深信服应用交付产品
技术白皮书

深信服科技股份有限公司

2016 年

地址：深圳市南山区学苑大道1001号南山智园A1栋
咨询热线：400-806-6868 | 服务热线：400-630-6430 | 邮箱：market@sangfor.com.cn | 网址：www.sangfor.com.cn

版权声明

免责条款

缩写和约定

英文缩写	英文全称	中文解释
ACL	Access Control List	访问控制列表
ADC	Application Delivery Controller	应用交付设备
BRAS	Broadband Remote Access Server	宽带接入服务器
DNAT	Destination NAT	目的地址 NAT
DNS	Domain Name Service	域名服务
DR	Direct Route	直达路由
FTP	File Transfer Protocol	文件传输协议
HA	High Availability	高可用性
HTTP	Hypertext Transfer Protocol	超文本传输协议
ICMP	Internet Control Message Protocol	因特网控制报文协议
ISP	Internet Service Provider	Internet 服务提供商
MAC	Media Access Control	介质访问控制
NAT	Network Address Translation	网络地址转换
OSPF	Open Shortest Path First	开放最短路径优先
RADIUS	Remote Authentication Dial In User Service	远程用户拨号认证系统
RIP	Routing Information Protocol	路由信息协议
RTT	Round Trip Time	往返时间
STP	Spanning Tree Protocol	生成树协议
SNAT	Source NAT	源地址 NAT
SNMP	Simple Network Management Protocol	简单网络管理协议
SOA	Service Oriented Architecture	面向服务架构
SSL	Secure Socket Layer	安全套接层
TCP	Transmission Control Protocol	传输控制协议
UDP	User Datagram Protocol	用户数据报协议
URI	Uniform Resource Identifier	统一资源标识符
URL	Uniform Resource Locator	统一资源定位符
VLAN	Virtual Local Area Network	虚拟局域网

目录

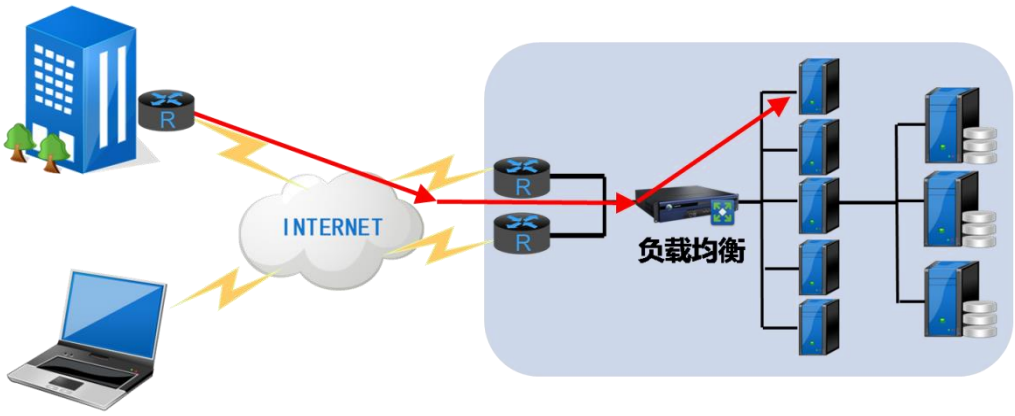
第1章 应用交付，后负载均衡时代的选择

1.1 背景概述

传统的负载均衡技术为用户提供了一种性价比不错的方法，通过扩展网络设备和服务器的带宽及吞吐量，进而提升网络数据处理的稳定性。而应用交付与负载均衡相比，在强调稳定性的基础上，增加了智能和优化方面的功能特性，以帮助用户应对来自于复杂应用环境中部署并交付服务的挑战。通过合理地部署应用交付设备，用户不仅能进一步改善其业务应用的性能与安全性，更可提高数据中心的基础设施效率，乃至应对未来部署虚拟化数据中心的趋势。

1.1.1 负载均衡专注的领域

服务器负载均衡原先被创造出来用以解决网络方面的问题，例如将访问请求在一组负责特定 Web 应用交付的服务器之间进行分配。最初是通过简单的 DNS 轮询来实现的，但是这种方法有其局限性。因此市场上出现了特定功能的负载均衡设备，通过分析入站的应用请求，将这些请求动态地映射到可用的服务器之上。



为了应对用户日趋复杂的访问需求，负载均衡方面的革新一直在持续，最初是集中于入站方面的问题，例如动态识别服务器的工作负荷和故障，以及确保用户访问操作不会丢失的会话保持功能。然而，市场很快又发生了演变，并开始着眼于其他问题，诸如应用和服务器的效率方面，最佳的例子便是 SSL 卸载技术的采用。

最后，关注点转移到了出站流量方面。市场上不断涌现出出一系列的技术和功能，用以

改善通过网络交付业务应用的效率。创新点也从专注于基础设施效率的单纯网络技术，转移到了业务应用的性能优化和安全方面，而负载均衡也开始从一个单纯的网络产品发展应用交付这样一个触及网络、服务器、业务应用乃至安全方面的全方位概念。

1.1.2 应用交付概念的诞生

先进的应用交付产品（ADC）能帮助用户缓解来自于当今复杂应用环境部署和交付的挑战。过去的十年，伴随着企业级应用以业务流程和用户生产力为目标，向基于浏览器模式的大量迁移，同时也见证了面向服务架构（SOA）、Web2.0 和现今云计算模型被广泛采用。

基于改善业务应用环境的理念，应用交付产品提供了一系列功能以应付复杂的网络环境。在业务持久性、终端用户访问体验、数据中心可用性等多个应用领域，均可通过部署应用交付产品予以优化。此外，应用交付产品能够帮助减少所需部署的服务器数量，提供实时的控制机制助力于数据中心虚拟化，降低数据中心的供电和冷却方面的要求，使用户的业务更好地顺应紧凑、节能、环保的趋势。

1.2 快速、智能的应用交付网络

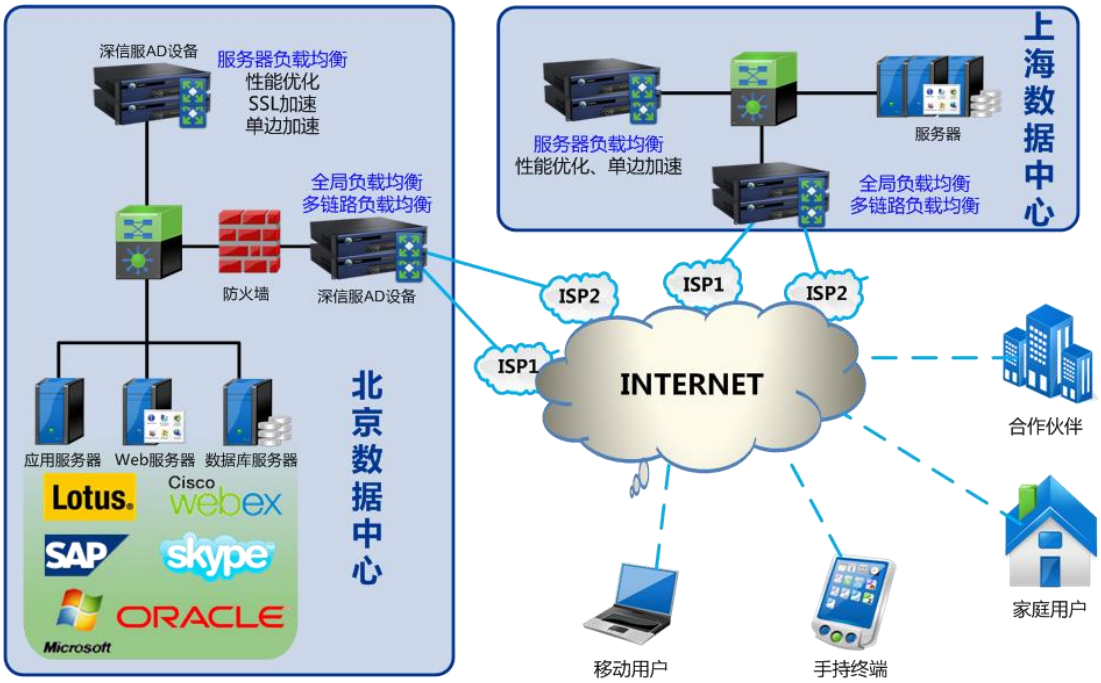
虽然应用交付产品能提供如此丰富的优化功能，可以改善用户的业务应用性能与安全性，但是国内用户在实际使用中还是以传统负载均衡的方面居多。然而在中国乃至整个亚洲市场，在线上业务和电子商务等方面都逐渐涌现出更复杂的应用交付需求，这对发布业务应用的服务商而言也意味着将面临更多的挑战。与欧美发达国家相比，国内在宽带普及率和线路质量等方面都还存在一定的差距，因此用户访问体验还存在很大的优化空间。换言之，应用交付层面的种种性能优化特性，在这种环境中会愈加地凸显出来，甚至被一定程度地放大。

应用交付技术发展至今已经十分成熟，各厂商的产品之间基本上都是大同小异，只是侧重点有所不同；有的走得是集中方案，有的强调整合，有的偏重于安全；而深信服致力于提供快速、智能的应用交付解决方案，在众多功能细节中都体现了深信服对于不同地域、不同用户群市场需求的深刻理解，旨在让客户能获得超出业界同类产品的投资回报，在改善用户访问体验的同时，提升自身的业务竞争力。

第2章 深信服应用交付解决方案

2.1 方案概述

深信服 AD 产品作为专业的应用交付设备，能够为用户的应用发布提供包括多数据中心负载均衡、多链路负载均衡、服务器负载均衡的全方位解决方案。配合性能优化、单边加速以及多重智能管理等技术，实现对各个数据中心、链路以及服务器状态的实时监控，同时根据预设规则将用户的访问请求分配给相应的数据中心、链路以及服务器，进而实现数据流的合理分配，使所有的数据中心、链路和服务器都得到充分的利用。不仅扩展应用系统的整体处理能力，提高其稳定性，更可切实改善用户的访问体验，降低组织的 IT 投资成本。



2.2 概念介绍

- **负载均衡算法** - 深信服 AD 设备根据设定的策略机制将来自业务访问的数据流量调度到不同的真实服务，例如对应服务器负载均衡中的服务器、链路负载均衡中的链路、以及全局负载均衡中的站点，用以调度的策略机制被通称为负载均衡算法。
- **健康检查** - 深信服 AD 设备对服务器以及链路进行主动探测，依据不同的健康性检测方法，可以判断服务器或链路的健康状况是否能正常提供服务。

- **就近性** - 在多链路负载均衡的场景中，深信服 AD 设备通过访问端的地理位置，链路隶属的运营商，以及实时探测的链路健康状况等因素进行分析比对，并根据判断结果选择最优链路，保证数据流量经由最优链路来传输。
- **虚拟服务** - 深信服 AD 设备对外发布的服务被称为虚拟服务。虚拟服务包含了服务类型（协议）、节点池（一台或多台服务器的 IP 地址和端口的集合）等配置属性。客户的访问请求通过网络到达深信服 AD 设备，并匹配到虚拟服务后，再由 AD 设备按照设定的负载均衡策略调度到真实服务器。
- **会话保持** - 深信服 AD 设备提供有一种称为会话保持（Session Persistence）的机制，可以识别客户与服务器之间交互过程的关连性，在实现负载均衡的同时，还可保证一系列相关连的访问请求会保持分配到同一台服务器之上。在这种特定情况下，AD 设备将放弃原有的负载均衡算法。
- **IP 地址库** - 深信服 AD 设备内置的 IP 地址库收录了不同运营商拥有的 ISP 地址段信息，以便链路负载均衡可以基于报文的源或目的 IP 地址查找 IP 地址库，并得到对应的运营商信息，再根据运营商信息为访问流量选择一条合适的 ISP 链路。

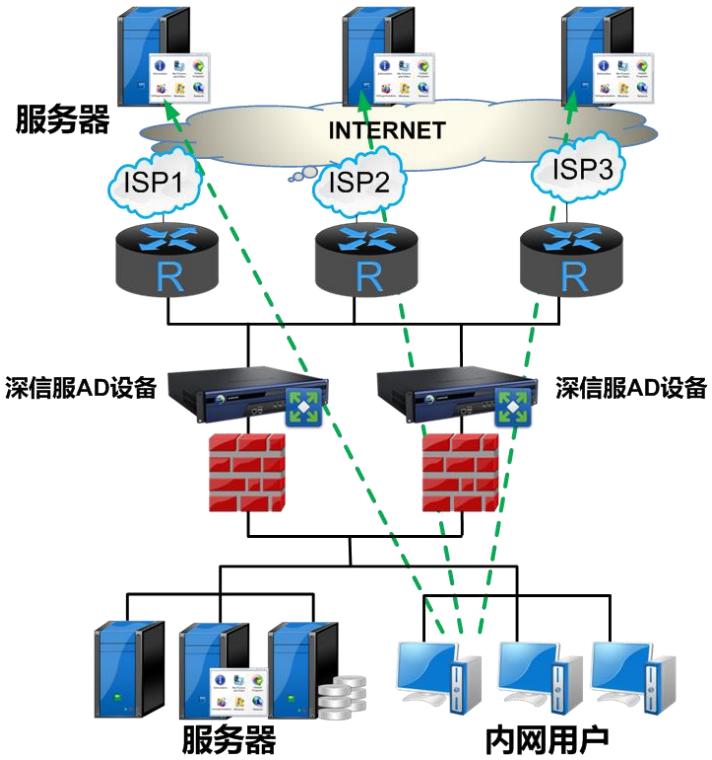
2.3 多链路负载均衡

深信服 AD 应用交付设备集合出入站智能 DNS 解析、轮询、加权轮询、静态就近性、动态就近性等算法，解决多链路网络环境中流量分担的问题，充分提高多链路的带宽利用率，节约企事业单位对通信链路的投资；并且通过为用户分配最佳的通信线路，使用户获得绝佳的访问体验。此外，深信服 AD 应用交付设备还利用链路健康检查及会话保持技术，实现了在某条链路中断的情况下任然可以提供访问链接能力，充分利用了多条链路带来的可靠性保障，使对于用户的访问达到了最全面的支持。

2.3.1 出站流量负载均衡

内网的用户访问互联网访问资源时，深信服 AD 接收到用户的访问流量后，通过预先设定链路负载策略将用户访问流量分配到不同的互联网链路之上，实现出站流量负载均衡，提

升互联网链路带宽利用率。

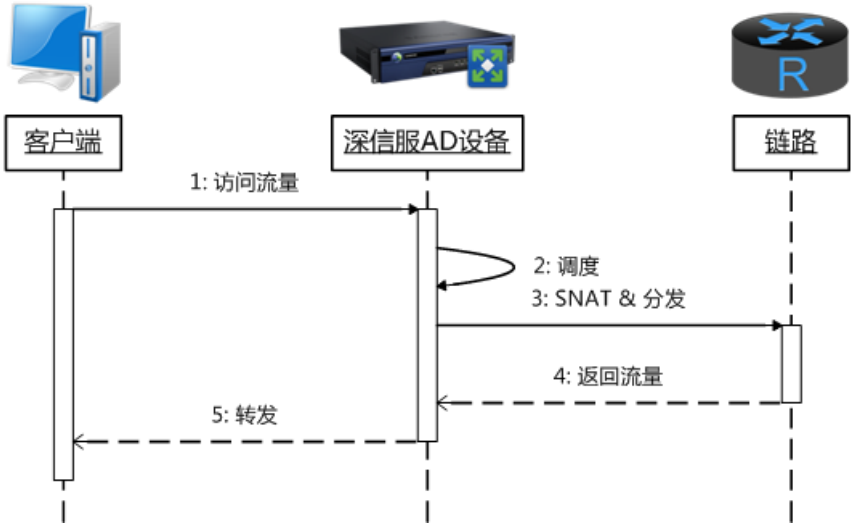


1. 实现方式

深信服 AD 接收到内网用户访问的流量以后，根据预先设定负载策略将访问电信的资源的出站流量分配到电信的链路之上，并做源地址的 NAT，(指定某一合法 IP 地址进行源地址的 NAT，或者用 AD 设备的接口地址自动映射)，保证数据包返回时能够正确接收；同理，其它的访问的流量会通过相应策略会被分配到其它的运营商链路之上。

2. 工作流程

出站流量负载均衡的工作流程如下图所示

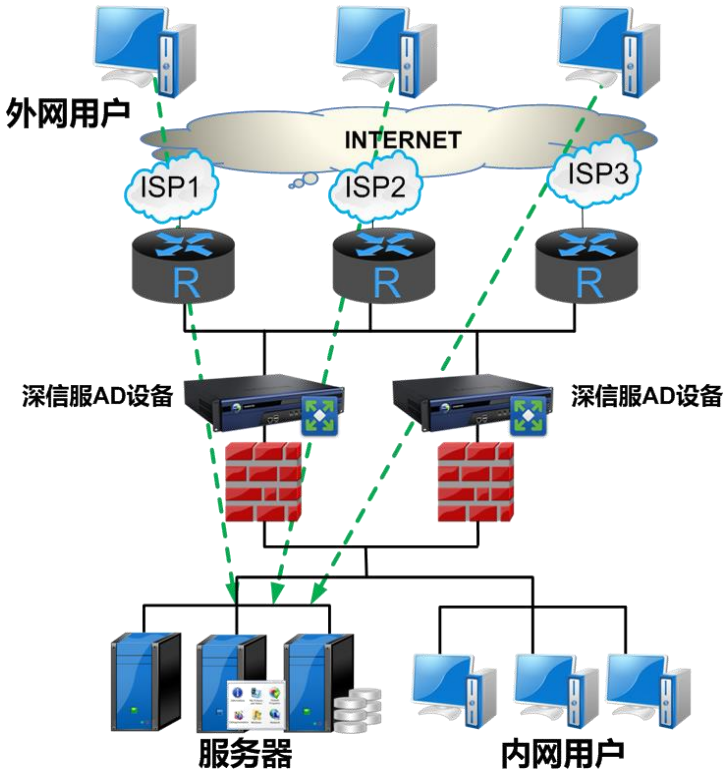


出站流量负载均衡的流程描述如下

步骤	说明
1	深信服 AD 设备收到来自内网用户的访问流量
2	深信服 AD 设备根据预先定义的负载均衡策略来选择出合适的出站链路
3	深信服 AD 设备按照链路选择的结果将流量分配给选定的出站链路，并做源地址的 NAT
4	深信服 AD 设备收到从外网返回的应答流量
5	深信服 AD 设备将流量转发给内网用户

2.3.2 入站流量负载均衡

当外部用户访问内部资源时，深信服 AD 通过智能 DNS 解析技术将一个域名绑定多个运营商的公网地址，负责解析来自不同运营商用户的域名解析请求；深信服 AD 根据不同负载均衡策略为不同运营商的用户返回最佳的访问地址，实现用户入站流量的负载均衡。



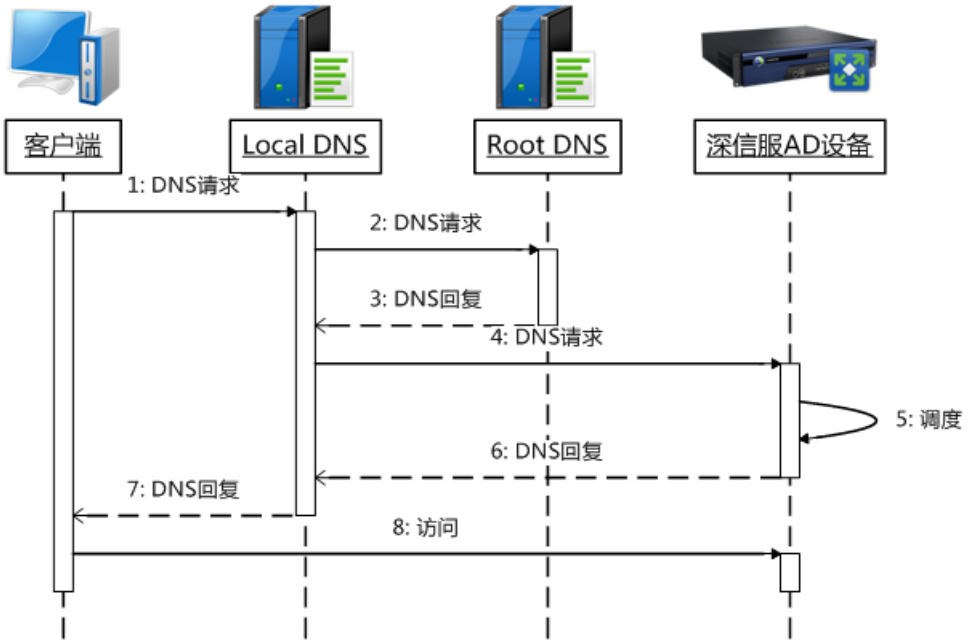
1. 实现方式

通过在域名注册提供商处修改域名 NS 记录，深信服 AD 设备获得域名解析权，深信服实现一个域名绑定多个运营商的公网地址，负责解析来自多个运营商用户的域名解析请求。

根据实现设定负载策略可以实现，如电信的用户通过电信的线路访问内部资源，联通的用户通过联通的线路访问内部资源；深信服 AD 还可以通过两条链路做反向查询，根据 RTT 时间判断链路的好坏，并且综合以上两个参数返回相应的 IP 地址。

2. 工作流程

入站流量负载均衡的工作流程如下图所示



入站流量负载均衡的流程描述如下

步骤	说明
1	外网用户的访问客户端向其本地 DNS 服务器发出域名解析请求
2	本地 DNS 服务器首先在本地搜索是否有相应的记录，如果没有就向根 DNS 服务器发起查询
3	根 DNS 服务器反馈本地 DNS 服务器，告知域名解析权授予深信服 AD 设备
4	本地 DNS 服务器会再次向深信服 AD 设备发出域名解析请求
5	深信服 AD 设备先判断链路的健康状况，再根据预先定义的负载均衡算法来选择出合适的 IP 地址作为域名解析结果
6	深信服 AD 设备将域名解析结果反馈给本地 DNS 服务器
7	本地 DNS 服务器将得到的域名解析结果转发给客户端
8	客户端根据得到的 IP 地址发起连接请求，对内网服务器进行访问

2.3.3 链路负载算法

深信服 AD 设备支持如下算法，管理员能够根据自身需求选择相应的链路分配策略，支持更多个性化的链路使用规则。

1. 轮询（Round Robin）

实现机制 - 将所有网络链路放在一个队列当中，按顺序依次返回给用户队列中下一个网

络链路的 IP 地址。

适用场景 - 拥有多条同一运营商的互联网链路，各条链路的带宽也相近。

2. **加权轮询（Weighted Round Robin）**

实现机制 - 由于各条互联网链路的吞吐量可能不一 ,因此可以为各条链路分配不同的加权值。根据这个比例，把数据用户请求轮询分配到每条链路。

适用场景 - 拥有多条同一运营商的互联网链路，但各条链路的带宽存在差异。

3. **加权最少连接（Weighted Least Connection）**

实现机制 - 根据事先为各条链路设定的权值 ,在调度新连接时尽可能地使各链路的已建立连接数和其权值成比例，把新的连接请求分配到当前比例最小的链路上。

适用场景 - 各条链路的带宽存在差异，并且不同用户发起的连接保存时长差异较大。

4. **加权最小流量（Weighted Least Traffic）**

实现机制 - 根据事先为各条链路设定的权值 ,在调度新连接时尽可能的使各条链路的实时流量与权值成比例，把新的连接请求分配到当前比例最小的链路上。

适用场景 - 拥有多条互联网链路，并且各链路之间的带宽差异较大。

5. **静态就近性（Static Proximity）**

实现机制 - 按照预先为某个目标定义的静态最佳路径来选择链路 ,或者根据设备内置的全球 IP 地址库来判断目标 IP 属于哪个互联网运营商，进而选择相应的 ISP 链路。

适用场景 - 拥有多条不同运营商的互联网链路，业务流量多为入站访问流量。

6. **动态就近性（Dynamic Proximity）**

实现机制 - 在选择链路时 ,通过综合考虑数据传输的延迟和链路的实时负载 ,准确计算出最佳路径。

适用场景 - 拥有多条不同运营商的互联网链路，业务流量多为出站访问流量。

7. **带宽比例（Bandwidth Ratio）**

实现机制 - 由于各条互联网链路的吞吐量可能不一 ,因此将为各条链路的带宽大小作为权值；根据这个比例（每条链路带宽大小的比值），把数据流量分配到每条链路上。

适用场景 - 拥有多条同一运营商的互联网链路，但链路之间的带宽差异较大。

8. **哈希（Hashing）**

实现机制 - 基于 LOCAL DNS IP 地址的哈希算法，将不同的用户访问调度到不同的链路之上。

适用场景 - 拥有多条互联网链路，需要保证来自同一个用户的请求分发到同一条链路。

9. **主备（Primary / Secondary）**

实现机制 - 即可以为网络设定主备链路，当主链路出现故障时，用户的访问请求才会被调度备用链路之上

适用场景 - 拥有多条互联网链路，对业务访问的持久性要求较高。

10. **首个有效（First Available）**

实现机制 - 即将用户的请求全部都调度第一条有效无故障的链路之上

适用场景 - 拥有多条互联网链路，对业务访问的响应时延比较敏感。

2.3.4 链路健康检查

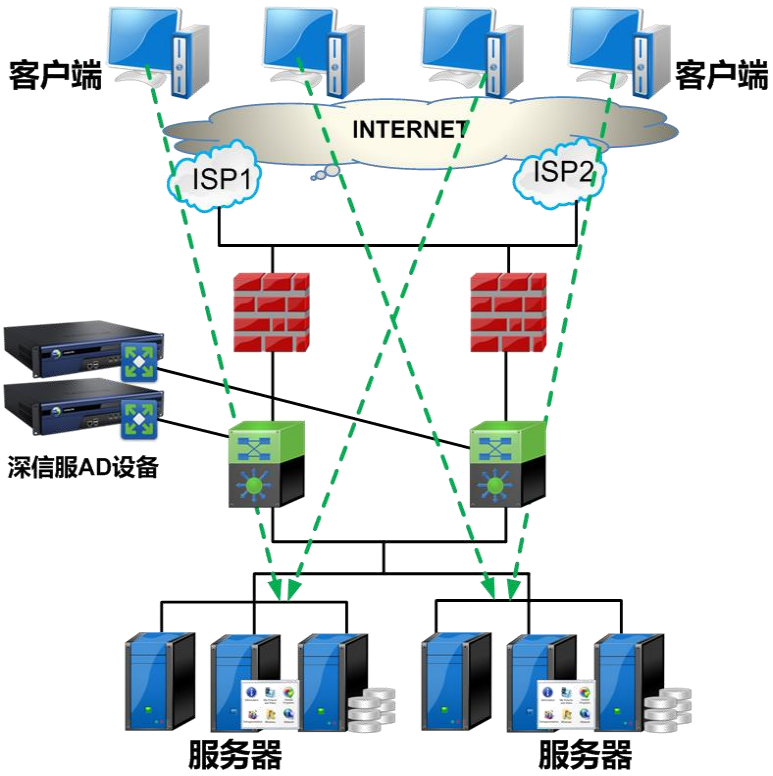
深信服 AD 通过多个 Internet 站点的可达性，来共同判断一条链路的状况。例如，通过电信线路检查 www.sina.com.cn、www.sohu.com、以及 www.qq.com 的 TCP 80 端口，并对检查结果做“或”运算。如此，只要其中一个站点可达，即可表明链路状态良好。该方法即避免了 ICMP 检查的局限性，也避免了单一站点检查带来的单点失误。

2.4 服务器负载均衡

运用多台服务器集群的机制，深信服 AD 应用交付设备能将所有真实服务器配置成虚拟服务来实现负载均衡，对外直接发布一个虚拟服务 IP。当用户请求到达应用交付设备的时候，根据预先设定的基于多重四、七层负载均衡算法的调度策略，能够合理的将每个连接快速的分配到相应的服务器，从而合理利用服务器资源。不仅在减少硬件投资成本情况下解决单台服务器性能瓶颈，同时方便后续扩容，为大并发访问量的系统提供性能保障。

通过对服务器健康状况的全面监控，深信服 AD 应用交付设备能实时地发现故障服务器，并及时将用户的访问请求切换到其他正常服务器之上，实现多台服务器之间冗余。从而保证关键应用系统的稳定性，不会由于某台服务器故障，造成应用系统的局部访问中断。

近年来随着 IPv4 地址的逐渐枯竭，以及 IPv6 对安全性和可靠性的增强，很多用户的网络正在逐渐向 IPv6 过渡。为了适应这种趋势，深信服 AD 应用交付设备不仅能对 IPv4 协议的应用系统进行负载均衡，也同样支持基于 IPv6 协议的 L4/L7 服务器负载均衡，以实现用户对 IPv6 服务的发布。



2.4.1 NAT 方式 L4 服务器负载均衡

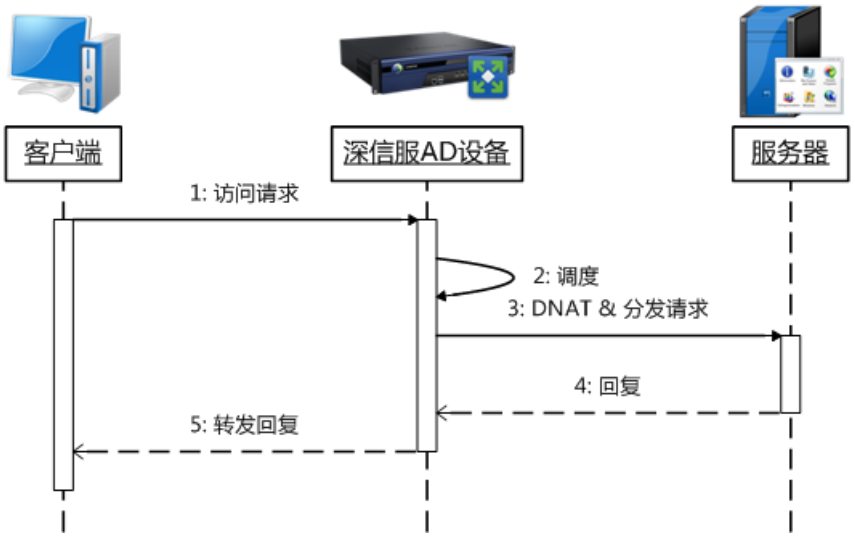
1. 实现方式

深信服 AD 设备支持基于 IP 地址、应用类型和内容等因素实现流量负载。通过这种方式管理员可以为不同类型的应用类型分配不同的服务器资源。应用类型调度支持基于不同协议上的多种应用，包括 TCP、UDP、IP、DNS、E-mail、FTP、HTTP、RADIUS 等等。在实现 L4 服务器负载均衡的场景中，AD 设备负责将客户端的请求转发给服务器，然后客户端与服务器之间建立 TCP 连接，此时 AD 设备所扮演的角色类似于一台路由器。在 NAT 方

式下，AD 设备调度访问请求时，先进行目的 IP 地址转换，再将访问请求转发给后端的每台服务器。

2. 工作流程

NAT 方式 L4 服务器负载均衡的工作流程如下图所示



NAT 方式 L4 服务器负载均衡的流程描述如下

步骤	说明
1	客户端发送服务访问请求，此时的源 IP 为客户端 IP、目的 IP 为虚拟服务 IP
2	深信服 AD 设备接收到访问请求后，根据预先定义的负载均衡调度算法判断出应该将访问请求分发给哪台服务器
3	深信服 AD 设备使用 DNAT 技术分发访问数据，此时的源 IP 为客户端 IP、目的 IP 为服务器 IP
4	服务器处理接收到的访问请求，并回复响应数据，此时的源 IP 为服务器 IP、目的 IP 为客户端 IP
5	深信服 AD 设备接收响应数据，转换源 IP 后再转发给客户端，此时的源 IP 为虚拟服务 IP、目的 IP 为客户端 IP

2.4.2 DR 方式 L4 服务器负载均衡

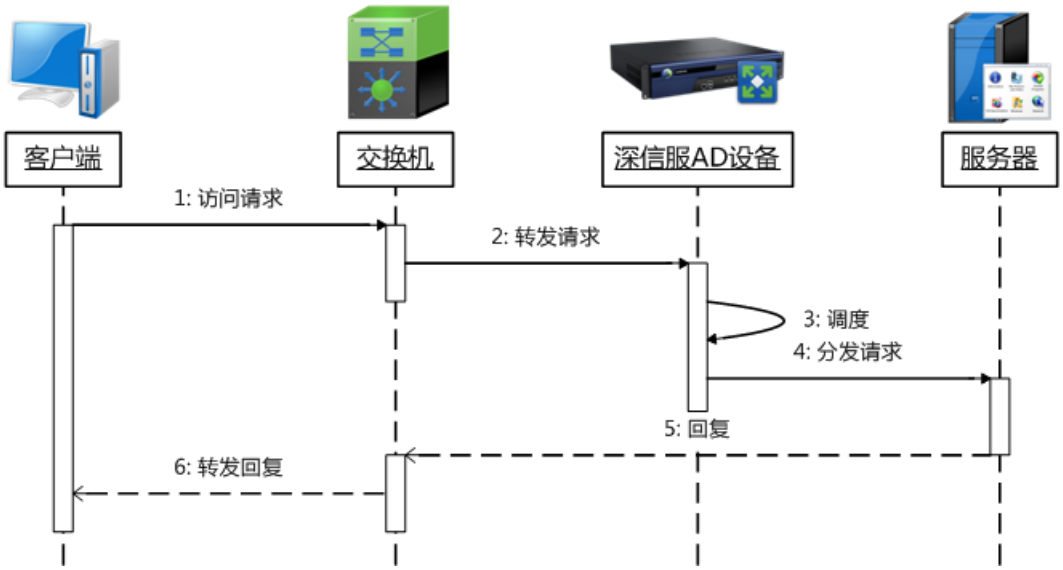
1. 实现方式

DR 方式又称为三角传输模式，在该模式下实现服务器负载均衡，需要后台服务器在环回接口上配置虚拟服务 IP。AD 设备在分发访问请求时不改变目的 IP 地址，而是将目的 MAC

替换为服务器的 MAC 后，再将访问请求转发给后端的服务器，服务器的回复会直接通过交换机返回给用户。由于服务器的响应回复无需经过 AD 设备转发，从而减少了 AD 设备的吞吐压力，有效的避免了整个业务系统的性能瓶颈。

2. 工作流程

DR 方式 L4 服务器负载均衡的工作流程如下图所示



DR 方式 L4 服务器负载均衡的流程描述如下

步骤	说明
1	客户端发送服务访问请求，此时的源 IP 为客户端 IP、目的 IP 为虚拟服务 IP
2	交换机将访问请求转发给深信服 AD 设备
3	深信服 AD 设备接收到访问请求后，根据预先定义的负载均衡调度算法判断出应该将访问请求分发给哪台服务器
4	深信服 AD 设备向服务器分发访问数据，此时的源 IP 为客户端 IP、目的 IP 为虚拟服务 IP，目的 MAC 为服务器 MAC
5	服务器处理接收到的访问请求，并回复响应数据，此时的源 IP 为虚拟服务 IP、目的 IP 为客户端 IP
6	交换机接收响应数据，直接转发给客户端

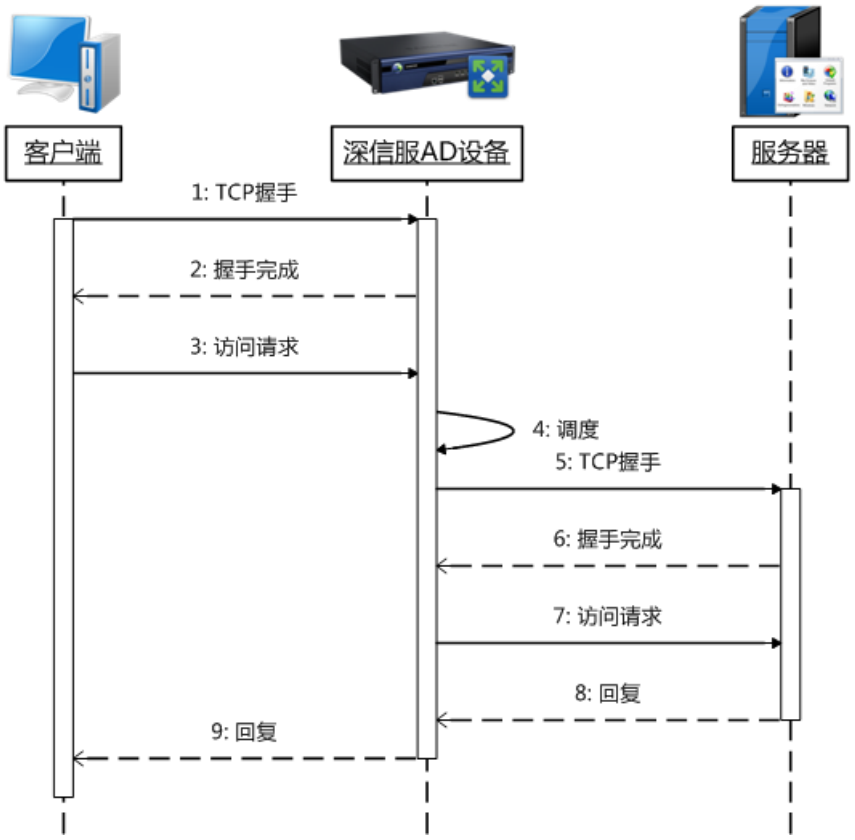
2.4.3 L7 服务器负载均衡

1. 实现方式

基于七层内容的调度机制，使得管理员可以通过应用层的内容交换来分配服务器资源，以实现用户请求调度的多元化和个性化，业务应用的场景十分广泛。例如，基于 URI、HOST、COOKIE、USER_AGENT 等 HTTP 头部内容的匹配策略来选择服务器，或者通过对 HTTP 头部进行请求改写和应答改写，执行页面跳转和丢弃等操作，实现不同业务系统之间的交互联动。在 L7 服务器负载均衡的场景中，AD 设备先与客户端建立 TCP 连接，获取到访问请求的报文之后，再根据报文中所带的应用层内容来选择合适的服务器并与之建立 TCP 连接，此时 AD 设备所扮演的角色类似于一台代理服务器。

2. 工作流程

L7 服务器负载均衡的工作流程如下图所示



L7 服务器负载均衡的流程描述如下

步骤	说明

1	客户端向深信服 AD 设备发起 TCP 连接请求，此时的源 IP 为客户端 IP、目的 IP 为虚拟服务 IP
2	客户端与深信服 AD 设备之间建立 TCP 连接
3	客户端发送服务访问请求，此时的源 IP 为客户端 IP、目的 IP 为虚拟服务 IP
4	深信服 AD 设备收到访问请求后，匹配虚拟服务调度策略，并负载均衡调度算法判断出应该将访问请求分发给哪台服务器，同时缓存该访问请求的数据
5	深信服 AD 设备向服务器发起 TCP 连接请求 ,其 SYN 序列号为客户端的 SYN 序列号，此时的源 IP 为客户端 IP、目的 IP 为服务器 IP
6	深信服 AD 设备与服务器之间建立 TCP 连接
7	深信服 AD 设备修改缓存的访问请求数据中的目的 IP 和 TCP 序列号，然后发送给服务器
8	服务器处理接收到的访问请求，并回复响应数据，此时的源 IP 为服务器 IP、目的 IP 为客户端 IP
9	深信服 AD 设备修改响应数据的源 IP 和 TCP 序列号，然后转发给客户端，此时的源 IP 为虚拟服务 IP、目的 IP 为客户端 IP

2.4.4 服务器负载算法

深信服 AD 设备支持多重负载均衡算法将所有流量均衡的分配到各个服务器，不仅充分利用所有的服务器资源，而且各个服务器均衡的承担流量处理任务，从而有效地避免服务器处理任务“不平衡”现象的发生。

1. **轮询（Round Robin）**

   实现机制 - 按照请求的先后顺序将用户请求循环地分配到每台服务器。一旦某台服务器出现故障将不在为其分配任务，直至服务器恢复正常。

   适用场景 - 服务器集群中各台服务器的性能相当。

2. **加权轮询（Weighted Round Robin）**

   实现机制 - 由于集群中混用了不同规格服务器，因此可以针对各个服务器的处理性能来分配不同的加权值。根据这个比例，把用户的请求分配到每个服务器。

   适用场景 - 服务器集群中各台服务器的性能差异较大。

3. **加权最少连接（Weighted Least Connection）**

   实现机制 - 根据事先为各服务器设定的权值，在调度新连接时尽可能的使服务器的已建

立连接数和其权值成比例，AD 把新的连接请求分配到当前比例最小的服务器上。

适用场景 - 各台服务器的性能存在差异，并且不同用户发起的连接保存时长差异较大。

4. **最快响应（Fast Response）**

实现机制 - 按照响应时间大小对服务器重新分配权值，响应时间小的服务器权值大，响应时间大的服务器权值小，从而响应时间小的服务器获得更多的连接请求，但又照顾一部分响应时间大的服务器，避免负载倾斜。

适用场景 - 服务器集群的拓扑分散，访问用户需要就近选择服务器。

5. **动态反馈（Dynamic Feedback）**

实现机制 - 通过检测服务器的 CPU、I/O、Memory 等影响业务的多个因素来确定负载均衡的标准，进而动态调整各台服务器的权值，在调度新连接时选择最合适的服务器。

适用场景 - 服务器集群中每台服务器的处理能力、访问业务存在差异，并且不易确定服务器之间的权值比。

6. **哈希（Hashing）**

实现机制 - 基于 URI、 HOST、SRC_IP、IP + PORT 的哈希算法，将包含不同元素的用户访问尽可能地平均调度到服务器集群中的各台服务器上。

适用场景 - 需要将包含相同元素的业务访问调度到同一台服务器。

7. **优先级（Priority）**

实现机制 - 将服务器按优先级分组，优先调度优先级高的，只有优先级高的服务器发生故障，才调度优先级低的服务器。

适用场景 - 服务器集群中各台服务器的性能、稳定性等因素存在差异。

8. **强行负载（UDP）**

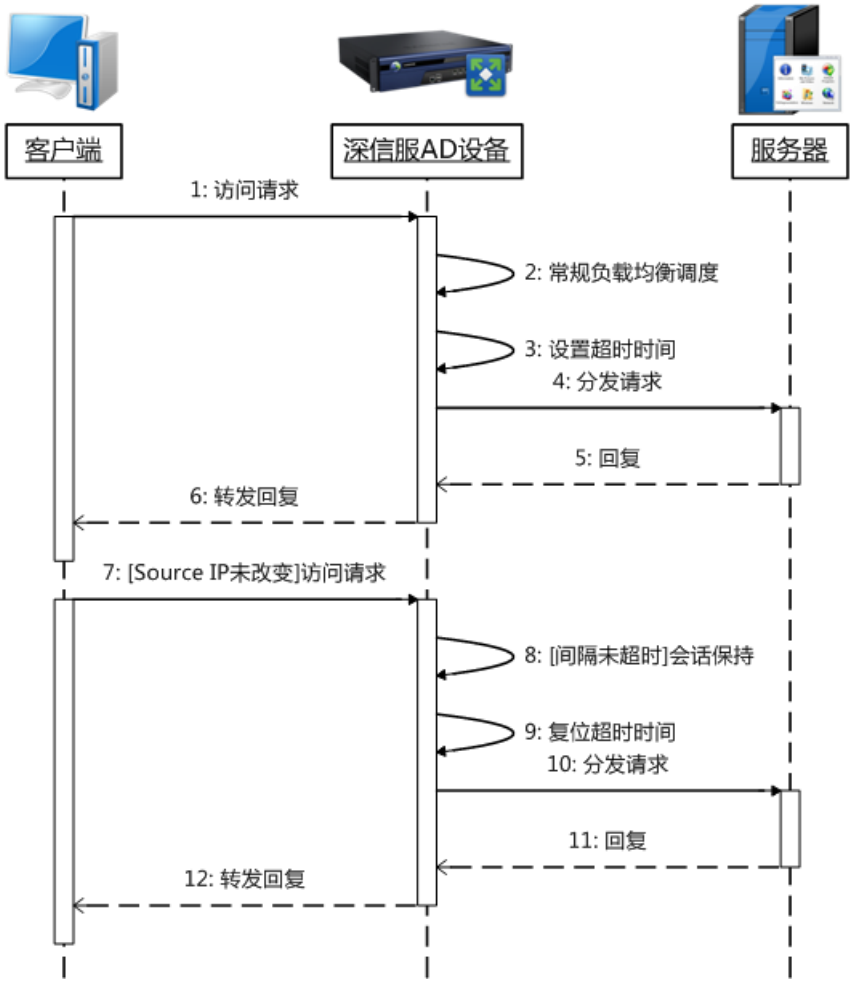
实现机制 - 对每个请求包做负载均衡，使得每台服务器收到的请求数量相当。

适用场景 - 诸如 RADIUS、DNS 等 UDP 协议的应用,用户发起的会话是一个请求包加一个应答包完成。

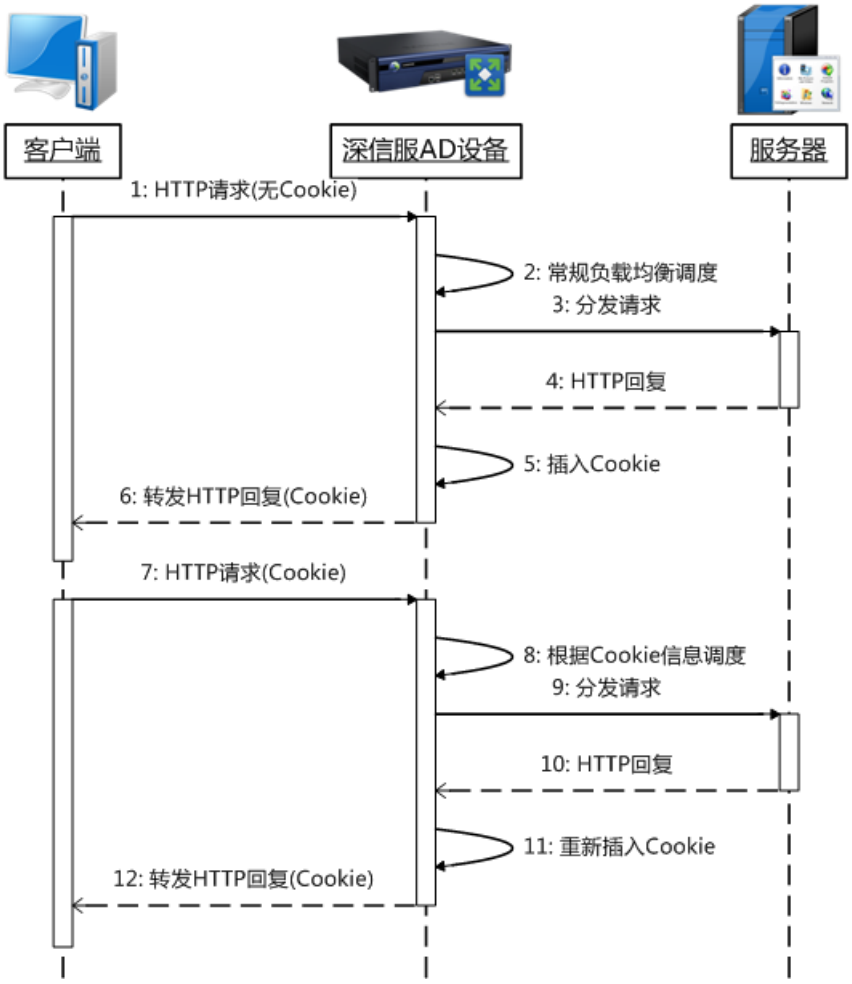
2.4.5 会话保持机制

通过深信服 AD 设备的会话保持技术，可以为访问用户选择曾连接上的特定服务器，实现无缝地处理用户请求；另一方面可以减少新建连接的数量，有助于减小负载均衡设备的系统开销。

- **基于 Source IP 的会话保持机制** - 也被称为基于简单会话保持，是指深信服 AD 设备在作负载均衡时是根据访问请求的源地址作为判断关联会话的依据，对来自同一 IP 地址的所有访问请求在做负载均时都会被保持到一台服务器上去。另外一个很重要的参数就是连接超时值，AD 设备会为每一个进行会话保持的会话设定一个时间值，从一个会话上一次完成到这个会话下次再来之前的间隔如果小于这个超时值，AD 设备会将新的连接进行会话保持，但如果这个间隔大于该超时值，AD 会将新来的连接认为是新的会话然后进行负载均衡调度，其原理如下图所示：

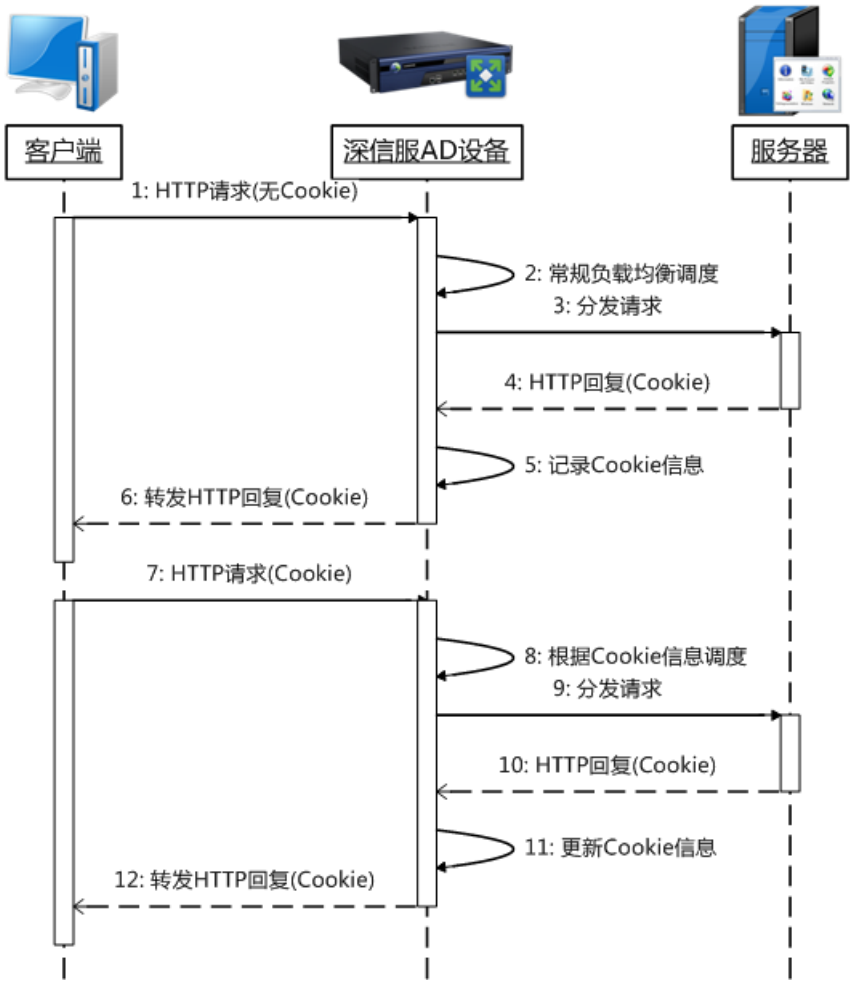


- **基于 Cookie Insert 的会话保持机制 -** 也被称为插入式 Cookie 会话保持，其原理是利用 Cookie 持续性，通过在客户端存储的 Cookie 信息来作为访问请求的调度依据。对于用户发起的 HTTP 请求，深信服 AD 设备会在转发服务器回复时插入用于识别的特殊 Cookie，待用户下次访问时便可将带有相同 Cookie 信息的请求始终调度到同一台服务器，以实现会话保持的效果，其原理如下图所示：



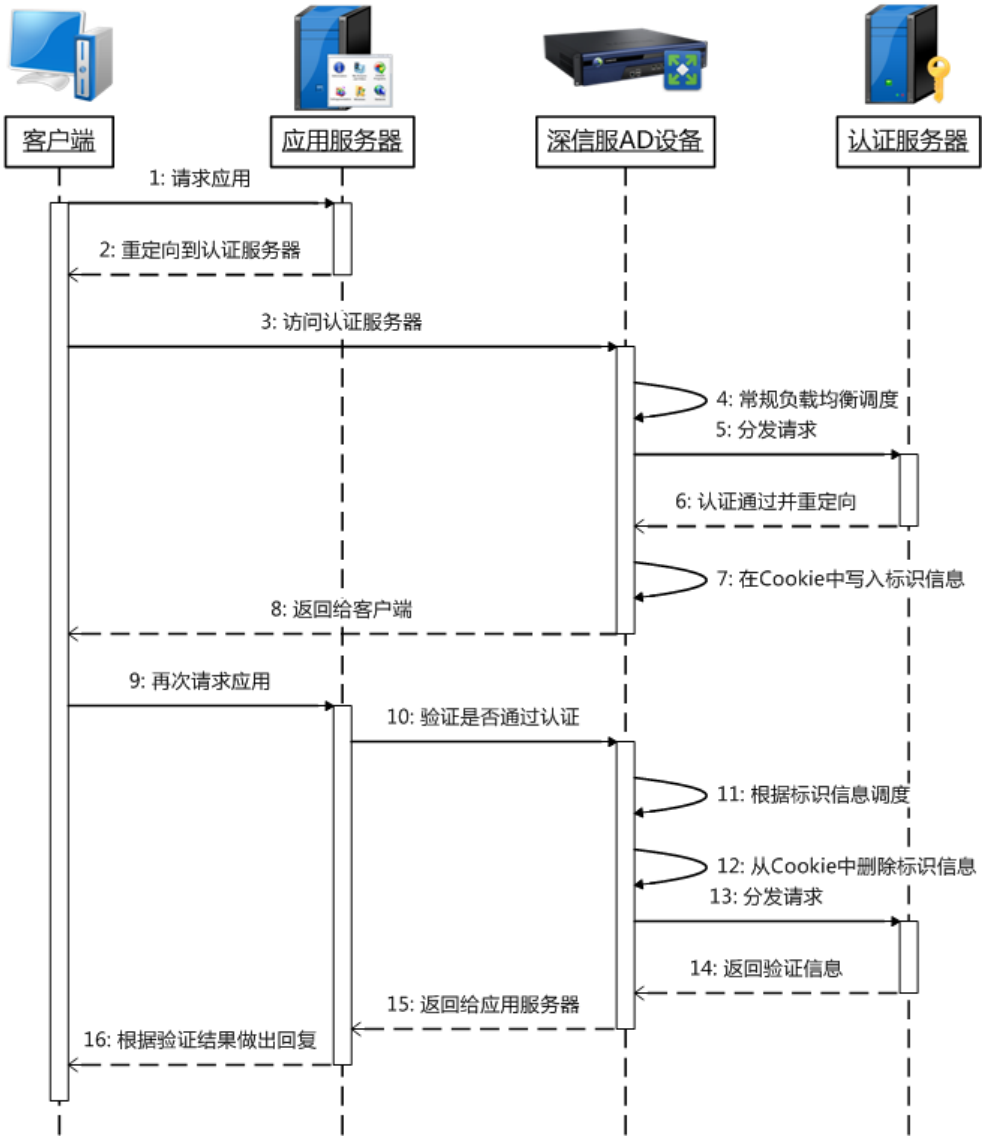
- **基于 Cookie Passive 的会话保持机制 -** 也被称为被动式 Cookie 会话保持，实现机制与 Cookie Insert 类似 ,都是通过客户端存储的 Cookie 信息来作为访问请求的调度依据；不同的地方在于，对用户发起的 HTTP 请求，会由后台服务器在回复时就写入相应的 Cookie，而深信服 AD 设备在转发回复时则记录该 Cookie 信息，待用户下次访问时便可将带有相同 Cookie 信息的请求始终调度到同一台服务器，

以实现会话保持的效果，其原理如下图所示：



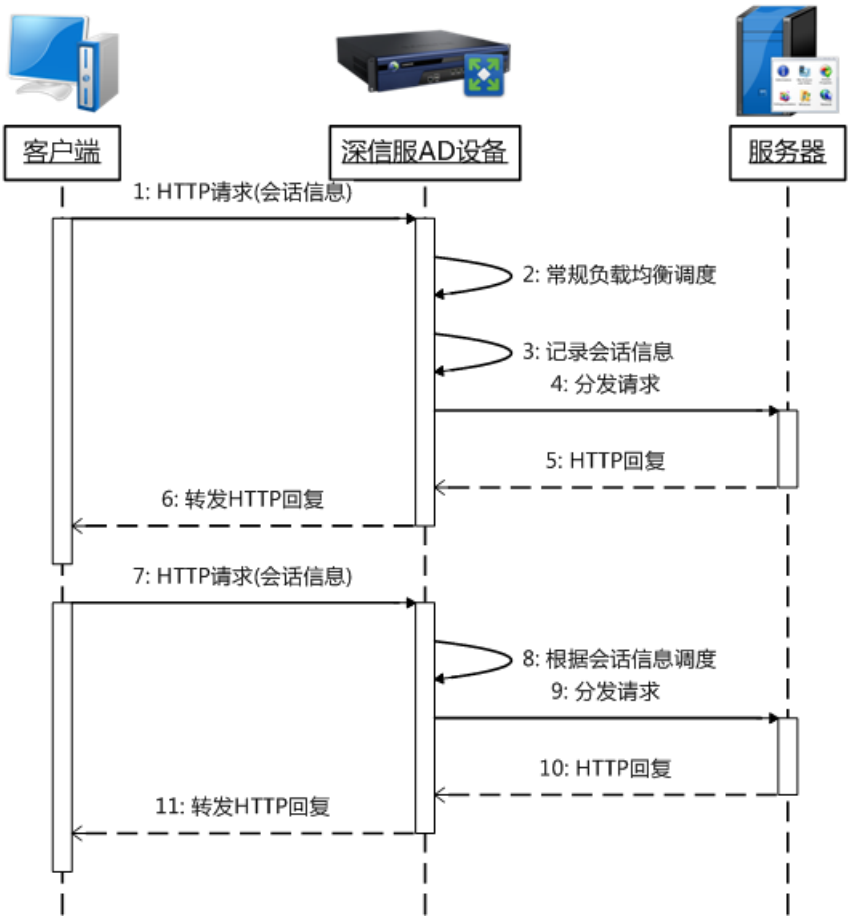
- **基于 Cookie Rewrite 的会话保持机制 -** 也被称为改写式 Cookie 会话保持，常被用于认证系统与应用系统交互的业务场景中。用户在首次访问应用服务器的时候，会被重定向到认证服务器进行登录认证，AD 设备在对多台认证服务器做负载均衡的同时，会在认证返回信息的 Cookie 中写入标识信息，以便于区分不同的认证服务器。应用服务器随后发起用户验证的时候，AD 设备先根据标识信息辨认出用户

先前登录的同一台认证服务器，并从 Cookie 中删除标识信息后对其转发验证请求，以实现会话保持的效果，其原理如下图所示：

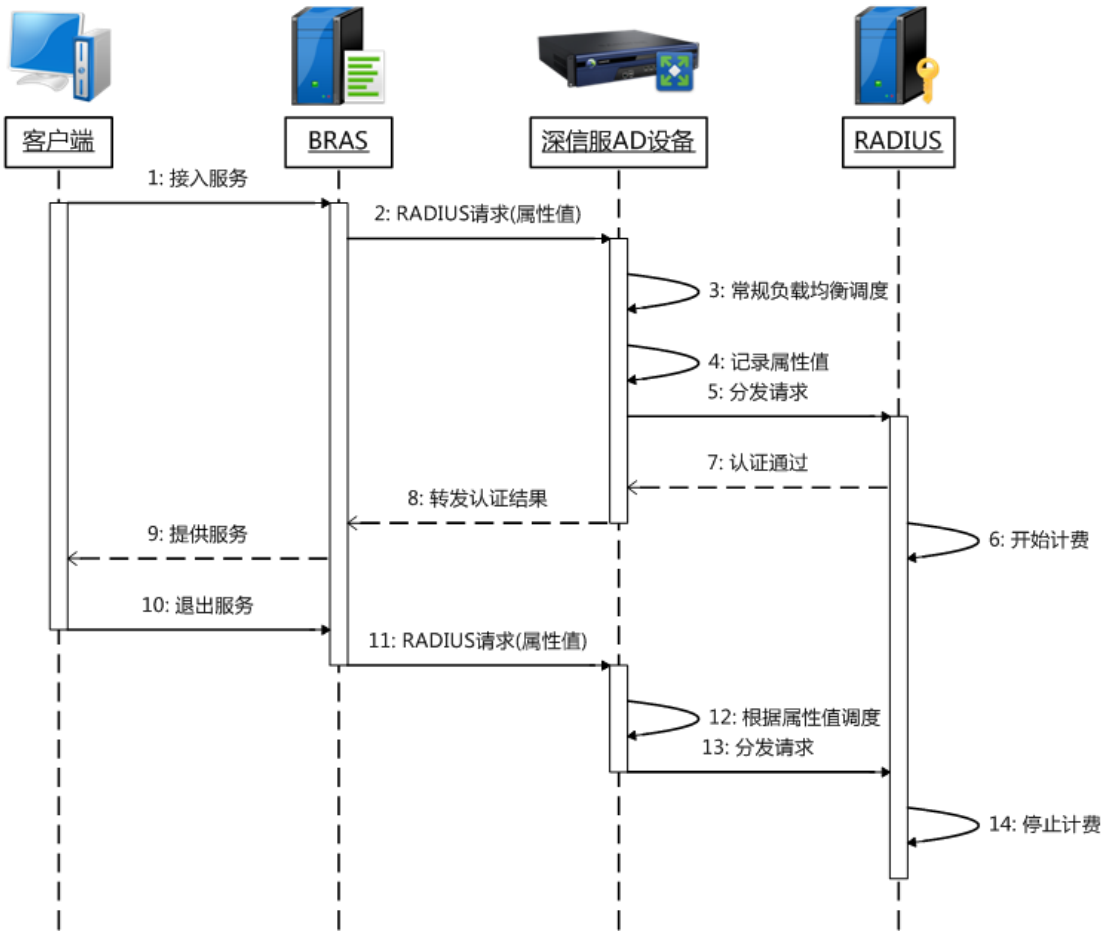


- **基于 HTTP Header 的会话保持机制** - 某些 HTTP 应用不支持或者不完全支持 Cookie 会话保持，但是用户的会话信息是可以带在 HTTP Header 里面的，例如移动用户进行 WAP 浏览时，使用 Call-ID 来代表用户的手机号码。在此种情况下，深信服 AD 设备会记录 HTTP Header 的会话信息，待用户下次访问时便可将带有相同会话信息的请求始终调度到同一台服务器，以实现会话保持的效果，其原理如

下图所示：



- **基于 RADIUS 的会话保持机制 -** 宽带用户的 RADIUS 请求通过 BRAS（宽带接入服务器）发送到 RADIUS 服务器，AD 设备在对 RADIUS 服务器实现负载均衡的同时，也会通过 RADIUS 属性来维护会话的一致性。同一用户的认证会话使用相同的 RADIUS 服务器，以确保 RADIUS 认证、计费的统一，其原理如下图所示：



2.4.6 服务器健康检查

深信服 AD 设备支持包括基于硬件运行状况的主动检查，基于应用类型的主动检查，基于观测方式的被动检查，以及自定义的健康检查机制。

- **基于硬件运行状况的主动检查** - 通过 PING、SNMP 等方式监控服务器的运行状况，一旦出现 ping 无回包、服务器资源消耗过高、死机等情况，都可以实时将访问请求分配到其他正常的服务器之上。
- **基于应用类型的主动检查** - 深信服 AD 支持根据不同应用类型交互机制设定相应的健康检查机制，如 HTTP、FTP、E-mail、DNS、RADIUS 等都可以通过相应的健康机制监控应用的运行状况，如果发现故障，用户即被透明地分配到其它正常工作的服务器上。
- **基于观测方式的被动检查** - 根据观测到的服务器连接数，上下行数据报文等参数，

判定服务器节点是否有效。例如，通过监控到业务流中的 HTTP403、404 错误等内容，或者感知到大量的 RST 关闭连接、零窗口等异常 TCP 传输行为，判断出服务器已经失效，进而将用户的访问请求分配到其他有效的服务器之上。

- **自定义内容检查机制** - 是通过预设自定义字符串，来判断服务器应用是否运行正常；如对某个应用用户通过预先设定该应用正常返回包中应该包含的字符串，深信服 AD 检验服务器返回数据包内是否包含了该特定内容，如果没有包含该内容，就认定该服务器故障，将用户的访问请求分配到其他健康的服务器之上。

2.4.7 服务器平滑退出

当需要进行系统维护或者服务器升级时，通过服务器平滑退出机制，深信服 AD 设备能够保证服务器退出时不会造成用户的访问中断。一旦管理人员选定某台服务器要从服务器组内退出服务后，深信服 AD 设备将不会把新的用户分配到该服务器。当该服务器处理完当前用户的访问之后，就可以开始进行对服务器的相应管理和维护了。

2.4.8 服务器温暖上线

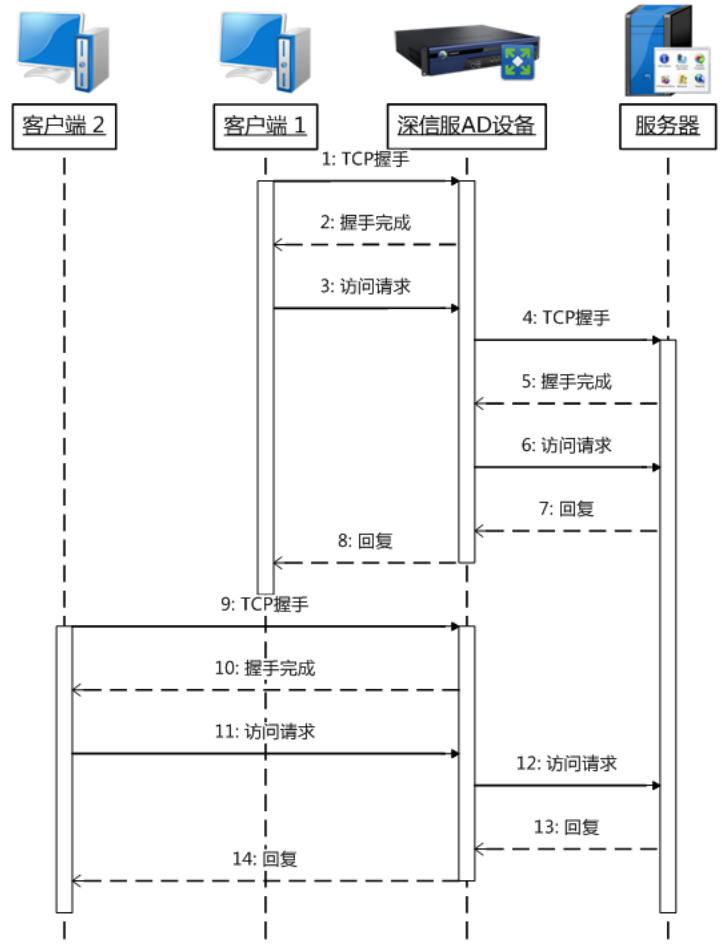
将新购置或维护后的服务器添加到服务器组时，深信服 AD 设备可以通过温暖上线方式，避免新服务器由于激增流量的冲击而导致系统故障，实现服务器的平滑进入。当新服务器上线后，在其恢复时间内，AD 设备不会向该服务器发送请求；而在随后的温暖时间以内，AD 设备则会逐渐地增加分配到该服务器的请求，使新服务器的压力缓慢增加到稳定状态，从而保证服务器在启动期间以及应用程序初始化时都能提供不间断的服务。

2.5 服务器性能优化

信息化建设对各个组织机构而言都是一项长期的工作，硬件投资成本与性能回报是每个IT部门需要权衡的问题。有别于传统的负载均衡设备，深信服AD应用交付产品除了能实现服务器负载均衡机制，提高服务器资源的利用率之外，还支持TCP连接复用、内存缓存、HTTP压缩、SSL卸载等众多性能优化技术。通过减少服务器的硬件资源消耗，缩短服务器响应时间，在节省了硬件投资成本的同时，有效地保障用户访问的速度和稳定性，进而提升用户的访问体验。

2.5.1 TCP连接复用

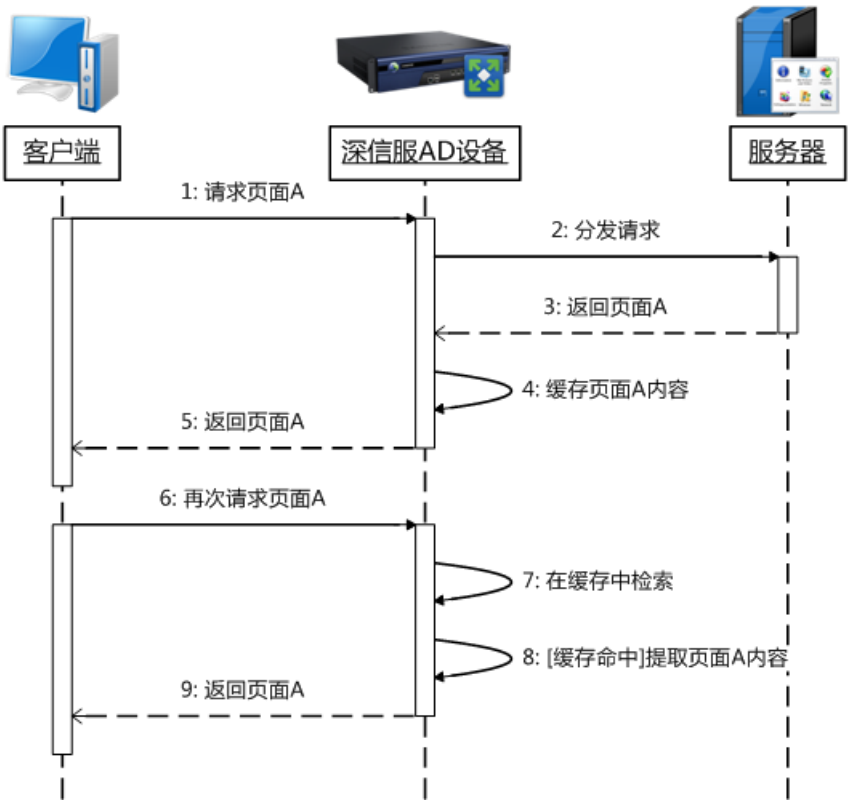
对于采用HTTP协议的应用系统，深信服AD设备可以将众多客户端访问请求捆绑处理，通过复用相对较少的服务器端TCP连接，将客户端请求依次转发到服务器，而不必通过一对一的方式将每一个客户端的HTTP请求通过专门建立的TCP连接转发到服务器。如此，在不需要改变任何网络构造也不需要增加组织的硬件投资成本的情况下，减少服务器的工作负荷，从而提高服务器的处理能力。



深信服 AD 设备保持先前用户访问时与后台服务器之间建立的 TCP 连接，以供后续访问使用，如此便显著减少了后台服务器需要处理的用户端连接数(减少量最高可以达到90%)，加快了客户端与后台服务器之间的连接处理速度，提高应用系统的处理能力，节省组织的硬件投资成本。

2.5.2 内存缓存

深信服 AD 设备基于内存的反向代理 Cache 功能，在内存中缓存网站等相关资源的页面内容；采用内存缓存和包存储结构的方式，通过动态调整缓存空间提供远比其它缓存产品更快速的响应速度。

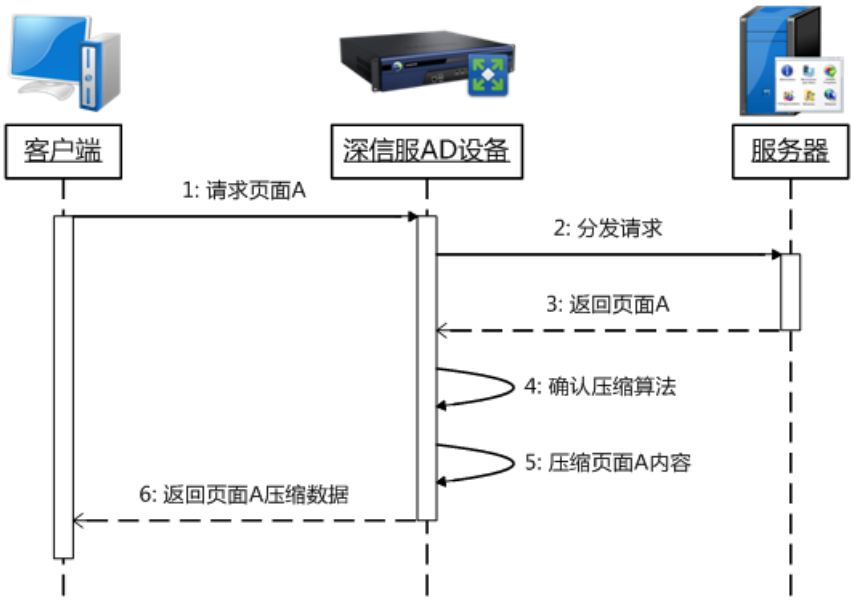


深信服 AD 设备缓存功能可降低用户访问对后台服务器的负载压力，在减少了后台服务器投资成本的同时，提高了系统的处理能力和用户的访问体验。

2.5.3 HTTP 压缩

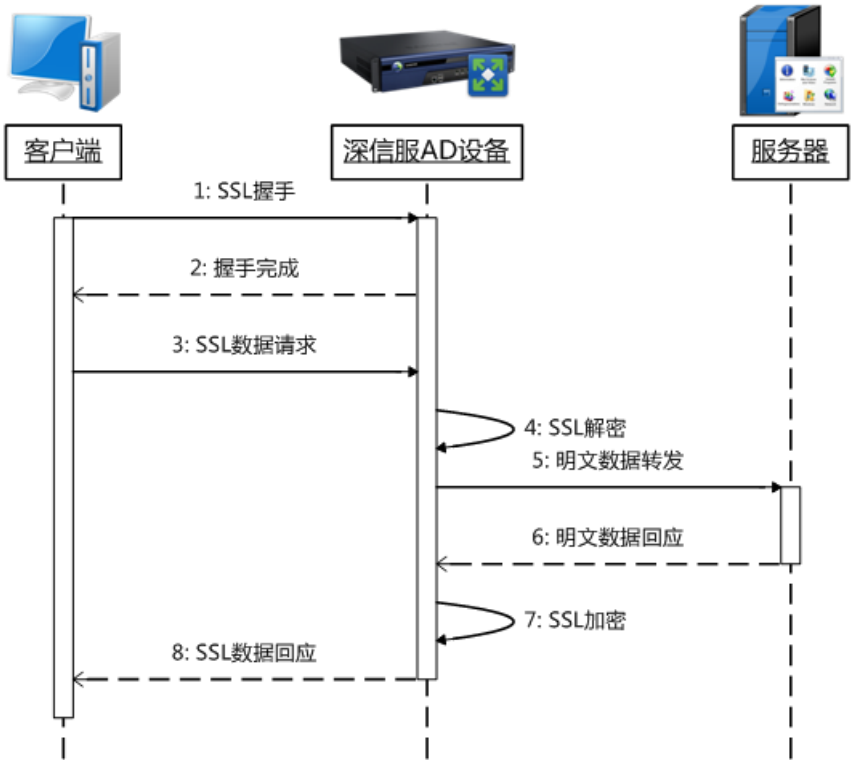
深信服 AD 设备的 HTTP 压缩功能，可通过标准的 HTTP 压缩规范自动识别客户端对 gzip 或 deflate 压缩算法的支持情况，并能够实现对数据动态压缩。

深信服 AD 设备的压缩功不仅能在最大程度上节省组织的互联网带宽，缩短用户下载内容的等待时间，更减轻了 Web 服务器的压力，节省硬件投资成本，提升用户的访问体验。



2.5.4 SSL 卸载

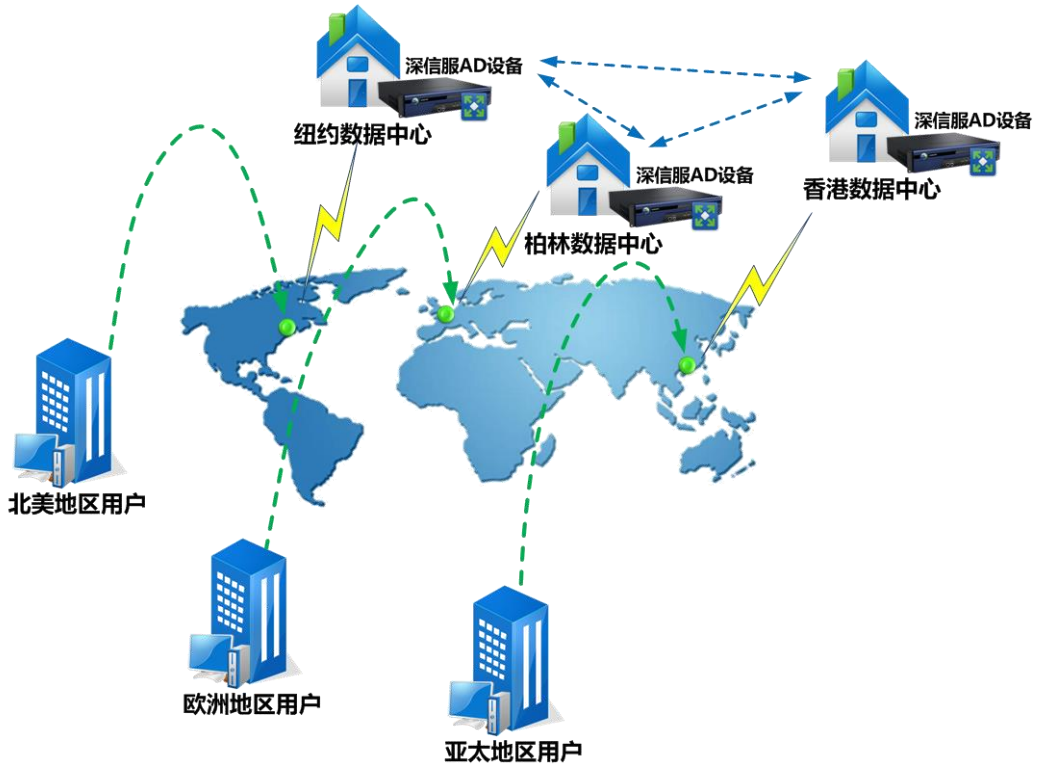
SSL 卸载技术是通过将应用访问过程中 SSL 的加解密过程转到深信服 AD 设备之上，从而减少服务器端的性能压力，提升客户端的访问响应速度。深信服 AD 设备具有强劲的 SSL 处理能力，不但能够实现端到端的 SSL 加密，同时支持全面的加密算法配置，并可管理服务器证书。



深信服 AD 设备通过对服务器的 SSL 卸载处理，在减少服务器性能消耗的同时，节省应用系统服务器数量，降低了业务系统的硬件投资，并大幅度缩短用户请求的响应时间从而极大提升了用户的访问体验。

2.6 全局负载均衡

深信服全局负载均衡解决方案能够帮助组织将相同应用系统的服务内容部署在不同地理位置之上，保证承载应用系统的多数据中心能够具备更高的持续性和可用性以及快速性，使得用户不管身处全球任何位置都能获得更快速、更稳定的访问体验。



通过全面的健康检查机制，AD 设备能够实时的监控各个数据中心的运行状况，及时发现出现故障的数据中心或者其内部服务器，从而保证将用户后续访问请求都分配到其他的正常数据中心或者服务器之上。不但使多站点之间形成冗余，保障用户访问稳定，还提升了各站点的资源利用率。

2.6.1 智能 DNS 方式多站点调度

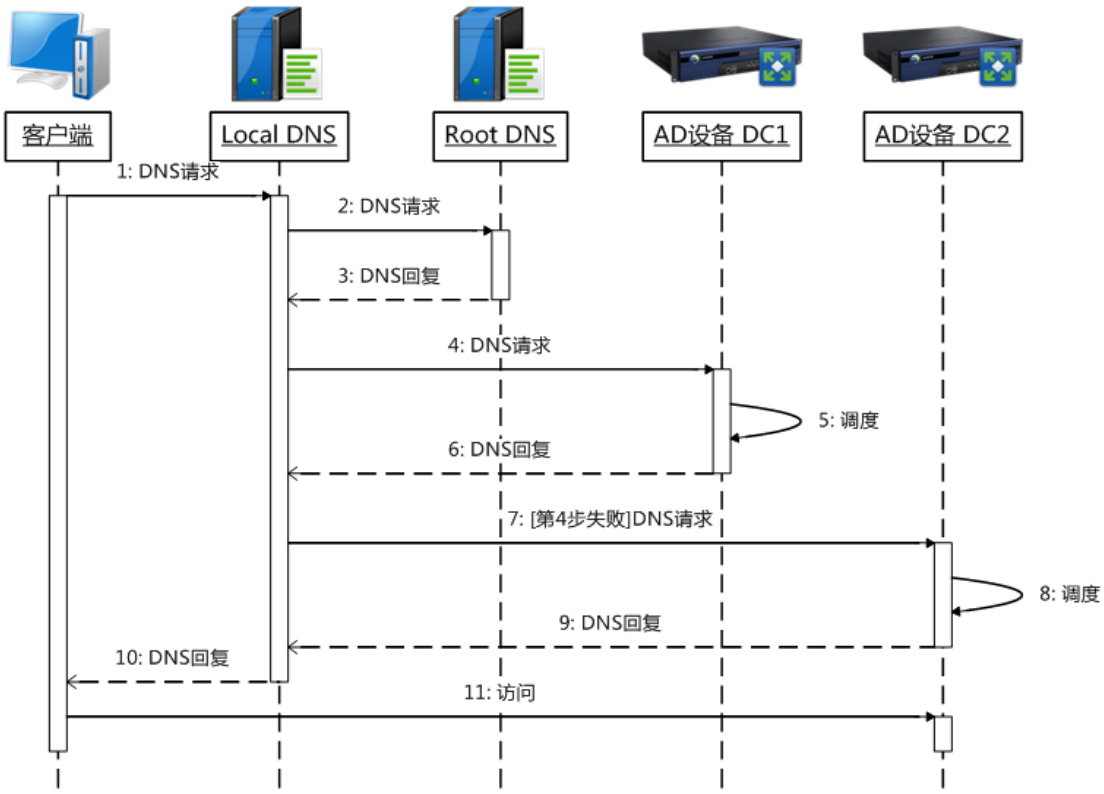
1. 实现方式

利用智能 DNS 解析技术，以唯一的域名的方式为所有发布相同服务的数据中心提供统一的入口，根据管理人员预先设定的负载策略将用户的访问请求分配到不同数据中心之上，从而实现多站点的负载均衡调度。

当用户通过域名方式进行访问时，可以根据用户使用的 Local DNS 位置进行就近性计算，将最佳站点的 IP 地址解析给用户。同时结合 AD 设备所配备的全球 IP 地址库，进一步提高用户请求就近分配的准确性，从而彻底解决用户跨国、跨运营访问速度慢的问题。

2. 工作流程

智能 DNS 方式多站点调度的工作流程如下图所示



智能 DNS 方式多站点调度的流程描述如下

步骤	说明
1	外网用户的访问客户端向其本地 DNS 服务器发出域名解析请求
2	本地 DNS 服务器首先在本地搜索是否有相应的记录，如果没有就向根 DNS 服务器发起查询
3	根 DNS 服务器反馈本地 DNS 服务器，告知两条 NS 记录，分别指向数据中心 DC1 和 DC2
4	本地 DNS 服务器向 DC1 的深信服 AD 设备发出域名解析请求
5	DC1 的深信服 AD 设备先判断链路的健康状况，再根据预先定义的负载均衡算法来选择出合适的 IP 地址作为域名解析结果
6	DC1 的深信服 AD 设备将域名解析结果反馈给本地 DNS 服务器
7	本地 DNS 服务器无法接收到 DC1 的深信服 AD 设备反馈的域名解析结果，再次向 DC2 的深信服 AD 设备发出域名解析请求
8	DC2 的深信服 AD 设备先判断链路的健康状况，再根据预先定义的负载均衡算法来选择出合适的 IP 地址作为域名解析结果

9	DC2 的深信服 AD 设备将域名解析结果反馈给本地 DNS 服务器
10	本地 DNS 服务器将得到的域名解析结果转发给客户端
11	客户端根据得到的 IP 地址发起连接请求，访问请求最终到达 DC2

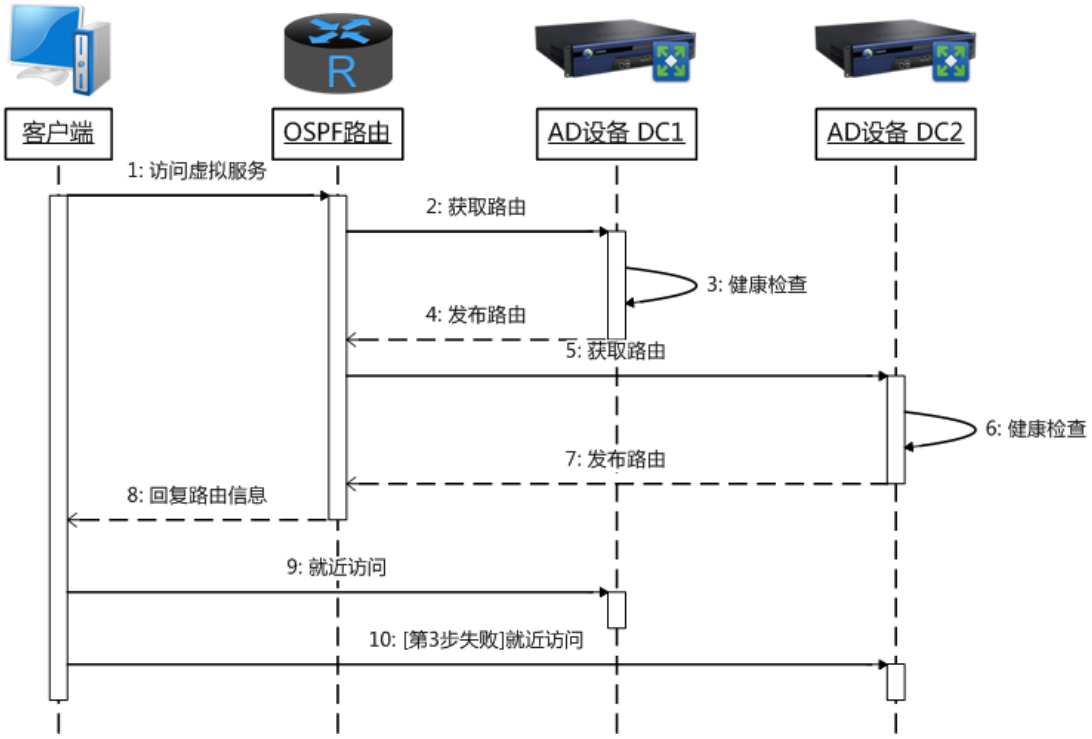
2.6.2 IP-Anycast 方式多站点调度

1. 实现方式

通过动态路由协议在多个站点发布虚拟服务 IP，利用网络中的路由信息实现用户的就近访问和站点冗余。当多个站点同时在线提供服务时，用户访问虚拟服务会根据路由信息来选择出最近站点。当某个站点出现故障时，OSPF 路由会删除到该站点的路由，因此用户访问虚拟服务时会将该站点排出选择，待该站点恢复以后，OSPF 路由则会重新将该站点的路由加入。路由通知的快慢取决于网络的规模，一般情况下在 1 分钟内。

2. 工作流程

IP-Anycast 方式多站点调度的工作流程如下图所示



IP-Anycast 方式多站点调度的流程描述如下

步骤	说明
1	外网用户的访问客户端发起对虚拟服务的访问请求
2	OSPF 路由向 DC1 站点的深信服 AD 设备获取该站点的动态路由信息
3	DC1 的深信服 AD 设备对虚拟服务进行健康检查，当虚拟服务可用时，发布到达虚拟服务 IP 的路由
4	OSPF 路由将来自 DC1 的深信服 AD 设备所发布的路由加入
5	OSPF 路由向 DC2 站点的深信服 AD 设备获取该站点的动态路由信息
6	DC2 的深信服 AD 设备对虚拟服务进行健康检查，当虚拟服务可用时，发布达到虚拟服务 IP 的路由
7	OSPF 路由将来自 DC2 的深信服 AD 设备所发布的路由加入
8	OSPF 路由将到达虚拟服务 IP 的动态路由信息回复给访问客户端
9	客户端根据路由信息来选择最近站点，访问请求最终到达 DC1
10	当 DC1 的深信服 AD 设备检测到虚拟服务不可用时，OSPF 路由则会相应地去除到达 DC1 虚拟服务 IP 的路由 ,当客户端根据路由信息来选择最近站点时，访问请求会最终到达 DC2

2.6.3 就近性判断机制

为了保障当全球范围的用户在访问资源时，能够被引导至“最优”的数据中心，全局负载均衡设备需要对用户到各站点之间的距离、延时及当前数据中心的负荷等众多因素进行分析判断。深信服全局负载均衡支持静态和动态两种就近性方法，两种方式可以并存使用。

- **静态就近性 -** 深信服全局负载均衡设备中都搜集了全球的 IP 地址形成地址库，并能够实现实时更新；当用户访问目标 IP 属于哪个运营商（或地区），就为用户选择这个运营商（或地区）的数据中心（或链路）。当用户请求没有包含在设备的地址库中时，深信服全局负载均衡设备将会主动查询该地址所属地区（或运营商），匹配之后再根据静态就近性为用户选择数据中心。如果上述两种方式都无法判别用户请求 IP 所属地区（或运营商），则直接使用动态就近性。
- **动态就近性 -** 当用户发起访问请求时 ,深信服负载均衡设备可以通过综合考虑各数据中心的传输延迟和数据中心链路的实时负荷，准确计算出最佳路径，将用户引导

至最佳的数据中心。

2.6.4 健康检查机制

深信服全局负载均衡设备的健康状况检查可以保证用户获得最佳的服务站点。

- **链路的健康检查** - 深信服全局负载均衡设备通过多个 Internet 站点的可达性，来共同判断一条链路的状况。例如，通过电信线路检查 www.sina.com.cn、www.sohu.com、以及 www.qq.com 的 TCP 80 端口，并对检查结果做“或”运算。这样，只要其中一个站点可达，即可表明链路状态良好。该方法即避免了 ICMP 检查的局限性，也避免了单一站点检查带来的单点失误。
- **虚拟服务的健康检查** - 深信服全局负载均衡设备在部署网络中，每台 AD 设备都会对所有数据中心的虚拟服务进行监测，这样不仅可以实时发现出现故障的数据中心，同时也可以监视虚拟服务在 IP、TCP、UDP、应用和内容等所有协议层上的工作状态。一旦发现某个数据中心或者服务器出现故障，用户即被透明地重定向到正常工作的数据中心或者服务器之上。

2.7 可编程功能 iPro

深信服 AD 应用交付系列产品提供脚本编程功能 iPro，通过 Lua 语言实现用户自定义的流量编排处理，支持流量调度、会话保持和 DNS 等子功能。它基于事件驱动，继承了 Lua 语言的基本特性，并加入了一定的扩展。通过 iPro 编写脚本，可提取应用数据包的不定址不定长特征码，实现 7 层内容交换，主要应用于未知协议的解析和基于消息的负载均衡，很多特殊场景下以往无法实现的用户需求，都可以通过 iPro 轻松搞定。

2.8 基于消息的长连接负载均衡

2.8.1 长连接的特点

TCP 协议通讯开始时 client 会跟 server 建立连接，待连接建立完成后才会传输消息，消息一旦传输完成，client 与 server 的连接会在短时间内断开。而长连接就是消息传输完成后，client 与 server 的连接会维持一段较长的时间不断开，后续消息可以复用该连接。长连

接一般用于大数据量的高性能处理，很多重要的 C/S 应用都会使用长连接，这种应用通常有相对固定的 client，它们会与 server 进行频繁通信，频繁创建短连接会极大的浪费 server 资源，而长连接有效的削减了连接建立和拆除的开销，极大的减轻了 server 的性能压力。

2.8.2 长连接的负载均衡

长连接的应用同样需要负载均衡，对于很多 C/S 应用来说，其协议私有，不对外公开，传统方式无法识别到协议内容，只能实现基于连接的负载均衡。这种方式缺点比较明显：

1.调度不合理。传统方式的负载均衡只能按连接调度，而长连接一旦建立就很难中断重连，同一个连接内的不同消息会发往同一台服务器，造成服务器压力过大，而这种情况下其他服务器节点可能处于空闲状态。

2.故障切换慢。当服务器发生故障时，客户端并不会立即重新发起新的连接，而是一直等待连接超时，通常来说这个时间非常长，造成故障切换非常慢。

深信服 AD 应用交付产品的 iPro 功能，可结合用户应用的具体特征编写脚本，分析出连接中单个消息的开始和结束位置，将同一连接中的不同消息调度到不同节点，实现基于消息的长连接负载均衡，使各节点资源利用更合理更平衡。同时，深信服应用交付 AD 产品会与各服务器节点建立长连接，并实时探测各服务器节点的健康状态，发现异常会将消息立即分配到正常节点，规避故障风险，大幅缩短故障恢复时间，提高业务访问的连续性。

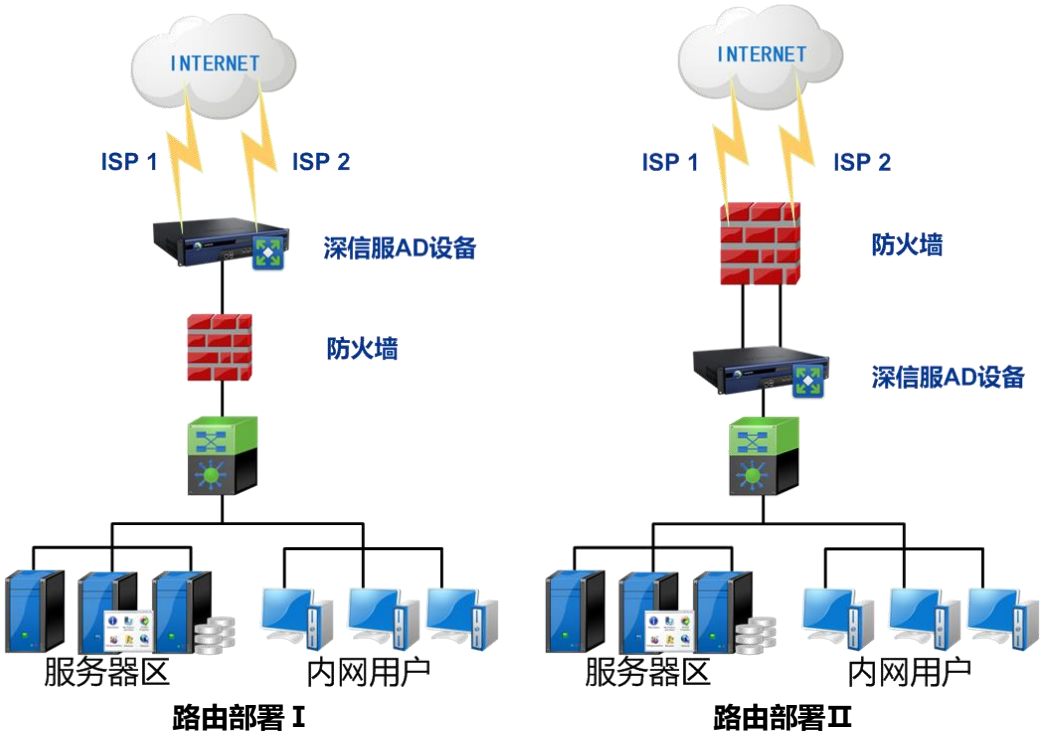
2.9 设备部署与管理

深信服 AD 系列应用交付产品支持路由、旁路两种部署模式，支持 VLAN、STP 等局域网协议，并可与动态路由协议（例如 OSPF、RIPv1、RIPv2）网络进行对接，以帮助用户应对各种复杂的网络环境。在业务持久性、终端用户访问体验、数据中心可用性等多个应用领域，均可通过部署 AD 设备予以优化。

在 AD 设备上启动 SNMP 服务后，用户可使用 SNMP 客户端软件查询到 AD 设备的 CPU 负载、内存占用、新建连接、并发连接、吞吐量等各种信息，方便地监控设备的硬件负载状况。此外，AD 设备也支持 ACL 功能，允许用户根据连接的五元组对访问进行精确的控制。

2.9.1 路由模式

将深信服 AD 设备通过网关模式串接在用户网络链路中，所有流量都通过 AD 处理，当用访问请求到达深信服 AD 时，深信服 AD 设备将根据预先设定的策略，将用户流量分配到最佳的互联网链路之上，同时在服务器群组中为用户提供性能最佳的服务器，保证用户访问体验，提高用户满意度。此种部署模式适合于链路负载、服务器负载以及全局负载需求。

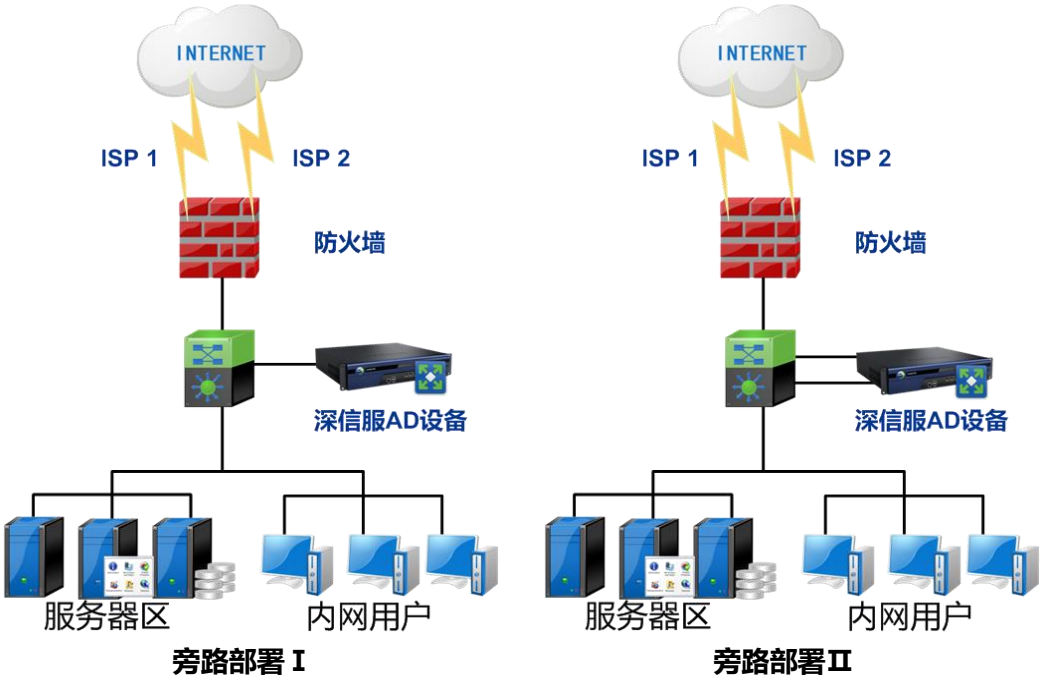


- **路由部署 I** - 常规的 AD 设备部署模式，可以同时实现服务器负载和多链路负载，AD 设备上做 NAT，防火墙做透明模式。
- **路由部署II** - 防火墙上做 NAT（WAN1-LAN1 / WAN2-LAN2），AD 设备做路由，可以同时实现服务器负载和多链路负载。防火墙需要同时接两条线到 AD 设备上，做 DNS 端口映射。

2.9.2 旁路模式

深信服 AD 设备以旁路模式部署到网络之中，此种模式不会改变客户的网络结构，同时设备上架时不会造成业务的中断，可以实现设备的快速简单部署。当用户请求到达深信服

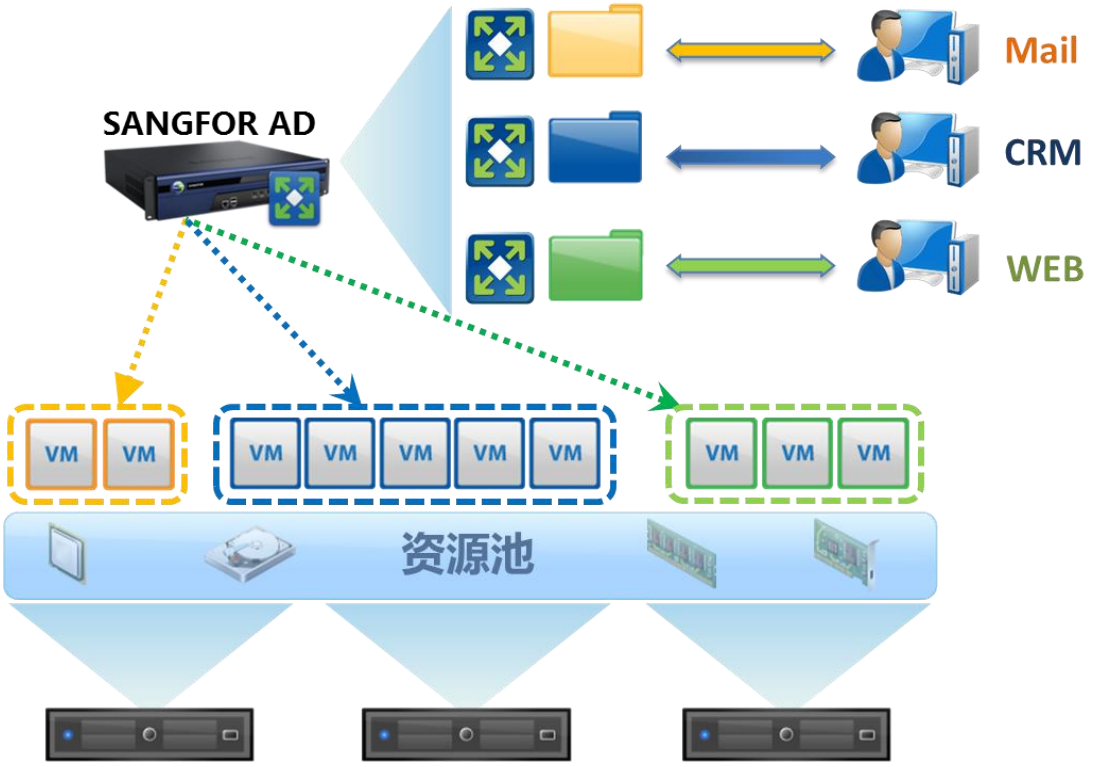
AD 设备，根据预先设定的负载据策略为用户选择最佳服务器，提升服务器利用率，避免服务器任务分配不均的情况出现，此种模式适合于服务器负载均衡需求。



- **旁路部署 I** - 防火墙做端口映射或直通到 AD 设备，只要 DNS 请求能够到达设备就可实现对入站流量的链路负载均衡，同时也可以实现对服务器的负载均衡。
- **旁路部署II** - 此种旁路多线的部署模式，要求前端防火墙支持策略路由功能，并针对 AD 的 WAN 口地址做策略路由，可实现对出站和入站流量的链路负载均衡，以及对服务器的负载均衡。

2.9.3 虚拟化分区

SANGFOR AD 产品支持在单一设备上划分配置多个虚拟 vAD 设备，按需分配给多个租户（部门）使用，在实现管理隔离和网络数据隔离的同时，让所有租户（部门）能够充分的共享并利用 AD 设备的计算资源。



- **多租户管理平台** - 解决客户的多租户使用需求，将一台 AD 设备的硬件资源划分成多个 vAD 提供给不同的租户使用，vAD 之间相互隔离，不受影响；结合服务器虚拟化架构，可快速构建高灵活、易管理的各类业务应用系统

  ✓ 可基于业务划分需要，自行创建创建、删除、启停单独的 vAD

  ✓ 支持为每个 vAD 分配 CPU、内存、网卡，并配置新建与并发性能

  ✓ 支持为每个 vAD 配置管理员密码与管理 IP

  ✓ 支持 HOST 机的 VLAN 划分，日志管理，用户管理

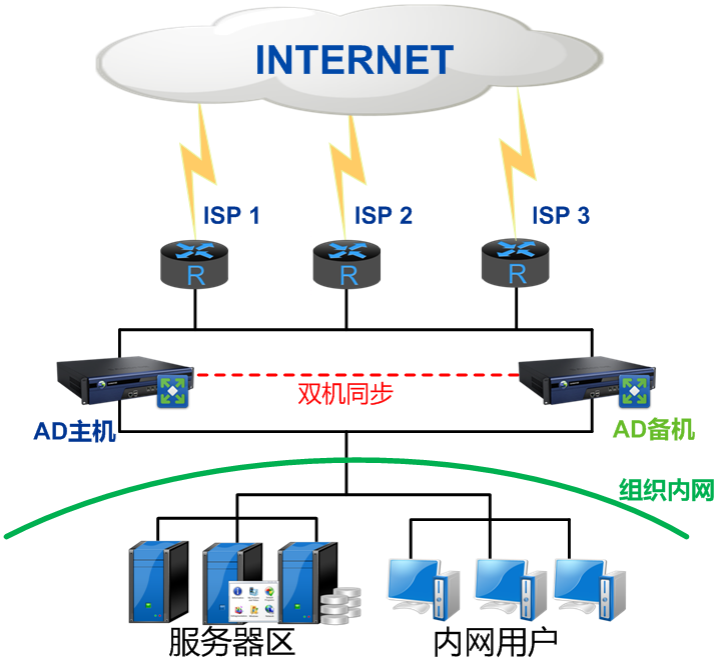
2.10 高可用性（HA）

无论是面向多数据中心的全局访问调度，还是针对多链路和服务器群的负载均衡，深信服 AD 设备都扮演着一个关键的控制节点角色，其设备的稳定性和安全性则直接影响到业务交付网络的可用性。为了避免单点故障的隐患，采用双机热备是保证业务连续性的一种有效解决方案，能够在很大程度上避免了网络业务的中断。

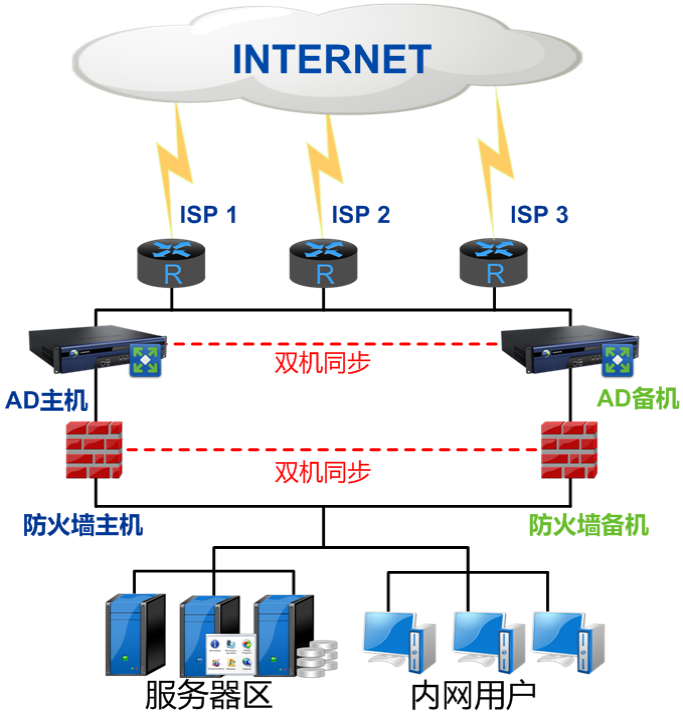
2.10.1 主备模式

两台深信服 AD 设备部署在网络中，在此种模式下通常把正在执行负载均衡调度的 AD 设备称为主机，而另一台处于待命状态的 AD 设备则称为备机。主机在处理业务的同时，会将业务产生的会话信息同步到备机，从而确保双机切换后，新发起的业务访问能继续得到响应处理，当前正在进行的业务访问也不会因此而中断。此种模式适合于大多数网络环境中避免单点故障的部署需求。

- **被动切换** - 当主机发生宕机而无法继续工作时，备机会检测到主机故障并立即接替主机来继续执行负载均衡调度。



- **主动切换** - 当主机检测到与其关联的网络故障（例如，链路健康检查发现 WAN 口线路中断、ARP 检测发现防火墙发生双机切换），此时备机就会主动切换为主机，以确保业务访问的连续性。

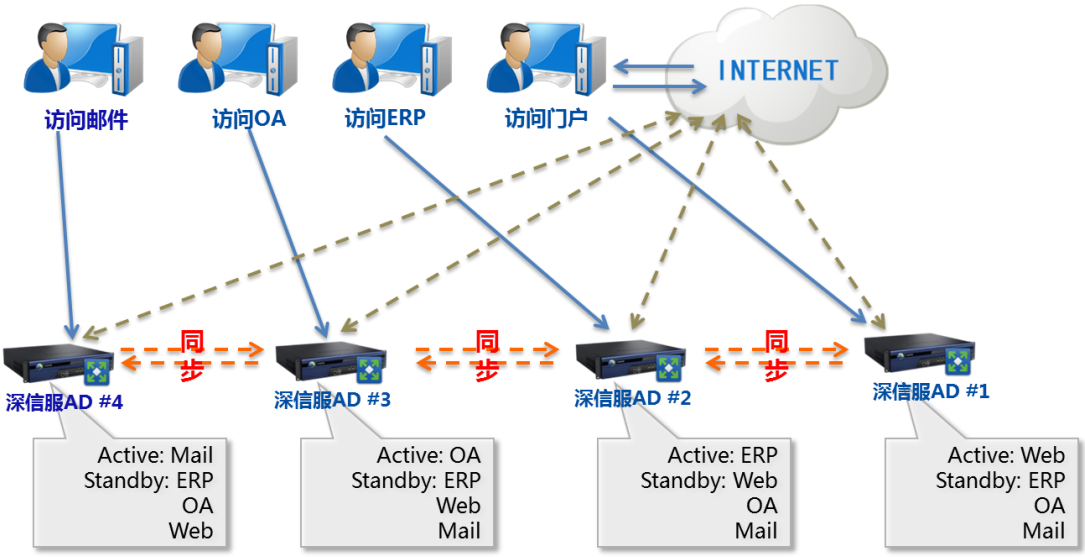


2.10.2 集群模式

通过多台 AD 设备组成集群的方式，即可以扩容整个系统的处理性能，并且充分利用集群中的每一台设备，发布不同的服务，同时又能互为备份，保证故障发生时的最小业务中断。即只要集群中仍有一台设备可用，就不会导致任何一个服务停摆。在此基础上，实现均衡的故障切换和动态切换机制。

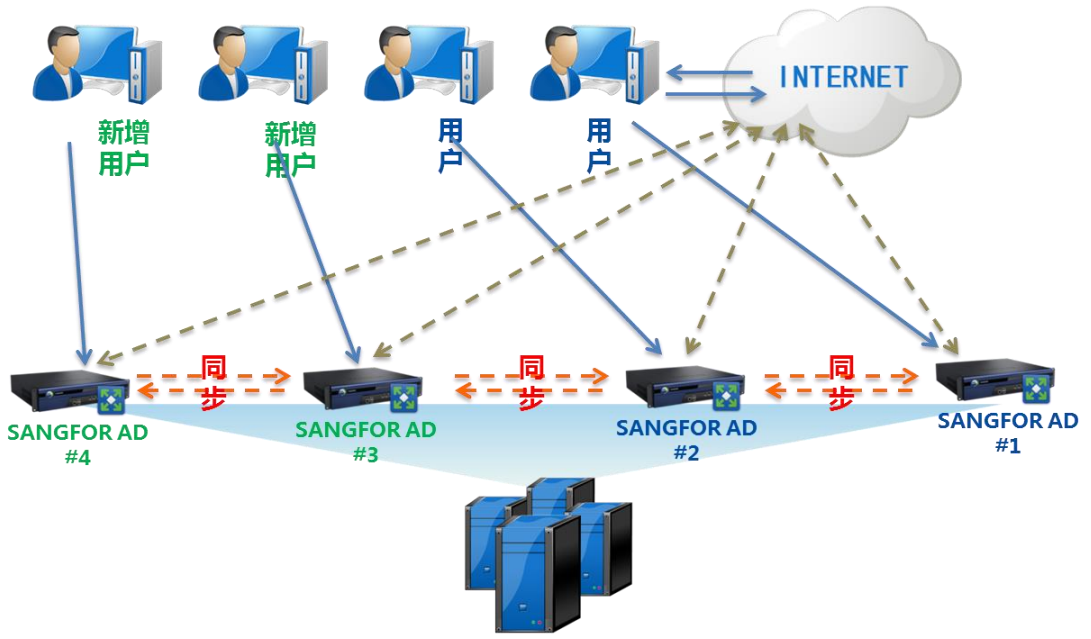
- **高可用性集群**

  - 多台 AD 设备组成高可用性集群，支持主备、双主、多主、M+N 等多种模式
  - 与双机模式相比，集群提供更高的冗余性（N 备一 vs 一备一）和更低的成本控制（1+N vs 2×N）
  - AD 设备之间互为备份，同步会话镜像（session mirror），发生切换不中断业务



➢ 高性能集群

- 支持多台 AD 设备组成高性能集群，即虚拟成一台逻辑设备
- 可应对 100Gbps 以上业务规模的性能扩容
- 支持与高可用性集群同时使用，使业务扩展性更加灵活



第3章 应用交付特色技术

3.1 单边加速技术

单边加速的核心技术理念在于，通过自动、实时、持续、动态地侦测网络路径中的延迟、丢包、重传的情况，改变传出机制和改善传输拥塞机制，避免数据报文的过度重发，从而减少应用响应时间并提升 TCP 传输效率。由于是针对所有 TCP 数据流进行优化，因此不仅文件、电子邮件、网站等应用可以通过该技术实现加速效果，而且只要运行于 TCP 之上的应用，都可以通过深信服 AD 单边加速技术提升用户的访问体验。

3.1.1 应用背景

在组织的业务系统建设的过程，不仅要解决如何保障系统运行稳定性和访问持续性的问题；与此同时，如何进一步保障用户访问速度，提高用户访问体验？很多的组织机构已经部署了多条运营商的链路，然而互联网用户访问内部资源的时候还是体验很慢，如手机炒股用户，运营商 3G 用户，跨国访问等尤为突出。究其原因在于，通过这类用户的网络在传输数据时都会存在一定的延时和丢包，必然造成访问速度变慢。

传统解决方案主要是通过提升网络带宽和部署多链路的解决思路，甚至于通过部署链路负载均衡设备，来提升链路的访问速度以及稳定性。然而从本质上而言，这是一种治标不治本的方法，它不仅增加了更多带宽费用，而且没有解决链路质量的问题，更重要的是没有减少应用响应的时间，所以在大部分情况下访问速度还是得不到有效的提升。

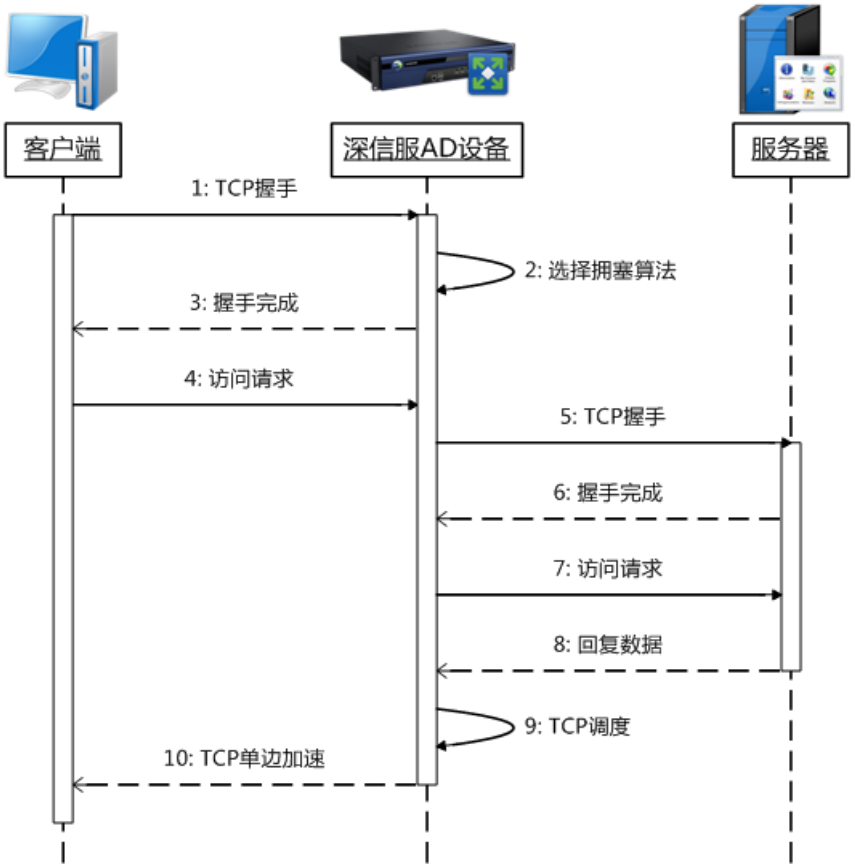
而传统的广域网加速解决方案，通过流缓存、压缩、协议优化等众多技术，能够在一定程度之上解决链路质量方面存在的问题，但这类解决方案往往需要通过对称部署相应设备来实现。这种对称部署的模式对于分支与总部之间的连接是比较适用的，但是对于访问用户分散、对外发布应用的网络而言，这种模式受客观条件限制就不利于实现，可行性比较低。面对这类应用发布的问题，无疑需要一种快捷有效并且方便实施的解决方案。

3.1.2 功能机制

传统 TCP 协议是基于局域网设计，一旦运用到互联网这种不稳定环境中的时候，就会

使得数据的传输效率降低，特别在延时比较大、有丢包的时候尤为明显。单边加速技术通过对拥塞算法做优化处理，解决一些 TCP 协议本身的缺陷，以实现加速的效果；其核心部分是对拥塞算法做优化，如慢启动，拥塞避免，快速重传，快速恢复等。

- **拥塞避免** - 能够快速的准确的预估出网络中可用带宽，并根据估计值确定拥塞避免窗口，从而最大限度的利用网络带宽。
- **快速重传** - 允许接收端通过使用 SACK TCP 选项指示最多四个接收数据的非邻接块。RFC 2883 定义用于确认重复的数据包的 SACK TCP 选项中的字段的额外使用。发送端可以通过此操作确定何时重传了不必要的段并调整其行为，以防今后不必要的重传。发送的重传越少，整体吞吐量越合理。
- **快速恢复** - 快速检测出丢包，并能快速准确重传该包，对时延较大，网络状况较差的情况能够有效的提升带宽利用率，通过更改快速恢复过程中发送端可以用来提高发送速率的方法，提供更大的吞吐量。
- **慢启动** - 避免发送 TCP 对等方拥塞整个网络的现有算法被称为“慢启动”和“拥塞避免”。在连接最初发送数据和还原丢失段时，这些算法可以增大发送窗口，即发送端可以发送的段数量。对于每个接收到的确认段或每个已经确认的段，“慢启动”算法会以一个完整的 TCP 段增大发送窗口。对于每个已经确认的完整窗口的数据，“拥塞避免”算法以一个完整的 TCP 段增大发送窗口。利用这些算法增大发送窗口的速度就足以充分利用连接带宽。



考虑到互联网的带宽是不断变化的，发送数据过快会导致数据拥塞，发送数据过慢会导致带宽利用率低，使网络速度减慢。单边加速技术也会实时根据网络延时调整发包频率，使发送数据量尽量接近实际网络带宽。而另一方面，标准 TCP 的算法在丢包的环境下重传数据过多，致使速度很慢。单边加速技术则会针对探测到的网络丢包现象，主动发送冗余数据，降低网络数据的重传率。

3.1.3 应用说明

有别于传统的加速解决方案，深信服 AD 系列应用交付设备的单边加速解决方案只需要在发布业务应用的中心端处部署 AD 设备，用户端无需安装任何的软件客户端或浏览器插件，对访问终端的设备形式，操作系统、浏览器种类等方面没有任何兼容性要求，对用户完全透明。而组织机构可以在不升级带宽的前提下，通过减少应用程序的响应时间，增强用户的访问体验，进而提升自身业务竞争力。

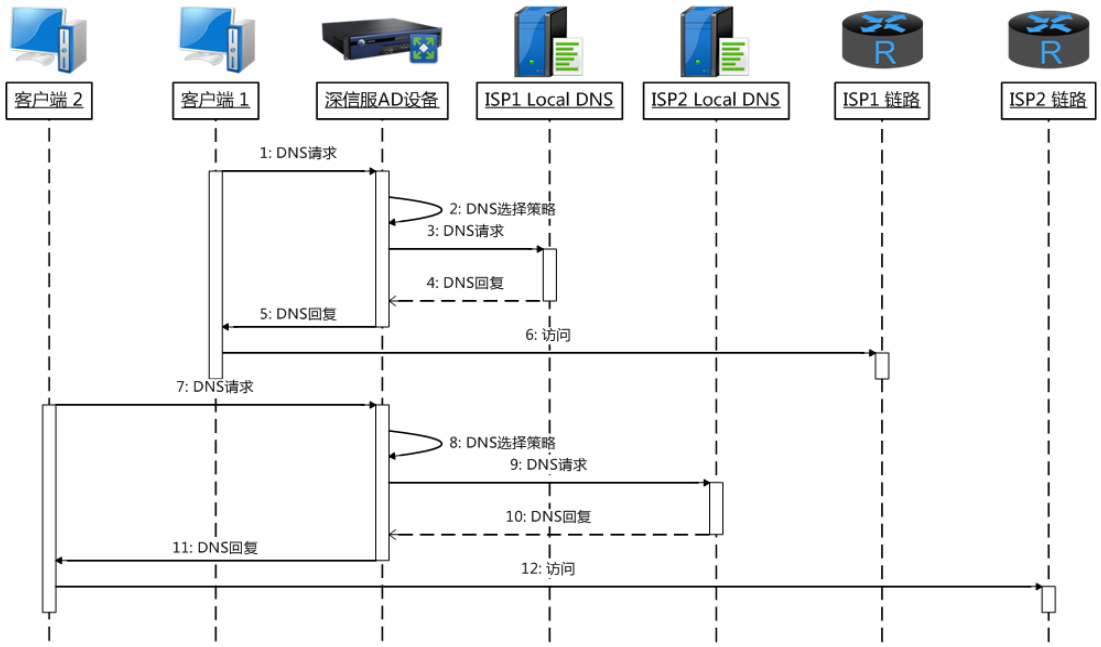
在诸如手机炒股，运营商 3G 上网，用户跨国访问等各类的网络环境之中，深信服 AD 的单边加速技术都能够提升用户的至少 30%访问速度，进而改善用户的访问体验，使组织机构在交付应用的同时，提升内部业务的生产力和外部业务的竞争力。

3.2 智能优化技术

随着业务系统对网络的依赖程度越来越大，对 IT 管理维护人员的要求也越来越高。为应对日趋复杂的用户访问需求，深信服 AD 提供智能路由技术、DNS 透明代理技术以及链路繁忙控制技术，实现基于链路的负荷情况、时间段、用户群体、访问对象等因素来分配链路的分配机制，进一步提升链路优化使用率。

3.2.1 DNS 透明代理

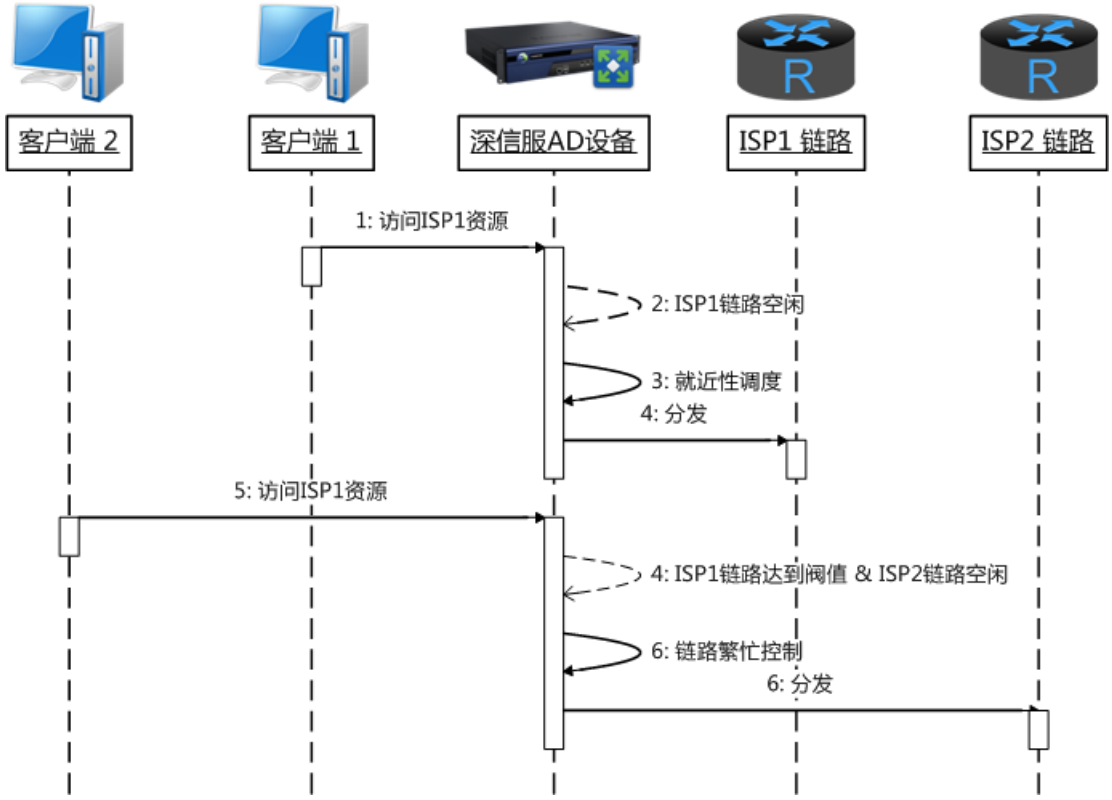
当网络之中部署了多条互联网链路后，内网用户在上网的时候由于都填写其中某一运营商的 DNS 服务器，于是大部分的用户都被分配到同一条出口链路上，使得该链路一直处于繁忙状态，进而导致该链路上的用户访问速度下降，而另一条链路却处于闲置状态。链路利用的不均衡，一方面造成互联网资源的浪费，另一个方面使得用户的访问速度得不到保障。



通过深信服应用交付产品 DNS 透明代理技术，不论内网用户填写哪家运营商的 DNS 服务器地址，都会经由深信服 AD 设备进行 DNS 请求转发。通过 AD 设备寻找合适的 DNS 服务器返回给内网客户端，再搭配相应的负载算法，就能按照预先设定的链路利用策略将流量分配到不同的链路之上，从而有效地实现对多条链路带宽的合理利用，避免带宽资源出现闲置的情况。

3.2.2 链路繁忙控制

随着访问用户对链路稳定性要求的提升，越来越多的网络将原有的宽带链路扩充至多条，通过多链路冗余的方式在一定程度之上保障数据访问的链路稳定性。然而，在进一步实现多链路的合理利用，以及提高用户访问的快速性方面，传统的流量控制设备可以基于应用实现对网络带宽的保护，常规的负载均衡设备也能够实现链路的就近性智能选择，但是两者都不能根据链路的带宽负载情况为用户选择相应访问链路。相比之下，深信服 AD 设备所提供的链路繁忙控制技术，恰恰能够帮助用户更加合理地使用多链路的带宽资源。



链路繁忙控制技术在于为特定链路设定相应的阈值，并结合深信服 AD 全面的负载均衡算法，以实现对链路的优化调度。当某条链路达到阈值之后，用户的访问请求将会通过事先设定的负载策略分配到其它链路之上，从而保证用户访问的快速性，并形成对链路资源的预留保障。

3.2.3 服务器弹性负载

面对愈加复杂的业务系统，主动探测式的服务器健康检查机制逐渐暴露出一定的局限性。尤其是在业务访问涉及到多层应用系统之间交互的情况下，当中途某个环节出了问题时，无法直接针对业务流程是否正常进行检测。此外，主动探测的方式对于未知协议、私有协议的应用支持往往表现欠佳，无法对服务器的健康状况做出准确判断。针对传统方法的不足，深信服 AD 设备提供了一套基于服务器健康度的弹性调控机制，可通过监控业务流中的 TCP 传输异常来衡量服务器节点的有效性，尝试对性能不足的服务器临时开启过载保护，动态调节服务器的负载，进而保障整个业务系统的高可用性。

- **智能感知 -** 深信服 AD 通过持续观测业务流中的 TCP 传输行为来进行判断，当某服务器节点的状态满足失效条件时，即将该节点标记为无效，并不再向此节点转发来自客户端的请求
- **过载保护 -** 当判断为服务器性能不足时，深信服 AD 只对该服务器进行会话保持，而不进行任务分配，一段时间后再恢复调度，实现对服务器压力的弹性调控。

3.2.4 智能路由

通过智能路由技术，管理人员即使对负载算法和策略不熟悉，只要选择自己对于链路利用所期望达到的效果，就可通过配置向导实现对负载策略的轻松配置。例如，用户的网络出口部署有多条的链路，管理者为保证各条链路的利用率，常常要指定专门的链路用于特定的业务访问，利用智能路由就可以实现根据访问的目的域名选择出站链路；或者配置基于时间段的负载均衡，即在不同的时间段采用不同的负载均衡策略，实现网络和服务器资源利用率的最大化。

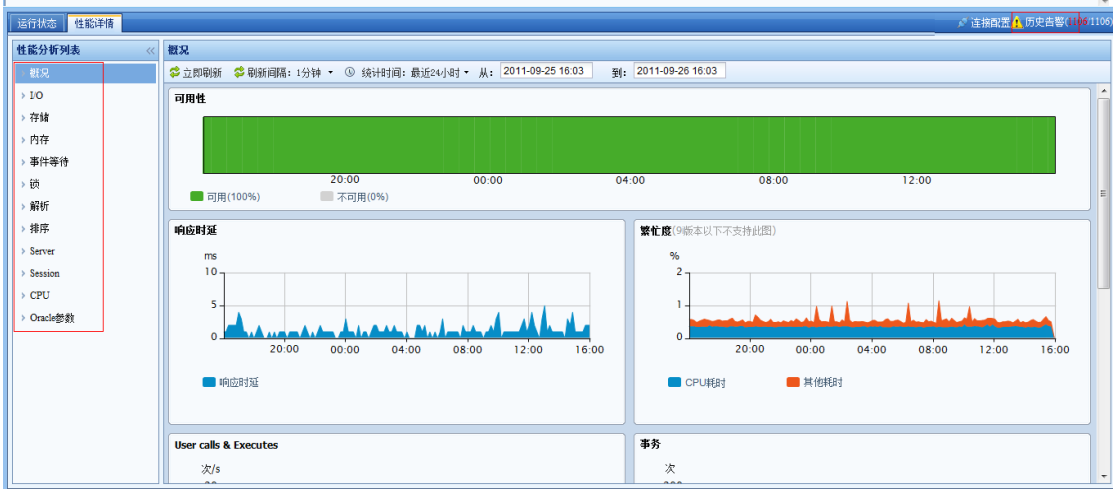
3.2.5 智能告警

部署负载均衡设备的用户，往往对业务系统的稳定性有着极高的要求，甚至需要对网络中发生的各种状况时刻关注，以便能及时排除故障隐患，从而避免发生业务停滞带来的巨大损失。深信服 AD 产品所特有的智能告警功能会基于管理员预先设定的关注信息，一旦网络或者业务出现故障的时候，将通过实时发送短信、邮件告警的形式通知相关管理人员进行维护；支持报警对象包括：链路、服务器、虚拟服务、双机切换、网络攻击等触发事件。

3.2.6 数据库中间件等后台应用组件深入分析

Oracle 数据库性能分析仅仅通过观察网络来定位系统问题，无法更确切地获知更深层次的原因，因为看不到除流量以外的信息。IT 管理者往往希望在出现性能问题的时候，能够定位到是网络问题、应用服务的问题还是后台数据库的问题，深信服 AD 还提供针对后台数据库的性能监控。

Oracle 版本	支持 Oracle 9i/ 10g/ 11g,Oracle RAC10g/11g 等版本无插件方式性能监控
监控优势	提供 Oracle 处理流程及关键性能指标的友好查看界面 可查看实时及历史信息 可以与数据库关联的硬件、应用、网络等信息联动，而非局限在数据库
性能指标	**包括**：数据库响应时间、当前总用户数、当前活跃用户数、繁忙度柱图、SQL*NET 接收速率、SQL*NET 发送速率、PGA 大小、PGA 使用、内存排序率、专用 Sever 数、共享 Sever 数、dispatcher 繁忙度、任务平均排队时间、redo 日志逻辑写速率、redo 日志物理写速率、SQL 执行速率、闪回日志写入速率、数据逻辑写速率、数据物理写速率、数据逻辑读速率、数据物理读速率、数据直接读速率、归档日志物理写速率、是否开启自动内存管理、SGA 大小、Redo Buffer 大小、Shared Pool 大小、Buffer Cache 大小、Large Pool 大小、Java Pool 大小、Redo 写入成功率、SQL 解析命中率、缓存命中率、死锁数、LGWR/RVWR/DBWR/ARCH 进程、数据库大小、重做日志组、归档日志、数据库 IO 时延等



创新优势：深信服 AD 无需在用户数据库服务器上安装插件，不会对用户当前 IT 架构造成任何影响。通过 Oracle/RAC 自身的接口读取相关信息，无风险。用户只需提供一个查

看数据库“系统视图”的普通管理员账号即可（例如 v$session,v$sysstat 这种视图的查询）。当添加正确的 Oracle 数据库后，系统即可自动识别 Oracle 版本及相关性能指标。

Weblogic 中间件性能分析 Weblogic Server 支持包括 EJB、JSB、JMS、JDBC、XML 和 WML 在内的业内多种标准，使 Web 应用系统的实施更为简单，并且保护了投资，同时也使基于标准的解决方案的开发更加简便。用户除了希望针对 WEB 做深入解码分析，还希望获知 Weblogic 的工作状况，深信服 AD 提供 Weblogic 性能监控。

Weblogic 版本	支持 Weblogic9.0/9.1/9.2/10.0/10.3/11g(10.3.1)等版本无插件性能监控
监控优势	提供 Weblogic 处理流程及关键性能指标的友好查看界面 可查看实时及历史信息 可以与数据库关联的硬件、应用、网络等信息联动，而非局限在数据库
性能指标	**包括**：weblogic 可用性、响应延迟等状态信息、JVM 信息、线程池使用、执行队列列表、等待处理的请求数、JDBC 信息、WEB 应用列表、会话数、Servlet 、EJB 应用缓存命中趋势、池的使用率、Bean 列表、JTA 活动处理数、时间、次数、JMS 消息数、等待消息趋势、server 详情等

创新优势：深信服 AD 无需在用户 Weblogic Server 上安装插件，不会对用户当前 IT 架构造成任何影响。通过 Weblogic 自身的接口读取相关信息，无风险。用户只需提供一个拥有 monitor 权限的用户账号即可。当添加正确的 Weblogic Server 后，系统即可自动识别 Weblogic 版本及相关性能指标。

3.3 安全技术

3.3.1 被动漏洞扫描

相比主动漏洞扫描工具或者是市场的漏扫设备，被动漏洞分析最大的优势就在于能实时发现客户网络环境的安全缺陷，且不会给网络产生额外的流量。此模块设计的初衷就是希望能够实时发现和跟踪网络中存在的主机、服务和应用，发现服务器软件的漏洞，实时分析用户网络中存在的安全问题，为用户展现 AD 的安全防护能力。实时漏洞分析功能主要可以帮助用户从以下几个方面来被动的对经过的流量进行分析：

Web 应用风险分析

针对用户 WEB 应用系统中存在的如下风险和安全问题进行分析：

1.SQL 注入、文件包含、命令执行、文件上传、XSS 攻击、目录穿越、webshell；

2.发现网站/OA 存在的设计问题，包括：

a)在 HTTP 请求中直接传 SQL 语句；

b)在 HTTP 请求中直接传 javascript 代码；

c) URL 包含敏感信息 :如 user、username、pass、password、session、jsessionid、sessionid 等；

Web 不安全配置检测

各种应用服务的默认配置存在安全隐患，容易被黑客利用，例如，SQL Server 的默认安装，就具有用户名为 sa，密码为空的管理员帐号。不安全的默认配置，管理员通常难以发觉，并且，随着服务的增多，发现这些不安全的配置就更耗人力。

AD 支持常用 Web 服务器不安全配置检测，如 Apache 的 httpd.conf 配置文件，IIS 的 metabase.xml 配置文件，nginx 的 web.xml 和 nginx.conf 配置文件，Tomcat 的 server.xml 配置文件，PHP 的 php.ini 配置文件等等

弱口令检测

支持 FTP，POP3，SMTP，Telnet，Web，Mysql，LDAP，AD 域等协议或应用的弱口令检查。

设备对弱口令的定义如下：

空口令

用户名和密码相同

a.长度小于等于 8 位的纯数字

b.长度小于等于 8 位的纯字母

c.长度小于等于 8 位的字典序

d.长度小于等于 6 位仅数字和字母

第4章 深信服科技简介

深信服科技股份有限公司成立于 2000 年，是专注于安全与云计算领域，致力于为用户提供更简单，更安全，更有价值的创新 IT 解决方案服务商。

目前，深信服在全球共设有 55 个直属分支机构，其中包括香港、新加坡、马来西亚、印尼、泰国、英国和美国等七个国际直属办事处和分公司，员工规模超过 3000 名。

随着企业规模的扩大发展，深信服也获得了多方认可。先后获得了“CMMI5 国际认证”、“第一批国家高新技术企业”、“国家规划布局内重点软件企业”“亚太地区德勤高科技高成长 500 强”等殊荣。同时，深信服还是 IPSec VPN 和 SSL VPN 两项国家标准的主要承建单位、并受邀参与制定《第二代防火墙标准》。在行业合作上，深信服是下一代互联网信息安全技术国家地方联合工程实验室、互联网应急中心应急服务支撑单位、国家信息安全漏洞共享平台 CNVD 成员单位、中国国家信息安全漏洞库 CNNVD 技术支撑单位和公共漏洞和暴露组织 CVE 认证合作单位。

目前，全球有近 40,000 家用户正在使用深信服的产品。其中，在中国入选世界 500 强的企业有 80%的企业都是深信服的用户。同时，凭借优秀的产品表现，深信服多款产

品入围了包括国家税务总局、国家电网、建设银行、工商银行、中国移动和中国电信在内的各行业集采，各款产品均得到了广泛应用。

时刻走在行业前沿，深信服始终保持着创新能力

多年来，深信服持续将年收入的 20%投入到研发，并在深圳、北京、长沙和硅谷设立了研发中心，研发人员比例达 40%。在对创新发展的持续投入下，深信服一直保持着每 1-2 年推出一款新产品、每季度更新 1 个新版本的研发速度。截至 2016 年 12 月，深信服共申请超过 400 项国内发明专利以及 20 项美国专利。此外，深信服是推出了全球第一台 IPSec VPN 和 SSL VPN 二合一 VPN，中国第一台上网行为管理和第一台下一代防火墙的厂商。

将产品和服务做到最好，深信服全情投入

深信服研发人员每月都会进行例行的客户拜访以收集产品需求，每年都能收到超过 1000 条有效需求，并在研发工作中将其迅速转化为产品新版本。同时，深信服在深圳、长沙、吉隆坡三地设有超过 100 坐席的 CTI 中心，提供 7*24 小时的电话咨询和远程调试服务。在全国范围内，深信服在 49 个城市设立了备品备件库，配有原厂工程师第一时间提供技术支持。

进入的每一个细分市场，深信服都会努力成为 No. 1

深信服的硬件 VPN、SSL VPN、上网行为管理、广域网优化等多款产品保持在市场占有率第一位；应用交付产品市场排名第二、也是排名第一的国产品牌。目前，深信服 SSL VPN、上网行为管理、下一代防火墙、广域网优化、应用交付、服务器虚拟化基础架构 6 款产品均入围了 Gartner 魔力象限，获得国际认可。

联系我们

咨询热线：400-806-6868

服务热线：400-630-6430 **（中国大陆）**

香港：(+852) 3427 9160

英国：(+44) 8455 332 371

新加坡：(+65) 9189 3267

马来西亚：(+60) 3 2201 0192

泰国：(+66) 2 254 5884

印尼：(+62) 21 5695 0789

官方微信，了解最新动态！