manually when poweron/off/reset/cycle !

poweron : to boot on the power of the host !
poweroff : to shut down the power of the host !
powerreset : to reset the power of the host !
powercycle : to reset the power of the host !
But only Dell's very few models need it ! You can ignore it !
powerstatus : to show the power's status of the host .
monitor : to check the oob's ping and username and password for monitor .
adduserpass : to add OOB user taobao/alibaba and password **** .

电源控制： powerreset/poweron/poweroff
作用：对远程机器进行 加电，关电，硬重启电源 的动作。

用法：
[geer@console-oob.cm3 ~]$console powerreset/poweron/poweroff 机器名/OOB_ip
可选选项是—force，以便批量执行时无需输入 yes。

##8.2 console 失败时的应急方案：
（当 console 脚本失效时的解决方案，特殊问题的特殊解决方法）

先把 OOB 分类：
只从硬件及其封装的协议来分：只有两种：IPMI 和 iLO。
（其实 iLO 是封装了 IPMI 的，更稳定更易用而已）
DELL： IPMI：IPMI 1.5，IPMI 2.0
HUAWEI： IPMI：IPMI 1.5，IPMI 2.0
IBM： IPMI：IPMI 1.5，IPMI 2.0
但是 IBM 的比较特殊。
HP： iLO：iLO 100，iLO 2.0，iLO 3.0
但是老机型：HP 180/185 G5，也是 IPMI。
各种机型或协议下的操作方式：
IPMI： --* ipmitool

一般遵循 IPMI 协议的机器，如 DELL 何 HUAWEI 和 IBM，
都可以用 ipmitool 工具来解决：
等同于 console 命令：
ipmitool –I lanplus –H oob 的 ip –U 用户名 –P 密码 –e ^ sol activate
可替换的变量：
oob 的 ip： 即 oob_ip
用户名： 即 root,admin 之类的
密码： 即…，可从/usr/local/bin/console 里获知。

实例：

由于有些 DELL 的 DRAC 卡不稳定，经常会遇到这种情况：

```
[magq@tool1.ops.cnz.alimama.com:~]$console ptest3.mm.cnz.alimama.com
node:    ptest3.mm.cnz.alimama.com
model:   Dell PowerEdge R710
ip:      10.10.14.199
Connect to ptest3.mm.cnz.alimama.com use IPMI2...use ^., exit to Escape.

Error: Unable to establish IPMI v2 / RMCP+ session
Error: No response activating SOL payload
^.^.[magq@tool1.ops.cnz.alimama.com:~]$
```

这个时候就很有必要直接用 ipmitool 命令行的方式来处理了：

先 ping 一下 OOB IP，通则继续下一步：

```
[magq@tool1.ops.cnz.alimama.com:~]$ping 10.10.14.199
PING 10.10.14.199 (10.10.14.199) 56(84) bytes of data.
64 bytes from 10.10.14.199: icmp_seq=1 ttl=64 time=0.577 ms
64 bytes from 10.10.14.199: icmp_seq=2 ttl=64 time=0.476 ms

--- 10.10.14.199 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1000ms
rtt min/avg/max/mdev = 0.476/0.526/0.577/0.055 ms
[magq@tool1.ops.cnz.alimama.com:~]$sudo ipmitool -I lanplus -H 10.10.14.199 -U [redacted] -P [redacted] power status
Password:
Chassis Power is off
[magq@tool1.ops.cnz.alimama.com:~]$
```

看来是可以直接用 ipmitool 正常连过去的， 加电，开机：

```
[magq@tool1.ops.cnz.alimama.com:~]$sudo ipmitool -I lanplus -H 10.10.14.199 -U [redacted] -P [redacted] power on
Chassis Power Control: Up/On
```

用 ipmitool 的方式进入 DRAC 的串口：

```
[magq@tool1.ops.cnz.alimama.com:~]$sudo ipmitool -I lanplus -H 10.10.14.199 -U [redacted] -P [redacted] -e ^ sol activate
[SOL Session operational.  Use ^? for help]

                                                    F2 = System Setup
Phoenix ROM BIOS PLUS Version 1.10 1.2.6          F10 = System Services
Copyright 1985-1988 Phoenix Technologies Ltd.     F11 = BIOS Boot Manager
Copyright 1990-2008 Dell Inc.                       F12 = PXE Boot
All Rights Reserved
```

现在能看见机器的自检状态了，这个跟 console 命令的效果一样：

```
tool1.ops.cnz.alimama.com | tool1.ops.cnz.alimama.com (1) | tool1.ops.cnz.alimama.com (2) | tool1.op

                                                    F2 = System Setup
Phoenix ROM BIOS PLUS Version 1.10 1.2.6          F10 = System Services
Copyright 1985-1988 Phoenix Technologies Ltd.     F11 = BIOS Boot Manager
Copyright 1990-2008 Dell Inc.                       F12 = PXE Boot
All Rights Reserved

Dell Syste9;1HTesting

        memory.  Please wait.
```

（!!! 慎重啊慎重 !!!）关闭电源则是：

```
[magq@tool1.ops.cnz.alimama.com:~]$sudo ipmitool -I lanplus -H 10.10.14.199 -U [redacted] -P [redacted] power off
Chassis Power Control: Down/Off
[magq@tool1.ops.cnz.alimama.com:~]$
```

ILO：--✳ ssh
如 HP 的：DL360G5,G7, DL380G5,G7,DL580G5 等。
（DL180/185 G5 是走 ipmi 的； iLO100 是另外一种，这种机器不多了）
一般可以走 ssh 协议：
ssh –vvv 用户名@oob 的 IP
输入 oob 的密码

实例：
（这台机器实际通了，只是举例，假如这个机器最后没能成功的 console）

```
[magq@tool1.ops.cnz.alimama.com:~]$console kwr3a19-old.kgb.cnz.alimama.com
node:   kwr3a19-old.kgb.cnz.alimama.com
model:  HP ProLiant DL360 G5
ip:     10.10.8.123
Connect to kwr3a19-old.kgb.cnz.alimama.com use ILO2...use Esc+(, exit to Escape
The authenticity of host '10.10.8.123 (10.10.8.123)' can't be established.
RSA key fingerprint is 80:7f:36:eb:bf:9f:88:46:80:38:59:fd:6a:b5:3a:1e.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added '10.10.8.123' (RSA) to the list of known hosts.
alimama@10.10.8.123's password:
Permission denied, please try again.
alimama@10.10.8.123's password:
```

那么试试能 ping 通 OOB IP 不：

```
[magq@tool1.ops.cnz.alimama.com:~]$ping 10.10.8.123
PING 10.10.8.123 (10.10.8.123) 56(84) bytes of data.
64 bytes from 10.10.8.123: icmp_seq=1 ttl=64 time=0.461 ms
64 bytes from 10.10.8.123: icmp_seq=2 ttl=64 time=0.446 ms
```

能通，下一步，通过 ssh 过去：

```
[magq@tool1.ops.cnz.alimama.com:~]$ssh [redacted]@10.10.8.123
alimama@10.10.8.123's password:
User:alimama logged-in to ILOCNG7275OMG.oob.cnz.alimama.com(10.10.8.123)
iLO  1.29 at 17:09:41 Feb 28 2007
Server Name: CNG7275OMG00
Server Power: On

</>hpiLO-> power status

POWER: unknown command status

</>hpiLO-> vsp

Starting virtual serial port
Press 'ESC (' to return to the CLI Session
```

能进入串口。注意 vsp 这 3 个字符相当于 ipmitool 里的 –e ^ sol activate，即进入观察状态。
查看电源状态：power
重启电源： （!!! 慎重啊慎重!!!） power reset 。
注意： 在从 vsp 的状态切换到 power reset 之间，
我其实同时按住了 Shife + ESC + (这 3 个字符，不然是没法切换的。

```
</>hpiLO-> power status

POWER: unknown command status

</>hpiLO-> vsp

Starting virtual serial port
Press 'ESC (' to return to the CLI Session

</>hpiLO-> Virtual Serial Port active: IO=0x03F8 INT=4
</>hpiLO->
</>hpiLO-> power

power: server power is currently: On

</>hpiLO-> power reset


</>hpiLO->
```

此时正在重启电源。。。。

输入 vsp 又能进入观察机器的状态：

```
</>hpiLO-> vsp

Starting virtual serial port
Press 'ESC (' to return to the CLI Session

</>hpiLO-> Virtual Serial Port active: IO=0x03F8 INT=4
```

正在自检 ing。。。。 ：

tool1.ops.cnz.alimama.com | tool1.ops.cnz.alimama.com (1) | tool1.ops.cnz.alimama.com (2) | t

```
                          8 GB Installed

ProLiant System BIOS - P58 (05/01/2007)
Copyright 1982, 2007 Hewlett-Packard Development Company, L.P.


Proc 1: Dual-Core Intel(R) Xeon(TM) Processor (2.00 GHz/1333 MHz, 4MB L2)
Proc 2: Dual-Core Intel(R) Xeon(TM) Processor (2.00 GHz/1333 MHz, 4MB L2)
Power Regulator Mode: OS Control

Advanced Memory Protection Mode: Advanced ECC Support
```

基本是这些，足够处理绝大部分的情况了。

如果以上这些应急方案依然无法解决，那只能：

找机房现场拔掉电源，以重置 OOB 硬件卡，再接上电源，然后开机。

这是终极解决方案，如果这个都解决不了，那就是 OOB 的硬件出问题了。

####9： 所有机房的 OOB 互通及全局网络拓扑图：

##9.1 OS 内的路由设置：各机房之间如何通过 OOB 网络互连（而非通过生产网）

以三淘的 CM4 的 Console 控制机为例：

[root@console-oob.cm4 ~]# **ifconfig 本机内网卡及 VLAN 配置（部分输出内容省略）：**

eth1 Link encap:Ethernet HWaddr 00:1E:0B:E9:87:76
UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1

vlan0031 Link encap:Ethernet HWaddr 00:1E:0B:E9:87:76
inet addr:10.14.4.243 Bcast:10.14.31.255 Mask:255.255.224.0

vlan0032 Link encap:Ethernet HWaddr 00:1E:0B:E9:87:76
inet addr:10.14.32.243 Bcast:10.14.63.255 Mask:255.255.224.0

vlan0064 Link encap:Ethernet HWaddr 00:1E:0B:E9:87:76
inet addr:10.14.64.243 Bcast:10.14.95.255 Mask:255.255.224.0

vlan0102 Link encap:Ethernet HWaddr 00:1E:0B:E9:87:76
inet addr:10.14.102.243 Bcast:10.14.102.255 Mask:255.255.255.0

vlan0128 Link encap:Ethernet HWaddr 00:1E:0B:E9:87:76
inet addr:10.14.128.243 Bcast:10.14.159.255 Mask:255.255.224.0

[root@console-oob.cm4 ~]# **route -v 查看本机内路由**

……省略……

10.14.128.0	*	**255.255.224.0**	**U**	**0**	**0**	**0 vlan0128**
10.14.64.0	*	**255.255.224.0**	**U**	**0**	**0**	**0 vlan0064**
10.14.32.0	*	**255.255.224.0**	**U**	**0**	**0**	**0 vlan0032**
10.14.0.0	*	**255.255.224.0**	**U**	**0**	**0**	**0 vlan0031**

……省略……

而 CM3 机房的 Console 控制机：

[root@console-oob.cm3 ~]# **ifconfig 本机内网卡及 VLAN 配置（部分输出内容省略）：**

eth1 Link encap:Ethernet HWaddr 00:1E:0B:1C:68:84
UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1

vlan0005 Link encap:Ethernet HWaddr 00:1E:0B:1C:68:84
inet addr:10.13.0.243 Bcast:10.13.31.255 Mask:255.255.224.0

vlan0006 Link encap:Ethernet HWaddr 00:1E:0B:1C:68:84
inet addr:10.13.32.243 Bcast:10.13.63.255 Mask:255.255.224.0

[root@console-oob.cm3 ~]# route -v　查看本机内路由

……省略……

10.13.32.0	*	255.255.224.0	U	0	0	0 vlan0006
10.13.0.0	*	255.255.224.0	U	0	0	0 vlan0005

……省略……

那么，如何让 CM4 能够和 CM3 机房的 10.13.0.0/16 段的 OOB_IP 通信呢？
(同理可以应用到其他两个机房的 Console 控制机之间)

用 route add 添加路由。

一条用来添加两个机房间的路由：

route add net 10.13.0.0 netmask 255.255.224.0 dev vlan0005

route add net 10.13.32.0 netmask 255.255.224.0 dev vlan0006

一条用来添加到所有其他机房间的路由：

route add net 10.0.0.0 netmask 255.192.224.0 dev eth1

即可实现 一对一 和 一对多 的路由通信。

如果要实现一台 console 总控制机即可控制阿里所有机房的服务器，则在总控制机上添加路由：

类似这种，每个机房一条：

如从兴义的某台总控机，到达三淘 CM3 的： ip route add 10.13.0.0/16 via 10.216.62.254

前提是物理链路都铺设完毕，及 OOB 核心路由上都设置正确。

##9.2 交换机内的路由设置：

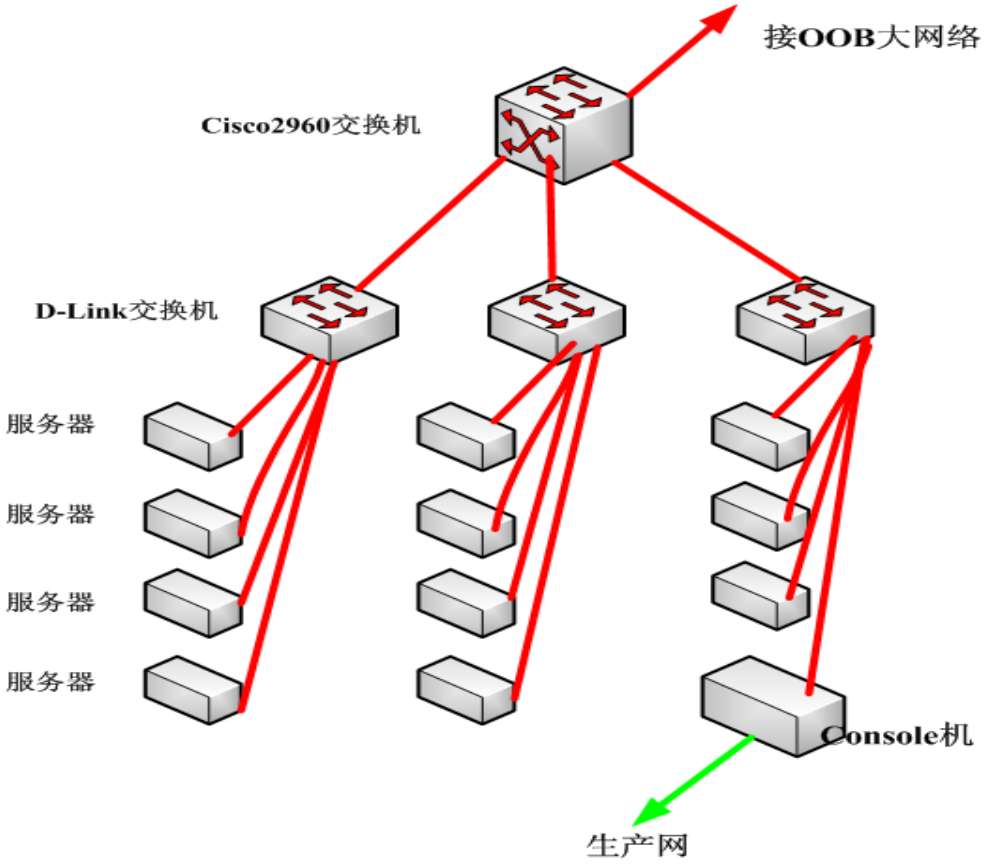
有 NETOPS 兄弟在 OOB 所用的核心交换机和路由器上设置。

##9.3 目前所有机房的 OOB 核心交换机的连接情况记录：

网段及 VLAN 划分情况：见本文嵌入的表格，里面有最完整的信息。

##9.4 各种级别的网络拓扑图

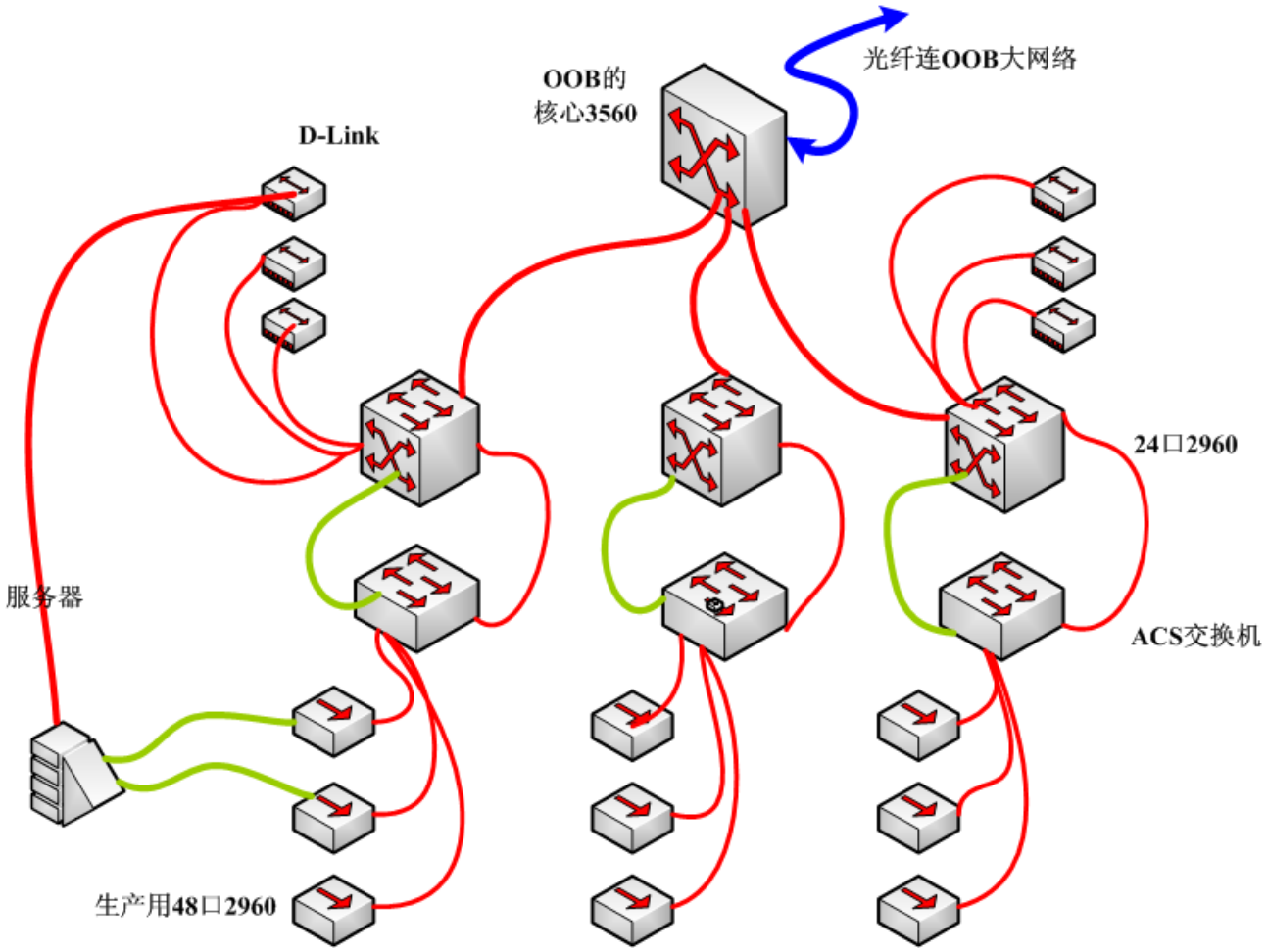
单个服务器的 OOB 网卡与上联二层交换机的网络拓扑图：
（红色线是 OOB 网卡接到 DLINK 的网线，绿色线是生产用的上联网线）



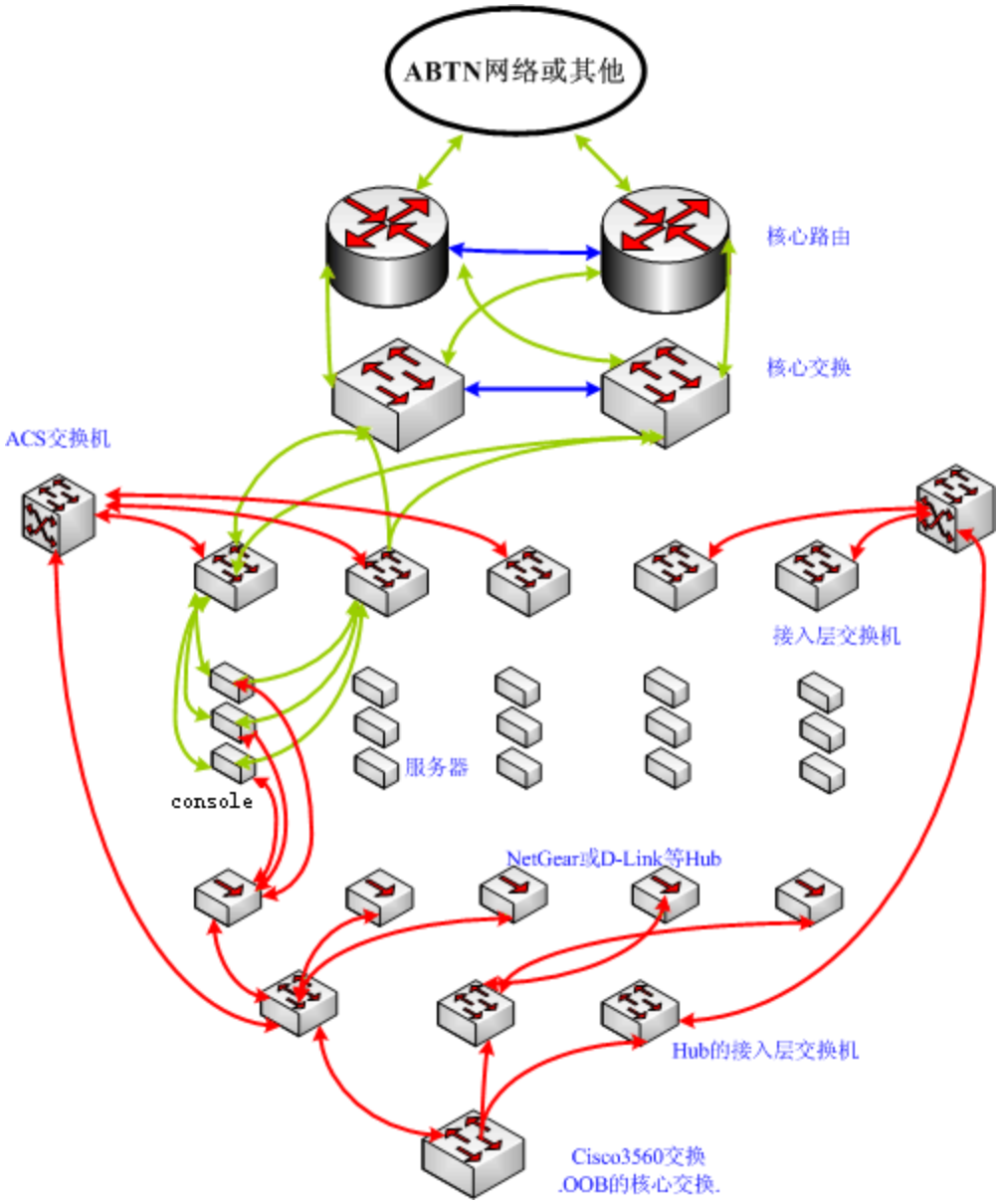
一个机房内的 OOB 拓扑：

服务器 OOB 网卡---->OOB 的 2 层交换机-–>OOB 的三层交换机--->出口交换机：

（红色线是 OOB 网卡接到 DLINK 的网线，绿色线是生产用的上联网线）

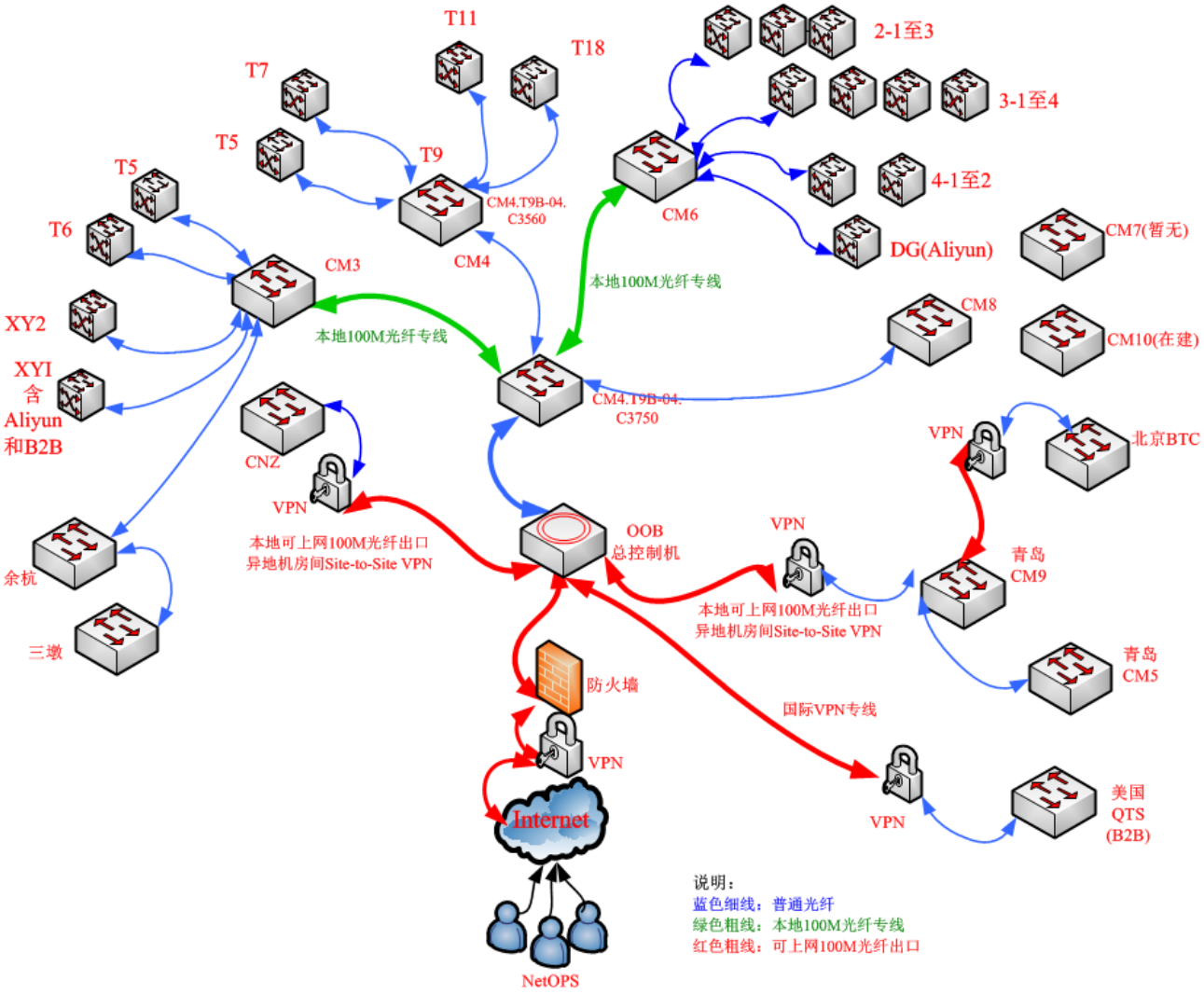


目前 OOB 的最终出口 走生产网：(把上面那幅图倒过来看即是)
（红色线是 OOB 网卡接到 DLINK 的网线，绿色线是生产用的上联网线）



所有机房连通后的全局拓扑图（OOB 的出口完全独立于生产网）：

（下图是截止到 20130201 时：三淘，阿里云，B2B 所有的机房）



上图说明：

本地 100M 光纤专线：

比如同为电信的机房，则可以走电信内部已经部署好的光纤专线。

如未部署好，可以协调电信重新布线。

可上网 100M 光纤出口（即 VPN）：

由于我们的机房即有电信的，也有联通的，所以跨了不同运营商。

跨运营商时，不同运营商之间是无法直接联网的，所以需要协调两方运营商，单独拉光纤专线，才能通信。

VPN（验证方式为：用户名+密码+RSA 令牌动态密码）：

因为这张 OOB 大网络是与生产网隔绝的，那么一旦生产网的总出口全部断掉（虽然这种天灾我们真的很不希望发生，但是万一发生了，还是得面对的），我们就可以从这个 VPN 入口进去紧急处理。如果这个故障是因为总出口交换机或路由器设备引起的话。

因为这个 VPN 是整张 OOB 网络的总入口，所以安全和权限控制非常重要，所以采用了**用户名+密码+RSA 令牌动态密码** 的双因素认证。

这个 VPN，NETOPS 兄弟最需要用到。

==========完==========