

P4及可编程芯片 在腾讯网络中的应用前景

TEG网络平台部
吕建超

腾讯需要敏捷高效的网络

精细
调度

- 互联网业务的高速增长，需要网络有很好的扩展性，需要网络具备灵活的、精细化的调度能力，高效利用网络资源的能力

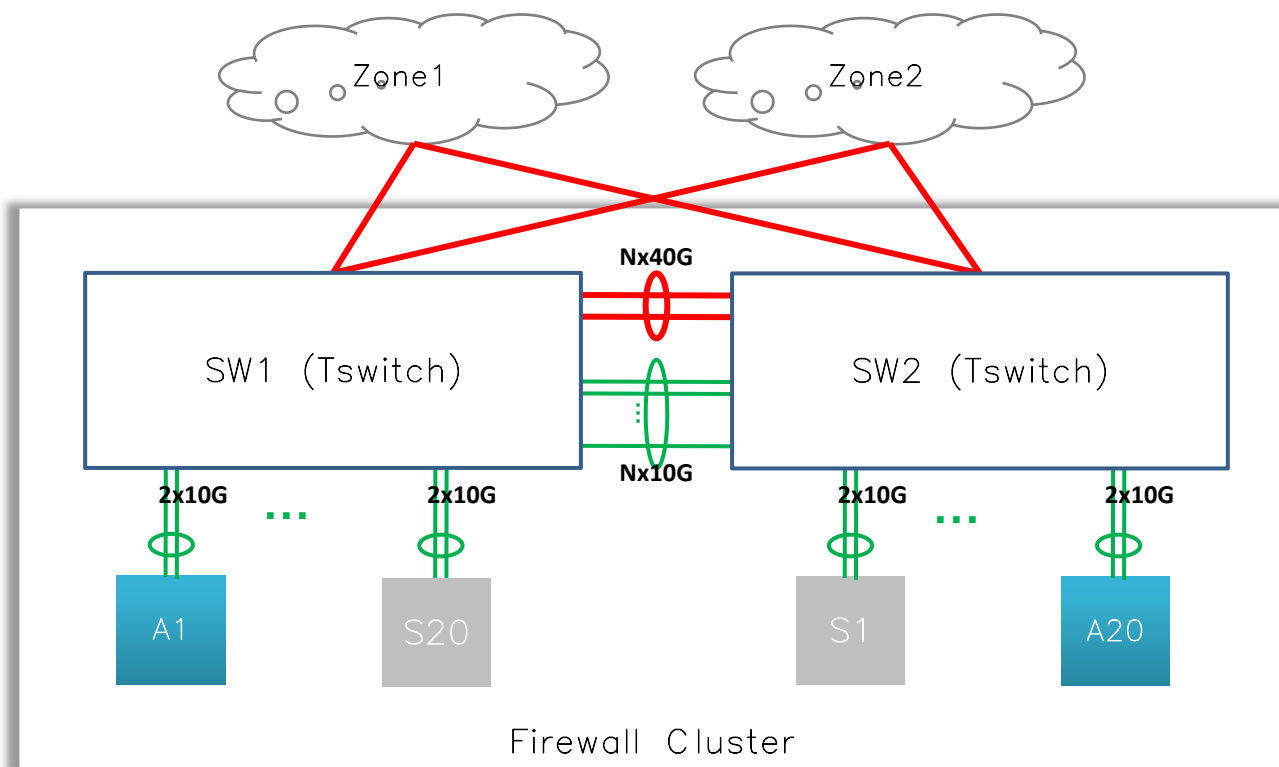
快速
迭代

- 互联网业务的多样性，以及业务在成长过程中的需求变化，都要求网络具备快速引入新特性、可编程和向前兼容的能力

高效
运维

- 从网络运维的角度看，要求网络具备端到端的全局的可视化的能力，在出现网络故障时，能够高效排查和快速定位，提升运维效率

防火墙集群的负载均衡



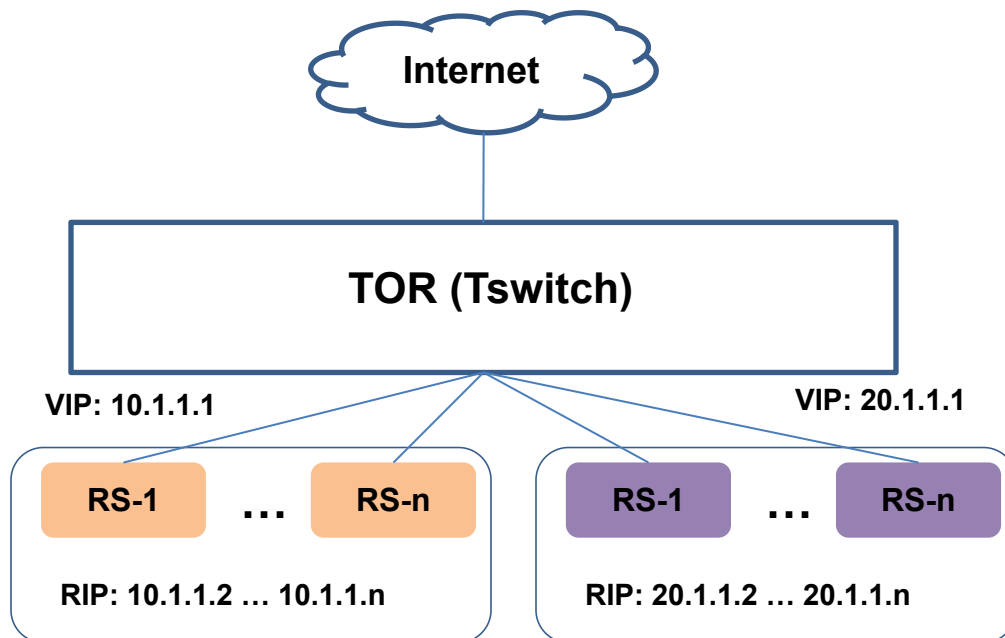
需求：

- ✓ 不同zone互访需要按需做安全访问控制
- ✓ 防火墙容量、性能的需根据实际应用横向扩展
- ✓ 同一会话的双向流量，并需保证同源同路

实现：

- ✓ 分流器交换机通过Hybrid（传统转发+Openflow转发）模式，实现和外部网络以及防火墙集群的通信
- ✓ 控制器预先下发流表（基于源、目的网段匹配）到负载均衡器进行流量调度。
- ✓ 交换机实时上报流统计信息到控制器，控制器通过相应算法调整优化交换机上的流表。

服务器集群的资源利用率



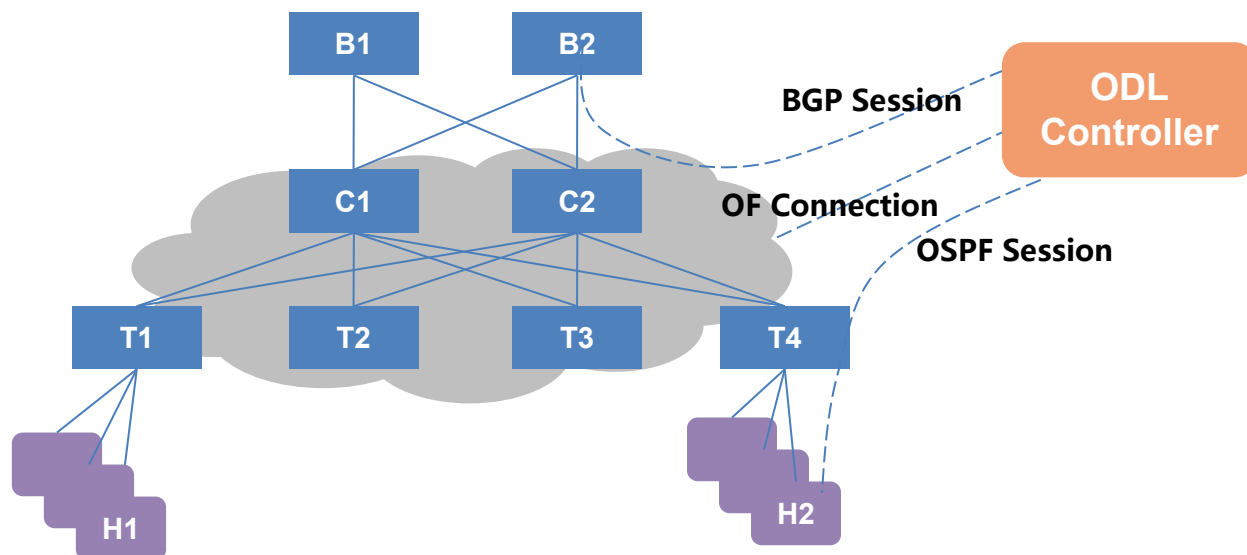
需求：

- ✓ 服务器以集群方式工作，若干服务器形成一个集群，每个集群对外呈现一个VIP
- ✓ 接入交换机需将流量负载均衡到多个服务器集群，提升服务器利用率
- ✓ 可视化的方式提供精确流统，服务器故障时自动检测和实时调度

方案：

- ✓ 服务器通过OSPF和接入交换机建立邻居，并通告VIP的路由
- ✓ 接入交换机通过OSPF路径计算，在转发平面利用ECMP实现负载均衡

CLOS架构的精细化运营



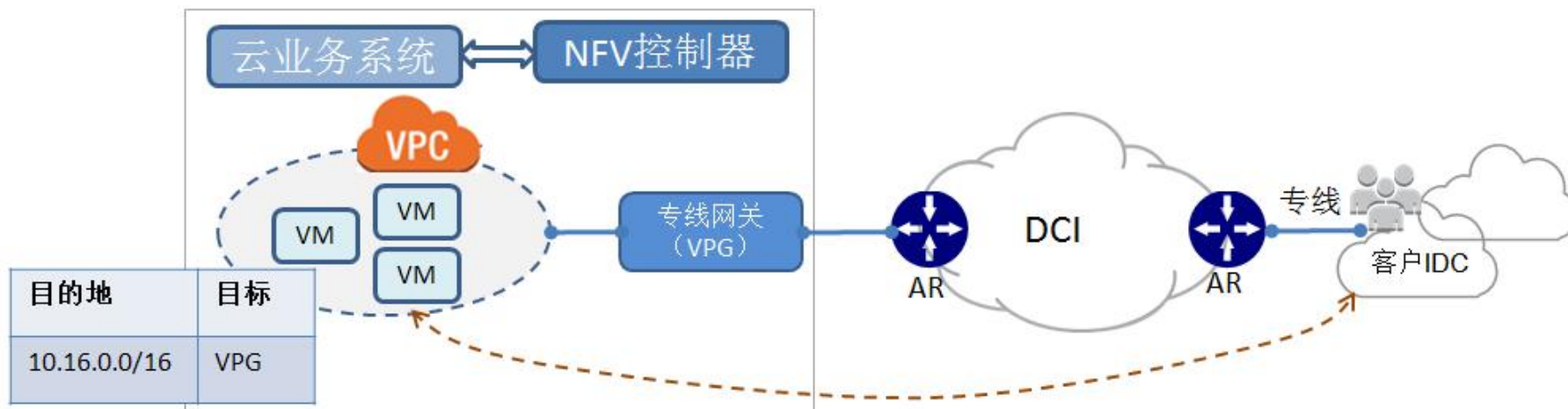
需求：

- ✓ 端到端流路径可视；
- ✓ 端到端丢包检测；
- ✓ 流的动态监控。
- ✓ 精准的流调度；
- ✓ 精细的业务流过滤、镜像、复制

方案：

- ✓ CLOS区域引入Openflow替代传统L2/L3转发，实现精细化流调度和监控
- ✓ 控制器通过LLDP协议计算网络拓扑，并和已知拓扑进行比较；交换机通过BFD进行快速链路检测
- ✓ 网络协议通过packt in/out和控制器交互，并转化为流表下发到CLOS区域；控制器同时支持ARP、DHCP代理

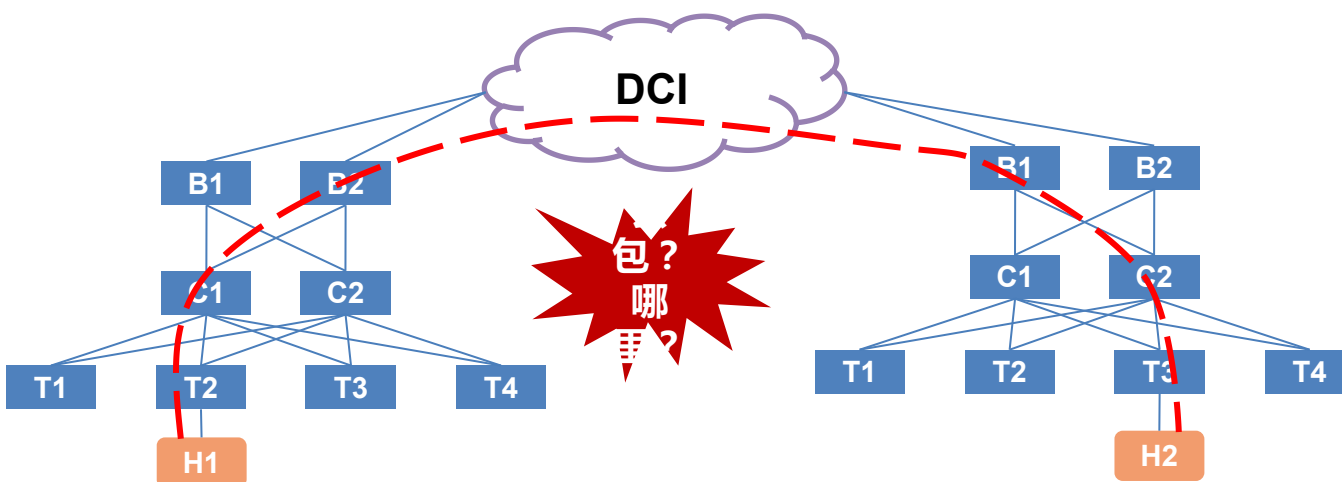
私有协议 + VPC网关集群的负载均衡



需求：

- ✓ VPC采用私有GRE协议实现overlay
- ✓ 专线网关VPG必须支持私有GRE协议才能与VPN对接，同时对专线网管的带宽要求很高

现网的故障定位



现有排障手段：

- ✓ Show、ping、Traceroute
- ✓ 通过ACL下发流统的配置信息
- ✓ SNMP采集接口收发包、CPU等信息

现有手段局限：

- ✓ 现有手段都是在建立在收集网络 设备的外围信息，网络设备对于运维同学来讲依然是黑盒子
- ✓ 当前的排障手段基本是单机调试，不具备端到端监控和检测的能力
- ✓ 出现故障时，网络设备关键组件的状态是否正常是运维人员迫切需要的

P4及可编程芯片可以使我們

- 精细化调度和运维，提升网络效率
- 快速响应业务变化，满足定制化需求
- 故障时快速定位，高效运维网络
- 共享开源软件社区将会极大丰富网络应用

End