Overlay网络可视化

April 2019

索引

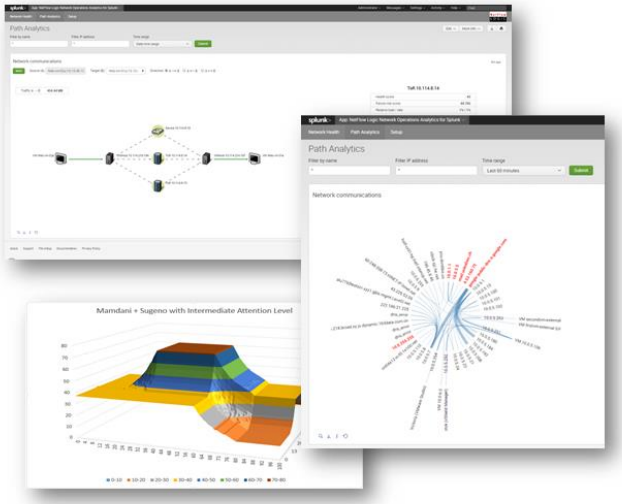
Overlay网络说明

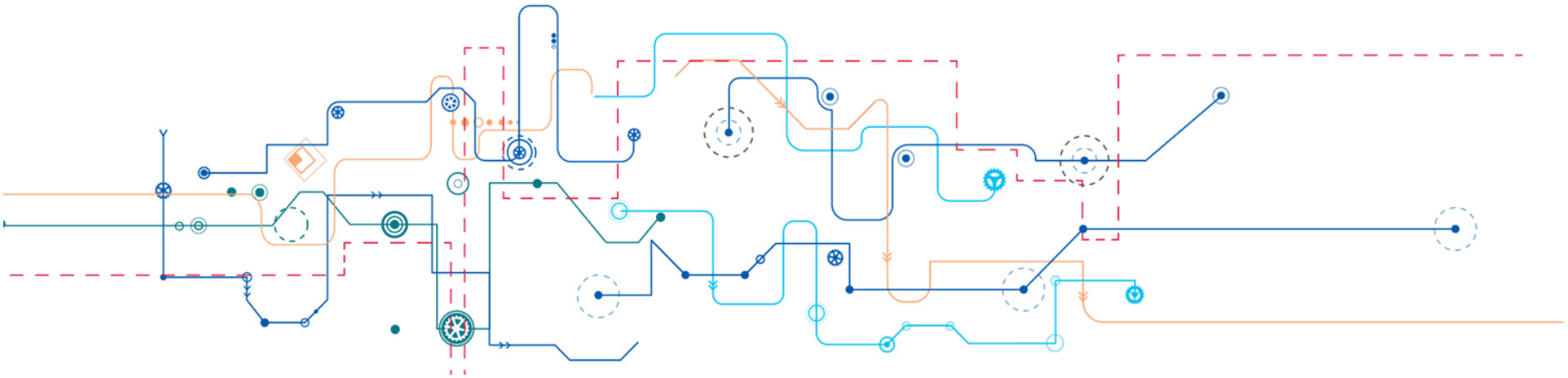
为什么以及如何进行监控

监控Overlay网络的问题

Cubro的解决方案

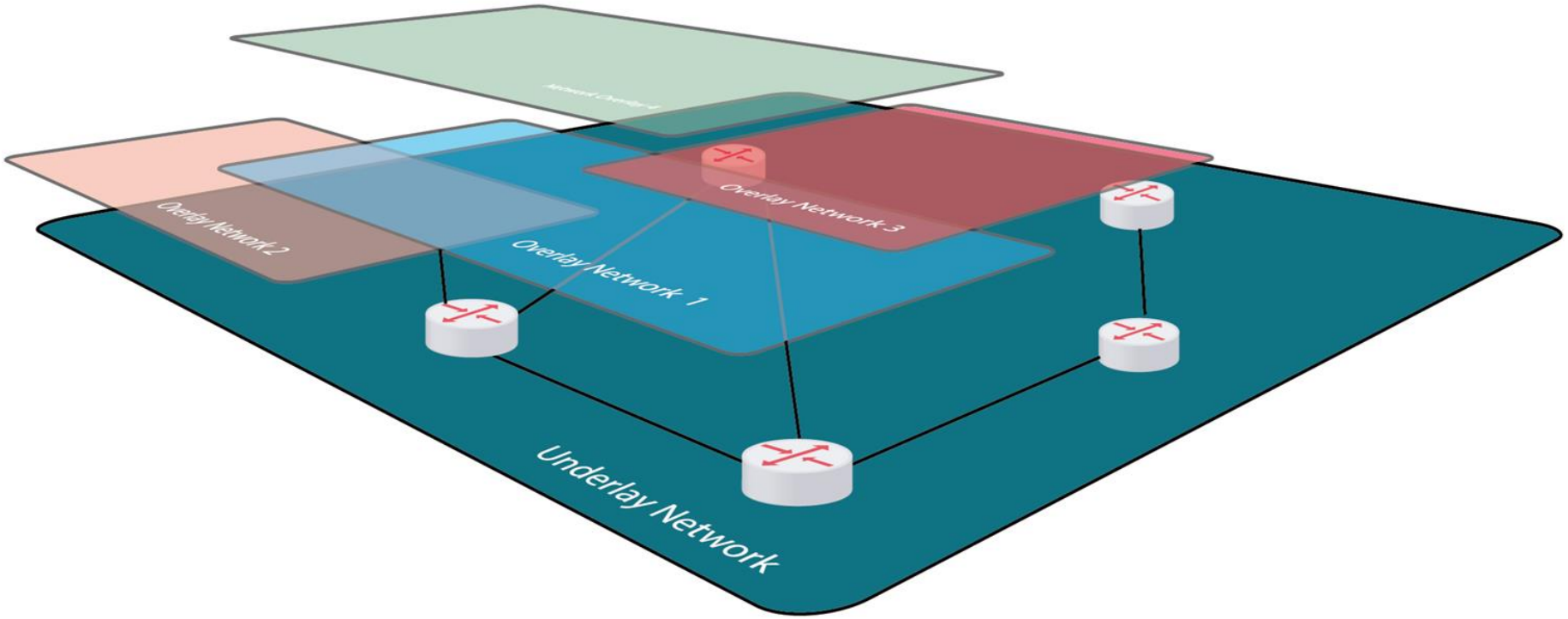
其他通常提供的解决方案



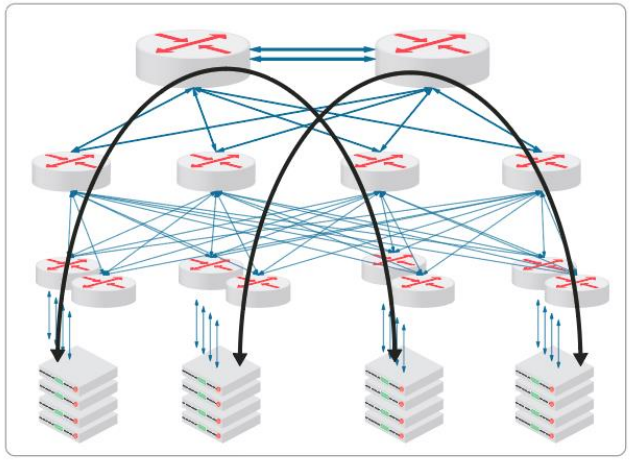


Overlay网络

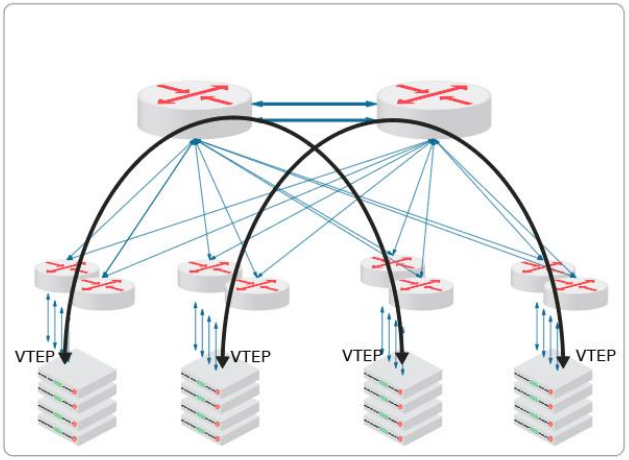
当今的Overlay网络在任何数据中心都是标准的，今天与过去的唯一区别是我们正在谈论每个数据中心有成百上千的overlay网络，最多可有数千个端点。此外，Overlay网络比过去更具动态性，因此不再可能收到配置可见性工具。



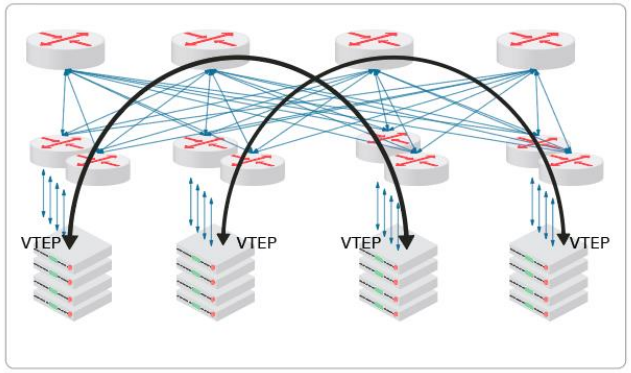
云数据中心网络的演变



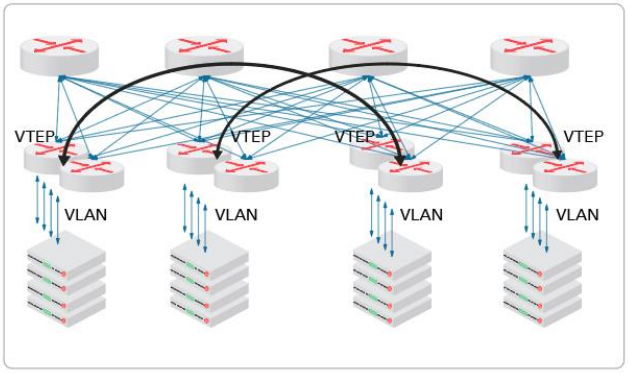
3-Tier Overlay Network



2-Tier Scale-Up Overlay Network



2-Tier Scale-Out Overlay Network



EVPN - Ethernet Virtual Private Network

□ 3层Overlay网络：与园区网络相同的3层架构（核心层、汇聚成和接入层），为计算服务器之间的（VXLAN[3]）重叠流量提供互联。

□ 2层纵向扩展Overlay网络：将网络体系结构从3层向下缩减为2层，即高spin层和leaf层，以减少用于计算服务器之间重叠流量的网络跃点数（并因此降低了端到端延迟）。

2层横向扩展Overlay网络：通过将单个的（通常更具扩展性）高spin交换机系统转变为多个（更具价值效益的）瘦spin交换机系统，从2层纵向扩展Overlay网络演变而来，以便提供：

- 更好的冗余：从高spin架构中的1+1冗余到瘦spin架构中的N+1冗余
- 更好的可扩展性：从只有几台瘦spin交换机的小型云数据中心，到拥有众多瘦spin交换机的巨型数据中心
- 易于物流：具有pizza-box外形的相同硬件平台允许瘦spin交换机和leaf交换机之间的互换；因此，降低了云数据中心资产管理和运营物流的复杂性。

以太网虚拟专用网（EVPN）[4]：由现有网络主流厂商提出的最新云网络架构将（VXLAN）虚拟隧道端点（VTEP）功能从计算服务器卸载到了leaf交换机，并降低资源消耗，从而实现了计算服务器中更高的链路吞吐量。

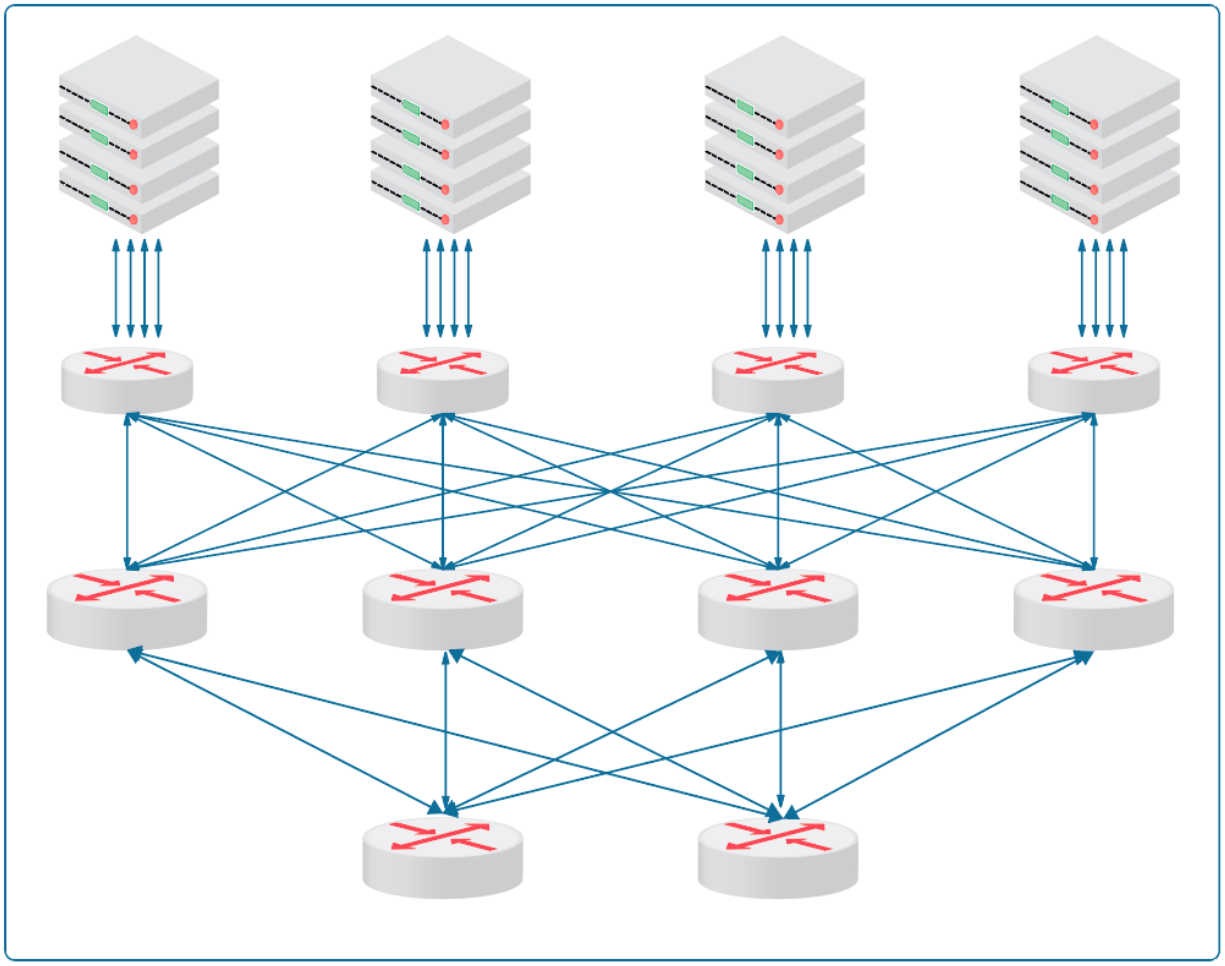
底层数据中心网络设计

Server farms

Leaf

Spine

Border Leaf



底层网络设计& L2 Overlay

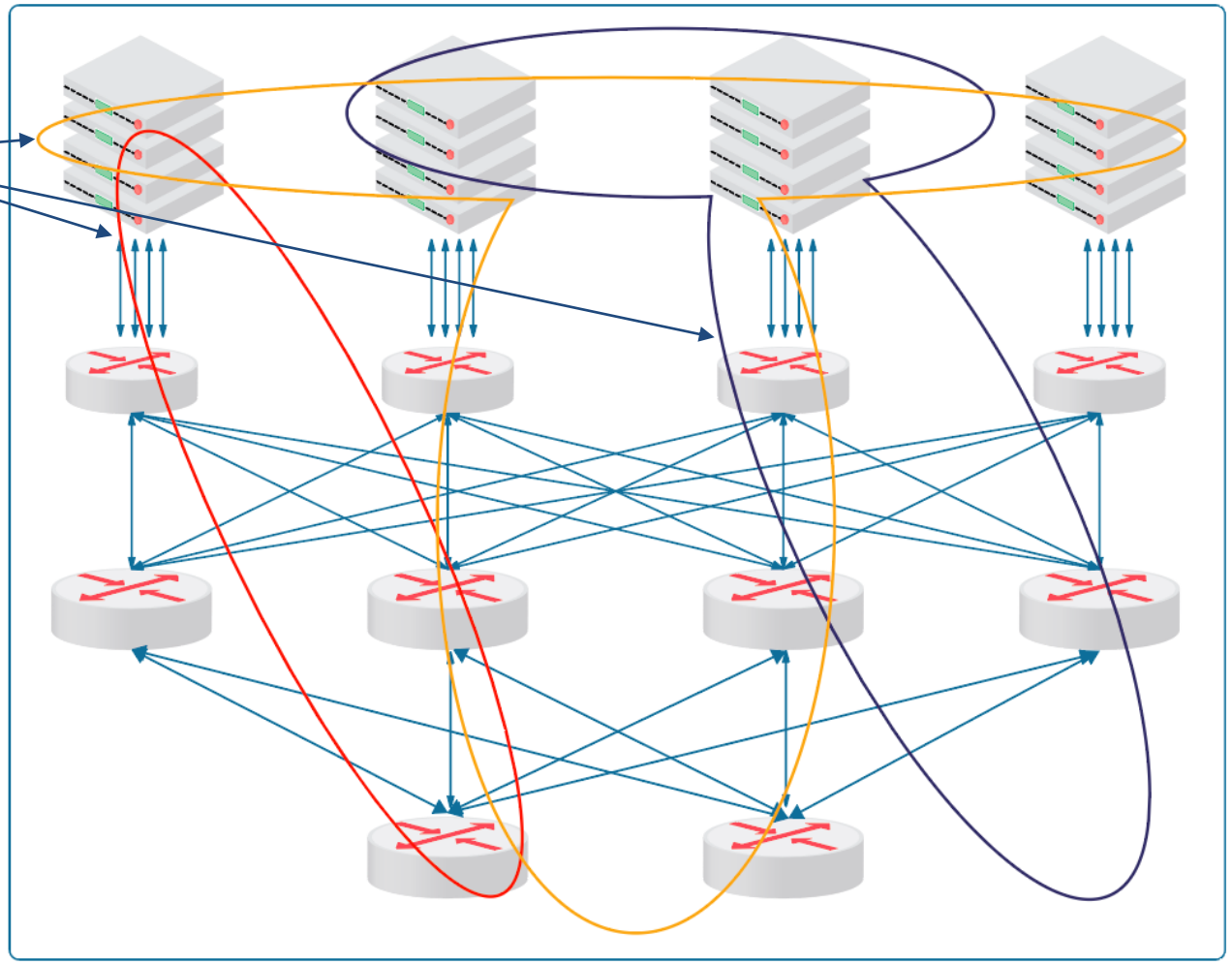
透明的L2 Overlay网络

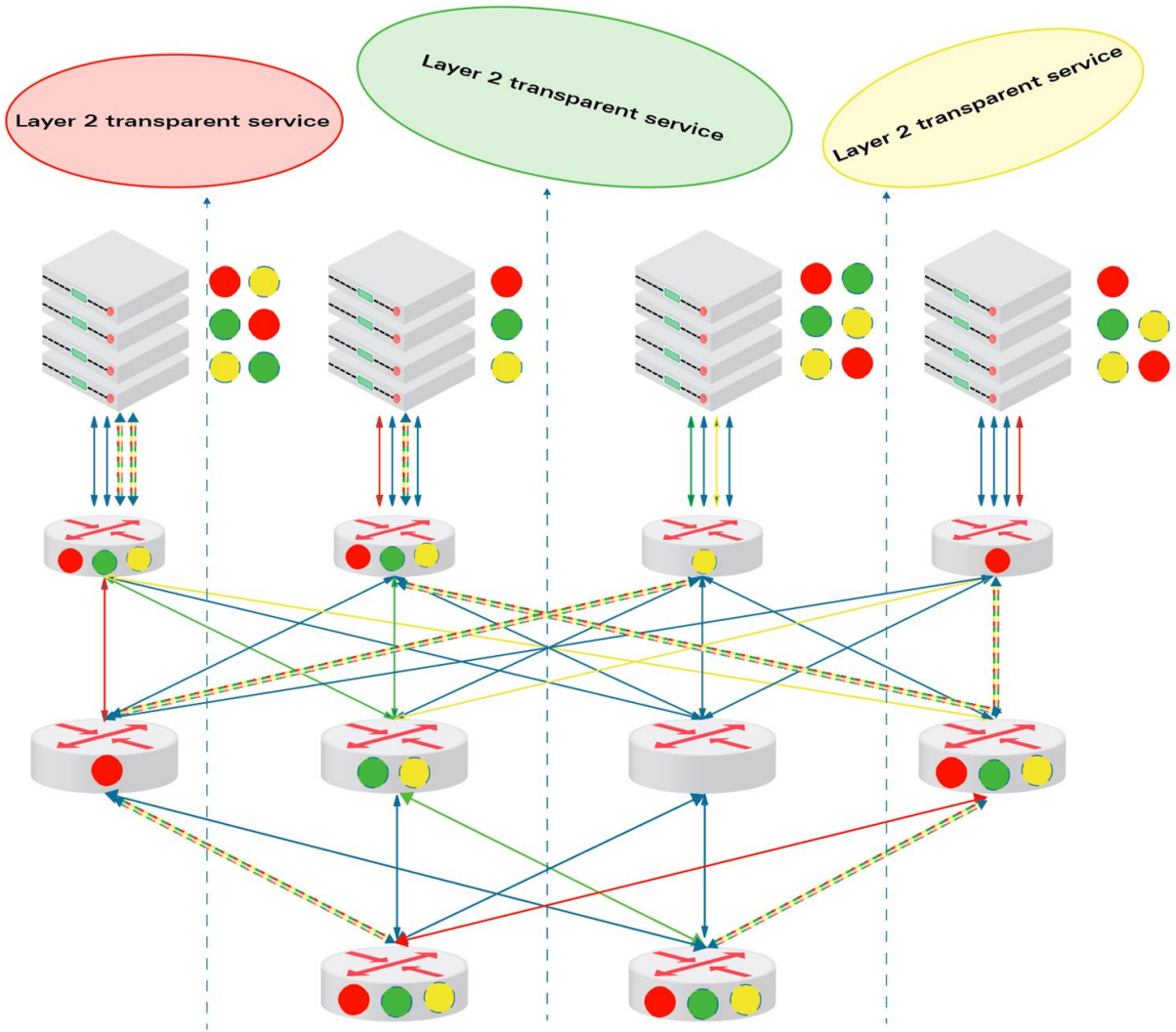
这些Overlay网络共享相同的底层网络，但对用户而言，这是一个完全透明的网络。

这对L2网络的用户来说是一件好事，因为他们可以做任何想做的事情。例如，使用任何IP地址或任何VLAN。

但是，缺点是监视底层网络，因为您还会看到Overlay网络。

另外，判断不同的Overlay网络也很复杂。





此图显示了普遍性问题
。

这些服务中每一个都可以使用相同的IP范围。

这显然是因为运行这些服务的人想让事情变得简单。

底层网络基础设施处理这些不同服务的分离。

这是通过隧道完成的。今天，这是典型的VXLAN；在过去，它是MPLS或VLAN。

不同的是今天它是动态的。

Overlay和底层网络的可见性方法

原则上，我们有几个可见性选项

- 底层网络的可见性
- 特定Overlay网络的可见性
- 所有Overlay网络的可见性
- 底层和Overlay层的同时可见性，一个“ 完整的端到端视图”

网络可见性 vs 端点可见性

网络监控和端点监控通常混在一起，但有很大的不同！

网络可见性

- 根据网络数据显示指标
- 大多数是被动解决方案
- 与设备与软件无关
- 低运营成本
- 端到端视图包括传输路径
- 有限的应用程序相关指标
- 安装阶段更复杂的方法（由于是硬件）
- 有助于故障排除

这是Cubro的主场**!**

端点可见性

- 显示基于日志或活动客户端的指标
- 通常不是被动的
- 不是不可知论的，通常需要采用每种设备
- 显示端到端性能，但不显示网络路径或网络参数。网络参数是端到端的间接派生
- 良好的以应用程序为中心的指标
- 昂贵且不可预测的运营成本
- 一开始易于安装
- 故障排除效率不高

网络可见性 vs 端点可见性

什么更重要？

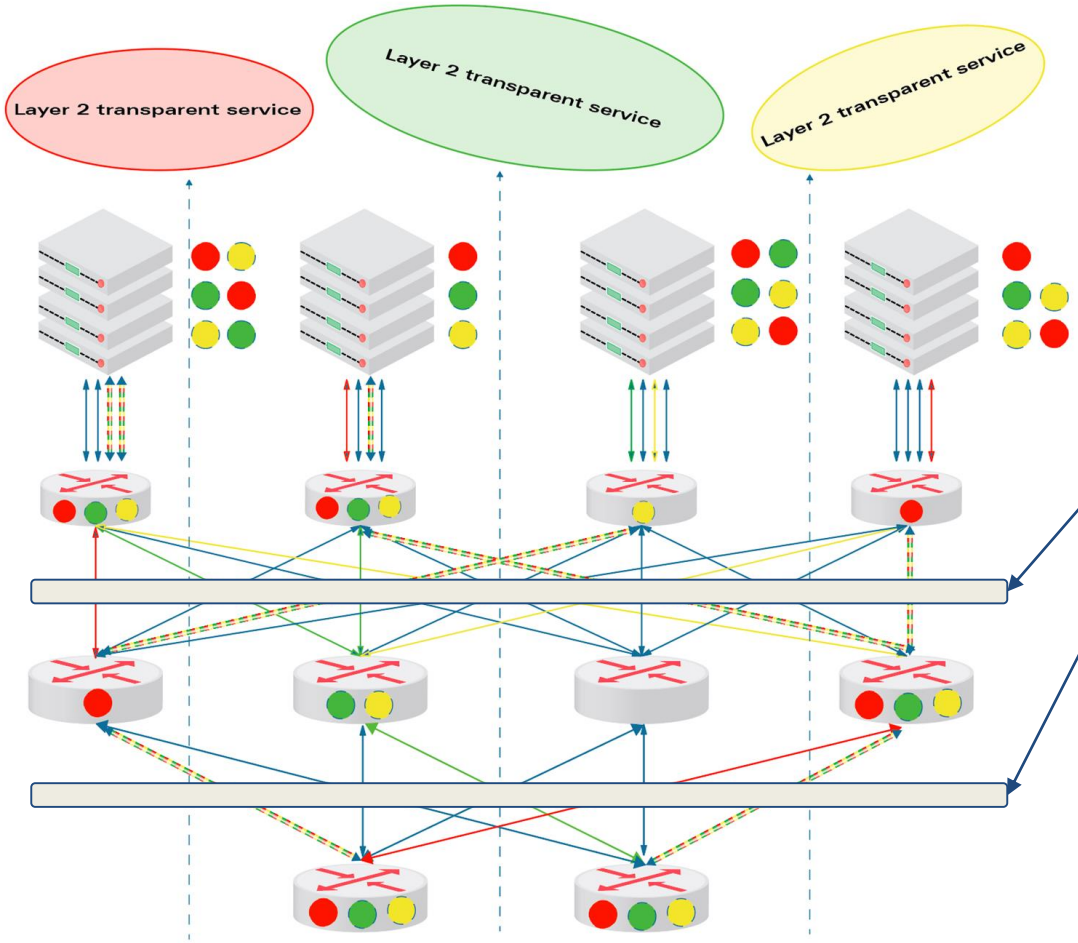
很难说，这取决于客户的需求... “ 客户业务案例”

- 服务提供商 = 80 %的网络 20 %的端点
- 数据中心提供商 = 90 %的网络10 %的端点
- 大型企业 = 40 %的网络 60 %的端点
- 拥有自己的云基础架构的企业 = 50 %的网络50 %的端点
- 拥有公共云基础架构的企业 = 10 %的网络90 %的端点

底线：这两个解决方案可以协同工作提供全面的可见性！检查您的需求，然后确定哪种方案最适合您！

https://www.egscyber.com/blog/what-would-be-better-out-of-network-monitoring-and-endpoint-monitoring/

Overlay网络设计 & L2 Overlay和监控



在此图片种可以清楚地看到问题如果您分流并监视以下几点：

我们看到2个问题：

可以看到几次相同的流量。

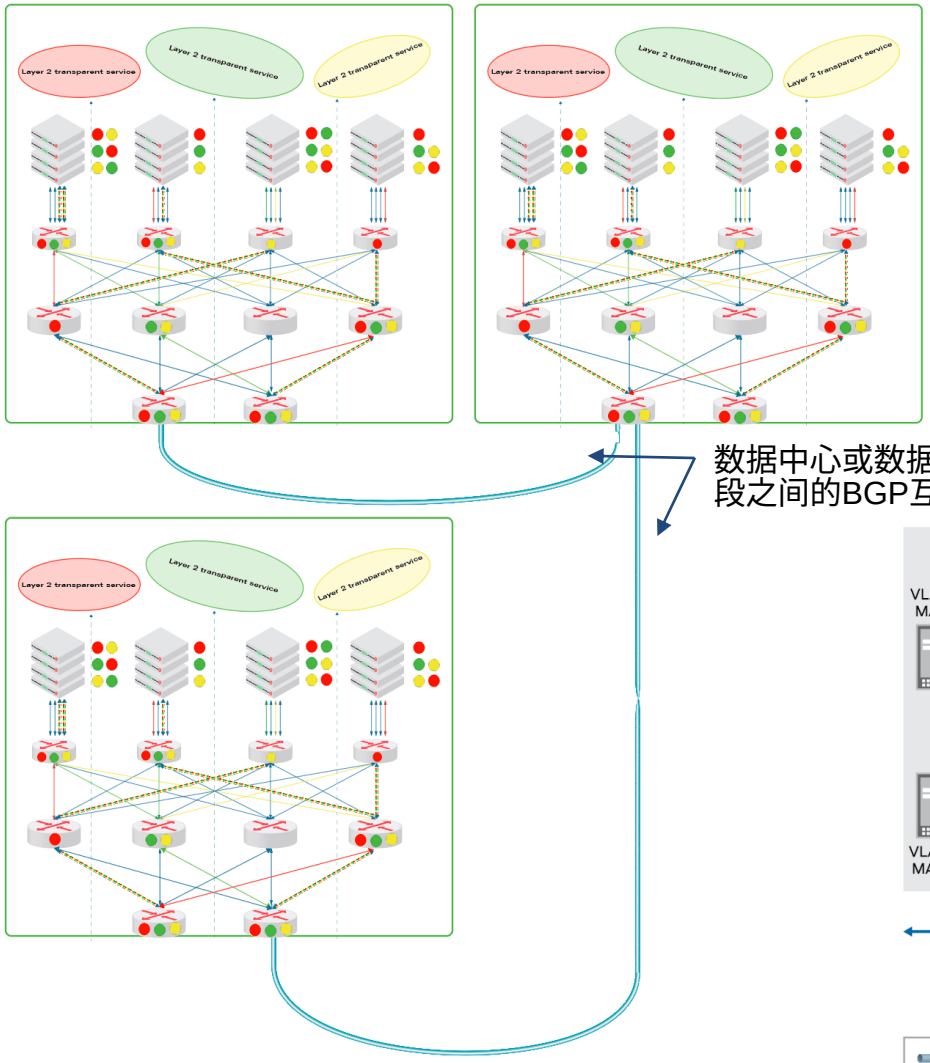
同时看到所有的Overlay网络。

“ L2网络可以运行相同的IP范围，因此，传统监控要将流量分离式非常复杂的，因为典型的监控解决方案使用IP地址来确定网络中的不同路径。”

典型的监控工具无法处理隧道流量。

几乎所有监控工具都涉及为仅在一个端口或一个逻辑网络层上处理流量。此工具通常无法关联流量。

问题 2

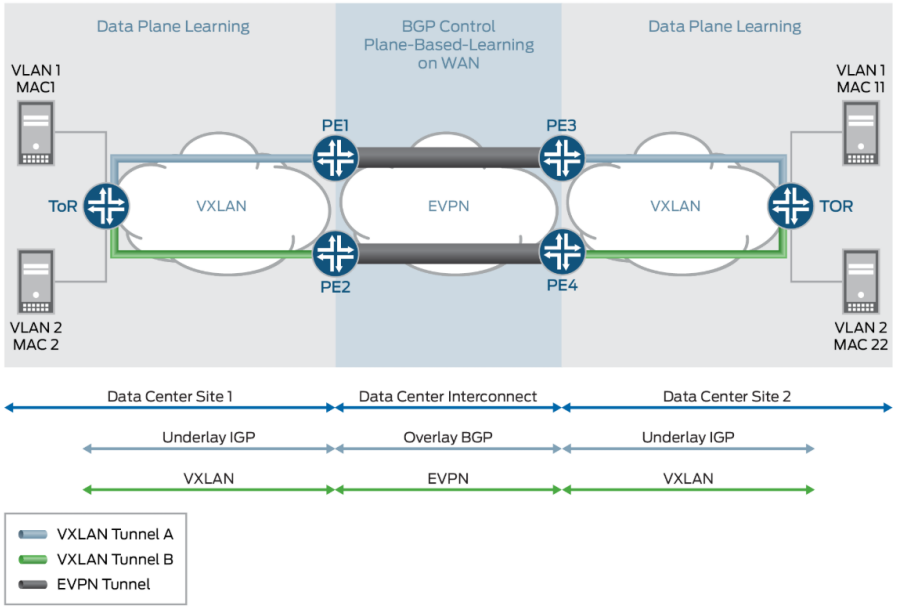


这个问题更加复杂。

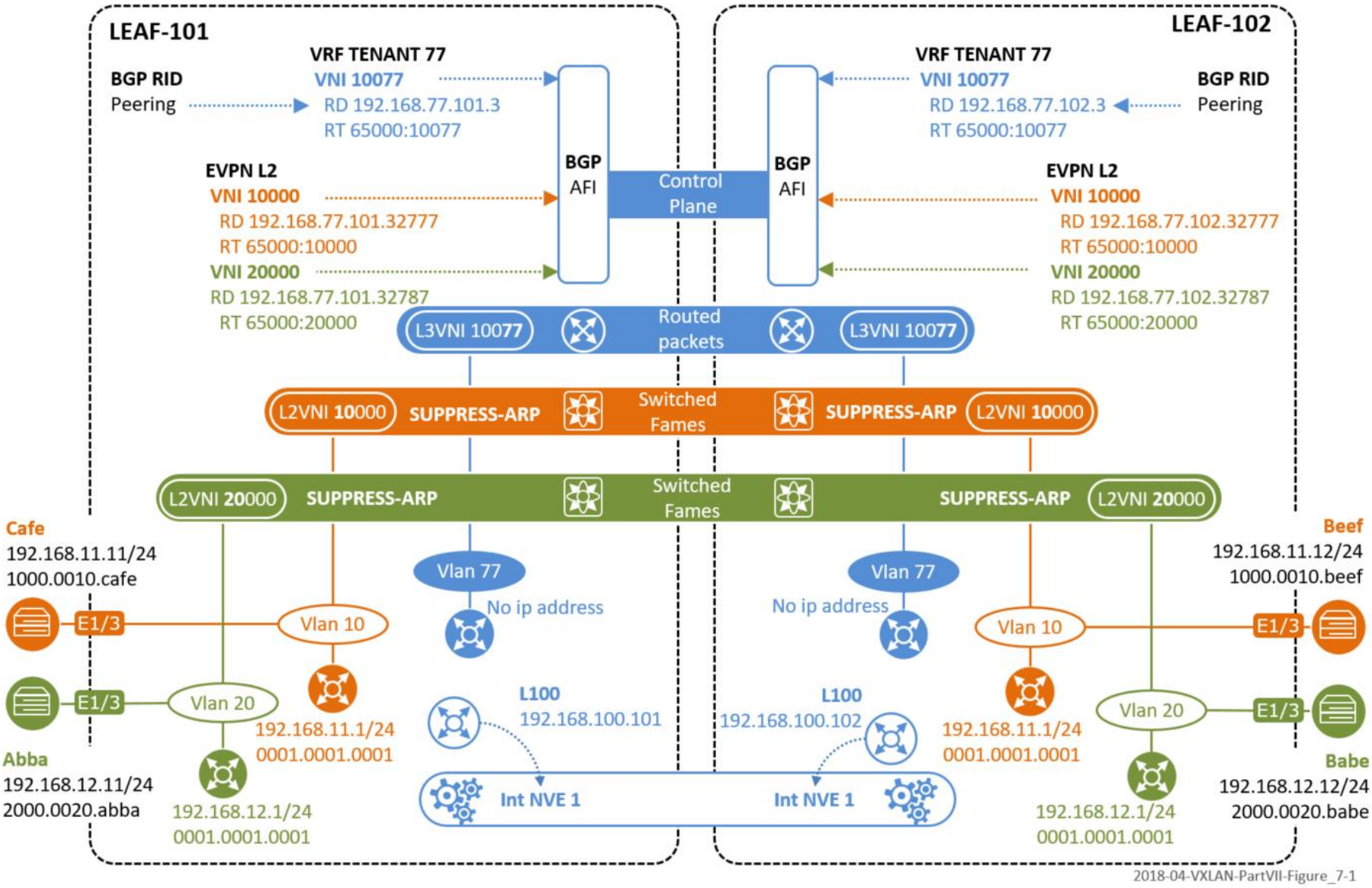
Overlay网络可以分布在不同的DC上。

这些不同的DC通常通过BGP链路连接。

在这种情况下，需要使用BGP关联才能产生有用效果。



VXLAN概念

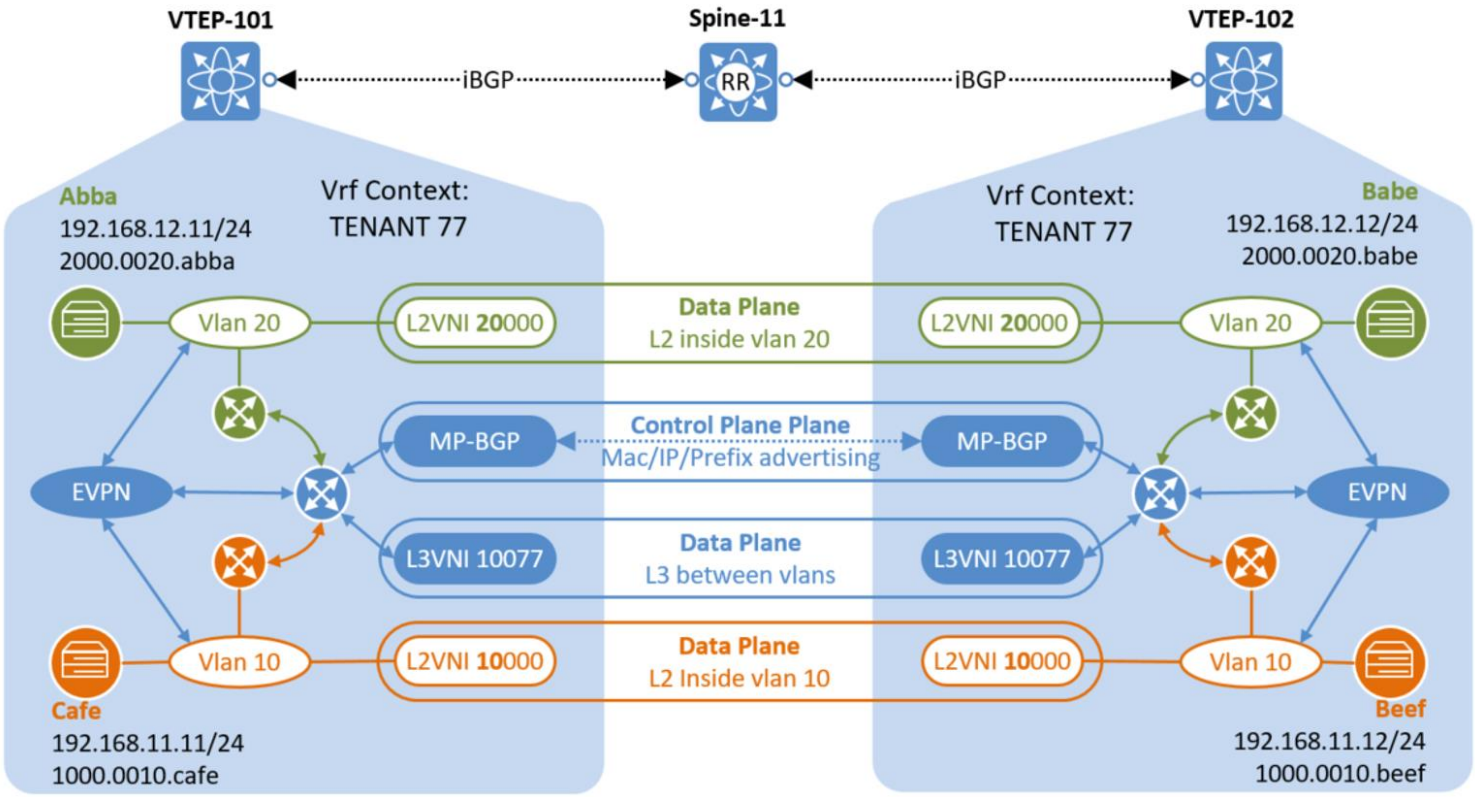


2018-04-VXLAN-PartVII-Figure_7-1

VXLAN概念

绿色和橙色是L2已交换
蓝色是L3已路由

请参阅上一张PPT

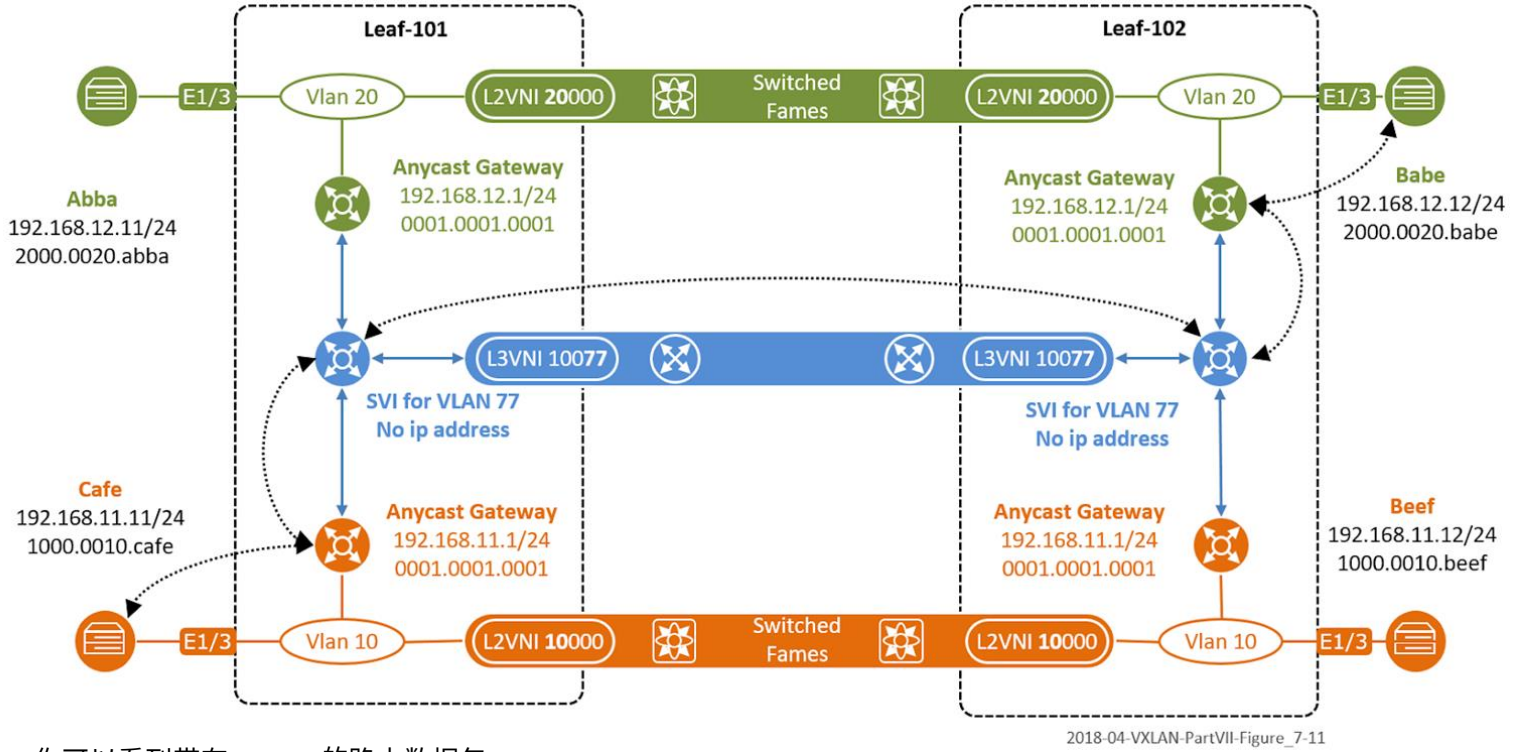


如果你不想监控来自Cafe的所有流量，仅在一个 VXLAN上过滤式不够的，在此示例种，您将会看到L2交换流量在VNI 10000 中，而路由流量在VNI 10077中.

现在的挑战是要知道哪个VXLAN属于同一个，因为当您有多个路由端点时，您就会有多个VXLAN ID。

连接是BGP！

VXLAN 概念



2018-04-VXLAN-PartVII-Figure_7-11

只有在 VXLAN ID 10077 你可以看到带有VLAN 77的路由数据包

在ID 10000 或 ID 20000 你只会看到发往GW带有 VLAN 10 和 VLAN 20的数据包

如果你删除VXLAN您将看到此数据包3次，带有不同的报头MAC和VLAN

此类流量无法通过重复数据删除功能删除，因为仅内容相同但报头不同*

```
> Frame 368: 164 bytes on wire (1312 bits), 164 bytes captured (1312 bits)
> Ethernet II, Src: 5e:00:00:00:00:07 (5e:00:00:00:00:07), Dst: 5e:00:00:02:00:07 (5e:00:00:02:00:07)
> Internet Protocol Version 4, Src: 192.168.100.101, Dst: 192.168.100.102
> User Datagram Protocol, Src Port: 60963, Dst Port: 4789
v Virtual eXtensible Local Area Network
  > Flags: 0x0800, VXLAN Network ID (VNI)
    Group Policy ID: 0
    VXLAN Network Identifier (VNI): 10077
    Reserved: 0
> Ethernet II, Src: 5e:00:00:00:00:07 (5e:00:00:00:00:07), Dst: 5e:00:00:01:00:07 (5e:00:00:01:00:07)
> Internet Protocol Version 4, Src: 192.168.11.11, Dst: 192.168.12.12
> Internet Control Message Protocol
```

VXLAN Concept timing

Basic connectivity test

We are going to test basic connectivity between the hosts with ping.

Ping from Café to Beef (L2VNI service over VXLAN fabric)

Figure 7: ping Café to Beef

```
Cafe#ping 192.168.11.11
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 192.168.11.11, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/2 ms
```

Ping from Café to Abba (Local routing)

Figure 8: ping Café to Abba

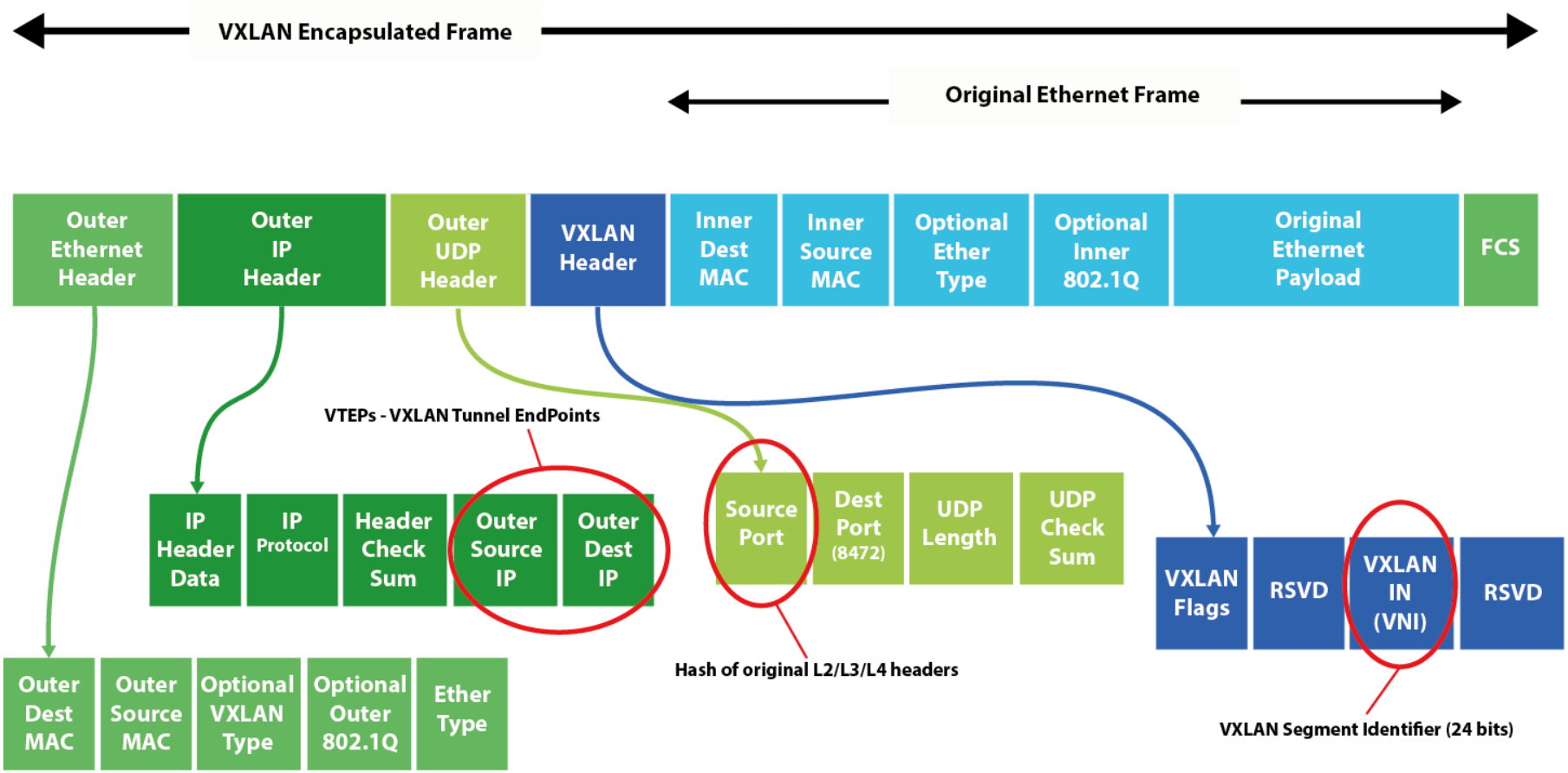
```
Cafe#ping 192.168.12.11
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 192.168.12.11, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/8/13 ms
```

Ping from Café to Babe (L3VNI service over VXLAN fabric)

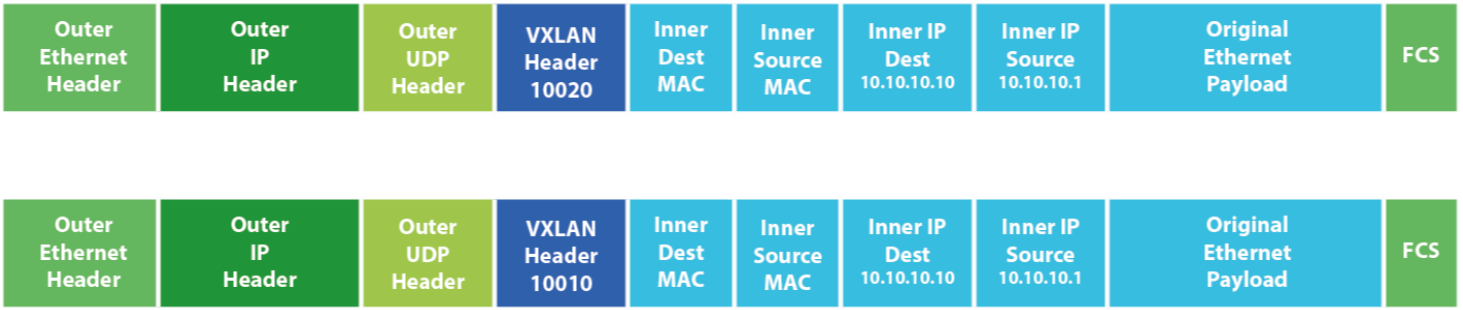
Figure 9: ping Café to Babe

```
Cafe#ping 192.168.12.12
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 192.168.12.12, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 20/23/29 ms
```

底层网络的帧结构


VXLAN Encapsulated Frame
Original Ethernet Frame
Outer Ethernet Header
Outer IP Header
Outer UDP Header
VXLAN Header
Inner Dest MAC
Inner Source MAC
Optional Ether Type
Optional Inner 802.1Q
Original Ethernet Payload
FCS
VTEPs - VXLAN Tunnel EndPoints
IP Header Data
IP Protocol
Header Check Sum
Outer Source IP
Outer Dest IP
Source Port
Dest Port (8472)
UDP Length
UDP Check Sum
VXLAN Flags
RSVD
VXLAN IN (VNI)
RSVD
Outer Dest MAC
Outer Source MAC
Optional VXLAN Type
Optional Outer 802.1Q
Ether Type
Hash of original L2/L3/L4 headers
VXLAN Segment Identifier (24 bits)

问题 1



这是常见的问题；相同的IP范围但不同Overlay。

通常情况下，标准监控工具看不到外部报头。

因此，内部IP测量的结果通常是错误的！

Overlay信息通常会丢失。

您得到了结果，但是错的！

使用SPAN端口的VXLAN监控场景

1:) 具有相同子网和虚拟机管理程序上的EVPN终端的leaf流量

2:) 具有相同子网和EVPN终端交换机端口的leaf流量

3:) 具有相同子网和EVPN终端一侧交换机端口另一侧虚拟机管理程序的leaf流量

4:) 不同leafs上的子网内流量和虚拟机监控程序上的EVPN终端（L2层之间）

5:) 不同leafs和EVPN终端交换机端口上的子网内流量（L2层之间）

6:) 不同leafs和EVPN终端一侧交换机端口另一侧虚拟机管理程序上的子网内流量（L2层之间）

7:) 具有不同子网和虚拟机监控程序上EVPN终端的leaf流量（L3层之间）

8:) 具有不同子网和EVPN终端交换机端口的leaf流量（L3层之间）

9:) 具有不同子网和EVPN终端一侧交换机端口另一侧虚拟机管理程序的leaf流量（L3层之间）

10:) 不同leafs和虚拟机管理程序上EVPN终端上的不同子网流量

11:) 不同leafs和EVPN终端交换机端口上的不同子网流量

12:) 不同leafs和EVPN终端上一侧交换机另一侧虚拟机管理程序的不同子网流量

使用SPAN端口的VXLAN监控场景

1:) 具有相同子网和虚拟机管理程序上EVPN终端的leaf流量

仅使用SPAN端口进行出口监控

没有重复项，但流量上有VXLAN和LAN标签

2:) 具有相同子网和EVPN终端交换机端口的leaf流量

仅使用SPAN端口进行出口监控

没有重复项，但流量上有VXLAN和LAN标签

3:) 具有相同子网和EVPN终端一侧交换机端口另一侧虚拟机管理程序的leaf流量

仅使用SPAN端口进行出口监控

没有重复项,

在交换机端口终端上，不使用VXLAN而使用VLAN

在虚拟机管理程序终端是VXLAN和VLAN

在这种情况下，请求没有VXLAN，而应答有VXLAN，反之亦然

使用SPAN端口的VXLAN监控场景

4:) 不同leafs和虚拟机管理程序上EVPN终端的子网内流量

仅使用SPAN端口进行出口监控

重复项，

数据包1:）在spin出口VXLAN和VLAN上

数据包 2:）在leaf出口VXLAN和VLAN上（两者都一样）

删除重复数据 ??

重复数据删除不支持VXLAN，因此必须首先删除VXLAN，这是一种非常昂贵的方法，无法通过过滤器实现。

问题是它不能基于链接来完成操作，因为重复项位于不同的链接上，所以首先我们必须删除VXLAN，而不是将全部聚合到一个大管道中，但是此大管道会使重复数据消除CPU（100 Gbit +）过载，因此需要为不同的重复数据删除CPU提供负载均衡，非常复杂！

(也不清楚网络层是否未受影响？“ *IP ID 字段*”)

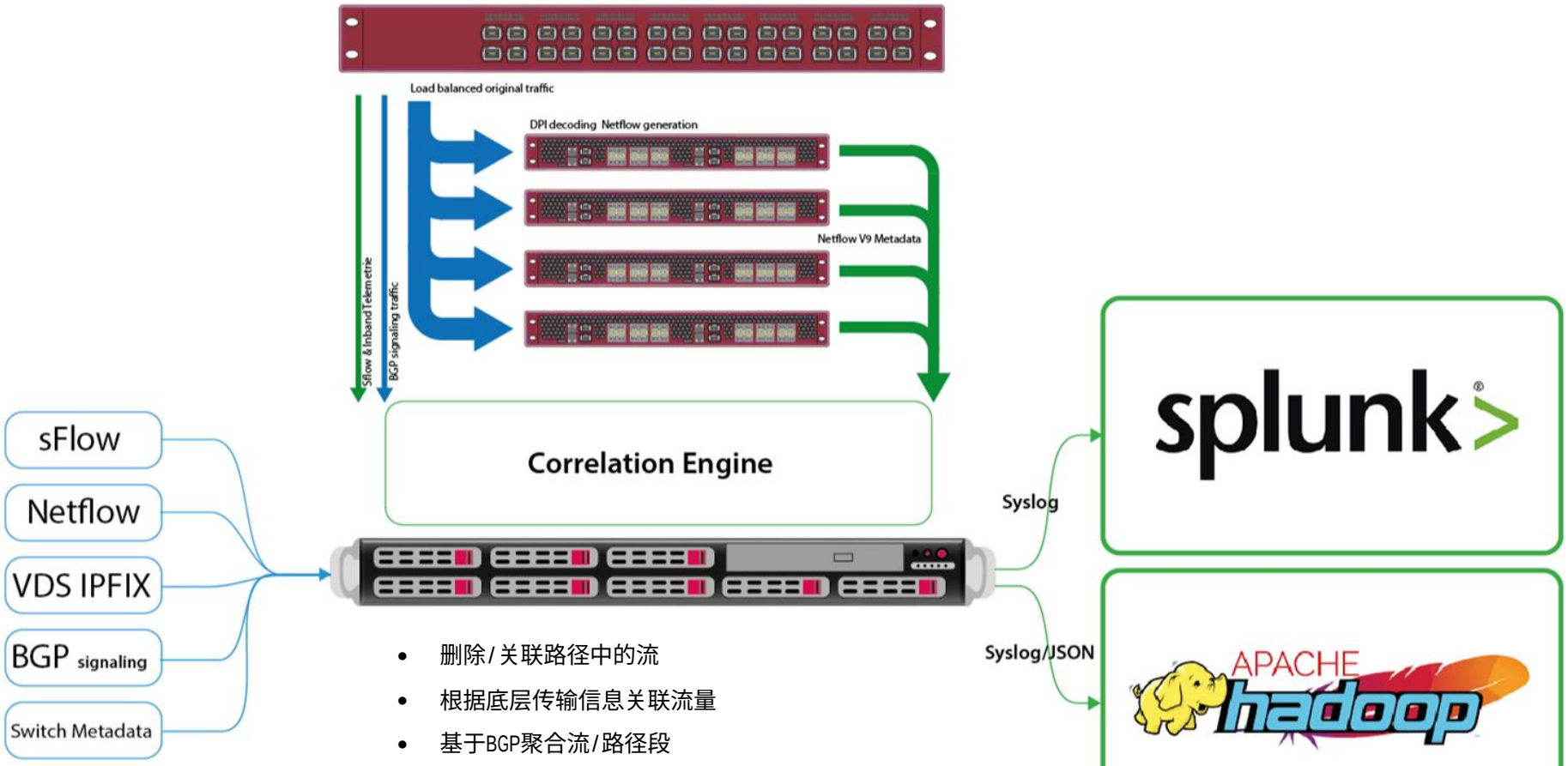
5:) 不同leafs和EVPN终端交换机端口上的子网内流量（L2层之间）

和第4点一样但没有VXLAN

6:) 不同leafs和EVPN终端一侧交换机端口另一侧虚拟机管理程序的子网内流量（L2层之间）

和第4点一样，但一个数据包有VXLAN + VLAN，另一个只有VLAN

Cubro解决方案设计 1



- 删除/关联路径中的流
- 根据底层传输信息关联流量
- 基于BGP聚合流/路径段
- 丰富来自交换机带内遥测的数据
- 通过交换机表信息丰富数据

Cubro 解决方案设计 1

网络路径

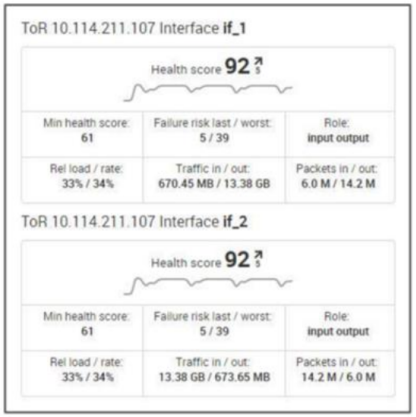
现在已经确定了受物理网络中断影响的应用程序和主机，SDDC管理员可以通过选择终端节点来查看VM到VM的流量被封装在哪里，并可以具体查看流量经过了哪些物理网络设备。

现在，SDDC管理员可以确定路径中的哪些网络设备是导致应用程序性能问题的原因。

下图显示了VM到CM通信中涉及的网络设备接口。

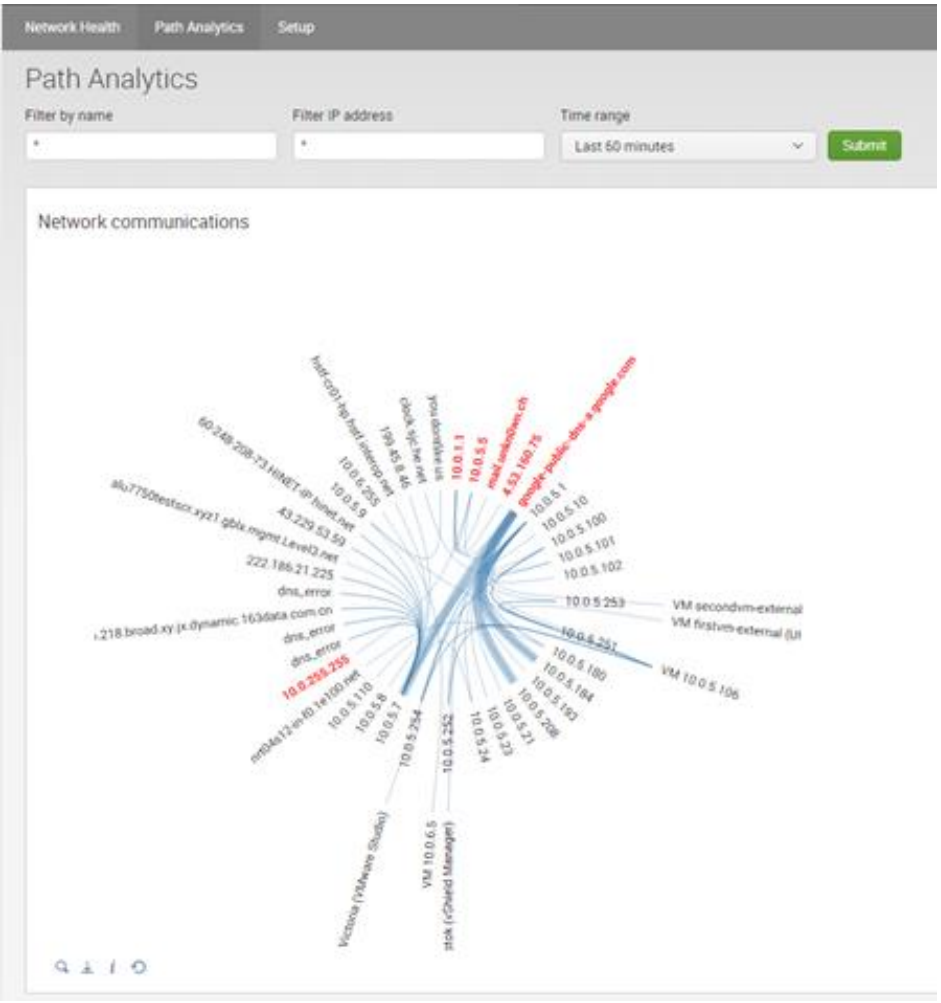
对于中继跟踪通信的接口，将显示以下信息：

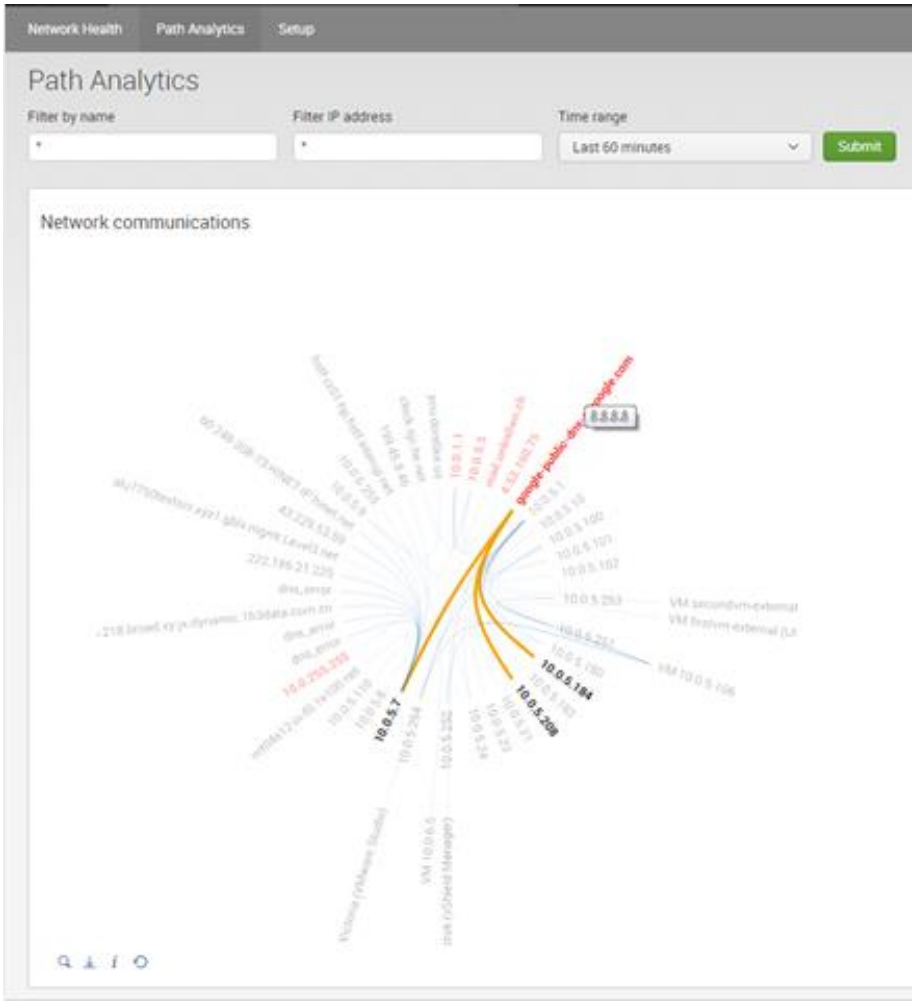
- 该接口上的相对业务量占其标称容量的百分比
- 在当前平均数据包大小下可维持的最大数据包速率接口上的相对数据包速率
- 在选定的时间间隔内，通过此接口在每个方向上通过的字节总数
- 在选定的时间间隔内，通过此接口在每个方向上通过的数据包总数



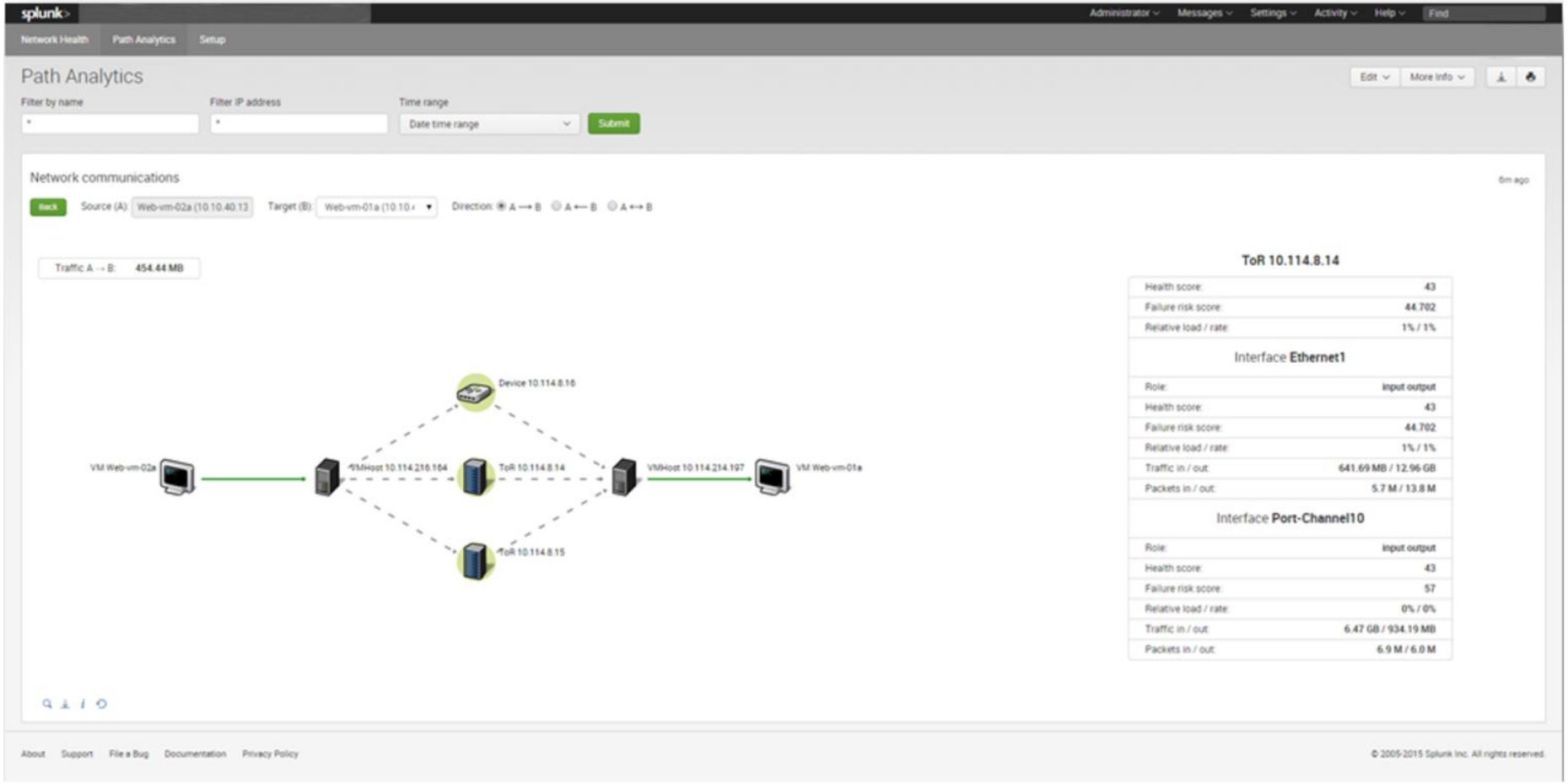
路径信息不仅适用于日期中心内的VM至VM（东西流量），也适用于VM至网关（南北流量）。此功能可用于识别网络拥塞和异常活动，例如数据泄露。

虚拟Overlay网络中的对话

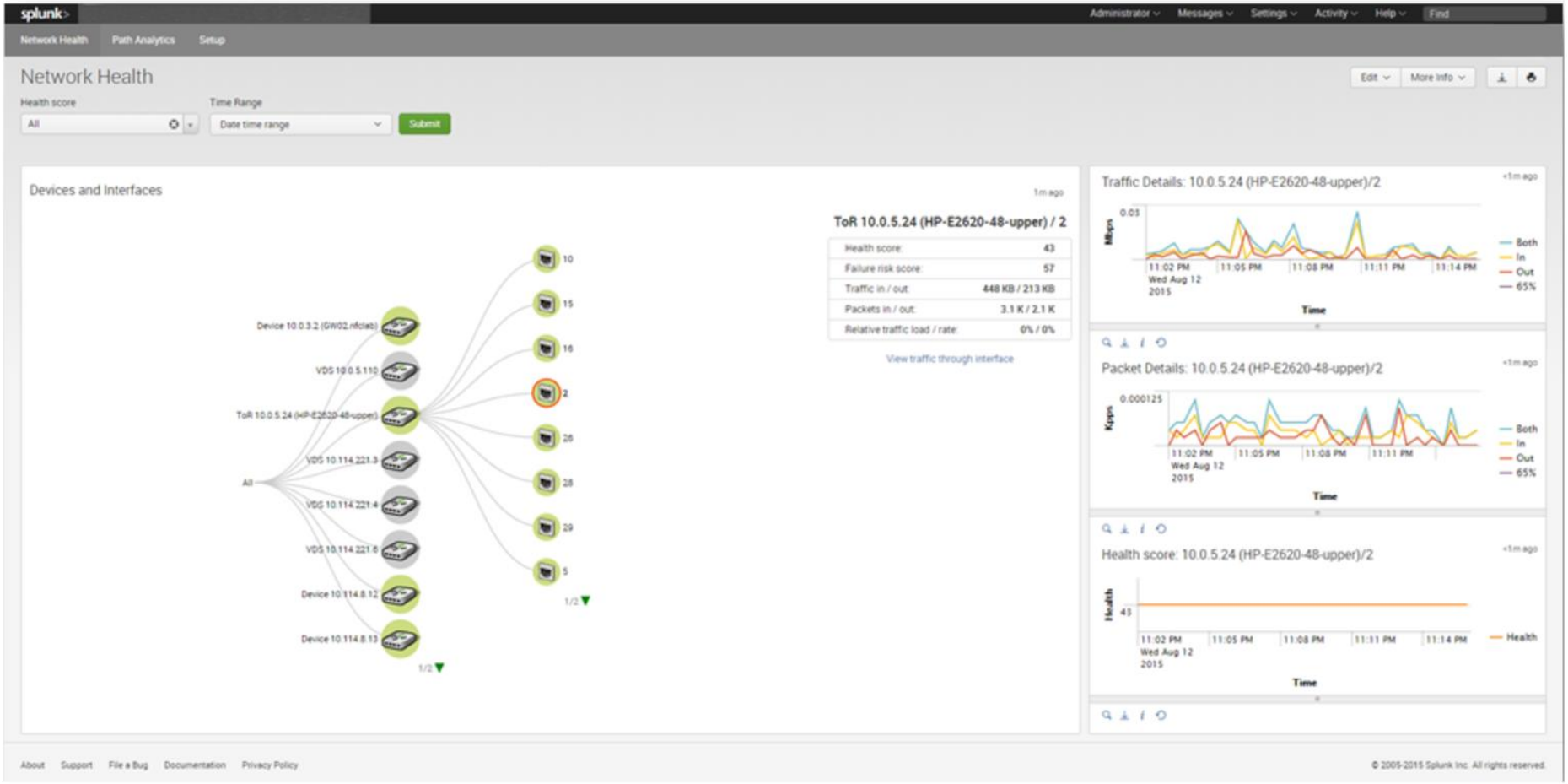




虚拟机VM会话的物理路径呈现



网络健康


splunk>
Administrator
Messages
Settings
Activity
Help
Find
Network Health
Path Analytics
Setup
Network Health
Edit
More Info
Health score
All
Time Range
Date time range
Submit
Devices and Interfaces
1m ago
Device 10.0.3.2 (GW02.nfclab)
VDS 10.0.5.110
ToR 10.0.5.24 (HP-E2620-48-upper)
VDS 10.114.221.3
All
VDS 10.114.221.4
VDS 10.114.221.6
Device 10.114.8.12
Device 10.114.8.13
1/2
10
15
16
2
26
28
29
5
1/2
ToR 10.0.5.24 (HP-E2620-48-upper) / 2
Health score: 43
Failure risk score: 57
Traffic in / out: 448 KB / 213 KB
Packets in / out: 3.1 K / 2.1 K
Relative traffic load / rate: 0% / 0%
View traffic through interface
Traffic Details: 10.0.5.24 (HP-E2620-48-upper)/2
+1m ago
Mbps
0.05
11:02 PM
Wed Aug 12
2015
11:05 PM
11:08 PM
11:11 PM
11:14 PM
Time
Both
In
Out
65%
Packet Details: 10.0.5.24 (HP-E2620-48-upper)/2
+1m ago
Kpps
0.000125
11:02 PM
Wed Aug 12
2015
11:05 PM
11:08 PM
11:11 PM
Time
Both
In
Out
65%
Health score: 10.0.5.24 (HP-E2620-48-upper)/2
+1m ago
Health
43
11:02 PM
Wed Aug 12
2015
11:05 PM
11:08 PM
11:11 PM
11:14 PM
Time
Health
About
Support
File a Bug
Documentation
Privacy Policy
© 2005-2015 Splunk Inc. All rights reserved.

其他特征

- Top tunnels ——按照流量显示Top隧道
- Top Flows ——显示隧道内的Top流
- 分布式防火墙（DFW）——即将到来
- 分布式逻辑路由（DLR）——即将到来

Top Tunnels, Top Flows

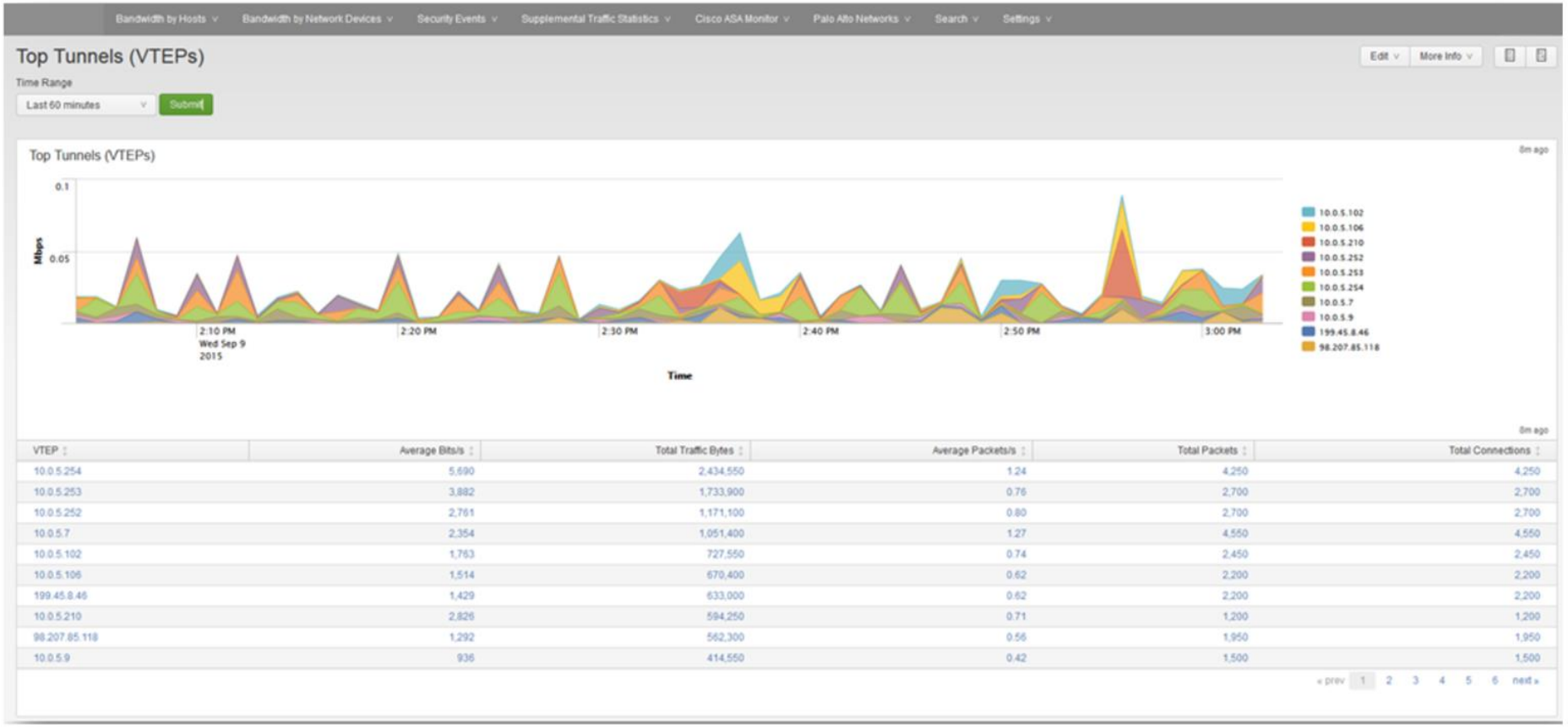
- 按流量排名的Top Tunnels(VTEP)；选择时间间隔；显示：

VTEP、平均Bits/s、总流量字节数、平均数据包/秒、总数据包数、总连接数

- 通过从上面的列表中选择VTEP进行深入研究； 显示：

VXLAN_ID、源VM IP、源VM名称、源VTEP、目标VM IP、目标VM名称、目标VTEP、平均Bits/s、总流量字节、平均数据包/秒、总数据包、总连接数

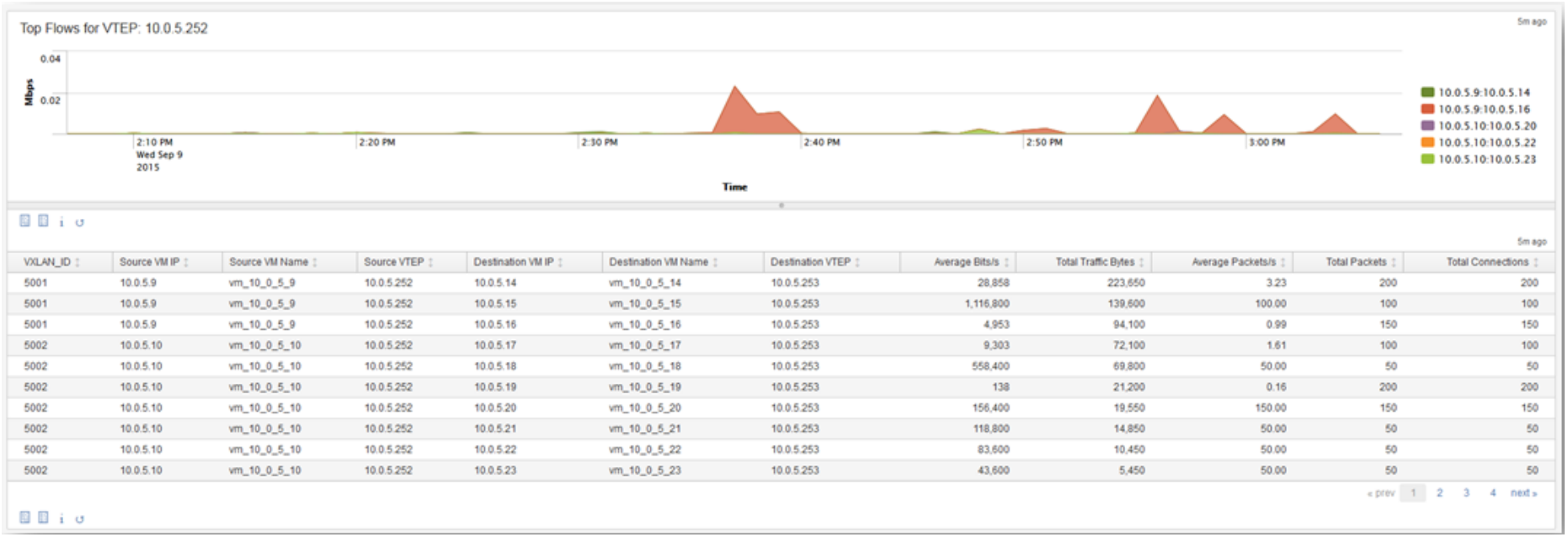
Top Tunnels (VTEPs)



VTEP	Average Bits/s	Total Traffic Bytes	Average Packets/s	Total Packets	Total Connections
10.0.5.254	5,690	2,434,550	1.24	4,250	4,250
10.0.5.253	3,882	1,733,900	0.76	2,700	2,700
10.0.5.252	2,761	1,171,100	0.80	2,700	2,700
10.0.5.7	2,354	1,951,400	1.27	4,550	4,550
10.0.5.102	1,763	727,550	0.74	2,450	2,450
10.0.5.106	1,514	670,400	0.62	2,200	2,200
199.45.8.46	1,429	633,000	0.62	2,200	2,200
10.0.5.210	2,826	594,250	0.71	1,200	1,200
98.207.85.118	1,292	562,300	0.56	1,950	1,950
10.0.5.9	936	414,550	0.42	1,500	1,500

« prev 1 2 3 4 5 6 next »

深入了解VTEP



5m ago

VXLAN_ID	Source VM IP	Source VM Name	Source VTEP	Destination VM IP	Destination VM Name	Destination VTEP	Average Bits/s	Total Traffic Bytes	Average Packets/s	Total Packets	Total Connections
5001	10.0.5.9	vm_10_0_5_9	10.0.5.252	10.0.5.14	vm_10_0_5_14	10.0.5.253	28,858	223,650	3.23	200	200
5001	10.0.5.9	vm_10_0_5_9	10.0.5.252	10.0.5.15	vm_10_0_5_15	10.0.5.253	1,116,800	139,600	100.00	100	100
5001	10.0.5.9	vm_10_0_5_9	10.0.5.252	10.0.5.16	vm_10_0_5_16	10.0.5.253	4,953	94,100	0.99	150	150
5002	10.0.5.10	vm_10_0_5_10	10.0.5.252	10.0.5.17	vm_10_0_5_17	10.0.5.253	9,303	72,100	1.61	100	100
5002	10.0.5.10	vm_10_0_5_10	10.0.5.252	10.0.5.18	vm_10_0_5_18	10.0.5.253	558,400	69,800	50.00	50	50
5002	10.0.5.10	vm_10_0_5_10	10.0.5.252	10.0.5.19	vm_10_0_5_19	10.0.5.253	138	21,200	0.16	200	200
5002	10.0.5.10	vm_10_0_5_10	10.0.5.252	10.0.5.20	vm_10_0_5_20	10.0.5.253	156,400	19,550	150.00	150	150
5002	10.0.5.10	vm_10_0_5_10	10.0.5.252	10.0.5.21	vm_10_0_5_21	10.0.5.253	118,800	14,850	50.00	50	50
5002	10.0.5.10	vm_10_0_5_10	10.0.5.252	10.0.5.22	vm_10_0_5_22	10.0.5.253	83,600	10,450	50.00	50	50
5002	10.0.5.10	vm_10_0_5_10	10.0.5.252	10.0.5.23	vm_10_0_5_23	10.0.5.253	43,600	5,450	50.00	50	50

« prev 1 2 3 4 next »

Top VMs

- 按流量排名Top VMs； 选择时间间隔； 显示：

VM_IP, VM_Name, Bytes_IN, Bytes_OUT, Packets_IN, Packets_OUT, 总连接数

- 通过从上面的列表中选择Bytes_IN进行深入研究； 显示（选定的VM为dest_VM）：

VXLAN_ID, src_vm, src_VTEP, dest_vm, dest_VTEP, Avg Bits/s, Bytes, 平均数据包/秒，数据包，连接数

- 从上面的列表中选择Bytes_OUT进行深入研究； 显示（选定的VM为src_VM）：

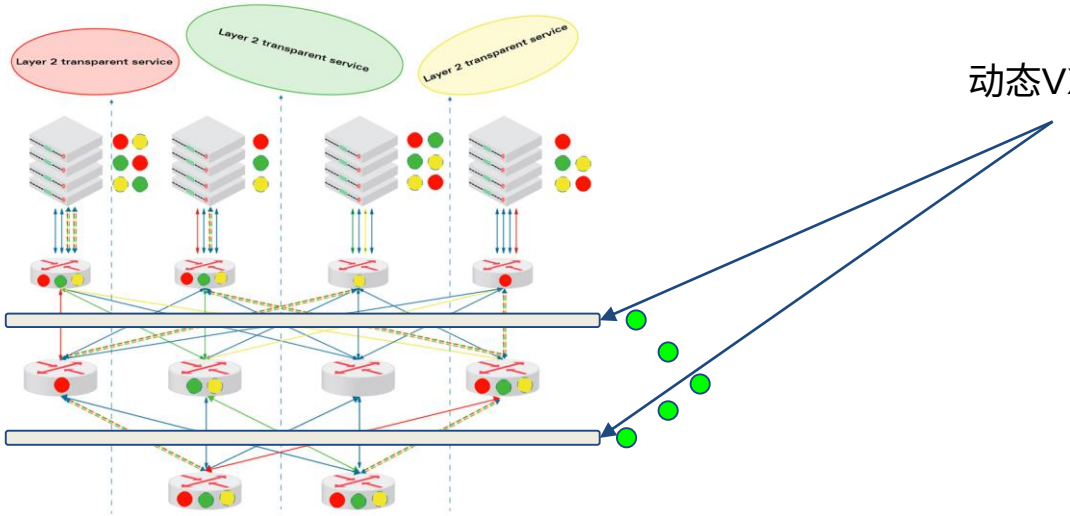
VXLAN_ID, src_vm, src_VTEP, dest_vm, dest_VTEP, Avg Bits/s, Bytes, 平均数据包/秒，数据包，连接数

Cubro解决方案设计 2

另一个可能的选项是动态VXLAN过滤。此解决方案适用于基于数据包的解决方案（例如Wireshark）和例如移动监控系统。

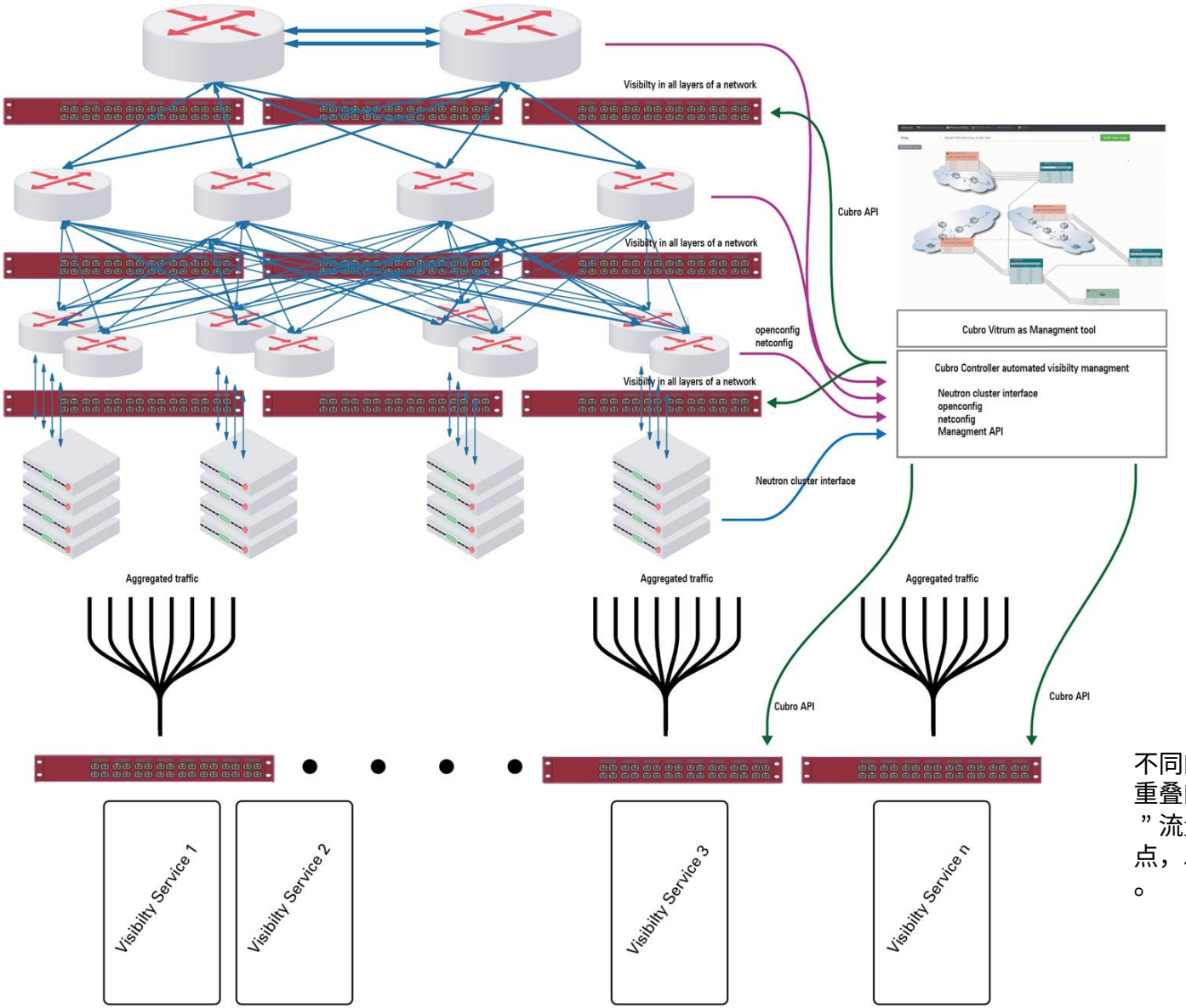
可以重新利用旧设备，因为动态VXLAN过滤将确保仅过滤出来自相关覆盖层的流量并将其发送到传统监视工具。

挑战在于，只有少数NPB能够进行VXLAN过滤。第二个问题是必须动态完成。因此，某些信令协议必须由包代理或外部设备解码。 -> Cubro云交换机！



动态VXLAN过滤请参阅下一张幻灯片

Cubro解决方案设计 2—多服务方法



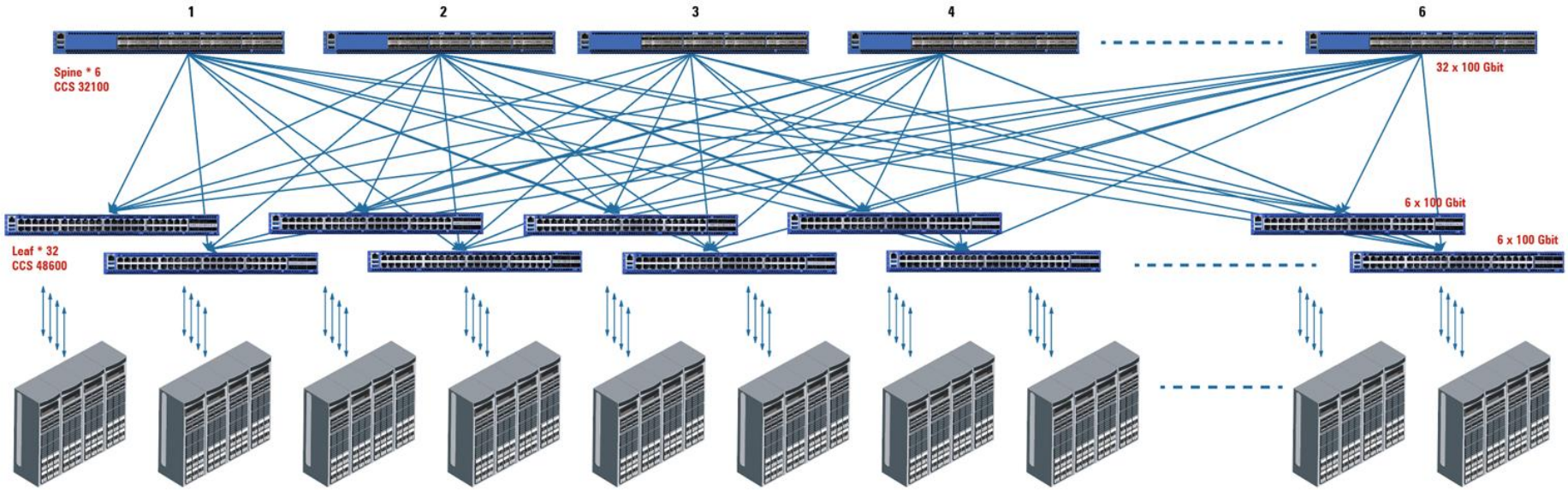
该解决方案的中心是Cubro可视化控制器，该控制器与Openstack和其他网络元素进行交互，并控制Cubro可视化节点（高级网络数据包代理）。Cubro可视化节点不是数据包推送器，它们具有高级API，可以满足当今网络的可见性需求！

不同的可视化服务，通常需要相同或重叠的流量。在这种情况下，“聚合”流量将发送到多个Cubro可视化节点，以便为请求的服务提供最终准备。

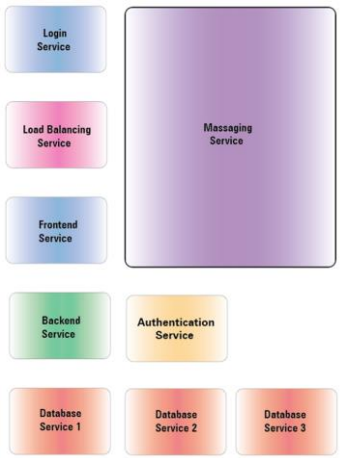
Cubro解决方案设计3

Cubro提供的另一种可选的高级解决方案是使用Cubro云交换机，因为Cubro云交换机（CCS）结合了高级交换结构和可视化结构。

在云中，可视化必须是网络基础架构的一部分！



The Cloud is Breathing

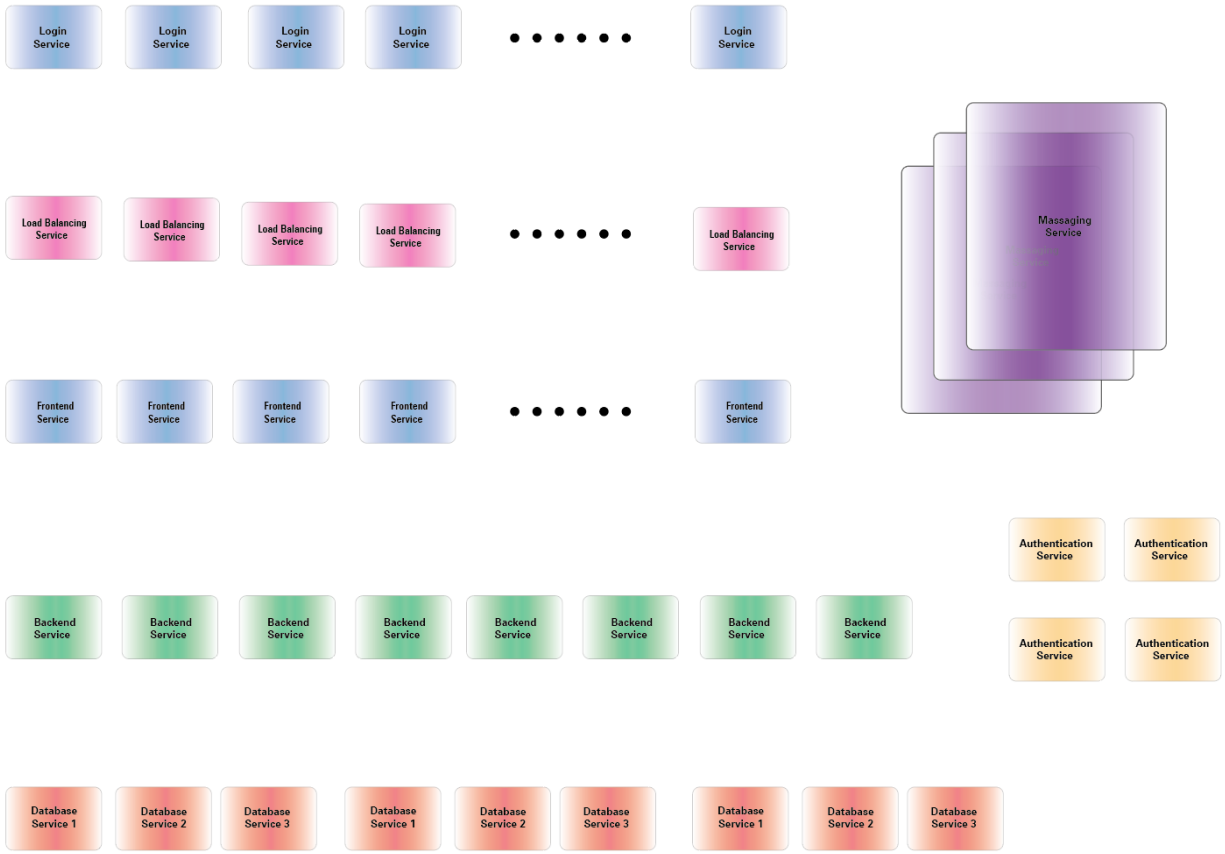


Application Breathing

云中的一大优势是，应用程序可以在需要更多资源时动态增长。

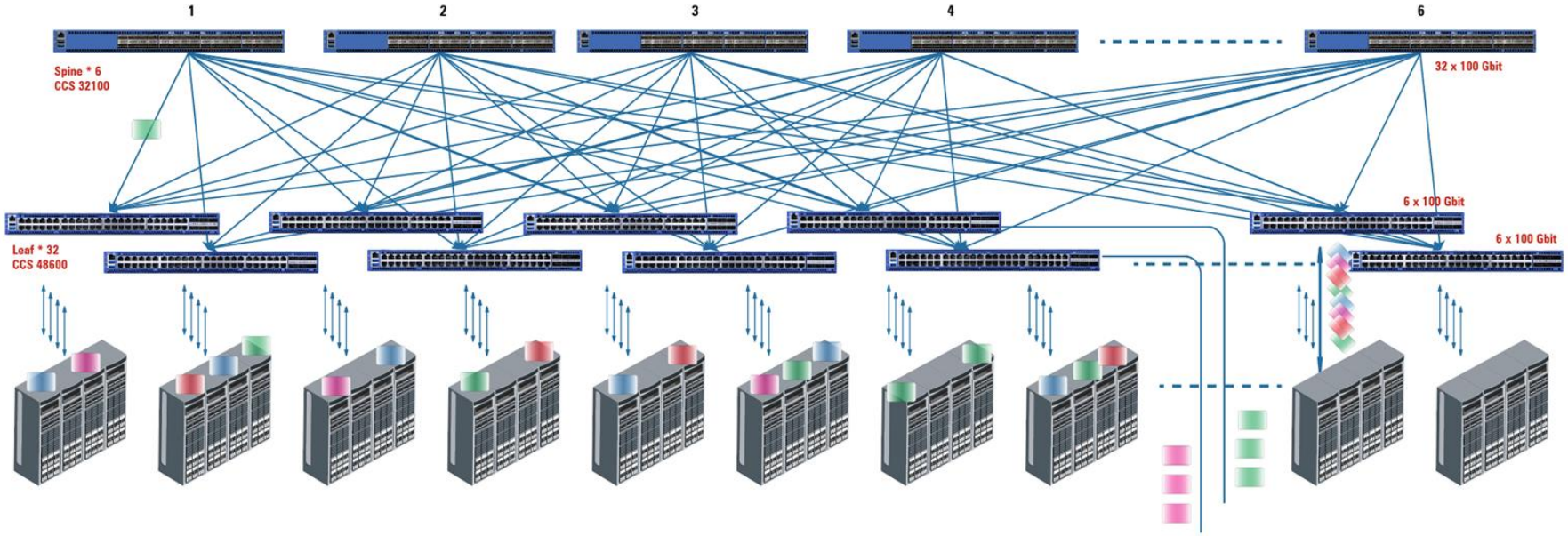
但是，同一应用程序也可以在遍布全球的不同数据中心中运行。

应用程序可以轻松地承受几千个微服务的负载。



Cubro解决方案设计3

可视化解决方案必须实时跟踪云的呼吸，如果可视化解决方案是网络基础结构的一部分并从网络结构接收相关的服务元数据，则这是可能的。



EX - EXA - CVN - CCS

EX = 具有L4功能的经典高端NPB

- 聚合
- 过滤
- 负载均衡
- .....

EXA = 具有L7功能的经典高端NPB

- 聚合
- 过滤到L7 “关键词搜索”
- 时间戳
- GTP负载均衡
- VXLAN元数据过滤
- ……..

静态方法 / 手动配置

CVN = Cubro可见性节点“自我组织”

这不再是NPB，因为它与Cubro可视化控制器交互并支持用于现代覆盖网络的动态数据包处理方法。

- 动态可视化服务指导
- 动态负载均衡
- 动态数据包修改
- ……

CCS = Cubro云交换机是“网络的一部分”

CCS当前是从L4 NPB到包含可见性功能的活动网络设备演进的终点。

- 带内动态可视化服务指导
- 以云为中心的可视化
- 应用程序呼吸支持
- ….

动态方法 / 自我组织可视化

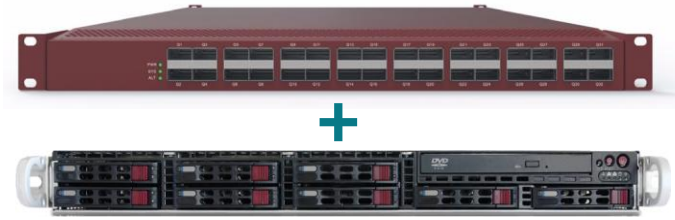
CUBRO NETWORK VISIBILITY
SIMPLE | SCALABLE | SUSTAINABLE

EX - EXA - CVN - CCS

EX

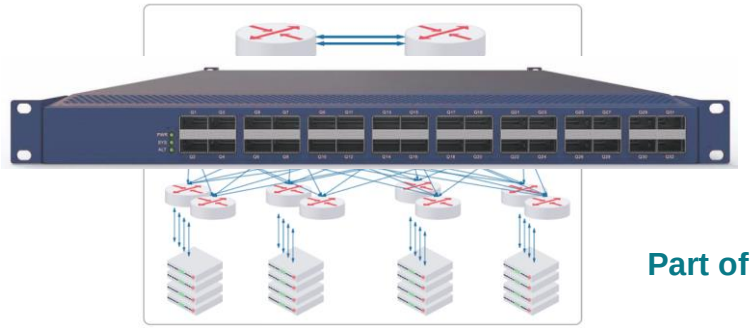
EXA

CVN



Cubro Controller

CCS

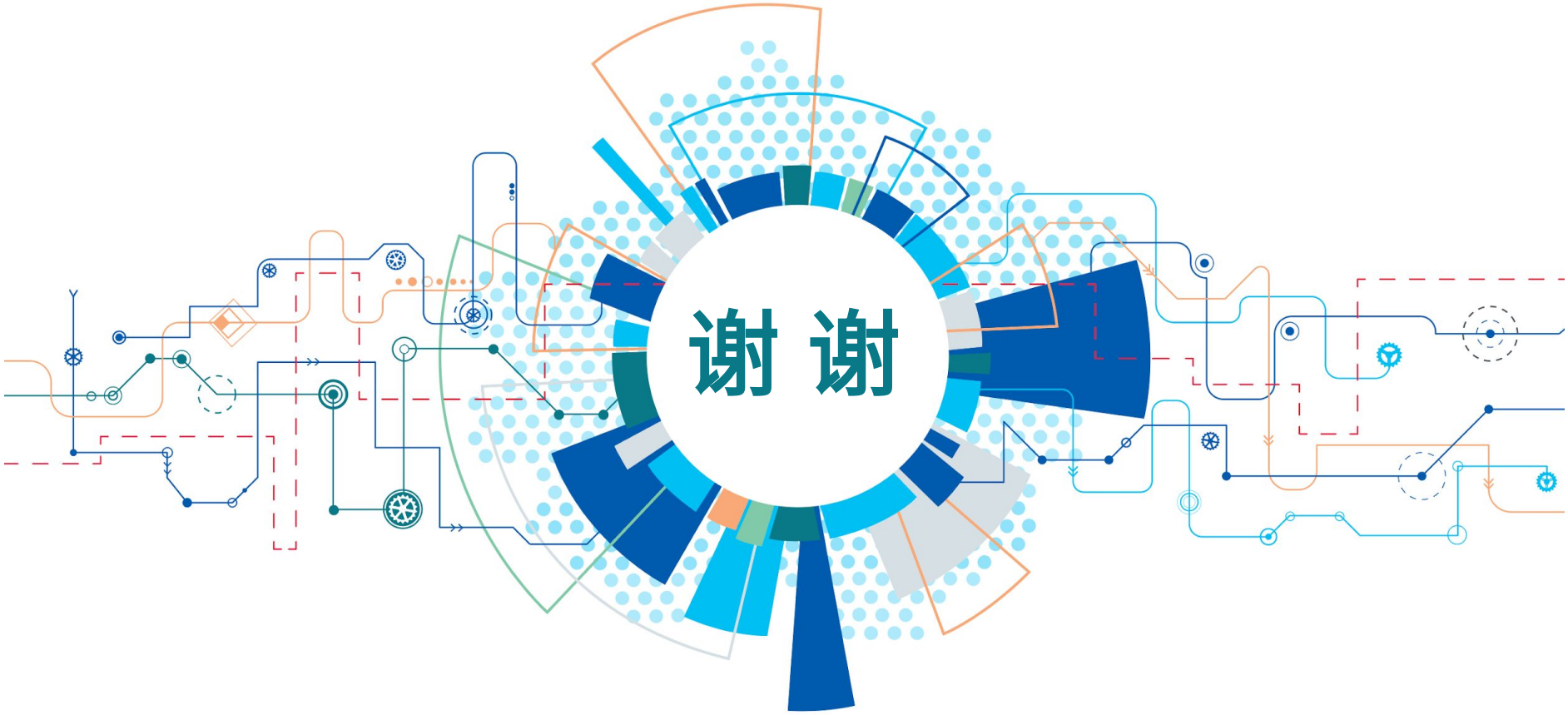


Part of the network infrastructure

EVOLUTION

CUBRO NETWORK VISIBILITY

SIMPLE | SCALABLE | SUSTAINABLE



HongKe 虹科

关注我们

hongwangle

广州虹科电子科技有限公司

需要详细信息？请通过sales@hkaco.com

联系我们 | 电话: 400-999-3848办事处：广州 | 北京 | 上海 | 深圳 | 西安 | 武汉 | 成都 | 沈阳 | 香港 | 台湾 | 美国