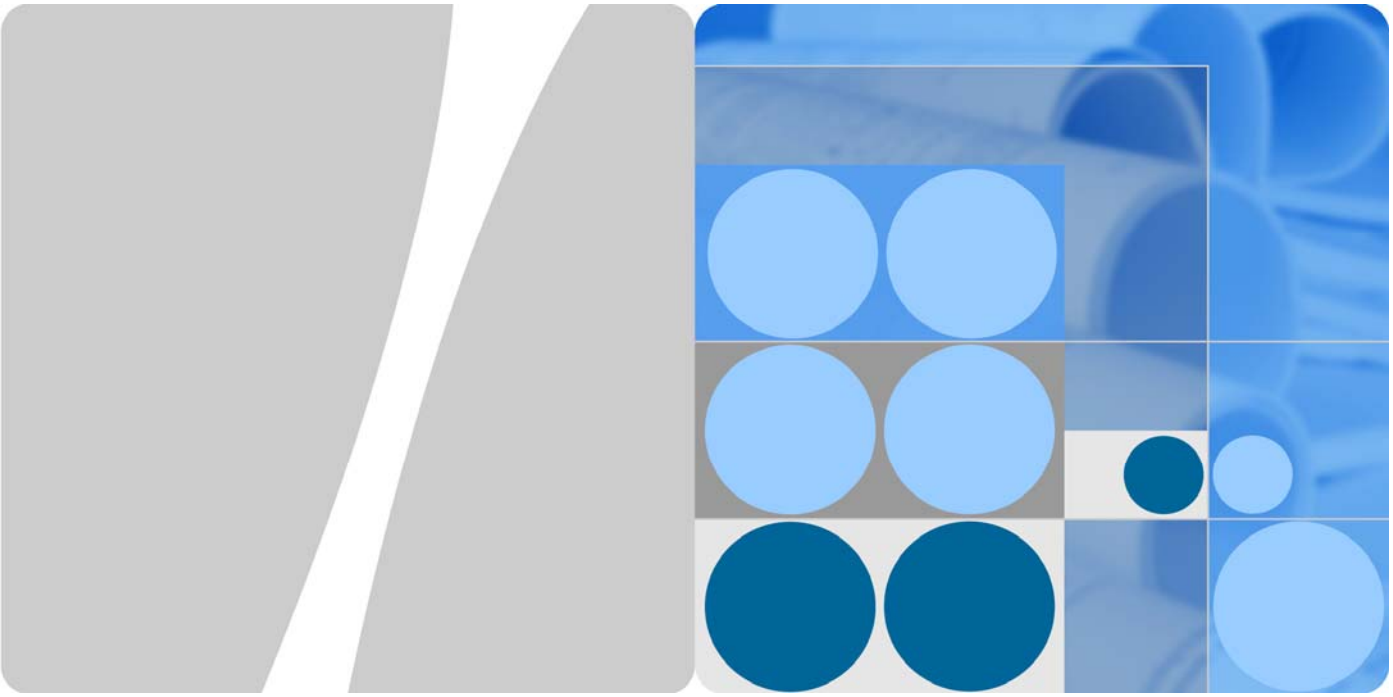


Internet Cache9200 互联网缓存系统
V100R002C00
技术白皮书

文档版本 V2.0

发布日期 2011-08-25

华为技术有限公司

商标声明


和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编：518129

网址： http://www.huawei.com

客户服务邮箱： support@huawei.com

客户服务电话： 4008302118

执行摘要

人类 5000 年文明的所有文字记载的信息量是 5EB（1EB=10 亿 GB），而在技术进步与想像力的持续驱动下，2006 年全球产生的数字信息量就已经达到 160EB，到 2009 年增长了 4 倍达到 800EB，随着数字媒体的市场规模不断扩大以及数字内容的无限提供，海量数字内容已经引发了互联网流量出现十倍甚至百倍的急速增长，2007 年整个互联网流量为 5EB，到 2013 年将达到 56EB，互联网数字洪水时代正式来临。

与此同时，运营商网络所面临的挑战和压力也与日俱增，市场调查显示，电信骨干网的流量每年正以 50%~80%的速度飞速增长，这直接导致运营商的带宽成本不断增加，而收入却没有得到相应提升，从而严重影响运营商的投资收益比。

华为 Internet Cache 解决方案在此背景下应运而生，以“存储换带宽”的设计理念将互联网视频、P2P、WEB 浏览等互联网主流应用产生的流量进行缓存，利用海量存储实现流量的本地化，从而有效缓解互联网出口流量压力，并大幅度提升用户体验，帮助运营商从容应对互联网数字洪水的冲击。

目 录

插图目录

表格目录

1 概述

近年来随着互联网的发展，互联网用户数、互联网应用种类、网络带宽等都呈现出爆炸式的增长，对社会和人们的生活产生了巨大的影响。在互联网应用中 P2P 技术、在线视频技术的发展给互联网带来的影响尤为广泛，据预测，到 2013 年，91%的个人消费者互联网流量为视频，在移动流量中，66%为视频。

P2P 应用的快速发展很大程度上促进了互联网的发展和宽带接入的普及化，但是其所带来的负面效应也随之显现，最突出的矛盾就是对网络资源的滥用。无论运营商如何增加网络带宽，P2P 应用总是会尽力抢占所有可用的资源。这种“黑洞效应”直接影响了运营商网络建设的积极性，同时也导致其用户满意度下降，进而间接导致运营商的商业利益受损。对 P2P 流量进行控制和管理已经变得非常必要和紧迫，目前主要采用的手段是部署流量分析及控制设备，即对 P2P 应用进行分析和识别，然后再进行阻断或限速。这种方案的优点是通过控制、管理可识别的 P2P 应用，来保证网络带宽和其他业务的 QoS，但其局限性也十分明显，因为通过阻止用户使用来缓解网络拥挤，会直接导致用户流失，甚至面临法律上的风险。显而易见，这种简单的“封堵”思路治标不治本。

在线视频的迅速发展源于用户上网习惯的转变，在 P2P 盛行的时代，受限于网络带宽等因素，在线视频的质量较差，难以满足用户对于高品质视频内容的需求，用户习惯于“先下载，后体验”的上网模式，而随着用户接入带宽的不断提升，在线视频正在迅速朝着高清化的趋势发展，在这种情况下，宽带用户更倾向于选择“即时体验”来获取视频等多媒体资源。

面对互联网流量的高速增长，运营商主要通过两种方法来应对：

- 扩容带宽
- 流量控制，包括技术上对用户流量进行整形、限速等，以及从资费套餐上进行流量限制等。
  - 扩容带宽的解决方案简单易实施，但是费用昂贵，并且新扩容的带宽很快就会被 P2P 等流量再次占满，投资收益极少。
  - 流量控制是立竿见影的解决方案，投资较小并可以迅速降低互联网出口的流量压力，但同时也极大的影响了用户体验，在较为成熟的竞争性市场容易造成用户流失。
  - 因此，上述两种解决方案在投资收益和用户体验上并不能兼顾，往往顾此失彼，运营商需要能在投资收益和用户体验上都能兼顾的解决方案，华为 Internet Cache 互联网缓存解决方案可以很好的满足运营商的需求。

2 解决方案

2.1 系统构架

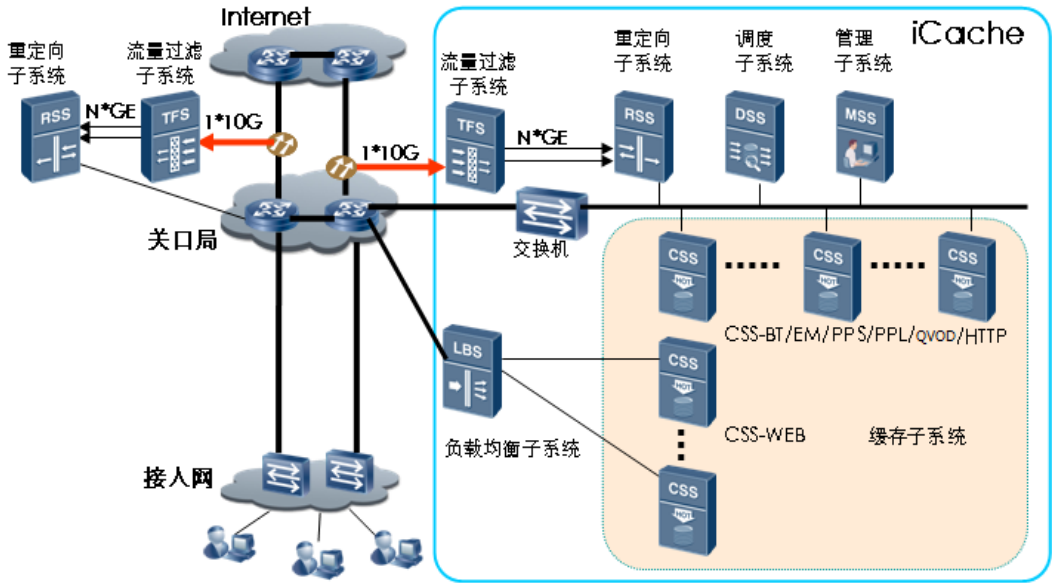
Internet Cache（Internet Cache System），是专门面向电信级运营商和宽带运营商的流量缓存、加速解决方案，用于 HTTP、在线视频、P2P 等大流量应用的网内缓存，通过流量本地化帮助运营商解决两大问题：

- 降低互联网出口压力，减少网间流量；
- 改善用户上网体验，提高用户满意度。

Internet Cache 系统由流量过滤子系统（TFS）、重定向子系统（RSS）、调度子系统（DSS）、缓存子系统（CSS）、负载均衡子系统（LBS）、管理子系统（MSS）组成。

系统构架如图 2-1 所示。

图2-1 Internet Cache 系统架构图



- 流量过滤子系统（TFS）

- Internet Cache 系统在没有 WEB Cache 需求的情况下，采用旁路部署，网络流量不需要经过 Internet Cache 系统，所以需要将经过骨干链路的网络流量复制一份到 Internet Cache 才能确保系统正常工作。
- 复制流量有两种方法：分光和镜像。如果采用分光，需要在骨干链路上增加分光器，分光器属于无源光学元器件，不会影响原有网络的可靠性。如果采用镜像，需要在路由器/DPI 设备上增加配置，将骨干网链路的上行流量镜像到 Internet Cache 系统。无论采用分光还是镜像，都不会影响现网的可靠性。
- 目前 Internet Cache 系统中担当骨干网流量接收和分析任务的设备 RSS 最大支持 8*GE（RJ45）或者 1*10GE(单/多模)流量镜像分析端口，而在大多数情况下，国内运营商的城域网出口或者互联网出口采用 10G/2.5G POS 高速链路，这个时候，需要将 10G 的流量转换成 n*GE 流量再输出到流量接收和分析设备上。
- 华为提供专用的流量过滤子系统，他具备两方面的功能：一是将 10G/2.5G POS 链路的流量转换成 N*GE 的流量；二是提供流量过滤功能，仅把缓存系统需要的网络报文如 BT 协议的 GET announce 报文等镜像给流量接收和分析设备 RSS。如果现网已经部署或者需要部署 DPI 设备，也可以由 DPI 设备充当流量过滤子系统的角色，只要 DPI 设备具备流量镜像功能。

- 重定向子系统（RSS）

  重定向子系统接收并分析流量过滤子系统传送过来的流量，识别出内网用户发向外网的资源下载请求等信息，然后通过发送重定向报文将用户请求重定向到内网的缓存子系统。

- 调度子系统（DSS）

  调度子系统是整个缓存系统的控制中心，具备以下三个方面的功能：

  - 对于需要缓存的资源，调度子系统会通知缓存设备进行下载并记录到数据库，然后将资源下载结果同步给重定向子系统。
  - 对部署的多台缓存子系统设备进行有效的负载管理。比如当部署了多台 BT 缓存设备后，调度子系统将把用户的 BT 下载请求在多台缓存设备上进行有效的负载均衡管理。
  - 对于 BT 应用，调度子系统实现了标准 tracker 服务器的功能，接受内网用户所有 GET announce 的请求，当调度子系统获得用户请求后它将含有内网用户 IP 和缓存子系统 IP 的 Peer-list 返回给用户。调度子系统在向用户发送 peer-list 的时候，优先分发内网的其他用户节点，引导内网的用户之间互连，降低网络资源占用，提高内网资源的利用。

- 缓存子系统（CSS）
  - 缓存子系统负责将互联网资源缓存到本地，然后为网内用户提供缓存加速服务。根据用户所需支持的协议的不同，缓存子系统又可以细分为 HTTP 在线视频及大文件下载、BT 下载、eMule 下载、PPStream、PPTV、QVOD 等多个模块，每个模块需要部署在独立的物理服务器上。

- 负载均衡子系统（LBS）
  - 当运营商需要 WEB 缓存功能时，需要部署 Internet Cache 中的 CSS-WEB 缓存模块。CSS-WEB 不同于其他缓存模块，它采用直路部署模式，通过策略路由（PBR）将 HTTP 流量从骨干路由器引流到 CSS-WEB 上处理。当多台 CSS-WEB 集群部署时，需要增加专用的负载均衡设备将 HTTP 流量平均分配到多台 CSS-WEB 设备上。

- 管理子系统（MSS）

  管理子系统完成系统参数的配置管理、状态监控、报表统计、日志管理和告警管理等功能。当与华为 DPI 设备 SIG 进行对接时，可以实现统一入口、集中管理功能。

2.2 系统功能

2.2.1 支持缓存的内容

- HTTP 网页浏览类内容的缓存，如：www.yahoo.com/www.sina.com/www.sohu.com 等。
- HTTP 文件下载类内容的缓存，如 Windows Update 操作系统补丁更新、软件下载等。
- HTTP 文件分享类内容的缓存，如 Rapidshare 、Megaupload 等。
- HTTP 在线视频内容的缓存，如 Youtube、Dailymotion、Tudou、Youku 等。
- 支持 BT 协议，持的典型客户端包括： BitSpirit、BitTorrent、uTorrent、BitComet 等，对于带 PassKey 的 BT 协议也能支持。
- 支持 eMule 协议，支持的典型客户端包括：eMule、verycd eMule 等。
- 支持 PPStream、PPTV、QQLive、QVOD 等国内主流的私有 P2P 协议或应用。

2.2.2 缓存内容更新机制

- WEB 缓存：
  - 如果网站资源自身定义了过期时间的，按定义的时间进行探测。
  - 如果网站资源未定义过期时间的，每隔 1 个小时进行探测。
  - 对于某些需要设置更新时间的内容，可设置规则，并在规则中设置更新探测时间。
  - 对于源网站不允许缓存的内容，也可根据规则进行强制缓存，并在规则中设置更新探测时间。
- HTTP 在线视频/下载：
  - 系统默认对 rar、zip、exe 这 3 种类型文件的过期探测时间为 2 个小时，用户可根据实际情况自定义过期探测时间，最小单位为秒。
  - 系统默认对其它类型的文件过期探测时间为 1 天，用户可根据实际情况自定义过期探测时间，最小单位为秒。
- PPStream/PPTV/QQLive/QVOD：

  跟原厂商深度技术合作，缓存内容的更新由原厂商统一完成。
- 缓存设备磁盘空间清理：

  当所使用数据空间占用率大于阀值时，进行磁盘空间清理，将数据空间占用率降低，保证有足够的缓存空间。在做磁盘清理的时候，优先删除最长时间没有访问的资源。

2.2.3 缓存内容管理

系统支持黑名单功能，可以根据运营商的实际情况将需要限制访问的资源放入黑名单，缓存系统对于放入黑名单内的资源不提供缓存服务。

- 对于通过 HTTP 协议获取的资源，可设置基于 URL、网站域名和文件 infohash 值的黑名单，对于通过 BT 和 eMule 等协议获取的资源可设置基于文件 infohash 值的黑名单。
- 配置黑名单后，如果该内容已经在缓存系统中，那么，缓存服务器把已经缓存的内容从服务器中清除掉，把文件缓存进度的数据从数据库中删除；如果该内容还没有缓存，那么缓存系统不对该内容进行缓存。
- 对放入黑名单的 HTTP 类的资源，缓存系统执行的处理动作有阻塞、重定向和不服务三种策略；对放入黑名单的 P2P 类的资源，缓存系统处理动作有阻塞和不服务两种策略.
- 缓存系统黑名单列表和前端 RSS 设备的数据库定期进行同步更新，同步更新频率为 2 分钟。

说明
黑名单策略:
- 阻塞：配置阻塞后，用户不能打开连接。
- 重定向：配置重定向后，可以把用户访问的连接重定向到管理员自定义的指定连接;
- 不服务：缓存系统不为用户提供缓存加速服务，用户仍然从互联网获取内容。

2.2.4 服务范围管理

系统可以通过设置缓存系统可服务的 IP 地址列表，只允许列表内的用户使用缓存系统，其他用户仍然从外网获取资源。

2.2.5 QoS 控制

系统可以根据时间段来控制缓存系统从外部下载资源时每个 Session 占用的带宽，从而避开上网高峰期缓存系统对互联网出口带宽的占用；同时，系统也可以控制网内用户从缓存系统下载资源时每个 Session 占用的带宽。

2.3 系统原理

2.3.1 WEB 缓存

- HTTP 基本原理
  - 超文本传输协议（HTTP，HyperText Transfer Protocol）是互联网上应用最为广泛的一种网络协议。所有的 WWW 文件都必须遵守这个标准。设计 HTTP 最初的目的是为了提供一种发布和接收 HTML 页面的方法。
  - HTTP 的发展是万维网协会（World Wide WEB Consortium）和 Internet 工作小组（Internet Engineering Task Force）合作的结果，（他们）最终发布了一系列的 RFC，其中最著名的就是 RFC 2616。RFC 2616 定义了 HTTP 协议中一个现今被广泛使用的版本——HTTP 1.1。

- HTTP 是一个客户端和服务器端请求和应答的标准（TCP）。客户端是终端用户，服务器端是网站。通过使用 WEB 浏览器、网络爬虫或者其它的工具，客户端发起一个到服务器上指定端口（默认端口为 80）的 HTTP 请求。
- 客户端与服务器之间的交互用到了两种类型的消息：请求（Request）和响应（Response）。其中请求最常见的包括 get 和 post 请求，而服务器的响应则包括 HTTP 版本号，后面跟着 3 位数字表示响应代码，如 200 代表请求成功，404 代表 not found。

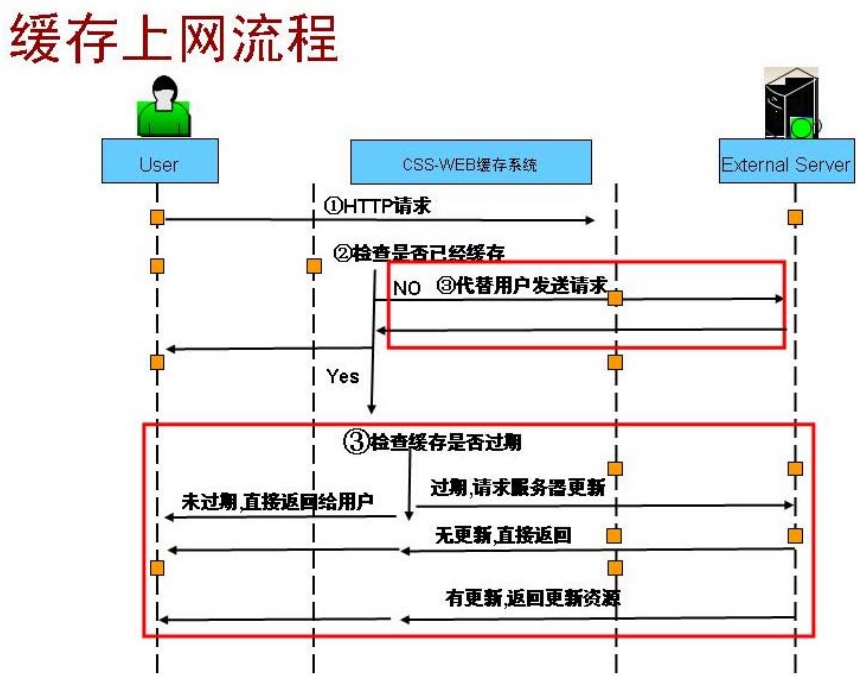
图2-2 HTTP 请求和应答模型

- CSS-WEB 缓存工作流程

  当客户机向服务器发送 GET 请求的时候，WEB 服务器将返回请求的资源，同时返回该资源的 ETag 标识以及 expire 时间，当没有超过 expire 时间，则客户机下次再次发送 GET 请求的时候，将直接从缓存中获取资源，而不需要再次从 WEB 服务器上下载资源。

  部署 Internet Cache 系统之后，缓存上网流程如图 2-3 所示。

图2-3 HTTP 缓存上网流程



1 当用户向 WEB 服务器发送请求的时候，流量会经过 CSS-WEB 系统，此时 CSS-WEB 会检测该资源是否已经被缓存，如果已经缓存，并且该资源没有过期，则直接将缓存内容返回给用户。

2 当缓存了该资源，但是该资源已经过期，则 CSS-WEB 会发送一个更新请求到 WEB 服务器，请求获得最新资源，如果 WEB 服务器返回“没有更改”报文，则直接返回缓存资源给用户，否则从服务器上下载资源返回给用户。

3 如果该资源没有被缓存，则 CSS-WEB 会代替用户去 WEB 服务器上下载，同时返回给用户，如果 CSS-WEB 定义的策略要求缓存该文件，则缓存该资源，并再次提供服务。

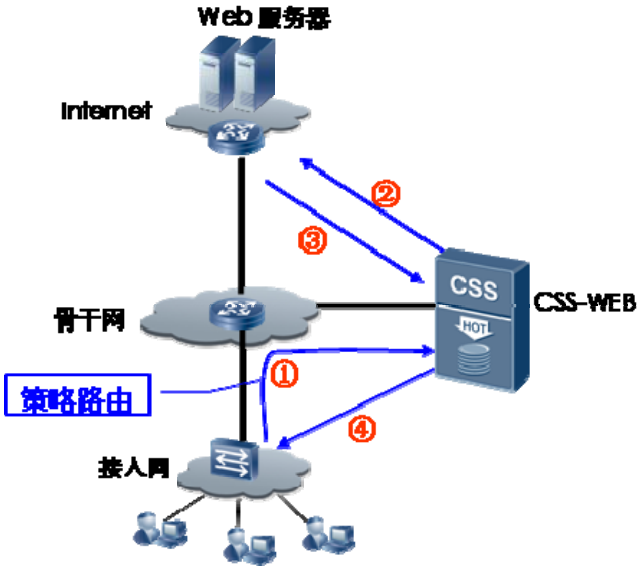
- CSS-WEB 工作模式

  WEB 缓存采用直路部署模式，通过在路由器上配置策略路由（PBR），将 HTTP 流量从路由器引流到缓存设备，它有 2 种工作模式：半透明模式、透明模式。

  - 半透明模式：需要在路由器/交换机上做一定的配置修改，将用户访问外网 WEB 服务器的流量通过策略路由（PBR）转向缓存服务器，然后由缓存服务器代理用户去外网下载内容再提供给用户。缓存系统对用户是透明的，但是 WEB 服务器看不到用户的 IP，只能看到缓存服务器的 IP 地址。

    半透明模式如图 2-4 所示。

图2-4 WEB 缓存半透明模式原理图

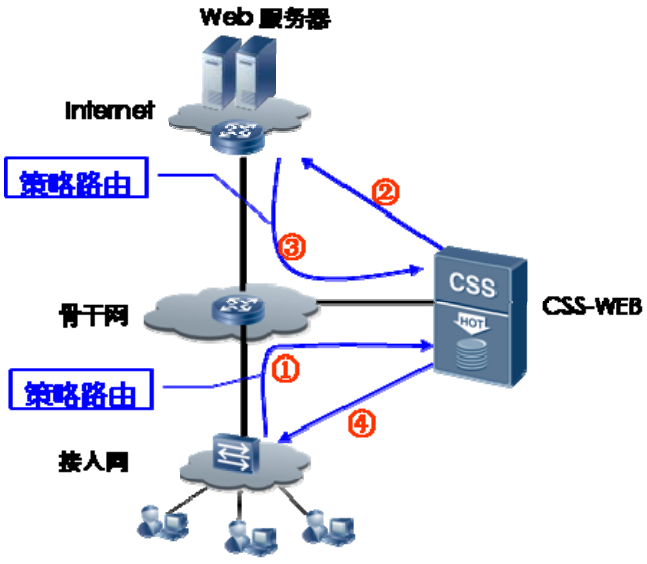


1 用户的 HTTP 访问报文通过策略路由（PBR）牵引到缓存设备。
2 缓存设备用自己的 IP 地址代理用户向 WEB 服务器发起请求。
3 WEB 服务器返回响应报文给缓存设备。。
4 缓存设备将响应报文返回给用户。

- 透明模式：需要在路由器/交换机上做一定的配置修改，将用户访问外网 WEB 服务器的流量以及 WEB 服务器返回的流量通过策略路由（PBR）转向缓存服务器，缓存服务器作为所有 HTTP 流量的中转站，缓存系统对用户和 WEB 服务器都是透明的。

  透明模式如图 2-5 所示。

图2-5 WEB 缓存透明模式原理图



1 用户的 HTTP 访问报文通过策略路由（PBR）牵引到缓存设备。
2 缓存设备使用用户的 IP 地址代理用户向 WEB 服务器发起请求。
3 WEB 服务器返回的响应报文通过策略路由（PBR）牵引到缓存设备。
4 缓存设备将响应报文返回给用户。

- CSS-WEB 两种工作模式的比较

  在半透明模式下：Internet 上的 WEB 服务器只看到 CSS-WEB 服务器的 IP，无法获取内网用户的 IP 信息。表现为一个 IP（CSS-WEB 的 IP）频繁的访问 WEB 服务器。这时，有可能存在两个问题：

  1. 如果 Internet 上的 WEB 服务器对用户 IP 的过于频繁的访问有限制的话，则会导致无法访问。这种限制在国内比较少，在海外有一些 WEB 服务器会有这样的限制。
  2. 由于 Internet 上的 WEB 服务器无法获取用户的 IP 信息，也会对该服务器的统计有一定影响。

  在透明模式下：CSS-WEB 不可见，Internet 上的 WEB 服务器以及用户端都感觉不到 CSS-WEB 的存在，因此，不存在半透明模式下的两个问题。

  在实际部署中，需要根据网络情况选用合适的方式，一般建议为透明模式，除非是客户存在多个站点，多个站点之间存在非对称路由的场景。此种场景下的部署方案选择，请参考：“非对称路由场景 Internet Cache9200 技术白皮书”。

2.3.2 Video 缓存

- Video 基本原理和发展趋势
  - 近年来随着互联网的发展，互联网用户数、互联网应用种类、网络带宽等都呈现出爆炸式的增长，对社会和人们的生活产生了巨大的影响。在线视频技术的发展给互联网带来的影响尤为广泛，据预测，到 2013 年，91%的个人消费者互联网流量为视频，在移动流量中，66%为视频。
  - 在线视频和传统的观看视频不同，视频资源全部保存在视频服务器上，用户同过浏览器点播视频，向服务器发送点播请求，服务器收到后查找用户要点播的视频并过网络将对应视频传到用户的电脑上，电脑经过处理通过显示器把视频呈现给用户。在线视频的迅速发展源于用户上网习惯的转变，在 P2P 盛行的时代，受限于网络带宽等因素，在线视频的质量较差，难以满足用户对于高品质视频内容的需求，用户习惯于“先下载，后体验”的上网模式，而随着用户接入带宽的不断提升，在线视频正在迅速朝着高清化的趋势发展，在这种情况下，宽带用户更倾向于选择“即时体验”来获取视频等多媒体资源。
- Video 缓存的部署方式

  Video 缓存采用旁路部署模式，不需要改变现有网络的组网结构，不会增加故障点。缓存系统并不是位于正常的网络流量经过的路径上，而是采用分光/镜像的方式复制了一份流量，然后通过对复制的流量进行分析后采用重定向技术将用户的请求引导到缓存系统。这种部署模式有以下好处：
  - 对于缓存系统不关心的流量，缓存系统的存在与否对这些流量没有任何影响。
  - 当部分缓存设备下线时，业务连续性不会受到影响，重定向子系统将检测到下线的缓存设备，并对位于这些设备上的资源停止发送重定向报文，用户仍然可以从外网获取资源。

– 即使整套系统下线，业务连续性也不会受到影响，就跟没有部署缓存系统时一样，用户直接从外网获取资源。

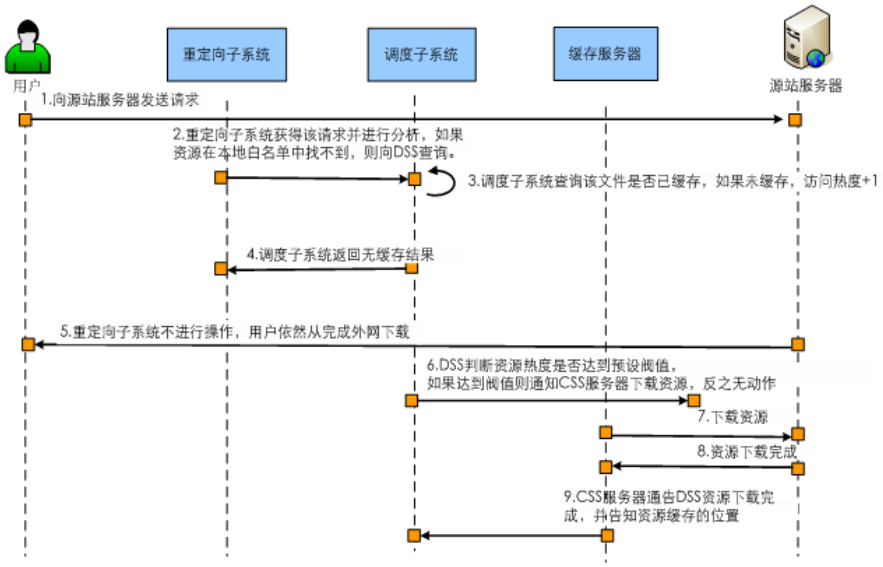
- Video 缓存工作流程：

  Internet Cache 采用热点探测和被动缓存技术，只有在内网用户访问同一资源达到热点阀值后，才会把资源下载到缓存系统并为后续的用户访问提供服务。

  在线视频缓存的具体工作流程分为两部分：

  用户请求的资源达到热度阀值前：

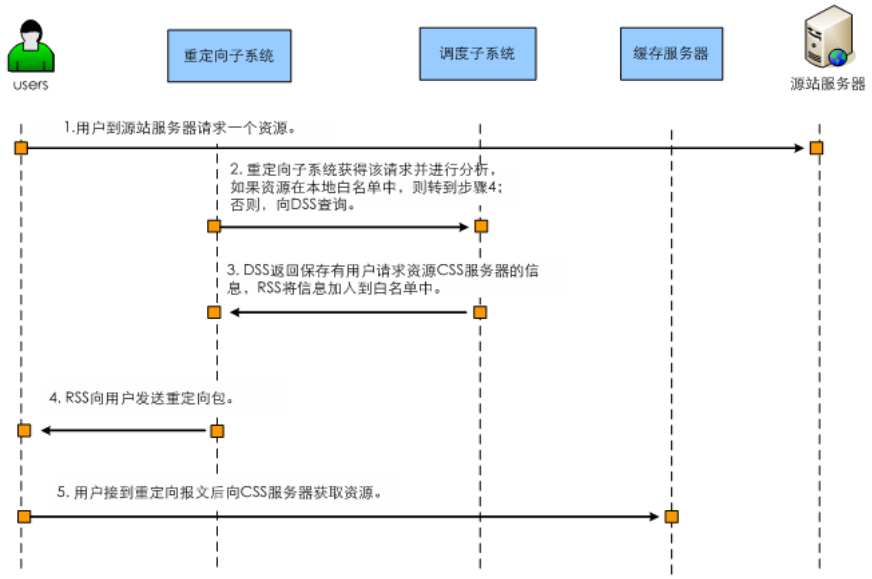
图2-6 Video 缓存进栈前工作原理图



1 用户发起 HTTP GET 请求。
2 RSS 监听到该请求报文符合缓存基本标准，RSS 查询内存中的白名单看资源是否已经被缓存，如没有，则查询调度子系统 DSS。
3 DSS 查询缓存服务器中是否存在用户请求的资源，如果没有，则将资源的热点阀值加 1；
4 DSS 通知 RSS 用户请求的资源没有被缓存。
5 重定向子系统不做任何操作，用户仍然从源站获取资源。
6 DSS 判断用户请求的资源是否达到热度阀值，如果达到热度阀值，DSS 挑选其中一台 CSS 服务器去外网下载资源。
7 CSS 服务器启动下载进程。
8 CSS 服务器完成资源下载。
9 CSS 服务器通知 DSS 资源已经下载完毕，并告知资源存储的目录，流程结束；

用户请求的资源达到热度阀值并被缓存到本地后：

图2-7 Video 缓存进栈后工作原理图



1 用户发起 HTTP GET 请求。
2 RSS 监听到该请求报文符合缓存基本标准，RSS 查询内存中的白名单看资源是否已经被缓存，如有，跳到步骤 4；如没有，则查询调度子系统 DSS。
3 DSS 返回保存有用户请求资源的 CSS 服务器信息给 RSS，RSS 将信息记录到本地的白名单中。
4 RSS 发送重定向报文给内网用户。
5 内网用户接到重定向信息后连接 CSS 服务器进行资源请求。

2.3.3 P2P 缓存

P2P 是 peer to peer 的缩写，就是数据不存放在统一的资源服务器上，而是存放在各个用户上的数据进行共享，每个用户下载的资源都是从其他用户下载，实现这种数据传输而制定的协议称着 P2P 协议，采用 P2P 协议进行用户间数据传输的软件称着 P2P 软件，常见 P2P 软件有 BT、eMule 等。

以往占据主导地位的互联网架构是客户端/服务器（Client/Server）结构。互联网以服务器为中心，各种各样的资源，包括文字、图片、音乐、电影等都存放在服务器的硬盘上，网民把自己的电脑作为客户端通过网络连到服务器上检索、下载、上传资料，在这种架构下，客户端和服务器存在着明显的主从关系，而 P2P 模式，也称为对等网络（PeertoPeer）结构，它的本质思想是，整个网络结构中不存在中心节点（或中心服务器），在 P2P 结构中，每一个节点(Peer)大都同时具有信息消费者、信息提供者和信息通讯等三方面的功能，在 P2P 网络中每一个节点所拥有的权利和义务都是对等的。

- P2P 缓存部署方式

  目前现网中网络设备部署方式主要有直路和旁路两种结构，在直路模式中网络设备需要串行连接到主干线路上，旁路部署则不需要将设备串进主干线路，只需要镜像、分光等方式将流量负复制一份给旁路网络设备。Internet Cache 系统采用旁路部署模式，与直路比具有以下优势：

- 不需要改变现有网络的组网结构，缓存系统并不是位于正常的网络流量经过的路径上，而是采用分光/镜像的方式复制了一份流量，然后通过对复制的流量进行分析后采用重定向技术将用户的请求引导到缓存系统。
- 不会增加故障点。由于在主干线路上没有增加任何设备，故不会增加故障点。
- 对于缓存系统不关心的流量，缓存系统的存在与否对这些流量没有任何影响。

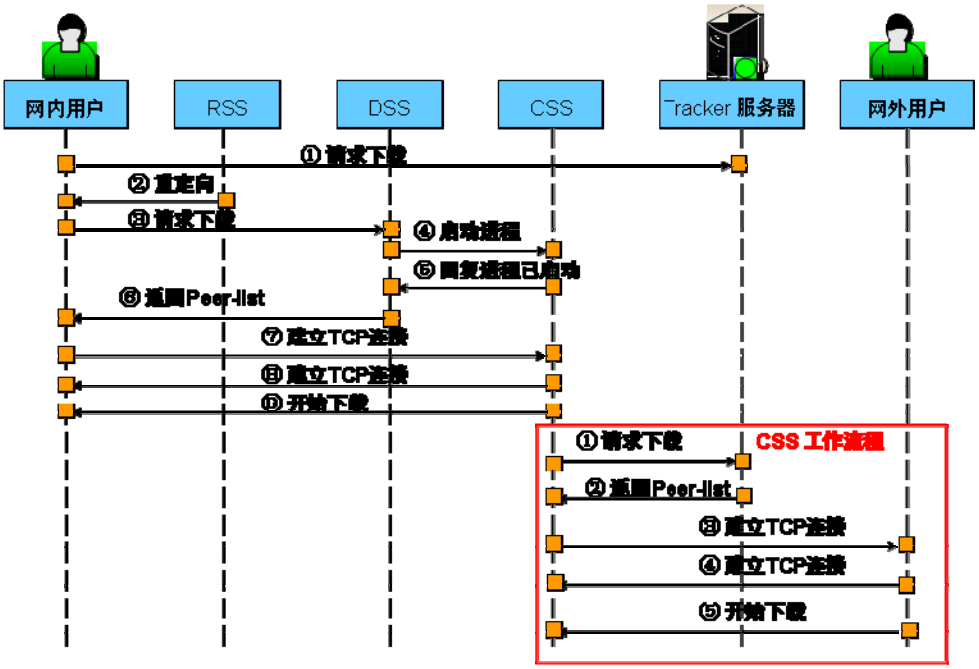
  当部分缓存设备下线时，业务连续性不会受到影响，重定向子系统将检测到下线的缓存设备，并对位于这些设备上的资源停止发送重定向报文，用户仍然可以从外网获取资源。
- 即使整套系统下线，业务连续性也不会受到影响，就跟没有部署缓存系统时一样，用户直接从外网获取资源。

- P2P 缓存工作流程

  Internet Cache 系统支持 BT、eMule、PPS 等多种 P2P 协议，由于每种协议特征的不同，所以对不同协议的资源提供缓存服务的原理也有所不同。
- BT Cache

  Internet Cache 在完成 BT 资源缓存时，由 RSS、DSS 和 CSS 共同完成，其热点处理机制与在线视频的不同：
  - 检测用户访问流量中包含的资源请求信息，一旦检测到 P2P 资源请求立即启动 P2P 缓存资源进栈。
  - 缓存设备边下载资源边为用户提供缓存加速服务，未缓存资源片段仍然通过外网下载。

图2-8 BT 缓存工作原理图



1 用户出口流量被复制到监控服务器 RSS 监控口，RSS 对出口流量进行深度分析，识别出网内用户对外的 BT 下载请求。

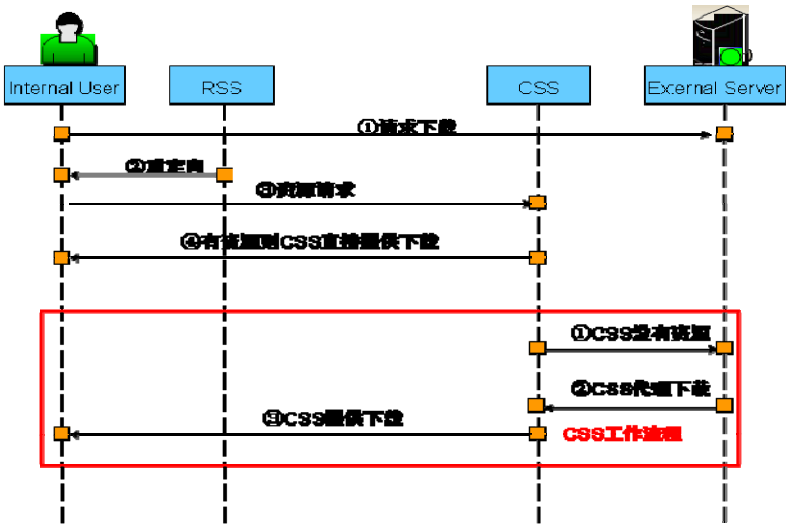
2 RSS 发送重定向报文给用户，将用户请求重定向到 DSS，DSS 将充当代理 Tracker 服务器的角色。

3 DSS 检索本地缓存系统，如果已经存在用户所需的数据资源，会将存有相关数据的缓存子系统 CSS 服务器的 IP 地址以及同时参与该任务的内网其他用户的 IP 地址列表，通过 PEER LIST 反馈给用户，用户根据所反馈 PEER LIST 中的 IP 地址建立连接，从 CSS 缓存服务器上和内网其它用户处获取数据；如果本地缓存系统里未包含用户所需要的数据资源，DSS 会调配某一台 BT 缓存服务器去外网下载该资源，并将该 BT 缓存服务器 IP 地址通过 Peer List 反馈给用户，BT 缓存服务器一边从外网下载一边为内网用户提供缓存服务。

- eMule Cache

  Internet Cache 在完成 eMule 资源缓存时，其热点处理机制跟 BT 类似，与 BT 不同的是，eMule 的缓存处理流程不需要 DSS 作为 Tracker 服务器参与。

图2-9 eMule 缓存工作原理图



1 RSS 对出口流量进行深度分析，识别出网内用户对外的 eMule 下载请求。

2 RSS 根据配置策略通过 UDP 报文向用户发送负责缓存该 eMule 资源的 eMule 缓存服务器的 IP 地址，普通 eMule 客户端收到 RSS 发送的带有缓存服务器 IP 地址的 UDP 报文后会自动识别该报文,并将报文中的缓存服务器地址增加到本机的资源用户列表中，同时会向这些 IP 请求下载数据。

2.4 系统优势

2.4.1 支持动态 URL

传统互联网一般使用固定的 URL 指向固定的资源，但是大量受欢迎的媒体网站使用复杂的动态 URL 来提供视频和音频文件，包括一次性或限时 URL，以此作为一种安全措施。对于这类媒体网站来说，一般情况下，同一个文件（即使该文件来自同一个用户）的两个视图不会使用相同的 URL。

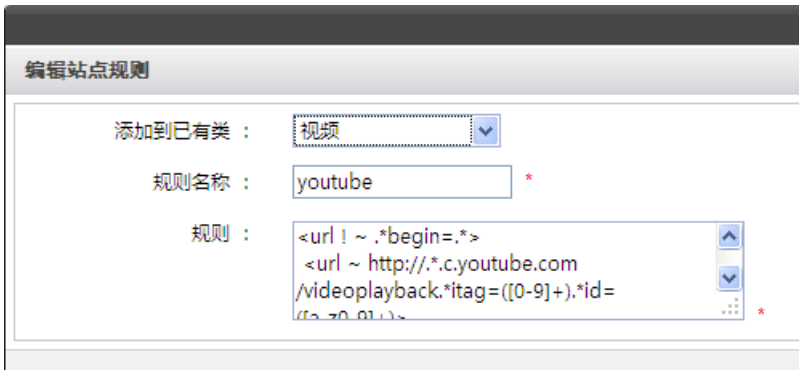
传统的缓存解决方案很大程度上依赖于对象的 URL，且许多这些解决方案根本不会缓存动态 URL。因此，这些解决方案只能有效地缓存网络上静态的对象，如 HTML 网页

和一些图像。缓存这些静态对象可以节省一些带宽，但这种方法无法解决在如今的网络流量中占很大部分的在线视频流量。

Internet Cache 专为媒体对象而设计的智能缓存策略，基于强大的正则表达式系统能够分析媒体对象的内容，为 YouTube、优酷网或土豆网等受欢迎的在线视频网站处理已知的 URL 模式，从而打破了上述局限。

同时互联网行业变化日新月异，每天都会涌现更多的采用动态 URL 技术的网站，通常的缓存设备对这样一个网站的支持需要 1 到 3 个月的时间，而 Internet Cache 采用先进的架构，能够通过配置正则表达式快速响应新网站需求，如图 2-10 所示。

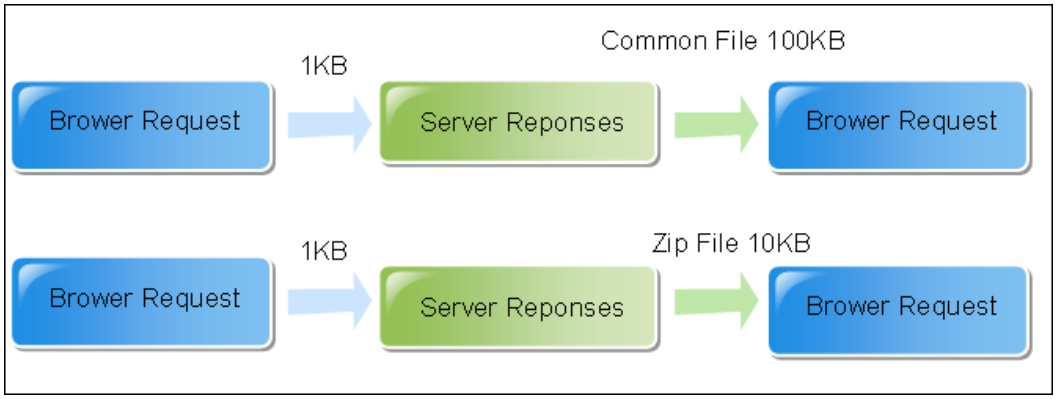
图2-10 动态 URL 规则配置图



2.4.2 支持压缩传输技术

压缩传输的原理是：服务器在进行 HTTP 传输之前，先使用 GZIP 压缩再传输给客户端，客户端接收之后由浏览器解压显示，这样虽然稍微占用了一些服务器和客户端的 CPU（Internet Cache 采用压缩传输功能后的负载会提升 5%左右），但是换来的是更高的带宽利用率和更好的客户体验：启用压缩后能有效减少网页传输数据大小，使得有限带宽能提供更多的请求，并在一定程度上提高了网页“显示”速度，提升客户的体验，如图 2-11 所示。

图2-11 传输压缩效果示意图



GZIP 压缩的原理是在一个文本文件中找出类似的字符串，并临时替换他们，使整个文件变小。这种形式的压缩对 WEB 来说特别适合，因为 HTML 和 CSS 文件通常包含大量的重复字符串，例如空格，标签，及样式定义。Gzip 支持的文件包括 HTML、JavaScript、CSS 和其他 TEXT 文件。在传输优化测试中，使用 GZIP 方式压缩的站点，其 TEXT 文件平均可以压缩 75%，压缩率是相当可观的。如果每个用户节约 50%的带宽，那么相同的带宽就可以服务多一倍的客户了。表 2-1 为 Google 网站的测试结果。

表2-1 HTTP Compression with Google (file size in bytes)

Home page	HTML Page Size (uncompressed)	HTML Page Size (compressed)	Percentagesavings
Google.com	3,873	1,412	63.50%
Google HTTP+Compression	26,321	5,505	79.10%
Note: These figures do not include HTTP header size, just the HTML size.			

2.4.3 支持 Safesearch 技术

谷歌 SafeSearch 是一种从搜索结果中过滤掉色情内容的技术。谷歌用户可据需要设定所需级别。

Internet Cache 支持设定是否开启 SafeSearch 的默认开启状态。例如，开启 Safesearch 功能之后，用户搜索的 URL 为“q=sex&safe=strict”，表示搜索启用了过滤功能，默认情况下该功能是关闭的，用户可以在“搜索设置”中来设置是否开启该功能。

管理员可以设置是否默认启用该功能，当启用该功能之后，用户使用谷歌搜索引擎的时候，默认的状态都启用了 Safesearch 功能，从而能够达到过滤色情图片和文字效果，净化上网环境。

2.4.4 支持 HTTP POST 缓存

很多文件分享网站，如 MegaUpload 等网站采用 HTTP POST 方式进行下载。传统的缓存解决方案无法处理这种方法。

Internet Cache 专为 HTTP POST 方式进行下载的内容而设计的智能缓存算法可以很好的支持此类流量。

2.4.5 支持缓存过期探测

Internet Cache 可自动检测和识别网站规定的资源过期时间，并完全遵从网站的过期时间，过期的资源自动清理。同时，如果网站没有规定过期时间，则可采用默认设置的过期时间进行清理。该时间，也可以由手工设置和更改。

2.4.6 支持缓存热度触发

Internet Cache 并不对所有的资源都直接进行缓存，而是维护了一个资源访问的计数，基于该计数，判断资源的热度，只有达到一定热度的资源，才触发缓存，从而提升缓存的命中率。

对 P2P 缓存，比如 BT/eMule 采用单次触发机制，即一旦监测到请求，CSS 立刻开始缓存进站，并作为 Peer 服务客户。

2.4.7 优化的存储技术

Internet Cache 凭借华为公司在存储领域的深厚积累对缓存存储技术做了深入的优化，包括多核多线程技术，存储调度技术等。

Internet Cache 采用了强大的 Intel Nehalem 平台 CPU，为充分利用处理器性能，对多线程处理进行了深度优化，保证充分利用每个核的处理能力。

Internet Cache 针对互联网缓存的特点，优化了存储调度技术，将小文件尽量保持在内存中，将大文件保存在磁盘中，并在高效的调度算法下，尽量做到顺序读写磁盘，以提高缓存设备的吞吐能力。

Internet Cache 支持缓存空间清理，当所使用数据空间占用率大于阀值时，可自动进行磁盘空间清理，将数据空间占用率降低，保证有足够的缓存空间。

2.4.8 支持丰富的统计分析

Internet Cache 保持近期客户访问记录，并进行数据挖掘，提供一系列统计分析报表，方便分析系统运行状况，分析网络状况。对于 WEB 缓存，支持包括：

- 所有网站进行 TOP100 统计；
- 访问记录查询；
- 支持所有网站基于不同视角的统计分析和排序，包含：
  - 网站名；
  - 总缓存数据大小；
  - 字节命中率；
  - 用户请求总数；
  - 对 Internet Cache 的请求总数；
  - 请求命中率；
  - 平均响应时间。

对 Video 和 P2P 缓存，支持：

- 热点统计

  将前 24 小时的最新热点信息统计为热点信息页面。热点信息显示显示资源的名称/URL、infohash、资源类型、缓存进度、访问次数、内容类型、文件扩展名、下载流量、流量放大比、系统内部下载链接信息。

- 热点查询：

  可查询查询 90 天任一时段内热点资源访问的历史信息。

- 内容查询

  可查询已缓存的 HTTP、BT、eMule 的资源信息，可按照资源的 infohash 值、URL 进行检索具体资源信息。

- 自定义 TOP 10 网站资源查询：

  支持自定义 10 个网站进行缓存服务资源统计，能够针对指定网站进行站内资源服务情况统计 TOP10/100 的热点资源内容。

2.4.9 强大的 URL Filtering 功能

支持黑白名单过滤

Internet Cache 支持设置黑白名单 URL，在设置的时候，支持通配符设置。在设置黑名单时，支持引入远程服务器的 URL 列表，便于对黑名单的维护和更新。

对识别出来属于黑名单的 URL，可以进行阻断访问或者重定向到指定 URL 的操作。

通过设置白名单，可以对只在白名单列表内的 URL 才允许访问。

支持基于正则表达式的 URL 过滤

Internet Cache 支持设置正则表达式来匹配 URL，可以对满足正则表达式匹配要求的 URL 进行过滤处理，实现对匹配的 URL 的不缓存、阻断访问或者重定向到特定 URL 的操作。

支持第三方 URL 库过滤

Internet Cache 与互联网著名的 URL 分类库 BrightCloud 合作。BrightCloud 公司是全球最大的网址分类导航和内容过滤的互联网服务提供商，其网络覆盖十数倍于其他同类服务提供商，提供了 80 多个详细分类，覆盖面广，命中率、准确率高。

通过和 BrightCloud 的合作，Internet Cache 支持只允许访问互联网特定分类的 URL，从而可以起到对访问互联网有效控制的作用。

2.4.10 缓存子系统功能的互换

华为 Internet Cache（Internet Cache System）解决方案充分站在用户的角度为用户考虑，为了可以使客户更有效的利用现有资源，节省客户的投资成本和维护成本，成功开发了缓存子系统功能互换的功能。该功能可以使不同的缓存子系统之间（P2P 和 Video）无限制自由切换，当客户设备故障或需求改变时，华为 Internet Cache（Internet Cache System）解决方案可以快速响应客户，将对应缓存子系统切换为客户需要的类型，快速满足客户需求，解决客户问题，并有效的节约了客户的投入成本，把可能问题对用户的影响降到最低。

2.4.11 支持限速控制

Internet Cache 支持基于时间段的限速策略，可以针对不同时间段，设置不同的下载速度控制上限。支持单任务下载限速、单用户服务限速、到外网下载总限速。

2.5 系统部署场景

在没有 WEB Cache 需求的情况下，Internet Cache 系统采用旁路部署模式，不需要改变现有网络的组网结构，不会增加故障点。缓存系统并不是位于正常的网络流量经过的路径上，而是采用分光/镜像的方式复制了一份流量，然后通过对复制的流量进行分析后采用重定向技术将用户的请求引导到缓存系统。这种部署模式有以下好处：

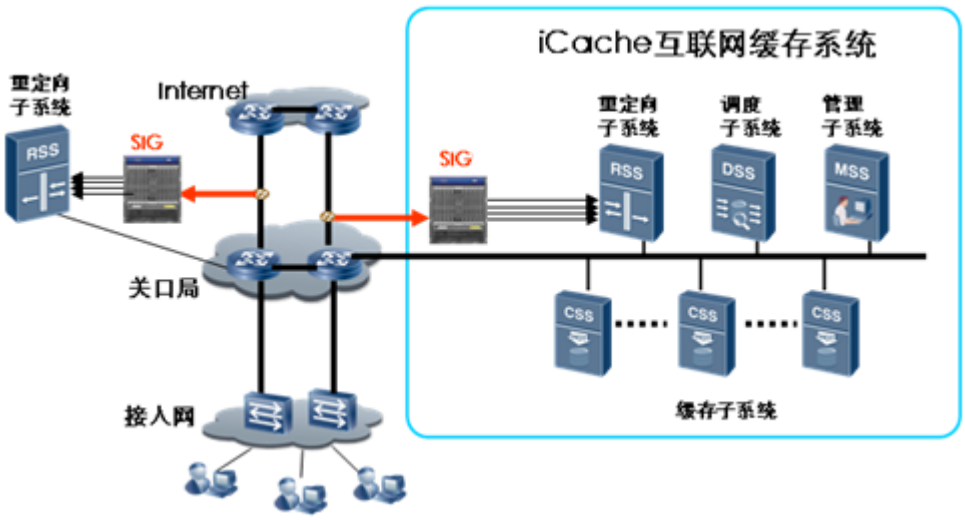
- 对于缓存系统不关心的流量，缓存系统的存在与否对这些流量没有任何影响。
- 当部分缓存设备下线时，业务连续性不会受到影响，重定向子系统将检测到下线的缓存设备，并对位于这些设备上的资源停止发送重定向报文，用户仍然可以从外网获取资源。
- 即使整套系统下线，业务连续性也不会受到影响，就跟没有部署缓存系统时一样，用户直接从外网获取资源。
- 当有 WEB Cache 需求时，WEB Cache 采用策略路由（PBR）的方式直路部署，一般情况下会部署多台 WEB Cache 设备，因此需要配套负载均衡器在多台 WEB Cache 之间进行负载分担。负载均衡器具备健康检测机制，可以实时监控 WEB Cache 设备的可用性，一旦发现某台 WEB Cache 设备故障，将停止向故障设备转发流量。

此外，负载均衡器也可以采用主备模式双机部署，进一步提高了可靠性。

2.5.1 与 DPI 集成部署

当现网已经部署了 DPI 设备或者计划部署 DPI 设备时，Internet Cache 可以与 DPI 设备集成部署，需要 DPI 具备流量镜像功能并有足够的 GE 接口将镜像流量转发到重定向设备。

图2-12 Internet Cache 与 DPI 设备集成部署拓扑图



Internet Cache 与 DPI 集成部署，可起到“疏堵结合”的效果，具备明显的优势：

疏堵结合：

缓存系统缓存并加速主流的协议，DPI 设备控制非主流的 P2P 协议。

流量过滤：

通过 DPI 分析，只把缓存系统需要的报文过滤转发给重定向设备，提高重定向设备的处理效率和性能。

负载均衡：

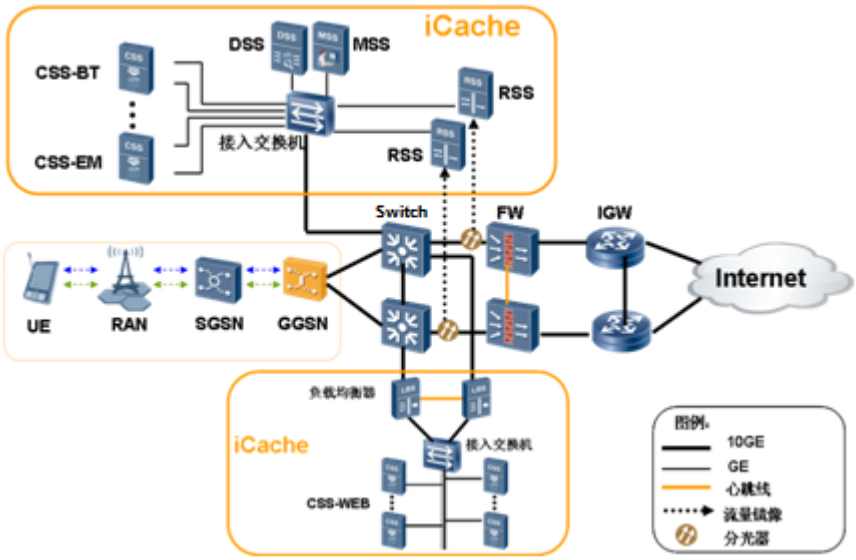
DPI 可以将流量均匀的分配到多台重定向设备或者多个接口上。

高可靠性：

DPI 设备可实时监控与重定向设备互联的接口状态，一旦发现故障，会将故障端口的流量转移到正常端口上。

2.5.2 核心网 PS 网络部署

图2-13 Internet Cache 与在核心网 PS 网络部署拓扑图



Internet Cache 系统可以在核心网 PS 网络中部署。建议部署在 GGSN 与防火墙之间，通过对 10G/2.5G 骨干链路的上行链路进行分光，然后将分光链路接到分流平台，由分流平台将 10G/2.5G 的流量平均分成 N*GE 流量输出到重定向设备；同时在交换机上进行策略路由，将 http 流量转发到 CSS-WEB 子系统。

2.5.3 系统扩展部署

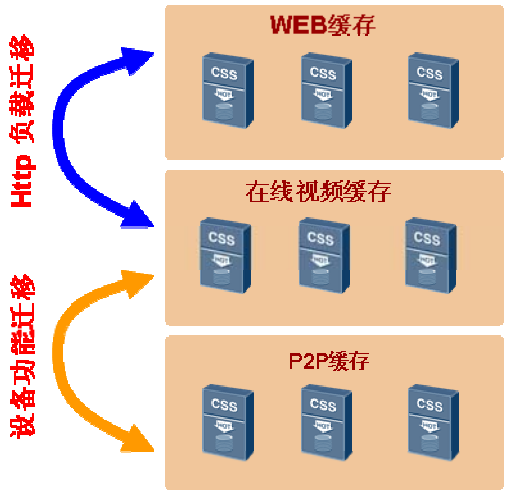
Internet Cache 采用模块化设计，模块之间可以灵活、自由的组合，所以具有非常好的弹性和扩展性。

- 灵活的配置保证最佳的缓存效率

  由于用户上网习惯的变化，互联网流量趋势会经常性的发生变化，比如前几年 BT 和 eMule 等 P2P 流量占据了互联网 50%以上的流量，那时候的用户上网习惯是“先下载，后体验”，但是目前用户的上网习惯已经逐渐变成“实时体验”，所以

P2P 流量占比逐渐下降，而在线视频流量发展迅速。为了充分保护运营商的既有投资，Internet Cache 系统提供了良好的弹性，如图 2-14 所示。

图2-14 Internet Cache 弹性部署图



- 在线视频和 P2P 缓存设备采用同一硬件平台，可通过更改软件 License 互换。
- 如果 HTTP Cache 设备超过正常负载，则可以通过手工配置将部分大文件下载类流量切换到在线视频缓存设备上。
- 如果在线视频缓存设备超过正常负载，则可以通过手工配置将部分在线视频流量切换到 HTTP Cache 设备。

- 良好的扩展性保护原有投资

  当运营商的网络规模和宽带用户数量增长后，为了继续保持较低的互联网出口带宽需求以及良好的用户体验，就需要相应的扩容缓存系统。Internet Cache 支持平滑的、线性的扩容升级，只需要增加相应的缓存设备即可。

图2-15 Internet Cache 扩容前拓扑图

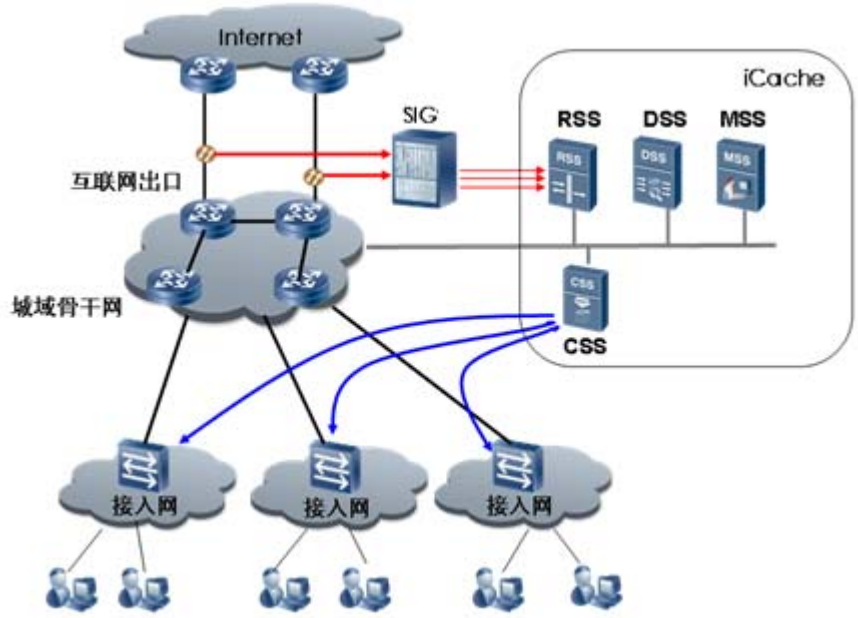
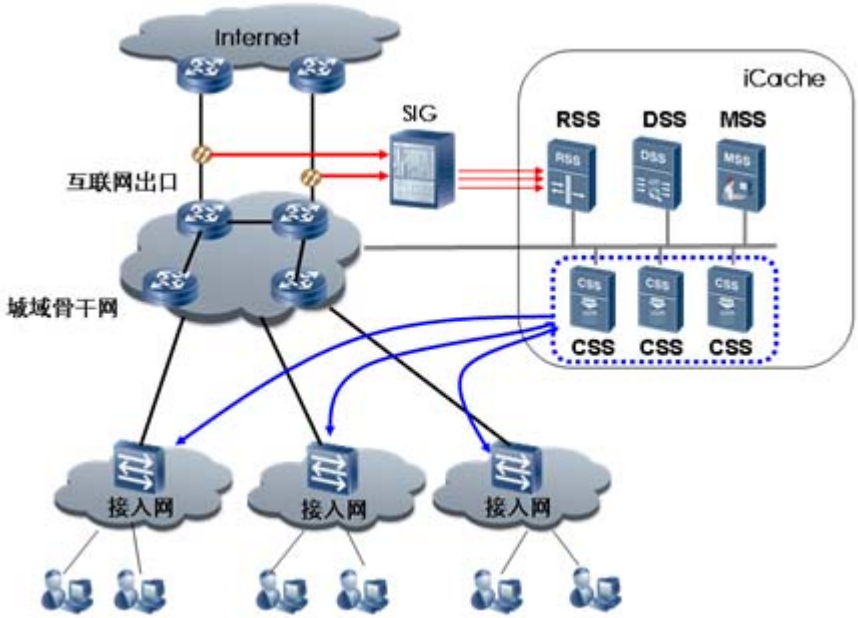


图2-16 Internet Cache 扩容后拓扑图

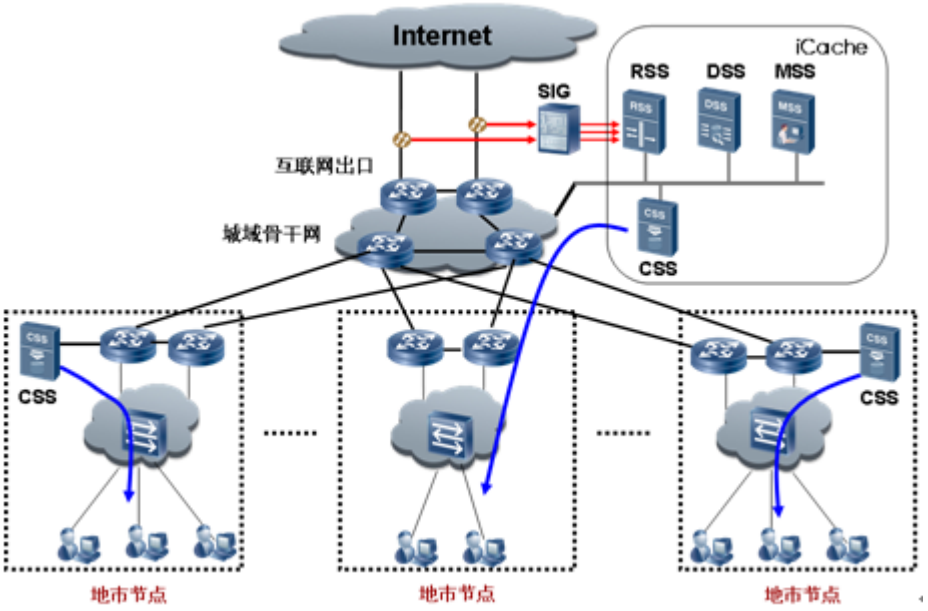


- 分布式部署

  上述部署方案都基于集中式部署。此外，Internet Cache 支持集中式和分布式两种部署模式，当网络规模较小的时候，可以采用集中部署的方式，不仅节省投资成本，也可以提高缓存效率。当运营商网络规模不断扩大时，仍然可以采用集中部署的模式，只需要增加缓存设备，但是这样有可能引起各个 POP 节点和骨干网之间的链路拥塞。这个时候，我们可以从集中部署模式向分布式部署模式迁移，在

流量比较大的 POP 节点部署新增的缓存系统，新增的系统只为 POP 节点内的宽带用户提供服务，如图 2-17 所示。

图2-17 Internet Cache 分布式部署图



2.6 系统可靠性

2.6.1 WEB 缓存可靠性

WEB 缓存采用策略路由（PBR）的方式直路部署，一般情况下会部署多台 WEB Cache 设备，因此需要配套负载均衡器在多台 WEB Cache 之间进行负载分担。负载均衡器具备健康检测机制，可以实时监控 WEB Cache 设备的可用性，一旦发现某台 WEB Cache 设备故障，将停止向故障设备转发流量。

此外，负载均衡器也可以采用主备模式双机部署，进一步提高了可靠性。

2.6.2 Video 缓存及 P2P 缓存

Video 缓存和 P2P 缓存采用旁路部署，不需要改变现有网络的组网结构，不会增加故障点。缓存系统并不是位于正常的网络流量经过的路径上，而是采用分光/镜像的方式复制了一份流量，然后通过对复制的流量进行分析后采用重定向技术将用户的请求引导到缓存系统。这种部署模式有以下好处：

- 对于缓存系统不关心的流量，缓存系统的存在与否对这些流量没有任何影响。
- 当部分缓存设备下线时，业务连续性不会受到影响，重定向子系统将检测到下线的缓存设备，并对位于这些设备上的资源停止发送重定向报文，用户仍然可以从外网获取资源。

- 即使整套系统下线，业务连续性也不会受到影响，就跟没有部署缓存系统时一样，用户直接从外网获取资源。

2.6.3 业务网络和管理网络分离部署

Internet Cache 的业务网络和管理网络可以工作在不同的子网，即使业务网络发生故障，管理网络仍然可以正常进行管理控制操作，便于及时、方便的实施故障恢复等操作。

2.7 系统主要指标说明

2.7.1 吞吐量

吞吐量（Throughput）：

缓存设备单位时间内成功地传送给客户的数据数量，单位为 Mbps，此指标为缓存设备最为重要的指标之一，直接反映了对运营商的价值。可以通过管理子系统查看缓存设备的吞吐量数据。

2.7.2 命中率

命中率（Hit Ratio）指标主要分为两种：

- 请求命中率（Request Hit Ratio）：
  - 按照用户每次请求命中与否的数量统计。（用户请求资源时，在缓存服务器中缓存了则称为命中<hit>，没有缓存则未命中<miss>。）
  - 请求命中率=请求命中数/总请求数
- 字节命中率（Byte Hit Ratio）：按照流量来计算命中的统计，单位是百分比。

  字节命中率=（Flowout—Flowin）/Flowout

  Flowout：缓存子系统流出流量

  Flowin：缓存子系统流入流量

2.7.3 连接数

连接数（Connection Number）指标主要有两种：

- 并发连接数是指缓存设备对其业务信息流的处理能力，是缓存设备能够同时处理的点对点连接的最大数目，它反映出缓存设备对多个连接的服务能力。单位都是个。
- 新建连接数：缓存设备在维持原有的连接数基础上，每秒最大能够新增的连接数。单位都是个每秒。

2.7.4 用户网络对系统指标的影响

目录	用户接入带宽	用户规模	用户上网习惯
吞吐量	用户接入带宽越大，缓存设备吞吐量越大	用户数越多，缓存设备吞吐量越大	访问网站越集中，缓存设备吞吐量越大
命中率	NA	用户数越多，缓存设备命中率越高	访问网站越集中，缓存设备命中率越高
连接数	用户接入带宽越大，缓存设备连接数越小	NA	NA

3 结论

Internet Cache 系统通过将海量的互联网资源缓存到运营商网内实现了流量本地化，它给运营商带来三方面的价值：

- 降低带宽成本

  降低网间流量，节省互联网出口带宽租赁费用。

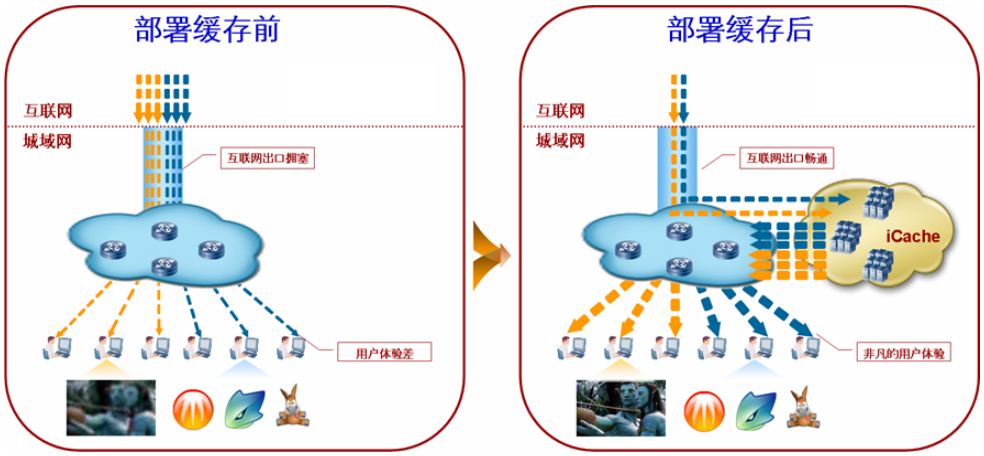
- 提高用户体验

  大幅提高用户上网体验，提升用户满意度和忠诚度，吸引新用户入网、减少老用户流失。

- 增加利润来源

  可基于缓存平台实现内容汇聚和用户行为分析，为差异化宽带服务等一系列增值业务提供运营基础。

图3-1 Internet Cache 的价值



目前，华为 Internet Cache 系统已经在中国移动、新加波 M1 等运营商大规模部署，成功的帮助运营商实现降低互联网出口压力和提升用户上网体验的双重目标，大大提升了运营商的市场竞争力。

A 缩略语表

Internet Cache	Internet Cache System	互联网缓存系统
RSS	Redirection Subsystem	重定向子系统
DSS	Dispatching Subsystem	调度子系统
MSS	Management Subsystem	管理子系统
CSS-WEB	Cache Subsystem for WEB Browsing	WEB 缓存子系统
CSS-HTTP	Cache Subsystem for HTTP	HTTP 下载/视频缓存子系统
CSS-BT	Cache Subsystem for BitTorrent	BT 缓存子系统
CSS-EM	Cache Subsystem for eMule	eMule 缓存子系统
CSS-PPS	Cache Subsystem for PPStream	PPStream 缓存子系统
CSS-QQL	Cache Subsystem for QQLive	QQlive 缓存子系统
CSS-PPL	Cache Subsystem for PPTV	PPTV 缓存子系统
LBS	Load Balancing Subsystem	负载均衡子系统
TFS	Traffic Filtering Subsystem	流量过滤子系统