SACC
2012中国系统架构师大会
SYSTEM ARCHITECT CONFERENCE CHINA 2012
架构设计 · 自动化运维 · 云计算

平台化 CDN 基础架构
-- ATS 缓存系统

赵永明

引言

CDN 技术发展经历了 10 多年的推演，相当多的领域已经有稳定的方案．而随着互联网的进一步发展，越来越多的 ugc 内容挑战着 CDN 系统的健壮性，同时新的 http 2.0 协议等新兴技术也开始挑战传统 CDN 的的方案．如何在新的形势下，让 CDN 系统健康向上发展？ 我们从运维的挑战为起点，介绍 ATS 系统方案．

赵永明 自我介绍

- 花名永豪，现就职于阿里巴巴集团技术共享平台系统部，负责 CDN 架构 .
- 多年机房建设系统运维经历 . 目前带领一个小团队参与 CDN 基础软件

  Apache Traffic Server 的开发工作，已经有 3 个人成为项目 PMC .
- 经过 2 年的磨练，正从一个写 shell 的码农转变成一个写 c++ 的大龄码农 .
- 任何 CDN 相关系统问题，都很乐于跟大家探讨
- 个人主要技术领域： 系统运维　CDN 架构
- 参与的主要开源项目 : ats tsar

提 纲

- 光纤时代 CDN 的挑战
- cache 系统的维度
- 性能指标
- 成本
- 可用性
- 平台扩展性
- 后 http 1.1 时代

光纤时代 CDN 的挑战

- 用户带宽大：很多城市已经推行 10-20M 带宽
- 在线时间长 :3G 手持设备如 Ipad 等可以提供更长的在线能力
- 交互数据多：页面丰富程度更高
- 移动客户端流行 :Adroid 和 IOS 设备大行其道
- CDN 发展： 容量更大，单个数据文件更大，命中率更低，业务逻辑需求增多

Cache 系统的维度

- 功能
  - 内容管理
  - 流量管理
- 性能
  - 连接管理能力
  - 本地 cache 的 RT
  - 动态 proxy 延迟
  - IO 效率
- 可用性
  - 磁盘，网络，源服务器等故障处理
  - 检测，统计，日志，报警数据支持
- 可扩展性
  - 能够支持多变的用户需求
  - 多种功能需求并存而不冲突
  - 扩展功能可快速开发
  - 核心框架和核心功能维持稳定

CDN 系统的性能指标

- 连接管理能力，并发连接数，维持长连接的能力
- 本地 cache 的 RT, 本机（本集群）已经缓存的内容响应时间（分内存和磁盘）
- 动态 proxy 延迟，对 TCP 网络优化的能力
- IO 效率：磁盘 IO 与网络 IO 的比率，磁盘 IOPS 与用户 QPS 的比率
- QPS/CPU=？决定了 CPU 的效率
- 处理复杂业务的能力：
  - Https
  - 大文件缓存
  - 甚至部分动态内容 :fast cgi, gzip 等
  - 混合环境表现（大小文件，快慢网络，丢包重传，超时等等）

成本总是问题，如何降低成本

Capex:

节点设备成本 / 实际计费流量

例如：某节点建设成本是 50 万，设计能力 10G,s 实际运行计费流量 10G，则成本 =5 万 /Gbps，按照 3 年折旧计算：则为 5/(3*12) =0.14 万 / 月 /Gbps

Opex:

节点带宽费用 / 实际计费流量

例如：某节点稳定流量大约 8G，实付带宽费用大约 40 万 / 月，则为：40 万 / 月 /8G=5 万 / 月 /Gbps

成本总是问题，成本的计算

节点设备成本 ↓

- 交换机
- 4-7 层设备
- Cache 服务器
  - 磁盘
  - 内存
  - CPU
  - 网卡

节点带宽费用 ↓

- 量大才能拿到批发价

实际计费流量 ↑

- 总流量 = 单机流量 * 机器数量
- 单机负载能力决定单节点能力，提高单机能力是降低成本关键

可用性

高压力下高可用性：

- 单盘损坏不影响机器服务能力
- 单机损坏不影响节点服务能力
- 快速启动，快速服务能力
- 能够处理客户端网络引入的各种问题
- 能够处理源端慢，宕机等引起的问题

可运维性：

- 内部数据可视，统计数据可定制
- 错误日志有利于问题分析.
- 能够报警程序故障
- 可配置，可查询，可快速调整
- 管理可自动化

平台扩展性

- 一个平台可以用几年？
- 一个平台可以 hold 住几个业务需求？
- 这个平台是开源的 还是私有的？
- 平台发展方向是否与我们需求一致？
- 谁来设计（调整）核心框架？

平台扩展性：Cache 系统的扩展方向

- 动态加速，回源加速， 链路选择，TCP 优化
- 动态合成技术： ESI
- 动态针对客户端进行压缩，以支持各种手持终端
- 复杂业务系统支持，如鉴权，视频点播直播
- 连接管理升级： websocks ， SPDY
- 安全过滤， 杀毒拦截

后 http/1.1 时代

- Ajax 的大量使用，促成了框架页面静态化 – – 有更多内容可以做 cache 啦
- Html V5 的普及，更多的视频等媒体展示 – – 文件 流量 都 变的更大啦
- http/2.0 是否会包括 SPDY? – – 不管怎么样，客户端的效率都要照顾到
- http/2.0 是否会包括 websocks? – – 我们需要一个双向通信的机制
- http/2.0 来了，服务器该咋支持呢？ – –V0.9 还在用哎，你用啥样的代码来支持？

我们的选择

Apache Traffic Server 简介

- Apache Traffic Server™ is fast, scalable and extensible HTTP/1.1 compliant caching proxy server
- ISP 级别的高性能 proxy/cache 服务器
- 缓存效率高，响应快
- 支持长连接、连接复用、过滤规则、映射、甚至 7 层 hash 和负载均衡、 Cluster
- API 很方便的支持各种环节的自由处理

Apache Traffic Server 对比其他软件区别

- **企业级目标专门设计：**
  - **全功能目标下，高性能要求**
  - **高压力下，高可用**
- **良好的社区及大企业支持：**
  - **Apache 基金会**
  - **阿里巴巴 雅虎 Akamai Comcast LinkedIn Cisco Apple...**

Apache Traffic Server 亮点 功能全，性能好

- 功能齐全
  - 支持正反向代理，透明代理
  - 设计巧妙的 Cache 系统
  - 面向管理员设计的管理界面
- 性能强大
  - 多线程异步全事件框架
  - 高压力下，服务表现很好 (RT 较低)

过去也曾经支持过 dns ftp 以及 Quicktime MMS Real 等流媒体

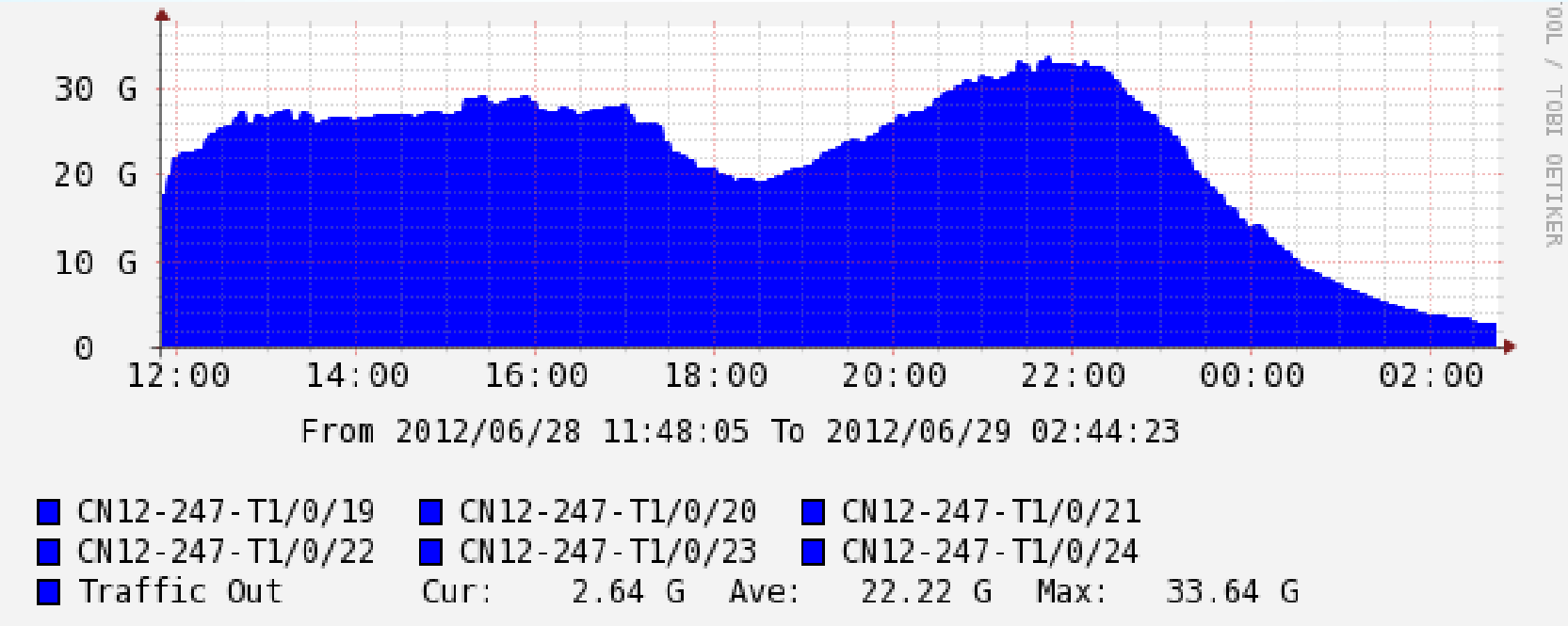
不绕过任何难题

Apache Traffic Server 亮点 功能全，性能好

- 我们配置 16core 8×600G 万兆网卡能跑出多少流量来？
- TS 能：
  - 实验室单机记录： 16gbps ， 112kqps
  - 线上试验记录： 7.7gbps （全硬盘命中）
  - 实际运行记录： 3.4gbps ， 17kqps（本机同时跑 haproxy)

Apache Traffic Server 亮点 功能全，性能好

- **我们线上运行节点某高峰流量，10 台 TS 在线服务**



Apache Traffic Server 亮点　可用性

- 可用性高
  - 多种管理界面，可管理程度高
  - 容灾设计，可消化常见硬件故障
  - 快速启动， 快速恢复
  - 配置可定制程度高，可控制核心参数 500+ 条目 .
  - Cache 控制能力强
  - 可定制化的日志采集汇总汇报系统
  - 可定制化的数据统计系统
  - 集群化管理能力

Apache Traffic Server 亮点 扩展性

- 高度可扩展性
  - 模块化程度高：核心 http 引擎只是 TS 的 2 个引擎之一（另一个是流媒体引擎）
  - 高度可编程核心插件设计，可以完成各式各样业务，如已有巨型插件：
    - ESI, Edge Side Include，雅虎贡献的代码
    - Metalink
    - Lua remap 插件，支持 lua 语言的 script
    - Gzip 插件，可以对 html 等文本文件进行深度压缩
  - API 扩展支持完善，插件开发介入门槛低

Apache Traffic Server 亮点 日志

- 完善的日志
  - 支持 binary 日志，高效
  - 支持 squid 日志格式，兼容
  - 可根据源站分离存储
  - 可将日志发送至日志服务器
  - 内建高性能、可定制化日志聚合功能，自动生成流量 / 性能报告
- 自带高性能日志收集功能，也有独立服务器

Apache Traffic Server 业界生态

- 用在私用 CDN 系统：
  - 阿里巴巴（淘宝）
  - 新浪
  - 雅虎
  - LinkedIn
- 用于商业 CDN 系统：
  - Akamai
  - Comcast
- 用于其他解决方案，做核心引擎：
  - Websense
  - Cisco

Apache Traffic Server 社区开发动态

- **20120907, V3.3.0:**
  - **更好的 SSL 支持，Lua 插件以及 Cluster 性能优化**
  - **RFC5861 插件**
  - **Gzip 插件**
  - **Metalink 插件**
- **即将发布的 V3.3.1:**
  - **SPDY 插件，支持 SPDY V1&V2**
  - **…**

Apache Traffic Server 技术方向

- 动态加速，回源加速
- 动态合成技术： ESI
- 动态针对客户端进行压缩，以支持各种手持终端
- 复杂业务系统支持，如鉴权，视频点播直播
- 连接管理升级： websocks ， SPDY, https, 甚至 cluster
- 高级 cache 平台 : SAS->SSD->MEM 3 级缓存设计，甚至多级
- 集群实践： 去掉 haproxy, lvs 等 4-7 层专用设备

扩展信息

- TS 相关资料：
  - TS相关专利, 学习 cache 系统核心设计的好地方
  - 我的个人收藏 http://people.apache.org/~zym/trafficserver/
- Join US here:
  - CU 服务器应用版：http://bbs.chinaunix.net/forum-232-1.html
  - IRC: #traffic-server on freenet
  - users@trafficserver.apache.org dev@trafficserver.apache.org

Q & A