Apache Traffic server

作者自述

赵永明， 07 年左右起使用 trafficserver 作 cdn 服务，国内最早的 trafficserver 用户。

Mail/Y!/MSN： ming_zym@yahoo.com.cn

http://zymlinux.net/trafficserver

Phone： 13910237531

随时欢迎沟通探讨 trafficserver 任何问题！

广而告之

- Trafficserver 需要大家的支持：
  - C++ 高手修复 bug
  - CDN 运营商的技术参与，扩展应用
  - 硬件 /OS 编程专家优化系统
  - 文档翻译
- 好用就帮忙一起推广啊！

http://trafficserver.apache.org/

Agenda

- trafficserver 是谁？
- trafficserver 历史与现在
- trafficserver 功能与特性介绍
- 如何用 trafficserver?
- trafficserver 的将来

Apache TrafficServer

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
- 代理支持长连接、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

Trafficserver 前世今生

- Inktomi
- 互联网泡沫
- ISP
- CDN
- 搜索
- Yahoo
- Apache

Inktomi & traffic server 的历史

- 1996 年， UC Berkeley 的 Eric Brewer 和学生 Paul Gauthier 成立 Inktomi 。以搜索算法起家，并开始开发 trafficserver 。
- 1999-2000 年间，收购多家 cdn 技术公司，成为 cdn 技术集大成者，一度掌控所有 cdn 技术市场，集成商甚至包括所有的主要硬件商如 IBM HP DELL Foundry 等。股价最高达 $241
- 2002 年，网络泡沫过后，为 yahoo 以每股 $1.63 收购
- 2003-2008 年， trafficserver 在 yahoo 内部逐渐发挥作用，并成为 yahoo cache/proxy 体系的核心。
- 2006 年， websense 购买并集成 trafficserver 。

traffic server

Apache Traffic Server 大事记

- 2009 年 7 月 13 日，进入 apache 基金会孵化器
- 2009 年 10 月 30 日，公开第一份完整代码
- 2010 年 4 月 30 日，成为 apache 基金会顶级项目
- 2010 年 5 月 4 日， trafficserver 稳定版 2.0 发布
- 2010 年 6 月 7 日， 2.1.1 测试版发布

TS 企业级功能与特性介绍

- 性能强劲
- 多 cpu ， 多线程支持
- 支持 raw 设备 / 分区
- 启动重启以秒记
- 64 位系统支持
- 内存 footprint 低
- IO 性能优化
- 适于中国网络
- 灵活的 map 机制
- 强大的 Cluster 支持
- 全方位管理界面
- 全方位监控接口
- 丰富的扩展接口

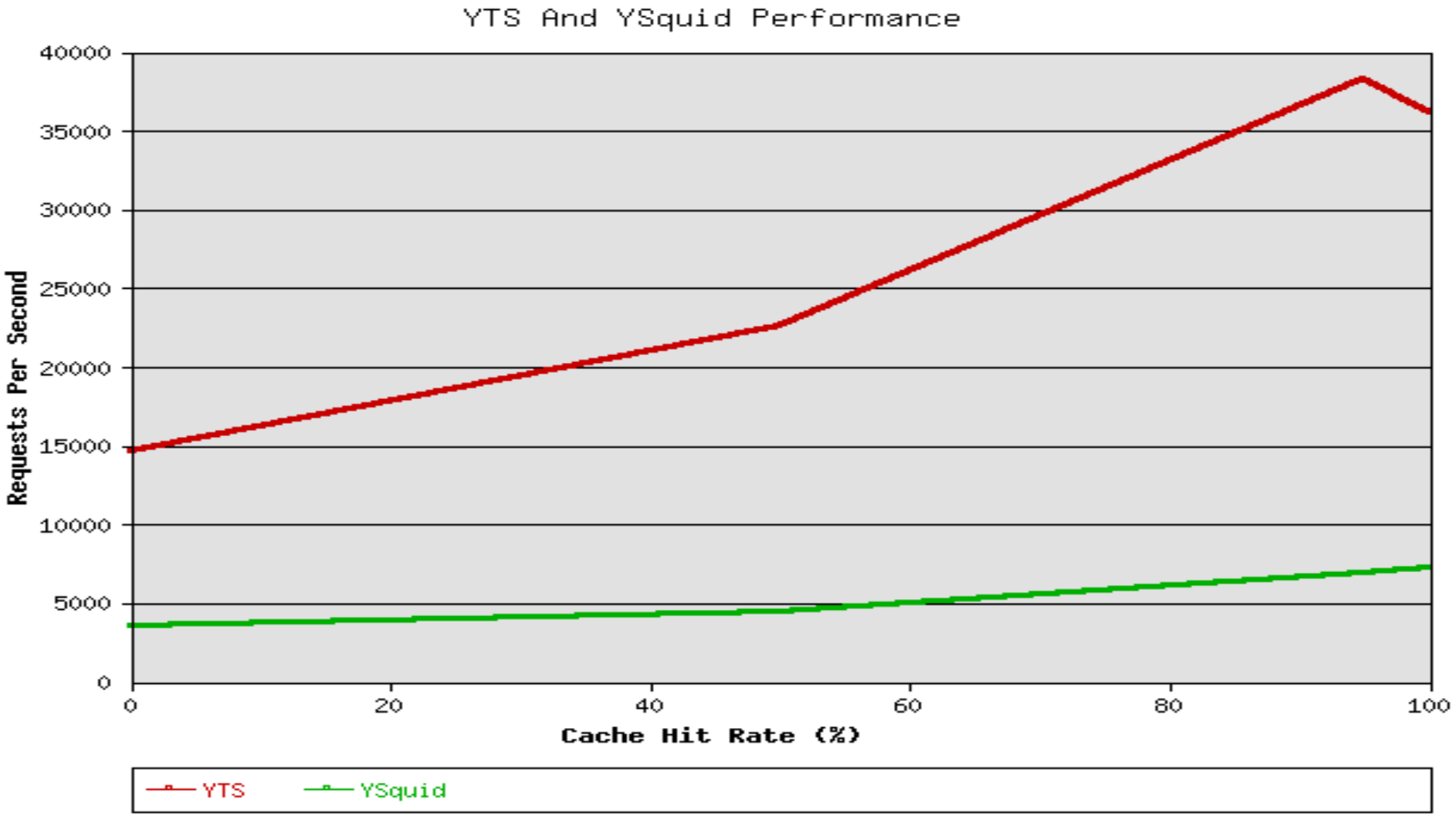
traffic server

TS 性能有多强（08 年 yahoo 内部测试）

- 硬件测试平台环境：
  - Dell PE2950, 2 x Xeon E5320 1.86GHz, 7.8GB / 8GB 667MHz / 6 x 147GB 15K SAS/3 Fujitsu MAX RAID-5
- 测试方法
  - Variable cache hit ratio percentages (0, 50, 95, 100)
  - 1,000 client connections
  - 1KB response from the origin
  - 4 Keep-alive requests per connection
  - 10,000 unique objects

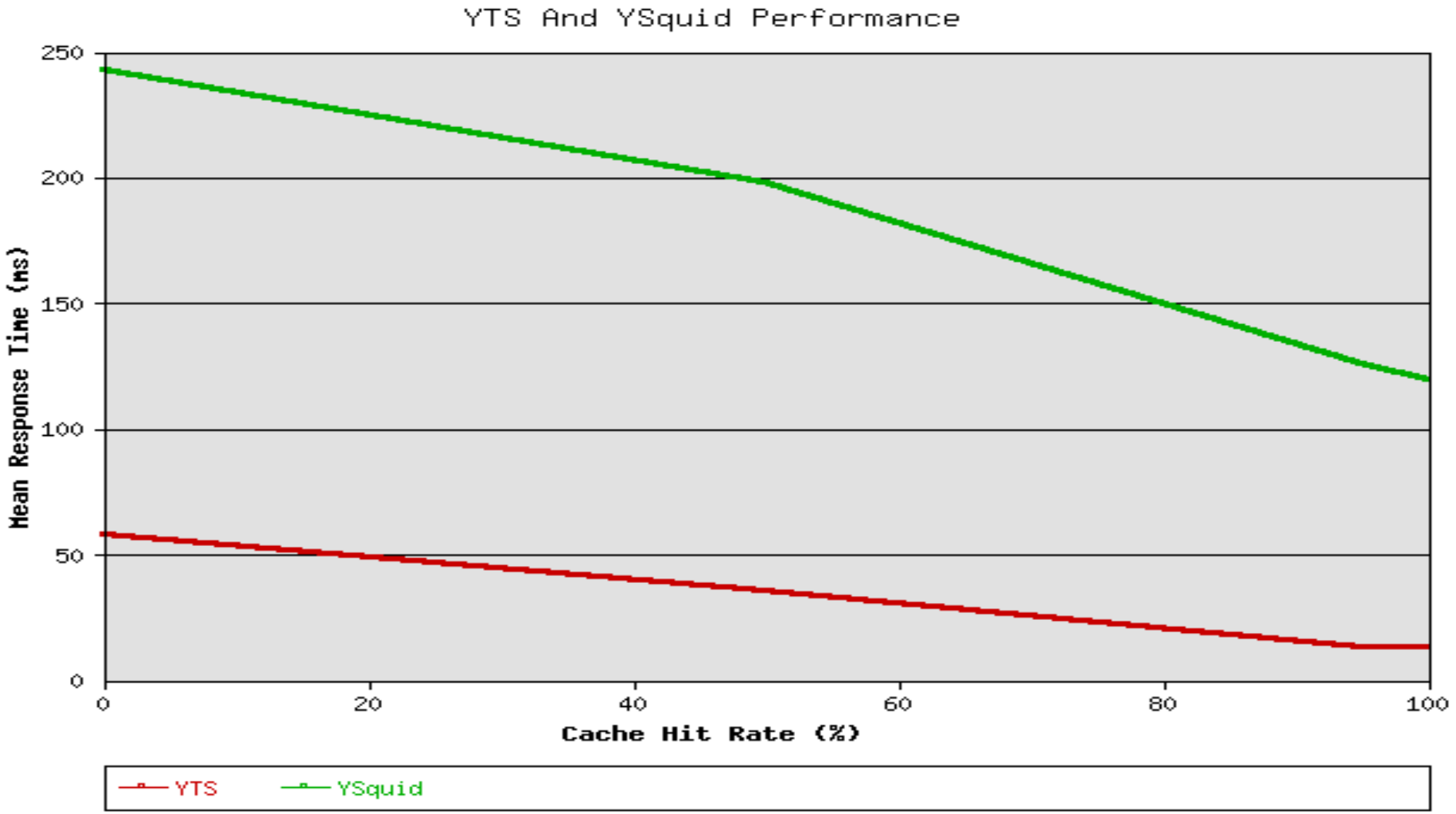
traffic server

测试结果 QPS



traffic server

测试结果 RT

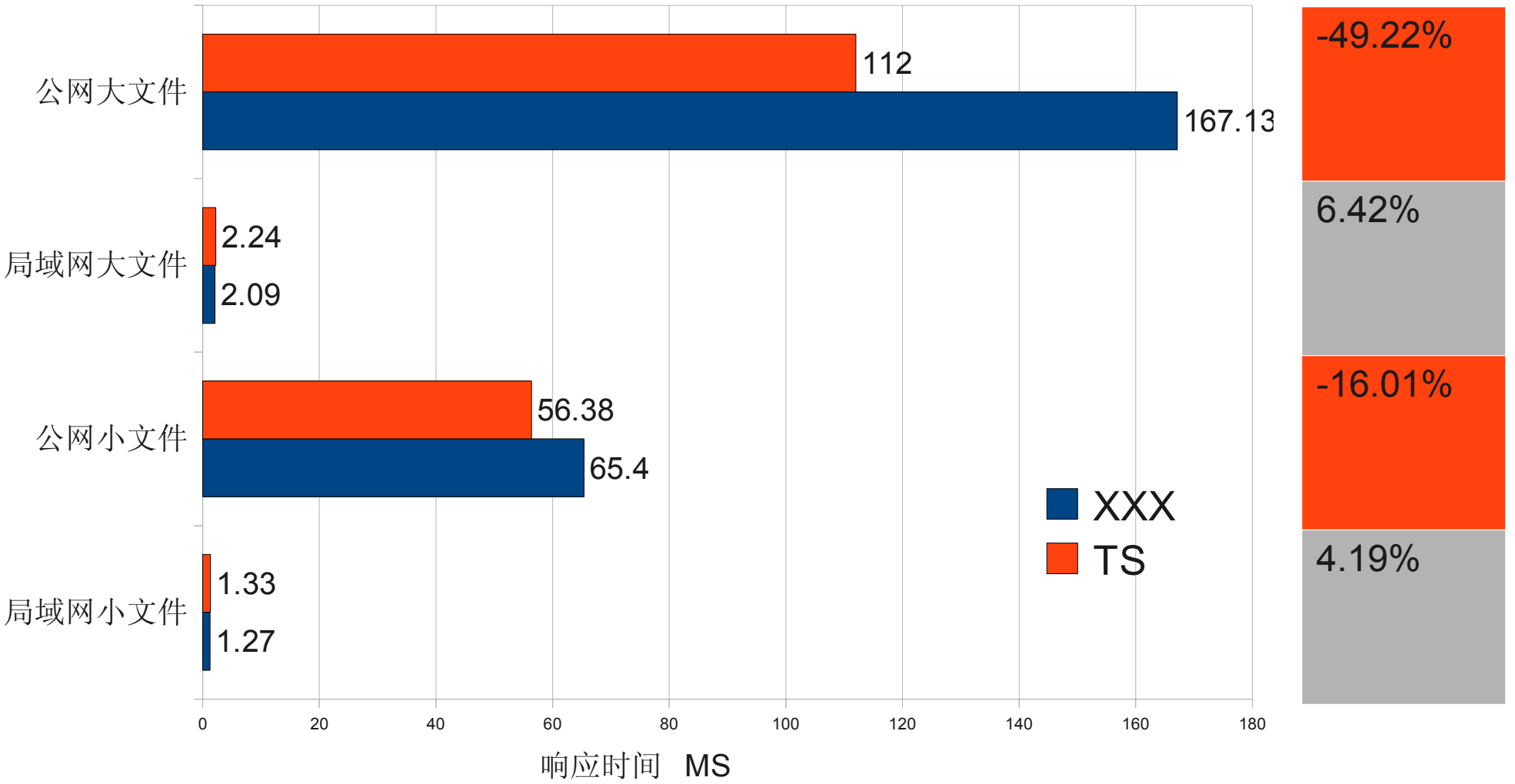


traffic server

与某 cache 软件对比

- 测试环境：
  - 内网测试：
    - 局域网千兆环境
  - 公网：
    - 北京到杭州间，中国电信网络间测试， ping 在 25 个 ms 左右。
  - 大文件： 20,833Byte 小文件： 2,226Byte
- 测试方法： nagios: check_http 抓 1000 次取平均响应时间。测 3 次取平均值。

另一个测试结果图



traffic server

SMP/ 线程 / 非阻塞 IO

- 支持多 CPU 多核心新硬件构架
- 支持多线程运行
- 支持非阻塞 IO （见后续框架设计说明）
- 支持一线程处理多个 http 连接

traffic server

存储系统

- 使用类似 squid 的 coss 存储
- 支持直接使用裸盘
- 支持对存储系统进行 partition， 以区别存放不同的内容。
- 支持对不同域名 / 协议进行匹配 partition 。
- 支持异步 IO 和 fastIO， IO 效率高。
- 硬盘采用 RRD 模式，写入（重写）压力小。

traffic server

启动 / 重启，进程监控

- traffic_server 主进程启动时间以秒记，存储空间大小对启动时间几乎没影响。
- 双保险：
  - traffic_manager 负责管理 traffic_server 的健康稳定运行，并在非正常情况下，负责重启 traffic_server 进程。
  - traffic_cop 监管 traffic_manager 运行
- 更新系统配置不需要重启进程。
  - 仅在变更如 cluster 、 storage 等基础设置，需要重启进程。

64 位系统支持

- 支持 64 位寻址内存空间
- 支持 64 位寻址硬盘空间

灵活的 map 机制

trafficserver 采用 map 的方法来进行反向代理的域名映射，有非常灵活多样的使用方法，其统计信息以源信息为准：

- 普通域名映射，对外域名是 cdn ，源域名是 source ：
  - map http://cdn.zymlinux.net/ http://source.zymlinux.net
- 普通域名映射，对外域名是 cdn ，映射到源 zymlinux.net 的 /source/ 目录下：
  - map http://cdn.zymlinux.net/ http://zymlinux.net/source
- 反向域名映射，用来修改如源的重定向结果里的 URL ：
  - reverse_map http://zymlinux.net/source http://cdn.zymlinux.net/

灵活的 map 机制

- 将 old 永久重定向到 new:
  - redirect //old.zymlinux.net/ http://new.zymlinux.net
- 正则匹配的映射，一把映射 N 个域名：
  - regex_map http://x([0-9]+).z.com http://real-x$1.z.com
- 还支持 map 插件，可自由编写插件，基于 cookie/header/ 源地址 / 账户等等，只要你想的出来，就没有做不到的。

traffic server

内存使用与管理

- 8byte/obj vs squid:56byte/obj
- 2TB 缓存内容下，如是 8K 大小的文件，大约需要 2.5G 的内存
- 2TB 缓存内容下，如是 128K 大小的文件，大约需要 0.5G 的内存

traffic server

IO 性能优化

- 支持 epoll/libevent 等技术
- 独特的 eventsystem/SM 设计
- 硬盘 IO 线程数可以根据应用环境设置
- 支持 directIO/ 异步 IO，降低系统开销

traffic server

适于中国网络

- 能够高效维持超量的网络连接。
- 能够在网络延时大的情况下，做到 RT 稳定。
- 能够进行各种 7 层转发，作 7 层路由
- 能够汇聚 http 连接、保持长连接，提高用户感受。

traffic server

Cluster

- 支持 3 种工作模式：
  - 全 cluster （ cache 内容 + 配置文件）
  - 管理 cluster （配置文件）
  - 独立服务器 （ cluster 关闭， -- 默认模式）
- 无 master 机制， cluster 成员可以自动加入。
- cluster 通讯可以采用独立网卡 / 交换机
- cluster 连接最少化， n-1 个 tcp 连接
- cache 内容 hash

traffic server

Cluster

- 内建 cluster RPC 协议，独立 cluster 通讯体系。
- 复杂配置文件更新机制，确保同步与 rollback 迅速有效。支持配置文件 snapshot/ 配置上传下载。
- 内建配置文件版本跟踪机制，确保配置文件不错乱。
- 能够区分全局设置与私有设置，不会错误同步信息。

traffic server

日常管理

- trafficserver 提供多种可以快速即时生效的配置管理界面：
  - 命令行
    - traffic_line -s
    - traffic_shell enable 模式
  - 配置文件，使用 traffic_line -x 刷新
  - Web 管理
  - 批处理文件　traffic_shell

修改系统参数的各种方法

- 例子：修改 TS 到源服务的 tcp 长连接在没有传输的情况下的 timeout 参数为 20 秒：proxy.config.http.keep_alive_no_activity_timeout_out
  - zymtest1 repositories # traffic_line -s proxy.config.http.keep_alive_no_activity_timeout_out -v 20
  - 修改 records 文件下的 CONFIG proxy.config.http.keep_alive_no_activity_timeout_out INT 30 行，改 30 为 20，保存后用 traffic_line -x 生效
  - 执行 traffic_shell，进如 enable 模式，执行 config:http inactive-timeout-out 20
  - 用 web 界面改之
  - 执行 traffic_shell 批处理文件改之

状态统计与监控

Trafficserver 支持多种状态查询工具：

- 功能强大的 traffic_logstats:
  - 按照域名分别统计
    - 请求结果
    - 返回码
    - 回源统计
    - HTTP Methods
    - Content Types
    - 响应时间

状态统计与监控

- 支持 mrtg 图的 web 管理界面
- 命令行工具： traffic_line -r
  - 查询 cache 使用的空间大小：
    - zymtest1 ts # traffic_line -r proxy.process.cache.bytes_used
      273408
  - 查询 cluster 中的服务器数量：
    - zymtest1 ts # traffic_line -r proxy.process.cluster.nodes
      2
- 支持邮件报警接口 records.config:CONFIG proxy.config.alarm*

traffic server

状态统计与监控

- 支持 SNMP 状态：
  - 独立 mib 库
  - 可控制读取权限
  - 64bit
- 支持 SNMP TRAP ： snmpd.cnf

traffic server

ISP 程度的计费能力

- 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

traffic server

cache 管理

- 支持 push
- 支持自动刷新、正则刷新
- 可以强制 cache
- 可以删除、批量删除、正则删除。

DNS

- 内建 DNS cache 服务，功能可以作为 cache resolver 啦
- 支持 split dns ，方便处理内外网。

SSL

- trafficserver 支持同时监听多个 IP，使用多个 ssl 证书。（非多进程模式）
- 能够仅在客户端与 TS 端使用 ssl 。
- 能够仅在 TS 与源端使用 ssl 。
- 能够同时在客户端与 TS 和 TS 与源端使用 ssl 。
- 能够支持 ssl 加速卡。

traffic server

强大的插件扩展能力

- 雅虎内部已实现：
  - 多种 7 层路由功能
  - 各种处理 header 的插件：
    - 添加 header
    - 删除 header
    - 验证 header …
  - 各种安全控制功能

由于事关内部业务，这些插件多数不会开源

traffic server

其他好用的小功能

- 源服务器拥塞管理：在源服务器发生拥堵的情况下，可以给用户一个特定页面，并发给管理员报警。
- 可配置的复杂的 enfreshness 机制
- Background 刷新、定时刷新
- 支持内容 2 次验证机制
- 日志文件自动清理，避免硬盘满

traffic server

透明代理服务

- 支持全透明代理（双向 + 反向）
- 支持客户端透明代理（ inbond ）
- 支持服务器端透明代理（ outbond ）
- 支持多端口单独定义，可以在一个机器上跑几个代理端口，并设置不同的类型。
- 详细设置参考 Tiphares

traffic server

怀疑到底？

- 单CPU的个人电脑跑出10万5千qps
- Websense 集成 inktomi trafficserver 至今。
- Yahoo 大部分流量已经跑在 trafficserver 上了
- 台湾无名小站用少量的机器即负载所有的视频。
- 世界最大的 CDN 运营商 Akamai 已经将 trafficserver 的 PM Leif Hedstrom 召入麾下。

试一下就知道啦

traffic server

为什 trafficserver 高性能？

- Event 系统设计
- 状态机
- 低内存占用，高硬盘 IO
- 均衡读写（异步 IO 、预读）
- 科学家的设计

当年的高人们

- Dirk Grunwald　现为知名教授，名下一堆的 phd，当年也为 ts 写了底层代码哦
- Brian Totty : HTTP: The Definitive Guide

代码量惊人

- 16MB 的纯代码。
- 历时约 7 年的设计开发
- 历时约 5 年的后期维护与验证
- OEM 客户

traffic server

原系统还支持过：

- Ftp cache
- Ftp2http cache
- Quicktime 视频
- REAL 视频
- MS 视频
- 病毒扫描
- 甚至有人准备 port 到 arm 啦

如果你的 CDN 体系有如下问题

- 内容多的不行，需要集群来处理
- 访问量大，得好多机器才能够抗住大压力
- 管理复杂，需要迅速响应，应对各种突发问题
- 提高用户访问速度，优化连接
- 应用需求越来越多，功能开发不过来
- 故障频发，疲于应付

上 trafficserver 啊

Perfect Match! 还免费的哦！捡了个宝啊！

http://trafficserver.apache.org

trafficserver 值得投资？

- 模块化极强、框架设计优秀，能够提供各种接口来完成各种业务需求！
- 雅虎多年应用开发的平台的锤炼！
- apache.org 的良好运作！
- 开源后核心代码更新明显加速！
- ISP 级别的同级系统无人能敌！
- Apache license 对企业用户友好！

Cluster CDN- 先

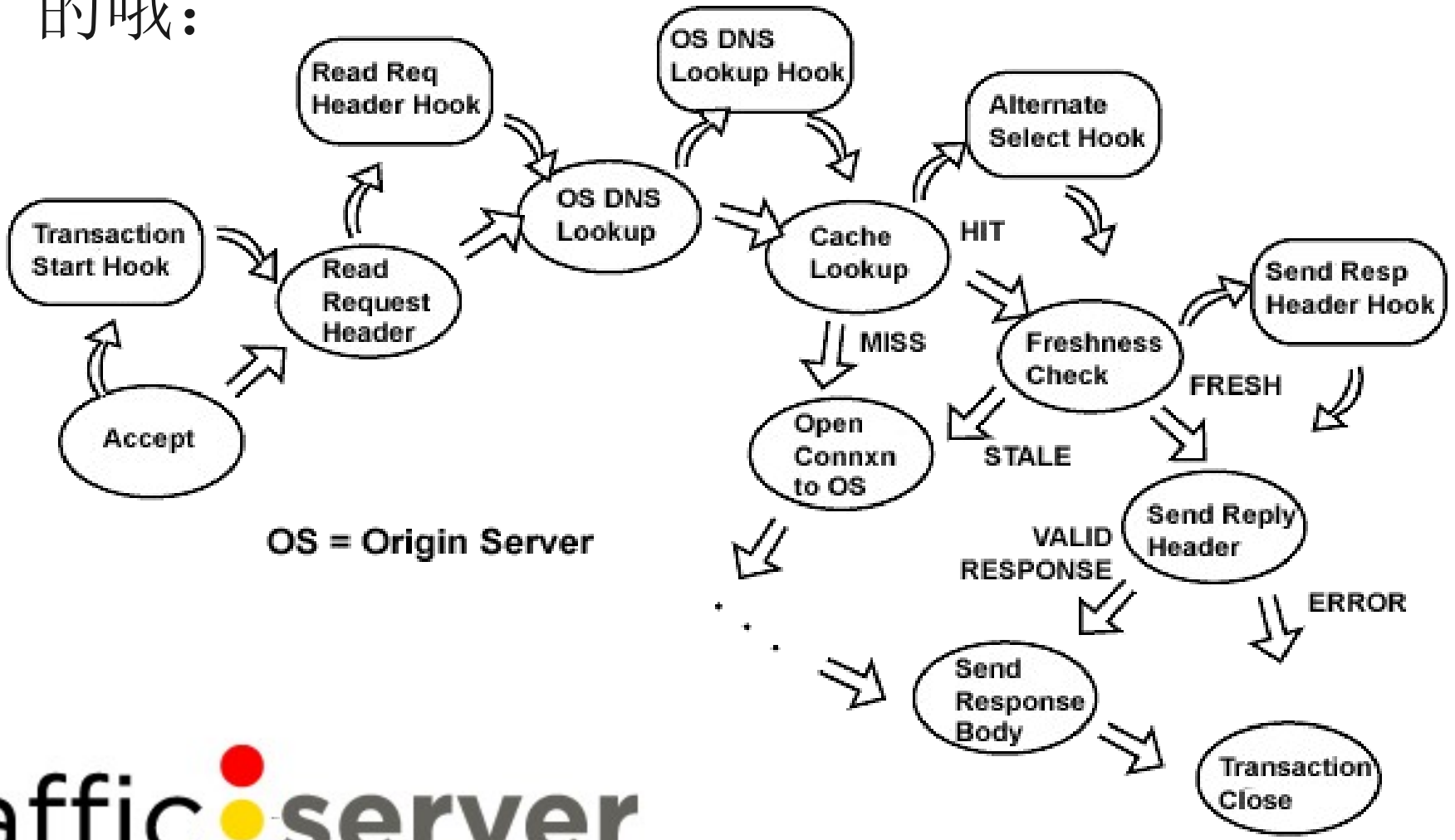
- 使用 trafficserver 构建 cluster 平台
- 提高管理行，运维效率
- 流量管理

Cluster Proxy- 后

- 7 层路由
- 连接管理
- SLA 管理

特殊应用 - 再

- 根据需求开发相关的插件 / 工具。好多钩子可用的哦：



traffic server

不远的 2.X 将会有如下新功能

- ipv6 集成
- 透明代理重新回归
- IO 层系统革新
- >2G 大文件支持
- solaris/osx/bsd port
- Cluster 采用一致性 hash
- 期待雅虎美国开源部分功能插件

已知问题

- web 界面老土，需要重新设计，目前版本默认不开启
- 设计文档少之又少
- QA benchmark 系统未开源
- 可选实用功能插件少（由于功能插件与业务关联太大， Yahoo 不太方便开源）

traffic server

QA

Do you really have Any Question?

20100810