

1 第 3 章 Cache 系统架构 3

近年来, 由于 Web 用户的快速增长和供用户访问 Web 内容的增加, 共同导致了网络上 Web 流量的爆炸性增长。根据国外一项最新研究报告显示: HTTP (Hyper Text Transfer Protocol) 在所有网络流量中占到约 46%, 这是 HTTP 网络流量在过去四年里首次超越 P2P (Peer-to-Peer), 成为互联网带宽流量消耗首位。因此, 人们越来越关注这个问题。

Web 缓存技术已被证明是最有效, 也是应用最广泛的 WWW 加速技术。它通过减少文件在网络上的重复传输来降低网络带宽的浪费, 从而减轻 Web 服务器的负载, 最终降低用户的等待时间。在分层网络的边界部分, 采用 Web 缓存技术是一种投入少、效率高的提高访问速度的方法。

3.1 Cache 系统组成

Cache 系统通常包含 3 个子系统: 重定向子系统、Cache 子系统和 管理子系统, 如图 3-1 所示。

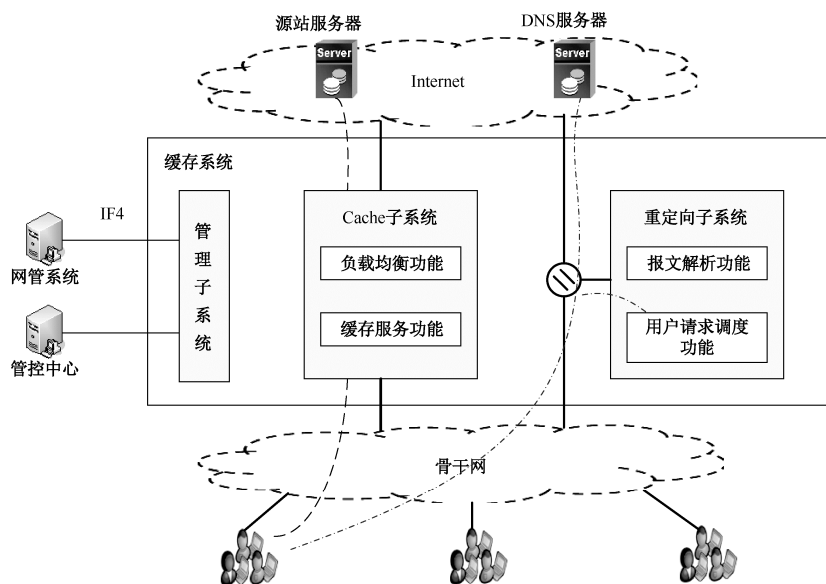


图 3-1 Cache 系统功能架构图

3.1.1 重定向子系统

重定向子系统的主要功能有时被称为调度子系统，一般包括数据监听模块、请求调度模块和负载均衡模块，是实现 Cache 能力的重要子系统之一。

1. 数据监听模块

数据监听模块能够根据应用层数据的特征字符串，对所监控链路的应用流量进行分析识别和过滤。从所有流量中识别出可 Cache 的应用会话，并转发给调度子系统。

数据监听模块支持对于满足设定规则匹配条件的 IP 报文进行过滤，将满足匹配条件的原始 IP 报文发送至用户请求调度设备。发送规则可以是指定协议类型、IP 地址、端口号、流量方向、应用层特征等组合。

数据监听模块能够分析链路中的业务原始流量数据，可基于 L3/L4 信息对数据流量进行分析和识别，可支持基于 L7 应用层特征（如应用层协议特征码）对报文进行分析和识别。

数据监听模块支持目前运营商网内主用的网络端口类型，如 GE、2.5G 或 10G 接口，后期能根据网络发展支持 100G 接口。

数据监听模块支持针对访问频次、带宽、流量、连接数等指标进行统计，支持输出外部资源的排名次序等统计数据，支持通过手动或自动方式上报至其他系统。

2. 请求调度模块

请求调度模块运用了负载均衡、Cache 记录搜索和热点内容管理调度等技术。作为整个 Cache 系统的调度控制中心，请求调度模块接受数据监听模块提供的用户请求，会判断资源是否已经 Cache，对于 Cache 的资源，将 Cache 服务器 IP 地址返回给用户，实现对用户请求引导至 Cache 的调度。

请求调度模块支持 DNS 重定向功能，采用 DNS 重定向或 DNS forward 等方式实现请求引导。同时也支持 HTTP 应用层重定向功能，监控用户请求，对于用户发起的 HTTP 请求，如果属于本地已 Cache 的资源，则由请求调度模块根据 Web Cache 系统的状态以及参数配置等情况，构造 HTTP 302 重定向消息发给最终用户。

请求调度模块支持通过多种方式监测 Cache 子系统的可用性及工作状态，能够根据 Cache 子系统的可用性及存储的文件内容进行调度。

请求调度模块具备本地系统的 Cache 内容管理功能，实时维护系统内部的 Cache 内容的情况，为请求调度提供基础数据支撑，同时支持与管理子系统进行本地 Cache 内容视图的共享。

3. 负载均衡模块

负载均衡模块的功能是将用户请求和 Internet 网络流量，按照配置的负载分担算法分发到不同 Cache 服务器进行处理。当用户请求到达 Cache 子系统时，负载均衡功能根据用户请求、服务器状态等策略，选择特定的 Cache 服务器，由该台服务器响应用户的请求，并提供

Cache 服务, 策略可以是选择性能最佳的 Cache 服务、存储该份内容的 Cache 服务器等。

负载均衡模块能够充分利用所有的 Cache 服务器的软硬件和网络资源, 将所有流量均衡地分配到 Cache 服务器, 尽量避免“不平衡”现象的发生。支持通过多种方式监控资源池中 Cache 服务器的健康状态, 如果某台服务器不可用, 则将该服务器从 Web Cache 服务队列中取出, 按照策略将负载分发到其他 Cache 服务器上。

负载均衡模块具备多台设备的冗余能力, 不同设备之间应当具有心跳监控及切换能力, 监控对端负载均衡设备的状态。备用设备在一定周期内没有收到主用设备的心跳信号, 支持主动将自己切换为主用设备。

如 Cache 系统未部署独立的负载均衡设备时, 负载均衡的功能可由请求调度模块兼做, 此时要求所有的 Cache 服务器均分配独立的公网 IP 地址, 直接面向用户终端提供服务。

3.1.2 Cache 子系统

Cache 服务模块提供对各类基于 HTTP 协议的互联网页面对象和文件的 Cache 能力, 对网内用户提供数据输出功能, 将 Cache 服务器内的数据传送至请求该内容的用户终端。接收到用户发送的访问请求时, Cache 服务模块需分析该请求是否在本地图中, 对于命中请求直接响应; 对未命中的请求, 则以代理方式将请求转发到源服务器, 下载源服务器的响应内容, 在本地图存储一个副本内容, 并同时返回给用户终端。

Cache 服务模块支持对网页的解析, 并对网页中的 html、xml、js、css、zip、mp3、图像等多种静态互联网 HTTP 对象进行 Cache。支持对不同静态对象的 Cache 策略配置, 管理员可通过修改配置达到不同类型文件是否 Cache 的目的。

Cache 服务模块支持对于 HTTP 的大文件的 Cache 加速功能, 如游戏软件、升级包等。Cache 服务模块接收用户的文件下载请求, 分析是否被 Web Cache 系统命中, 对命中的请求直接返回。

Cache 服务模块支持对 HTTP 视频文件的 Cache, 主要针对各种在线视频网站, 文件类型包括 FLV、MP4、MOV、WMV、RMVB、F4V 等文件格式。Cache 服务模块接收用户的媒体访问请求, 分析请求是否被 Cache 服务模块命中, 对命中的请求直接返回, 对不命中的请求, 转发到源站获取相应内容进行 Cache 后, 再将视频对象返回给用户。

对于不能 Cache 的动态内容, Cache 服务模块不在本地图进行 Cache, 而是不 Cache 或采用代理方式向源网站下载数据并传送给用户, 以避免访问失败。

Cache 服务模块支持多维度的 Cache 策略管理功能, 系统管理员可以严格、清晰地定义 Cache 服务模块的访问控制策略, 将通过 ACL 访问列表, 支持对用户的请求进行访问控制, 并根据不同策略能够灵活组合使用。

Cache 服务模块支持对本地图 Cache 文件的管理和存储, 并根据热度策略实现对 Cache

文件的更新替换。在硬盘存储到一定比率时，能够根据对象的访问频率和文件大小核算对应的热度值，访问频率越高权值越高，而热度值较低的 Cache 对象将会被优先删除。

3.1.3 管理子系统

管理子系统面向系统管理员提供网络管理功能和系统管理功能，支持功能如下：

状态监测：实现本地网管功能，对调度子系统和 Cache 子系统的可用性、设备性能、网络指标进行实时监测，实时获得业务系统的资源使用情况和健康状态。

系统配置：提供 Web 管理界面，对系统的工作模式、参数、策略进行配置管理，可以依据用户、URL 等维度实施区分处理，支持对 Cache 内容的监管配置管理，提供热点管理相关策略等参数的配置。

报表日志：对监控产生的各种数据进行记录和分析，将自动生成常规报表和个性化报表，以支撑各类分析管理需要。

告警管理：对系统中产生的告警信息进行统一收集，并根据告警级别调用相应的策略进行处理。

内容管理：支持分析从其他模块获取的流量及用户访问数据，以实时维护 Cache 系统的内容分布信息，形成内容资源视图，并支持将内容视图数据通过接口传送至其他设备。

用户管理：支持设置多个不同权限的管理员用户，支持对管理员操作记录日志审计，支持短信/邮箱方式通知。

字典管理：支持由管理员自定义地域字典和网站字典，地域字典中存储特定 IP 地址范围，网站频道字典存储特定资源 URL 的集合，频道字典支持通配符*、?，支持冲突检查，支持导入和导出操作。

节点管理：支持对于各 Cache 节点的手工增加、删除、修改操作，支持对重定向系统和 Cache 系统的可用性、设备性能、网络性能等指标进行实时监测。

Cache 策略管理：支持定义多级 Cache 策略，支持下发至所有 Cache 系统，优先级可灵活调整，支持引用地域字典和网站字典，支持冲突检查，支持导入和导出操作。

黑名单管理：支持定义 IP 地址、网站、URL 级别的黑名单，能够下发至所有 Cache 系统，要求黑名单中的资源不提供 Cache 服务。

关键字管理：支持定义、配置全局文件名称的敏感关键字，能够下发至所有节点的 Cache 系统，对于包含敏感关键字的 Web 对象，要求 Cache 系统不提供 Cache 服务。

备份还原：支持系统配置信息的自动/手动备份，支持操作回退。

接口管理：支持定义与 Cache 节点管理服务服务器接口的绑定 IP 地址、访问端口、用户名、密码、IP 地址访问列表等信息，支持记录接口同步日志，支持超时和失败告警。

3.2 Web Cache 业务流程

3.2.1 Web Cache 组网方案

Web Cache 系统通常部署在网络出口附近，采用串行方式或并行方式，接入到互联网中，典型部署方案主要分为 DNS 引导模式、HTTP 重定向模式和策略路由模式。

DNS 引导模式下的缓存系统，是由本地 DNS 配合缓存系统进行用户访问的引导，一般分为两种方式：DNS 劫持和 DNS Forward。DNS 劫持是通过用户访问本地 DNS，由出口路由器侧的 DPI 设备监测用户请求，把域名解析引导至 Cache 系统的调度模块，由重定向子系统反馈域名解析结果，从而指向 Cache 缓存模块，以实现用户的缓存系统访问。而 DNS Forward 方式则可以在本地 DNS 中直接进行配置，将用户的访问请求引导至缓存系统，如图 3-2 所示。

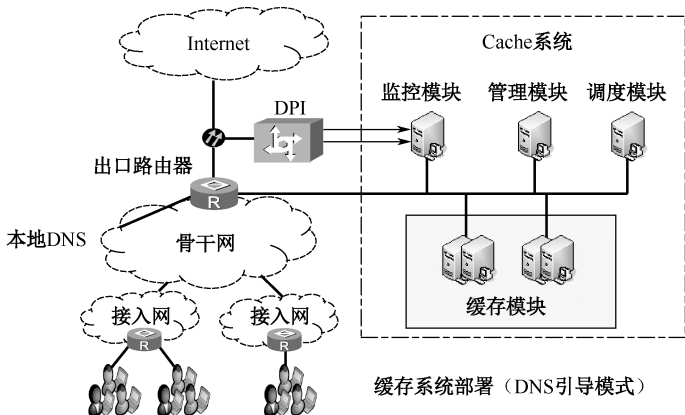


图 3-2 互联网缓存系统部署示意图——DNS 引导模式

HTTP 重定向模式下的缓存系统，组网部署与 DNS 引导模式一致，区分在于用户获取授权 DNS 解析结果后，在访问源网站时，由重定向子系统识别 HTTP 数据，将用户请求跳转至缓存系统进行服务，如图 3-3 所示。

策略路由模式下的缓存系统，与 HTTP 重定向模式类似，都是用户在获取授权 DNS 解析结果后开始执行调度，区分在于重定向操作是由路由器识别 HTTP 数据，再应用策略路由的方式，将请求重定向到缓存系统进行服务，如图 3-4 所示。

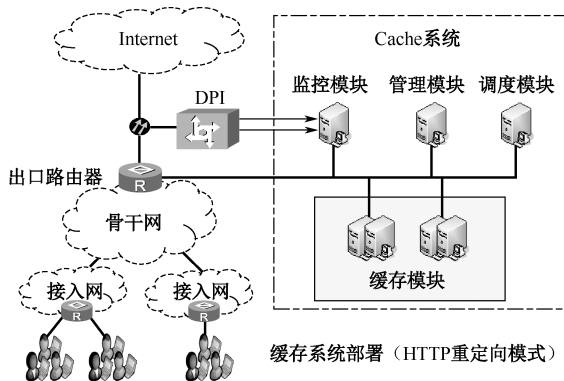


图 3-3 互联网缓存系统部署示意图——HTTP 重定向模式

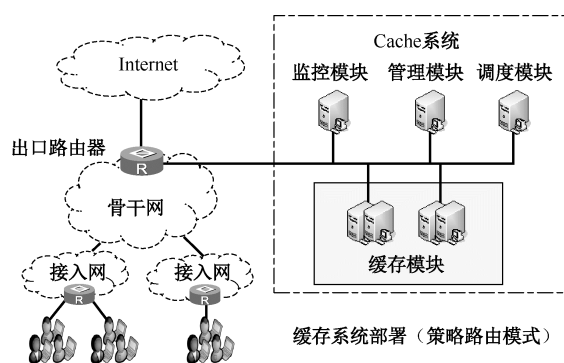


图 3-4 互联网缓存系统部署示意图——策略路由模式

3.2.2 Web Cache 系统技术实现方式

目前业界传统的 Web Cache 系统技术实现方式主要包括 DNS 重定向、HTTP 重定向和策略路由 3 种方式。

(1) DNS 重定向

DNS 重定向方式是指 Web Cache 系统主动或被动缓存网站的内容,通过配置相关策略劫持 DNS 解析请求,如果用户访问某网站 (Web Cache 已缓存),直接返回 Web Cache 系统的 IP 地址,由 Web Cache 系统响应客户请求。DNS 重定向的配置策略主要包括 2 种途径:

① 在 DNS 服务器上配置,由 DNS 服务器判断用户的 DNS 解析请求是否是访问某网站。若是,则 DNS 服务器向用户直接返回 Web Cache 系统的 IP 地址;若不是,则返回源服务器的地址。

② 通过在网间出口路由器上旁路部署 DNS 重定向系统,通过 DPI 设备复制用户的 DNS 解析请求,由重定向子系统判断用户是不是访问某网站。如果是,则向用户返回 Web Cache 系统的 IP 地址,如果不是,则不处理,由外网 DNS 服务器返回源站地址。

(2) HTTP 重定向

HTTP 重定向方式 (HTTP 302 重定向) 主要针对 HTTP GET 应用层请求进行解析,在

网间出口链路处部署深度分组解析 (DPI) 设备, DPI 深度分析旁路复制的数据流量并把 HTTP 请求报文复制给重定向子系统, 重定向子系统对 HTTP 请求进行判断。若用户请求的内容命中, 即已经缓存在指定的 Web Cache 系统中, 则将 Web Cache 系统相应服务器的 IP 地址返给用户, 用户重新发起至 Web Cache 的 HTTP 请求, 并丢弃正常 HTTP 请求返回结果 (网内速度快于网外返回的正常 HTTP 请求)。若用户请求的内容没有命中, 则不对用户请求进行任何操作, 用户等待正常 HTTP 请求并按照原来的路径向外网获取内容。

(3) 策略路由

策略路由方式是指 Web Cache 系统旁挂在骨干网网间出口路由器上, 当用户请求的数据到达出口路由器时, 路由器通过策略路由将用户请求转发到 Web Cache 系统。如果用户请求的数据已经在 Web Cache 系统中缓存, 即命中, Web Cache 系统会直接将数据返回给用户。如果用户请求的数据未在 Web Cache 系统中缓存, 即未命中, Web Cache 系统会主动到互联网请求数据, 将响应数据发送给用户, 同时在 Web Cache 系统中保留一份数据, 从而达到一次缓存多次服务的目的。

综合考虑以上几种 Web Cache 技术, 其中策略路由方式和 HTTP 重定向方式有以下不足之处:

① 策略路由方式为静态路由时, 不能根据网络状态变化动态改变路由策略, 将在网络链路故障或设备故障时出现路由环路, 引起整个网络的震荡, 导致业务产生故障。

② HTTP 重定向方式大部分采用对 HTTP 大文件或 HTTP 下载进行重定向, 很少对 Web 页面的小文件 (小于 1M) 进行 HTTP 重定向, 且用户 IE 浏览器地址栏显示的为 Cache 系统的 IP 地址, 不为源服务器的域名, 有可能会引起用户投诉。

综上所述, 运营商一般均会采用 DNS 重定向方式进行 Web Cache 系统的部署。而对于两种 DNS 重定向方式, 由于修改全国各省的 DNS 配置, 需要对省 DNS 进行统一管理和维护, 会增加工程的复杂度。所以, 运营商在全网部署 Web Cache 系统时, 一般会采用在网间出口路由器上旁路部署 DNS 重定向系统的方式, 此种方式对省 DNS 无改造需求, 能够更快速地进行部署。

另外, 在实际部署中需根据业务场景需求, 还要结合实际网络条件选择使用。在面向互联网网站、Web 小文件场景进行缓存加速时, 建议 Web Cache 系统优先采用 DNS 重定向模式。策略路由引导模式应在面向城域网出口部署的 Web Cache 系统中使用, 主要面向小文件及网站对象的缓存加速。对于省网及一级 Web Cache 系统, 考虑到策略管理的复杂度及网络可靠性, 建议谨慎使用。

3.2.3 DNS 重定向模式

Web Cache 系统通过检测用户侧发出的 DNS 解析请求, 如果用户访问的站点域名属于

系统配置的白名单,则在 DNS 解析响应消息中向用户返回 Web Cache 系统的 IP 地址,引导用户的 HTTP 业务请求发送至 Web Cache 系统,由 Web Cache 系统响应用户访问请求。

在面向互联网网站、Web 小文件场景进行 Cache 加速时,建议 Web Cache 系统优先采用 DNS 重定向模式。

对于文件下载、视频播放类的应用场景,也可选择采用 DNS 重定向模式。但对于以 IP 地址标识的对象,由于在下载过程前没有 DNS 解析流程,故无法通过 DNS 重定向模式进行处理。

在 DNS 重定向模式下,Web Cache 系统可通过分光镜像和 DNS Forward 两种方式获取用户发起的 DNS 解析请求,在实际部署中建议根据网络部署情况选择使用。

分光镜像方式要求在需监测的链路上部署无源分光设备,将链路中的信号通过分光处理后发送至 Web Cache 系统的报文解析功能中。

报文解析功能中应根据需要配置过滤流量的条件,如 DNS 协议类型、端口号等,将满足条件的用户请求转发送至后端的请求重定向功能模块。

在报文解析功能中不开启流量过滤功能时,报文解析功能会将全量的上行请求均发送至 Web Cache 系统。此时由 Web Cache 系统需根据本地配置的域名白名单进行匹配,如满足一致性条件,再向用户返回对应的重定向响应消息。

由于分光镜像模式需要在源网站的授权 DNS 返回响应结果之前对用户进行重定向,故要求重定向子系统发送重定向报文应当满足时延要求。

DNS 劫持采用分光镜技术,其业务流程如图 3-5 所示。

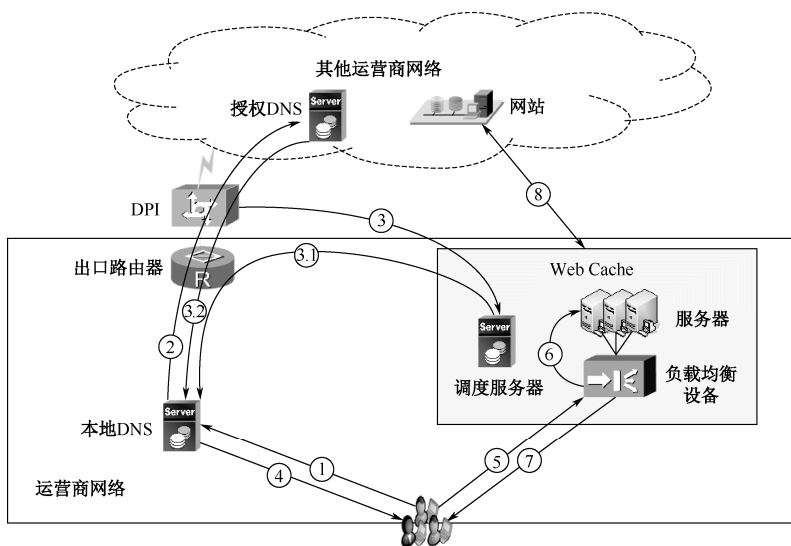


图 3-5 DNS 劫持模式的业务流程

流程中各步骤说明如下：

- (1) 用户发起网站域名的 DNS 请求。
- (2) 本地 DNS 查询本地 Cache,如果有,则返回结果给用户;如果没有,则向外网发起域名的 DNS 请求。

- (3) 网间 DPI 将 DNS 请求复制给 Web Cache 调度器。
- ① 调度器抢先响应 DNS 请求，回复 Web Cache 系统地址。
- ② 外网 DNS 响应较晚，被丢弃。
- (4) 用户获得域名解析结果，地址为 Web Cache 系统地址。
- (5) 用户访问 Web Cache。
- (6) Web Cache 负载均衡器将用户请求分发到一台 Cache 上。
- (7) Web Cache 命中则直接回应用户；不命中则从外网取回数据后回应用户。
- (8) 对于不命中的内容，Web Cache 服务器从外网取回数据。

对于用户发起的访问请求，当 Web Cache 系统未命中时，Cache 服务器可通过查询本地 DNS，或使用系统内置的 DNS 功能来实现面向互联网网站的域名解析功能。

本地 DNS 可有效减少 Web Cache 系统对外的大量 DNS 解析请求，以提高 DNS 响应速度，也可避免向内网 DNS 服务器查询时产生的 DNS 环路。

对于 Web Cache 系统发起的 DNS 解析请求，应当通过 IP 地址配置避免被系统再次拦截。

DNS 转发模式要求将省内的 Local DNS 与 Web Cache 系统相连接。由管理员在 Local DNS 上开启 Forward 功能，并在转发策略中将需要加速的域名列表(白名单)配置为 Forward First 模式，目标为 Web Cache 系统的调度服务器的 IP 地址。在 Forward First 模式下 DNS 服务器会优先选择转发目标 DNS 返回的解析结果，当 Forward 目标因故无法返回结果或返回结果延迟较大时，Local DNS 本身则会继续递归解析以获取解析结果。

当用户终端发起的 DNS 解析请求发送至 Local DNS 服务器后，DNS 服务器将目标域名与本地配置的 Forward 名单进行匹配，如满足一致性条件，则将该条 DNS 解析请求前传至 Web Cache 系统的请求调度功能模块。

Web Cache 系统的请求调度功能模块在接收到该条 DNS 解析请求后，应向 Local DNS 返回对应的 DNS 解析响应消息，解析结果的目标地址为 Cache 子系统的 IP 地址。

DNS 转发模式的业务流程如图 3-6 所示。

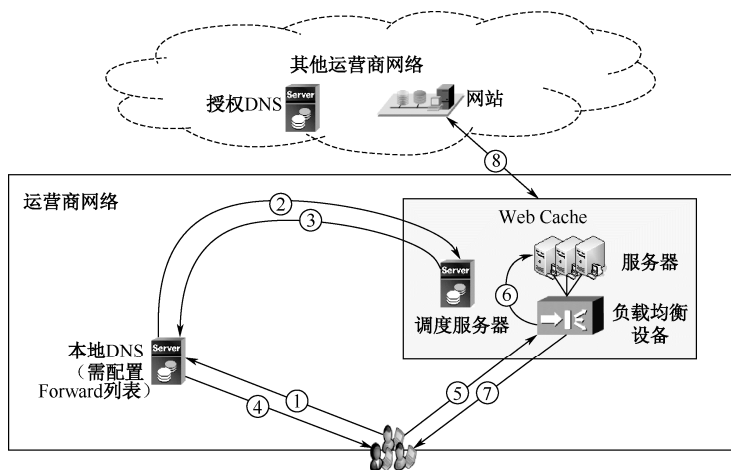


图 3-6 DNS 转发模式的业务流程

流程中各步骤说明如下：

- (1) 用户发起网站域名的 DNS 请求。
- (2) 本地 DNS 查询本地 Cache, 如果有, 则返回结果给用户 ;如果没有, 则检查 Forward 列表, 如果有该域名, 说明该域名下的内容已被 Web Cache 系统 Cache, 则将 DNS 请求转发至 Cache 调度器 ; 否则向外网 DNS 进行查询。
- (3) Cache 调度器根据本地调度策略, 回复 Web Cache 系统地址。
- (4) 用户获得域名解析结果, 地址为 Web Cache 系统地址。
- (5) 用户访问 Web Cache。
- (6) Web Cache 负载均衡器将用户请求分发到一台 Cache 上。
- (7) Web Cache 命中则直接回应用户 ; 不命中则从外网取回数据后回应用户。
- (8) 对于不命中内容, Web Cache 服务器代理用户从外网取回数据后, 再返回给用户。

3.2.4 HTTP 重定向模式

HTTP 重定向模式包含非代理方式和代理方式两种处理机制。

非代理处理机制：仅当 Web Cache 系统监测到用户发出的 HTTP 访问请求属于本地已 Cache 的资源后, 才由 Web Cache 系统向用户返回 HTTP 302 重定向报文, 响应报文的目标地址为 Web Cache 系统的 IP 地址, 用户终端接收到该条 HTTP 302 响应消息后, 将向 Web Cache 系统发起请求下载数据。

代理处理机制：HTTP 重定向模式也支持工作于代理模式, 对于满足域名匹配条件的请求均引导至 Cache 子系统, 如已 Cache 时, 则直接向用户提供服务 ; 如该内容在本地未 Cache 时, 则由 Web Cache 系统作为代理向外网下载并传送给用户侧。

在面向大文件下载、视频播放类的场景进行 Cache 加速时, 建议 Web Cache 系统优先采用 HTTP 重定向模式及非代理处理机制。

在 HTTP 重定向机制下, Web Cache 系统应采用分光镜像方式来获取用户发起的 HTTP 请求。通过在需监测的链路上部署无源分光设备, 将链路中的信号通过分光处理后发送至 Web Cache 系统的报文解析功能中。报文解析功能中应根据需要配置过滤流量的条件, 如 HTTP 协议类型、端口号 (80/8080)、关注域名或 IP 地址列表等维度, 将满足一致性要求的 HTTP 请求转发送至后端的请求重定向功能模块。

由于分光镜像模式需要在源网站返回响应结果之前对用户进行重定向, 故要求重定向子系统发送重定向报文应当满足时延要求。

在特定场景下, 也可通过端口镜像方式来获取用户请求。端口镜像模式主要面向网络流量较小的场景, 通过在路由器或交换机上开启端口镜像功能, 将所有流量均通过特定端口转发至 Web Cache 系统, 并由报文解析设备过滤出所需的 HTTP 流量。端口镜像方式可

进行全流量镜像，也可依据端口号、目标地址、源地地址等进行镜像。

HTTP 拦截的业务流程如图 3-7 所示。

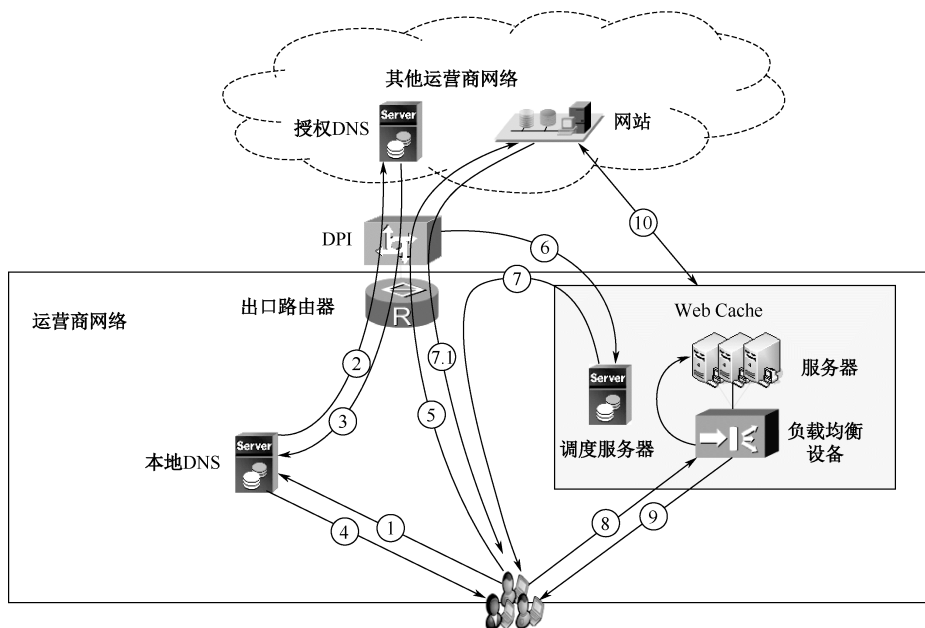


图 3-7 HTTP 拦截的业务流程

流程中各步骤说明如下：

- (1) 用户发起网站域名的 DNS 请求。
- (2) 本地 DNS 查询本地 Cache，如果有，则返回结果给用户；如果没有，则向外网发起域名的 DNS 请求。
- (3) 外网 DNS 进行响应。
- (4) 本地 DNS 收到响应后，转发给用户。
- (5) 用户根据 DNS 响应，向网站发送 HTTP get 请求。
- (6) 省网出口 DPI 设备将 HTTP get 请求转发至调度器。
- (7) 调度器查询本地 URL 列表，如果没有找到则不进行任何响应；如果找到对应的 URL，则模拟网站进行响应，通过 HTTP 重定向指示用户访问 Cache 系统。
- (8) 网站的 HTTP 响应不在网内，响应较晚，被丢弃。
- (9) 用户根据调度器的指示访问 Web Cache 系统。
- (10) Web Cache 命中则直接回应用户；不命中则从外网取回数据后回应用户。
- (11) 对于代理模式，不命中内容，Web Cache 服务器代理用户从外网取回数据后，再返回给用户。对于非代理模式，若未命中内容，会放行用户请求，重定向子系统累计请求次数，当热度达到阈值后，为后来的相同请求服务。

3.2.5 策略路由引导模式

策略路由模式主要应用于省网内部署的 Web Cache 系统，通过在核心路由器上开启策略路由（PBR）功能，将满足条件的用户流量通过路由转发至 Web Cache 系统，判断依据可以是端口号、目标地址、源地址等，后续相关的上行、下行流量均会经由 Web Cache 系统进行处理，而无须向终端侧发送重定向报文。策略路由模式下具备双向透明的 Cache 处理能力，在此过程中，用户端、服务器端的 IP、端口、URL 和方法均不发生任何改变。

对于用户请求，如果 Web Cache 系统本地命中，则直接向用户返回数据；如果本地未命中，则由 Web Cache 系统代理用户向源网站请求数据，并返回给用户。对于热点内容，由 Web Cache 系统在本地存储一份副本。

在策略路由引导模式下，当网络中断时，路由器应能够自动调整为直通模式，以避免可能出现的访问故障。

在策略路由引导模式下，也可采用发布 BGP 路由的方式，将需加速的流量从核心路由器汇聚转发到 Web Cache 系统专用的路由器，然后再在该台路由器上面配置 PBR 策略进行流量引导。

策略路由引导模式应面向城域网出口部署的 Web Cache 系统中使用，主要面向小文件及网站对象的 Cache 加速。对于省网及一级 Web Cache 系统，考虑到策略管理复杂度及网络可靠性，建议谨慎使用。

策略路由方式的业务流程如图 3-8 所示。

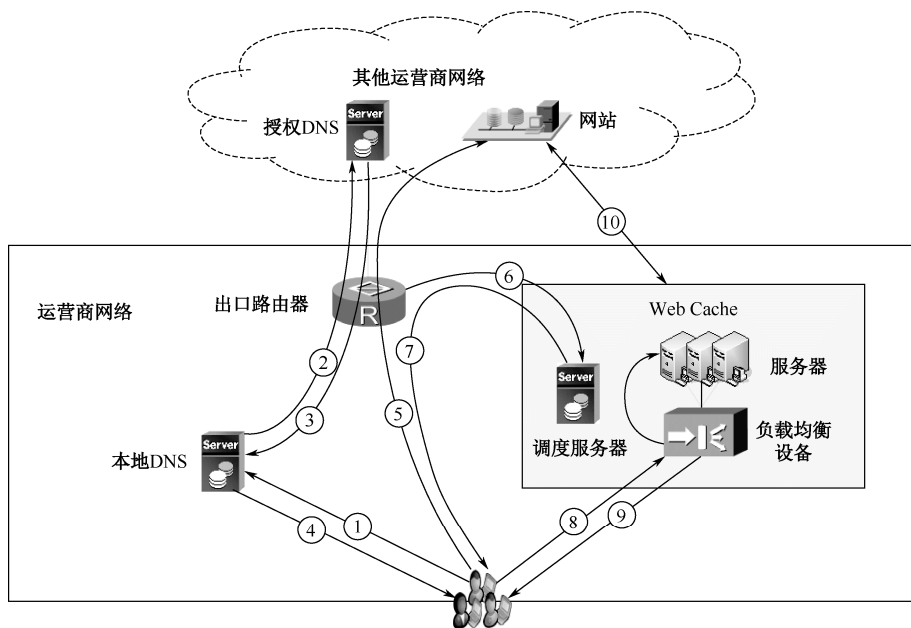


图 3-8 策略路由方式的业务流程

流程中各步骤说明如下：

- (1) 用户发起网站域名的 DNS 请求。
- (2) 本地 DNS 查询本地 Cache，如果有，则返回结果给用户；如果没有，则向外网发起域名的 DNS 请求。
- (3) 外网 DNS 进行响应。
- (4) 本地 DNS 收到响应后，转发给用户。
- (5) 用户根据 DNS 响应，向网站发送 HTTP get 请求。
- (6) 出口路由器根据策略，将全部 HTTP 请求转发至调度器。
- (7) 调度器查询本地 URL 列表，并模拟网站进行响应。
- (8) 用户根据调度器的指示访问 Web Cache 系统。
- (9) Web Cache 命中则直接回应用户；不命中则从外网取回数据后回应客户。
- (10) 对于不命中内容，Web Cache 服务器代理用户从外网取回数据后，再返回给用户。

以上 4 种实现方式中，DNS 劫持及 HTTP 拦截方式（采用分光技术）主要的优点是采用旁路监听方式，不影响承载网络构架，不需要在现网路由器做任何数据配置工作，且灵活支持对 HTTP 小文件、大文件等流量的优化；在分光旁路部署时，支持 HTTP 302 和 DNS 两种重定向方式，分光重定向缺点是部署方案较复杂。而 DNS 转发方式与现网 Web Cache DNS 系统关联部署方案是较为常用的一个方式，具有部署方案简单，成本低，能够实现对 HTTP 小文件很好控制，缺点是需要 local DNS 进行配置重定向的域名，且不支持以 IP 地址开头的 URL 资源。策略路由方式需要在出口路由器上进行策略路由配置，较为复杂，Web Cache 系统上线时需对现网路由器配置进行较大改变，目前网内按照这种方式实施的不多。

3.3 P2P Cache 业务流程

3.3.1 P2P Cache 系统原理

P2P Cache 系统节省带宽的原理主要是分为 3 个技术：缓存、主动挂断和内网 peer 优先。

(1) 缓存技术节省入流量。通过部署一定数量的 PPC（数量和配置根据网络出口流量确定），达到一定的命中率后，开启 PPC 缓存功能，开始缓存网络数据流量，开启 PPS 重定向功能，将用户对外部的 BT 请求重定向到 PPC 上，不再允许用户直接通过运营商出口获取 BT 流量，而通过 PPC 分发给用户的流量 `local_snd` 必然大于 PPC 从外部获取的流量 `remote_rcv`，以达到节省运营商入流量的目的。

(2) 主动挂断节省出流量。虽然用户对外部 BT 的请求被重定向到 PPC 上,但是由于 PPS 属于旁路设备,用户的第一次 BT 请求报文还是会正常传到外部 Tracker 上,而外部 Tracker 记录下用户的联系地址和端口以及 infohash 后,外部 peer 请求相同 infohash 时就会获取内部用户的联系地址和端口,并从内部用户获取数据流量。因此,PPS 的另一个重要功能就是可以挂断外部 peer 的请求,阻止外部 peer 对内部用户的流量请求,以达到节省运营商出流量的目的。

(3) 内网 peer 优先节省出入流量。BT 协议中没有对内网 peer 的优化技术,因为 Tracker 服务器并不能判断哪些 peer 属于内网,应该优先互相传输。使用 P2P Cache 系统的 LBE 模块,只要运营商指定 IP 地址段,LBE 模块就会将内网保存有相同 infohash 内容的 peer 联系地址和端口优先传送给请求用户,而不需要向外网 peer 进行请求,以达到节省双向流量的目的。

3.3.2 P2P Cache 系统特点

P2P Cache 系统特点如下:

(1) 部署简单,无须更改运营商网络

P2P Cache 系统的各个模块都无须运营商对网络进行大的调整,只需要将各个模块运行的服务器配置相应的地址即可。PPS 由于必须捕获出口流量,需要与出口路由器部署在一起,在出口路由器上做一次端口镜像或者进行分光,将镜像端口或分光端口连接至 PPS,不增加网络故障点(有源故障点)。因此,P2P Cache 系统部署时比较灵活,割接也比较方便,即使出现故障,也不会对运营商现有业务造成任何影响。

(2) 效率高

由于用户对相同文件的请求重复率比较高,根据实际统计数据,7 天可以达到 2:1 左右,也就是 P2P Cache 系统对 P2P 等流量的缓存如果达到 7 天,理论上可以节省一半的流量。从实际部署的情况来看,P2P Cache 上线后基本上可以节省运营商 P2P 流量的 30%。流量峰值期间,P2P Cache 的命中率达到 70%以上,缓解 P2P 流量压力达到 60%以上。P2P Cache 系统平均命中率为 70%以上,缓解 P2P 流量压力大于 30%。

(3) 各模块耦合度低,部署适用范围广

P2P Cache 系统的各个功能模块,如 PPS、PPC 和 PPM 相互之间的耦合度比较低。PPC 服务器作为缓存代理服务器,可动态添加或者减少机器数量配置,并且不要求配置在一个物理或者逻辑位置,只要用户和 P2P Cache 系统的各个服务器模块能够访问到即可。

(4) 安全性高

在 P2P Cache 中,P2P Cache 系统的各个服务模块均进行了冗余部署设计。PPS 对用户的连接封闭和转发功能在服务器发生严重故障时可以自行停止服务,不影响网络的正常使

用。某些 PPC 服务器不能为用户提供服务时，其他的服务器可动态重新组合，接受用户的所有请求。

(5) P2P 用户行为分析

P2P Cache 系统可准确地对 P2P 内容进行分析，得知用户具体下载行为，真正做到 4W (who/where/what/when)，使网络管理者有针对性地对网络进行管理。

(6) 用户体验提高

P2P Cache 系统不仅可以提高内网用户 P2P 应用体验。同时由于缓解出口的 P2P 流量，提高了其他关键应用的用户体验，如 HTTP、game 等。

(7) 强大的数据处理能力

P2P Cache 系统的关键设备 PPS 服务器部署在网络出口，对进出网络的流量进行分析，并对 BT、迅雷、迅雷看看、HTTP 下载、HTTP 视频、PPLIVE、电驴等请求报文进行重定向功能。目前单台 PPS 最高处理能力达到 8G（双向 4G）。

(8) 管理方便

P2P Cache 系统提供的 PPM 管理模块，可灵活统计运营商出口流量的分布情况，对目前常见 P2P 流量可进行分析，并可统计 PPC 缓存效率，以及各个服务器的系统状况。对于运营上加强出口流量管理具有重要意义。

3.3.3 P2P Cache 组网方案

P2P Cache 服务器，通过新建核心交换机与骨干网相连接，获取流量分析与资源引导服务器过滤后流量，对用户请求进行响应。对于 Cache 文件阈值设置，可在资源引导服务器中进行设置，对于达到阈值热点资源，Cache 服务器从外网进行下载。P2P Cache 组网方案如图 3-9 所示。

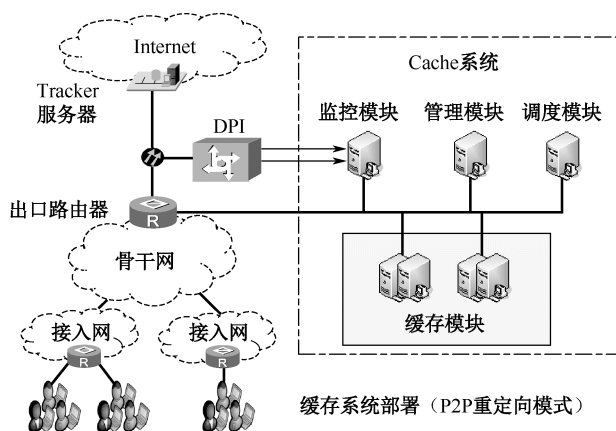


图 3-9 P2P Cache 组网方案

网络内容服务商（ICP，Internet Content Provider）非合作模式下的 P2P 缓存系统，由

管理子系统对未命中的资源进行热点排序，调度模块则负责判断资源是否达到阈值，通知缓存模块向外网 Tracker 服务器进行资源缓存，后续用户访问将通过重定向方式跳转至缓存模块。

ICP 合作模式下的 P2P 缓存系统，由缓存服务器事先获取 ICP 网络的 ACL，用户访问将通过 ICP 网络分析用户的 IP 来源，跳转指向对应的缓存节点来提供服务。P2P Cache 组网方案（ICP 合作模式）如图 3-10 所示。

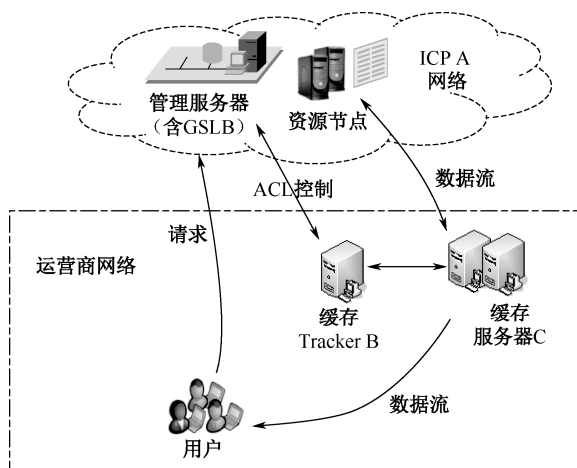


图 3-10 P2P Cache 组网方案（ICP 合作模式）

3.3.4 通用 P2P Cache 调度模式

对于公开协议和采用破解类私有协议的 Cache 业务，其流程如图 3-11 所示。

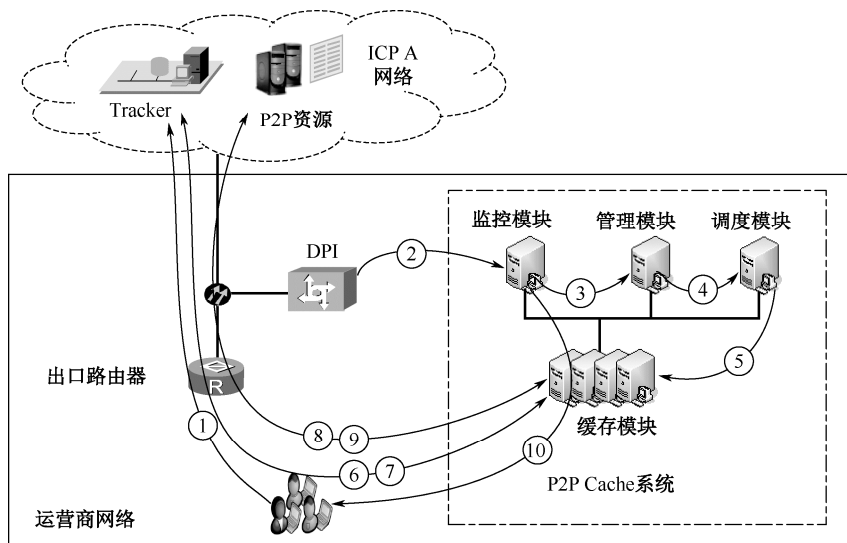


图 3-11 P2P Cache 的业务流程图

流程中各步骤说明如下：

- (1) 用户向外网 Tracker 服务器发起 P2P 请求。
- (2) 监控子系统判断资源是否缓存，没有缓存则上报资源信息给管理子系统。
- (3) 管理子系统对热点进行排序。
- (4) 管理子系统定时上报 TOP 热点资源给调度子系统。
- (5) 调度子系统判断该资源是否达到热点阈值，如果达到热点阈值则通知缓存子系统下载该资源。
- (6) 缓存子系统收到调度子系统该资源已经达热点阈值，并开始下载该资源，向 Tracker 服务器发起 P2P 请求。
- (7) Tracker 服务器响应请求返回 P2P 资源列表。
- (8) CSS（缓存子系统）向外网源站及用户请求资源。
- (9) 外网资源节点提供资源给缓存子系统，在下载完成后，向调度系统上报缓存资源信息。
- (10) 监控子系统发送缓存子系统信息给客户端，客户端访问缓存子系统并获取资源。

3.3.5 ICP 合作调度模式

ICP 合作调度模式的业务流程如图 3-12 所示。

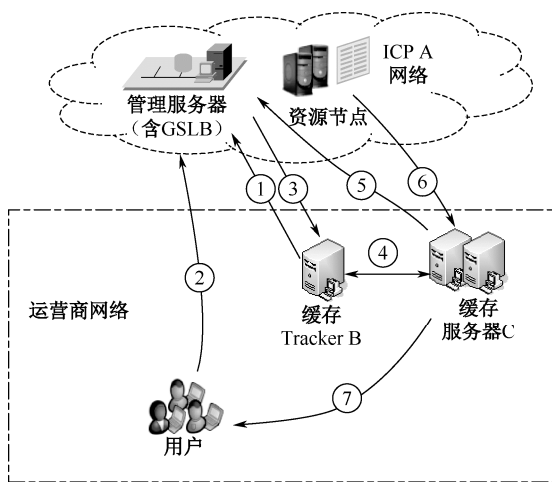


图 3-12 P2P Cache 业务流程图 (ICP 合作模式)

- (1) 运营商向 ICP A 网络提供网内用户 IP 的 ACL（访问列表）。
- (2) 运营商中的用户向 ICP A 网络请求 P2P 资源。
- (3) ICP A 网络分析用户的 IP 来源，若满足运营商 ACL，即向用户提供运营商网络内的资源信息，并请运营商的缓存 Tracker B 服务器提供数据。

(4) 缓存 Tracker B 与其互联的缓存服务器 C 数据交互。若缓存服务器 C 没有资源, 则服务器 C 操作步骤 (5); 若缓存服务器 C 有请求的资源, 则服务器 C 操作步骤 (7) (见图 3-12)。

(5) 缓存服务器 C 向 ICP A 网络中的管理服务器和资源节点发起缓存请求。

(6) 缓存服务器 C 进行资源的高速缓存, 缓存结束向上通知 ICP A 网络。

(7) ICP A 网络在收到新的资源信息后, 向用户更新资源信息, 用户即可向缓存服务器 C 请求资源的下载。