系统也智慧

百度系统部

刘宁

Agenda

1、我的地盘我作主

2、无创新，不系统

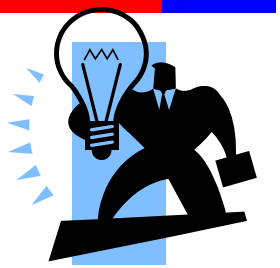
Agenda

1、我的地盘我作主

2、无创新，不系统

Baidu 百度

我理解的系统工程师

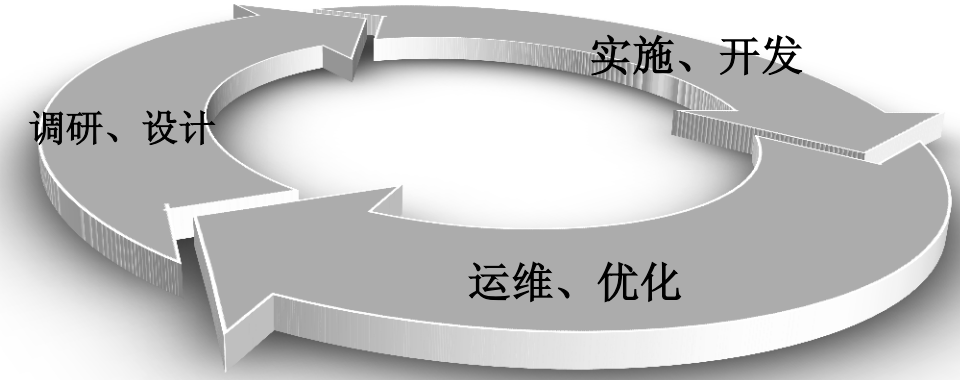


个人理解

系统：

相互关联的单元，按制定的规则进行运转成为整体。

业务单元、计算、存储、数据、网络、IDC 。



系统工程师：

PM+RD+OP，制定规则，给予系统以智慧。

调研、设计、开发、实施、运维、优化。

合格的系统工程师：

苦练内功，博采众长

视角不能受限

Baidu 百度

我们要解决哪些问题

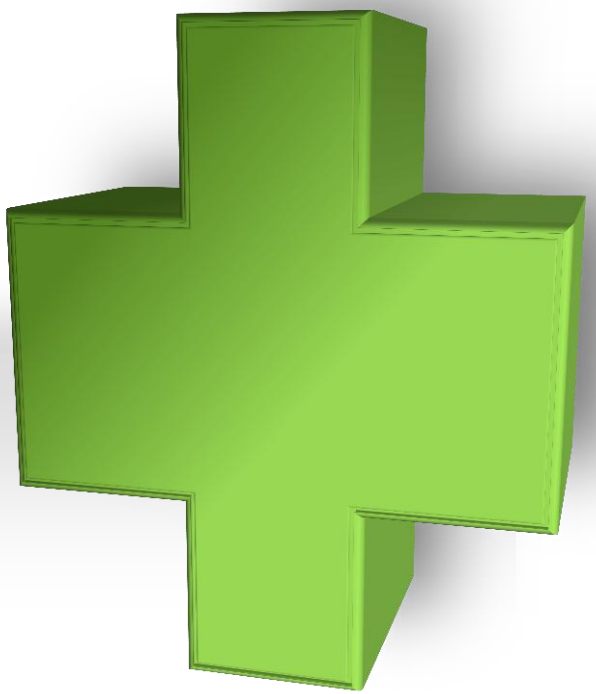
- 搭建适合业务的平台
- 简单、可依赖的系统
- 易于运维、快速扩展

Baidu 百度

Trouble 1

Load balancer

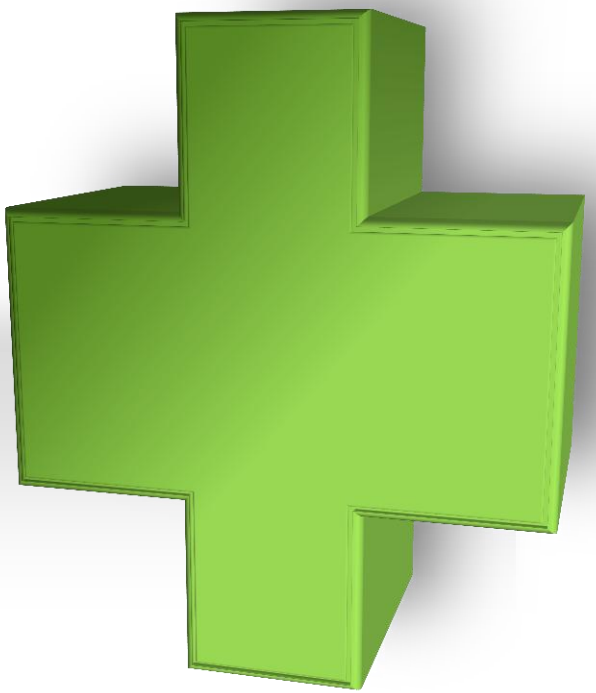
- 网络、集群设计受限于商用设备
- 功能特性是否能快速升级
- 业务多样，操作繁多
- DIY？性能、稳定性如何保证



Trouble 2

GSLB

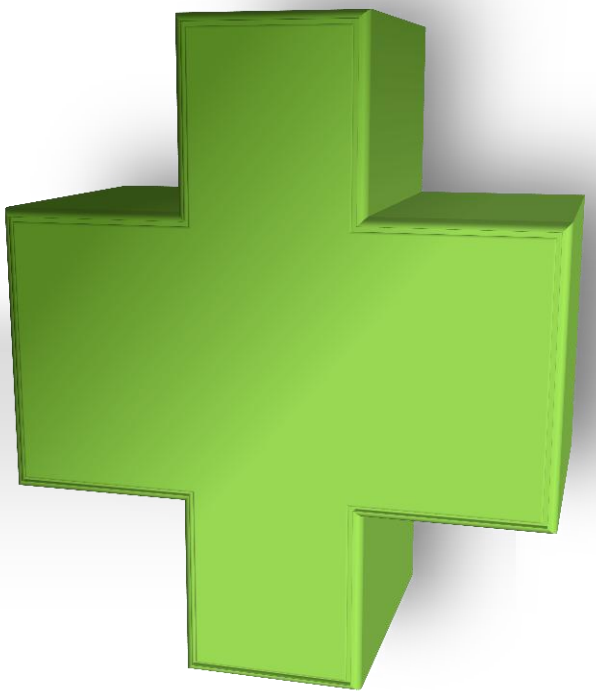
- 多出口、数百G业务流量如何调度
- 怎样提升用户体验



Trouble 3

大规模DDoS攻击

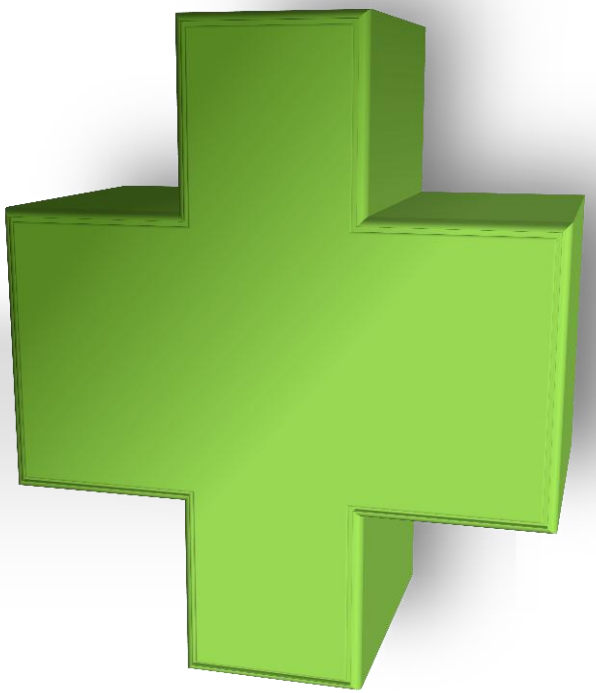
- 业务大量资源及代码逻辑进行攻击防御
- 如何能动态学习攻击特征，而非固定策略
- 多个层面的攻击，如何防御



Trouble 4

专家系统

- 如何精准、精简的进行告警
- 报警系统能否分析告警，学习告警

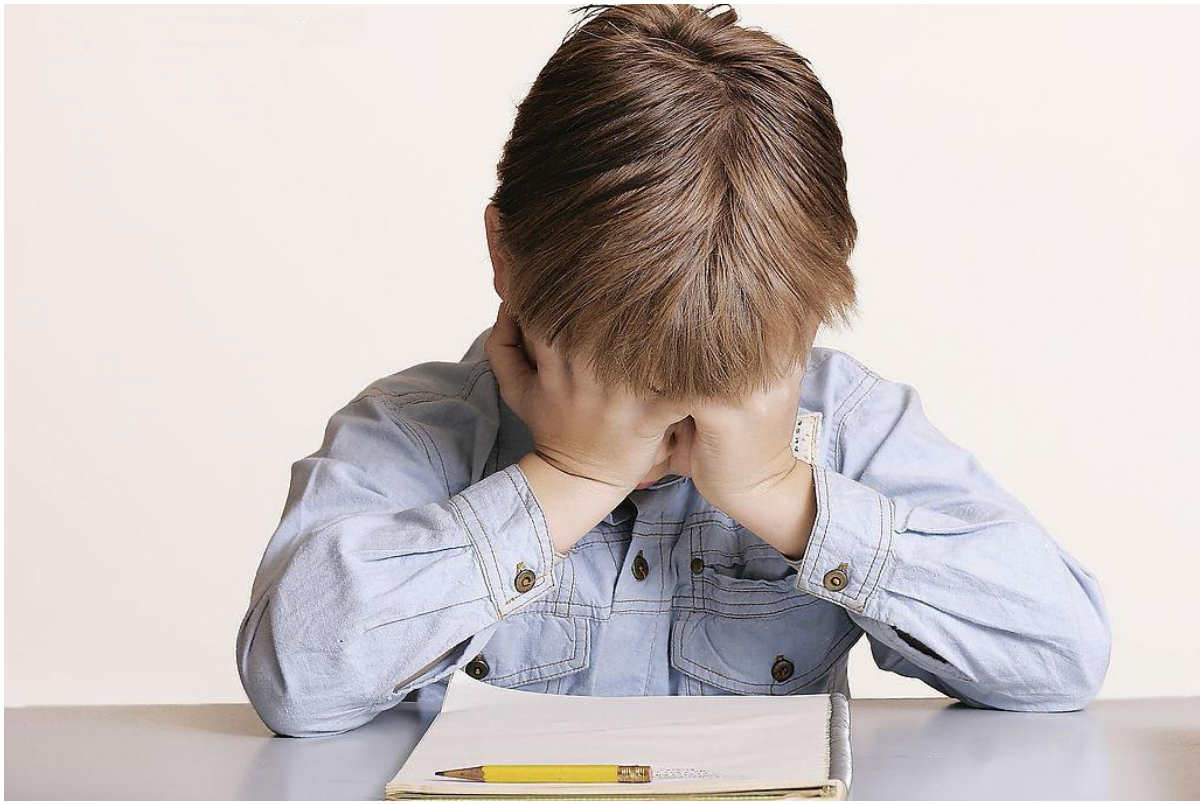


Baidu 百度

思考

if you don’t trouble trouble,

the trouble will always troubles you



Agenda

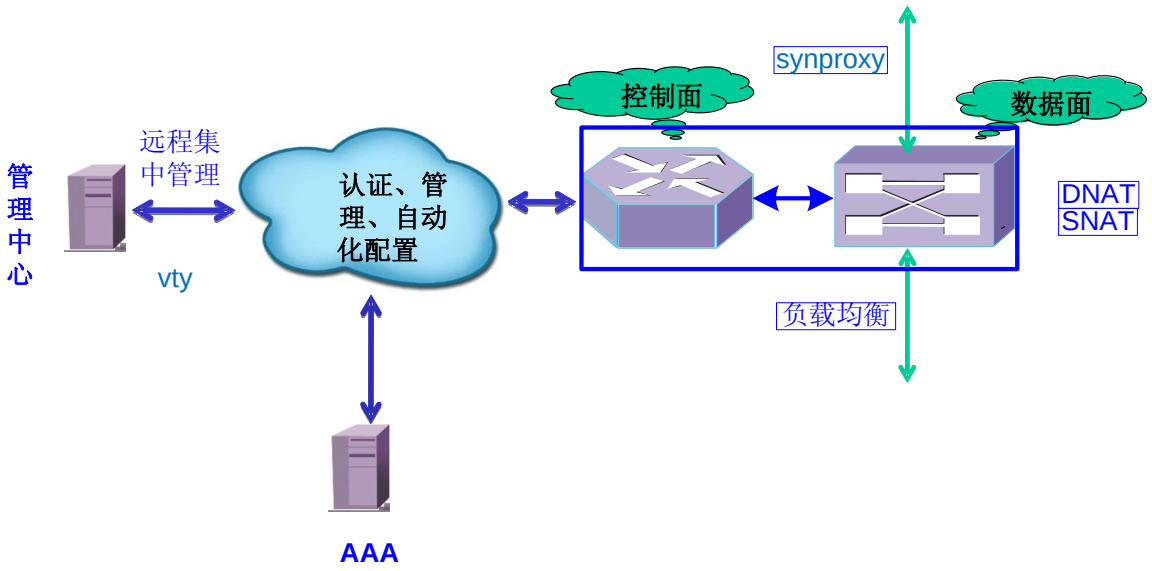
1、我的地盘我作主

2、无创新，不系统

CASE 1

百度智能网关BGW

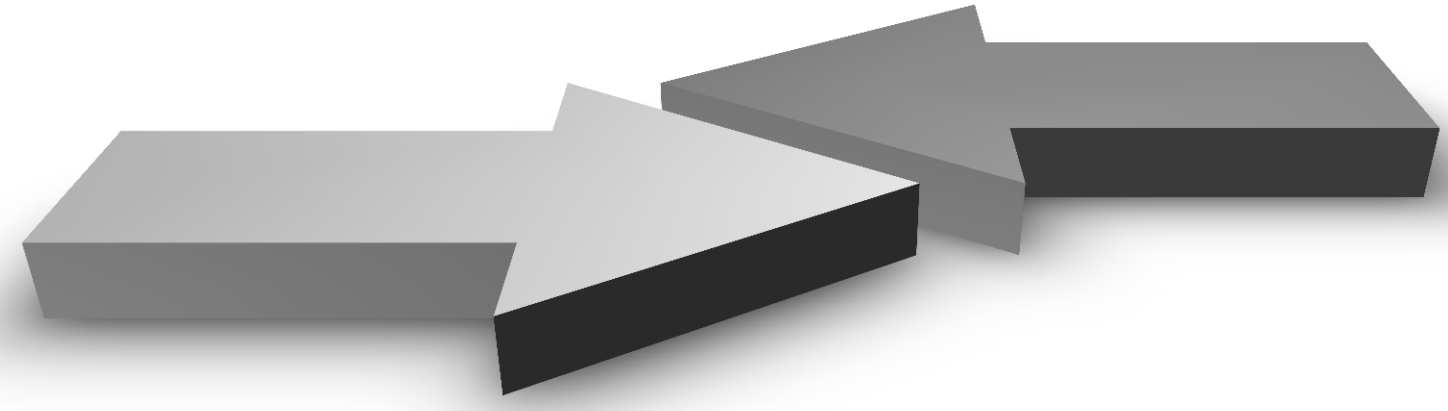
- 主备/集群方式
- 众核设计
- D、C分离
- Freecore无锁技术
- 自动配置管理
- 跨IDC RS调度
- SNAT client IP carry
- 丰富的调试信息



CASE1

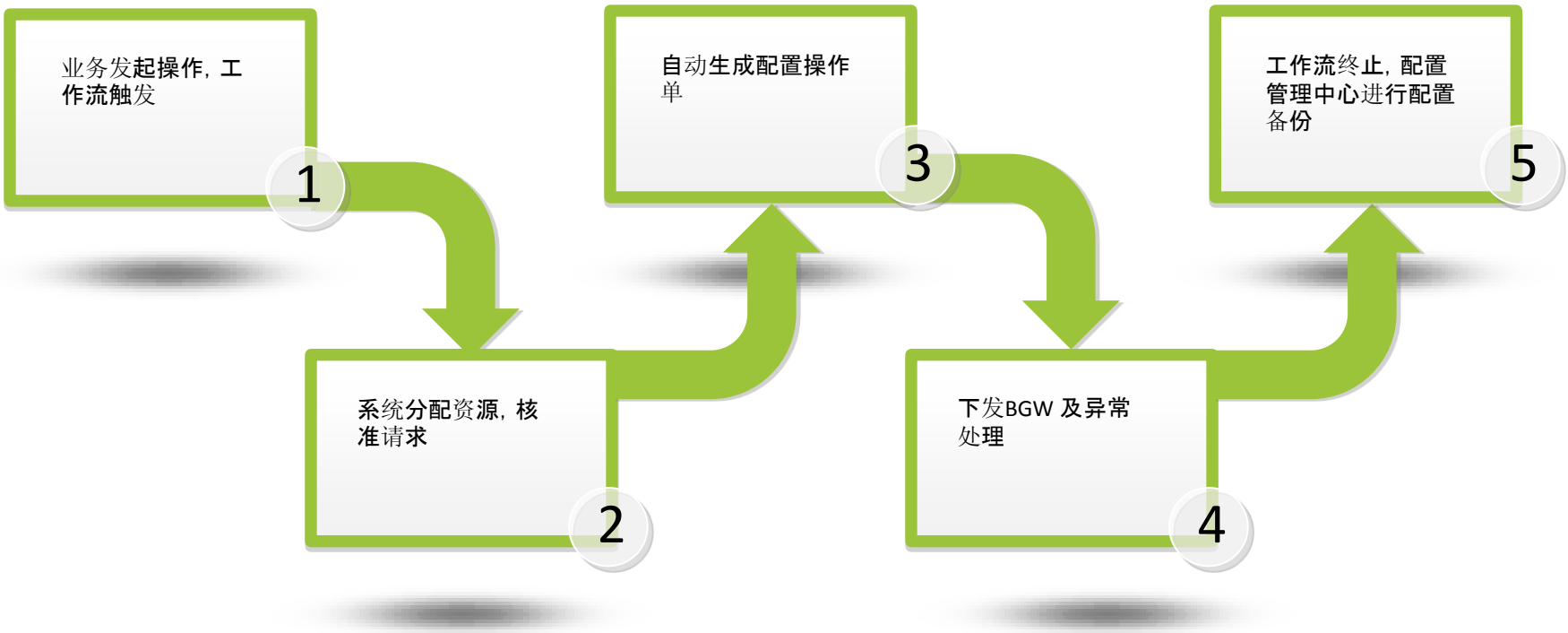
Single Box perfmance PK

BGW	X86 LVS
▪ 转发处理 10G线速	▪ 转发处理300w
▪ Synflood防御 10G线速	▪ Synflood防御200w
▪ Session 取决内存	▪ Session 取决内存
▪ 新建连接3Mcps	▪ 新建连接10wcps



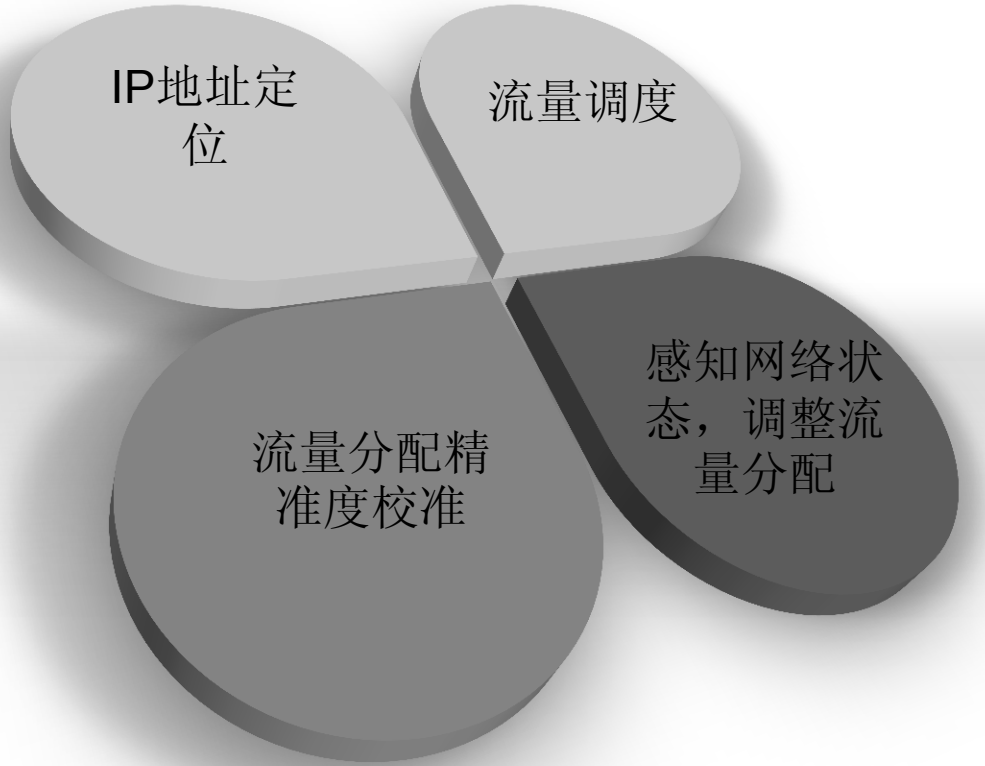
CASE1

BGW自动配置管理



CASE2

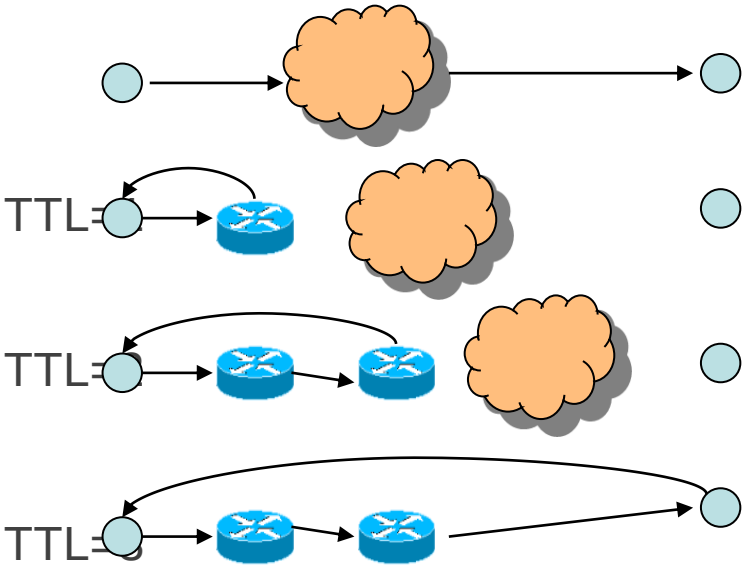
百度GSLB系统



CASE2

哥伦布——广域网IP、拓扑定位

- 邻接表
  - destIP:IP1|x ms:IP2|y ms: IP3|z ms….
    - IP1=>IP2,IP2=>IP3
- 骨干网提取
  - 延时
  - 节点度
  - 节点承载量
- 可视化
  - Dot,Neato,FDP



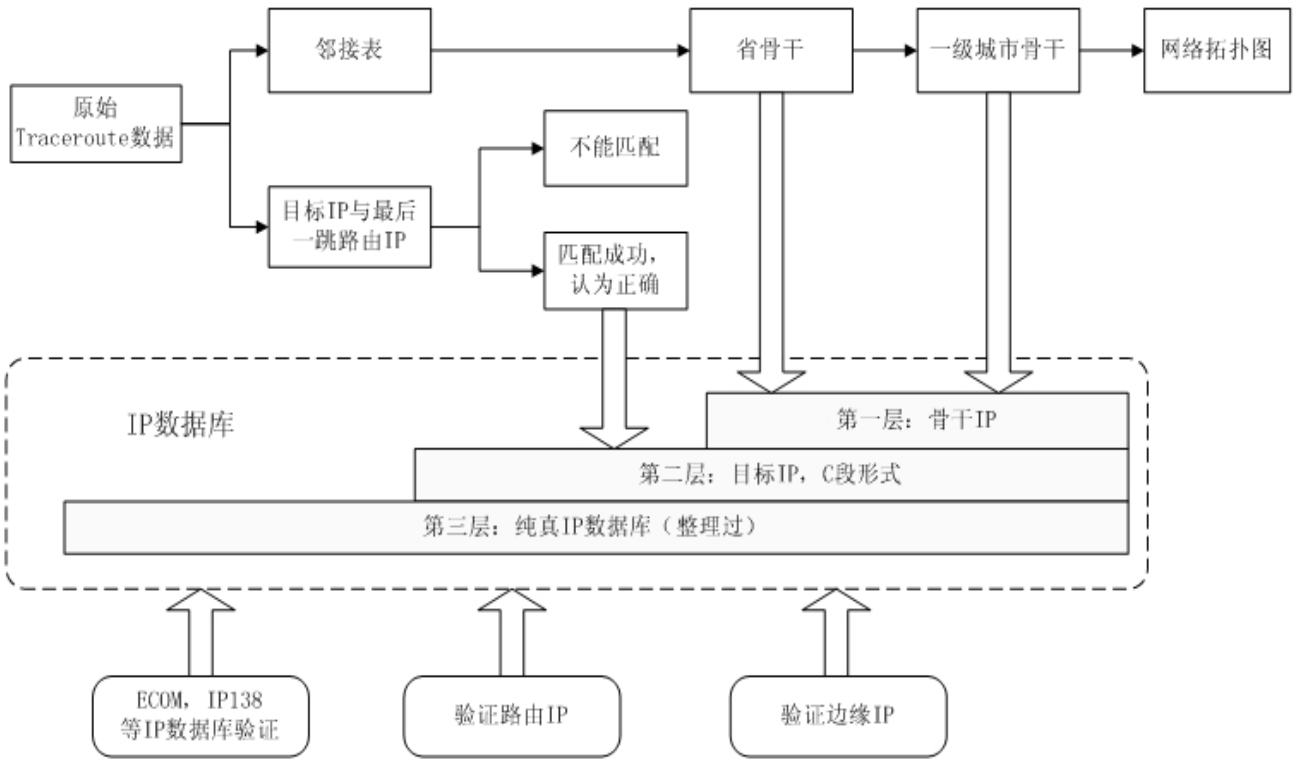
CASE2

教育网拓扑定位

Baidu 百度

CASE2

建立基础ip库

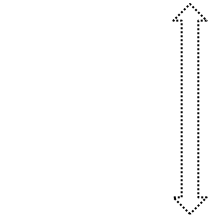


CASE2

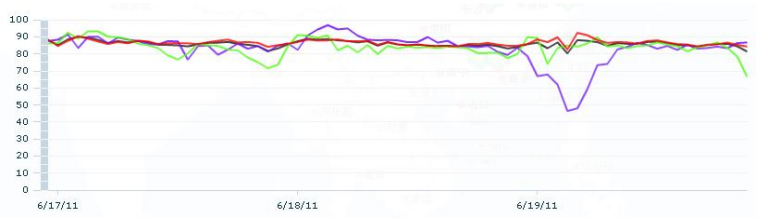
某地区用户

感知服务质量，调整分配策略

Baidu GSLB



Colombo+Galileo

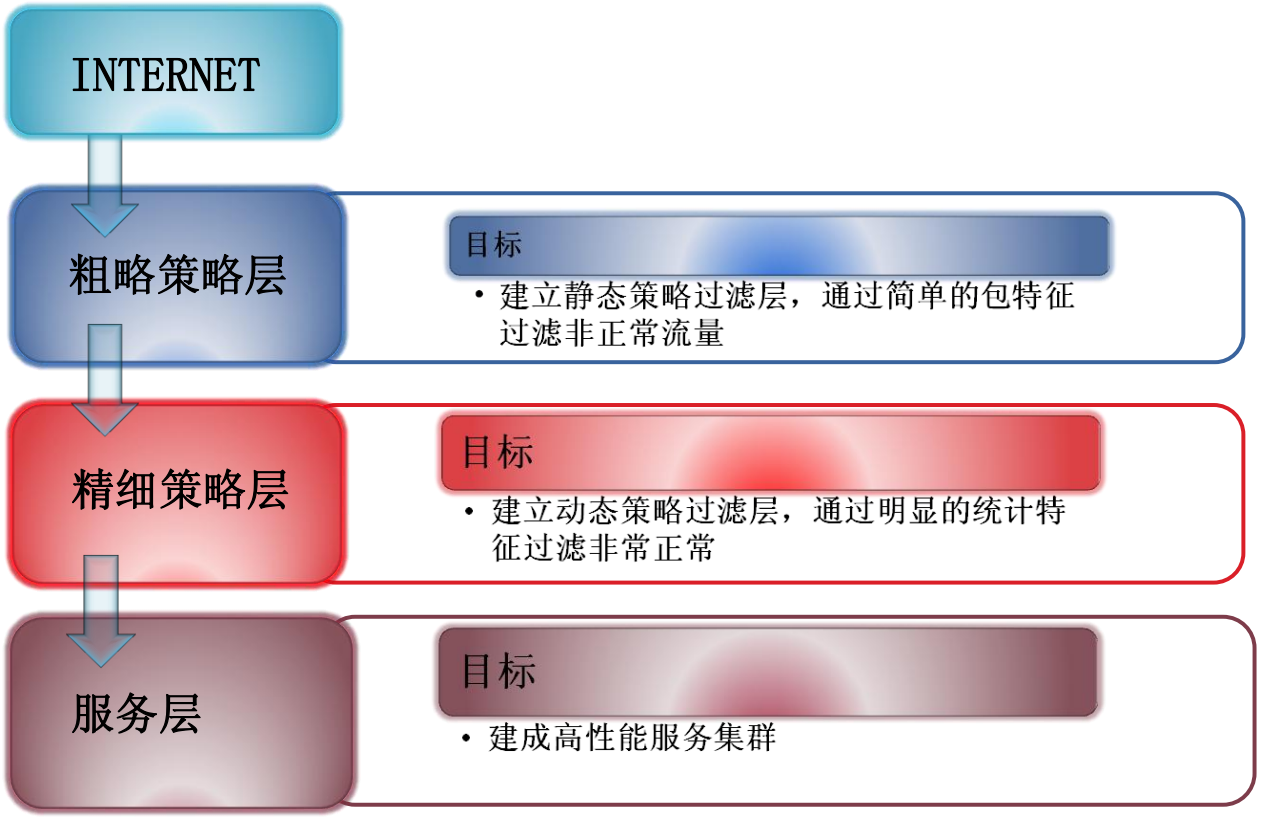


Baidu 百度

CASE3

DDoS防御系统

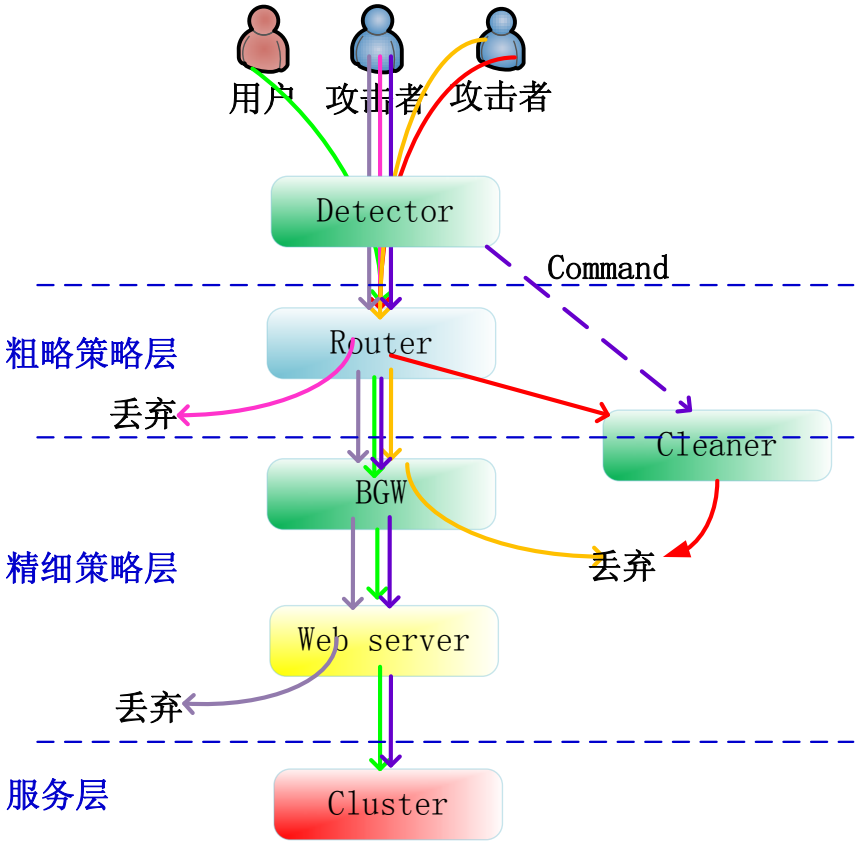
体系结构

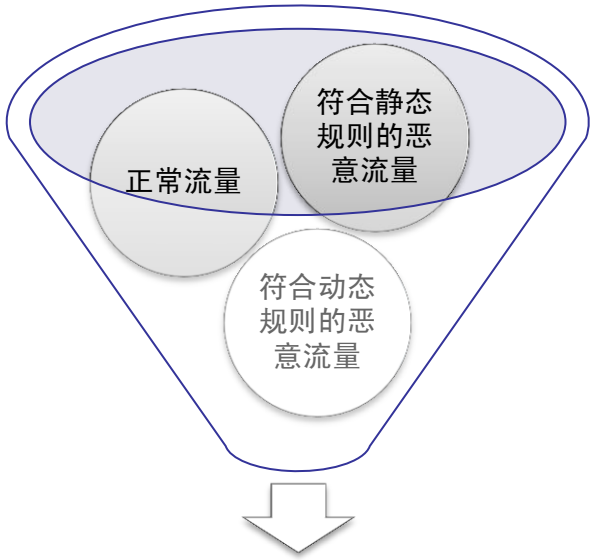


Baidu 百度

CASE3

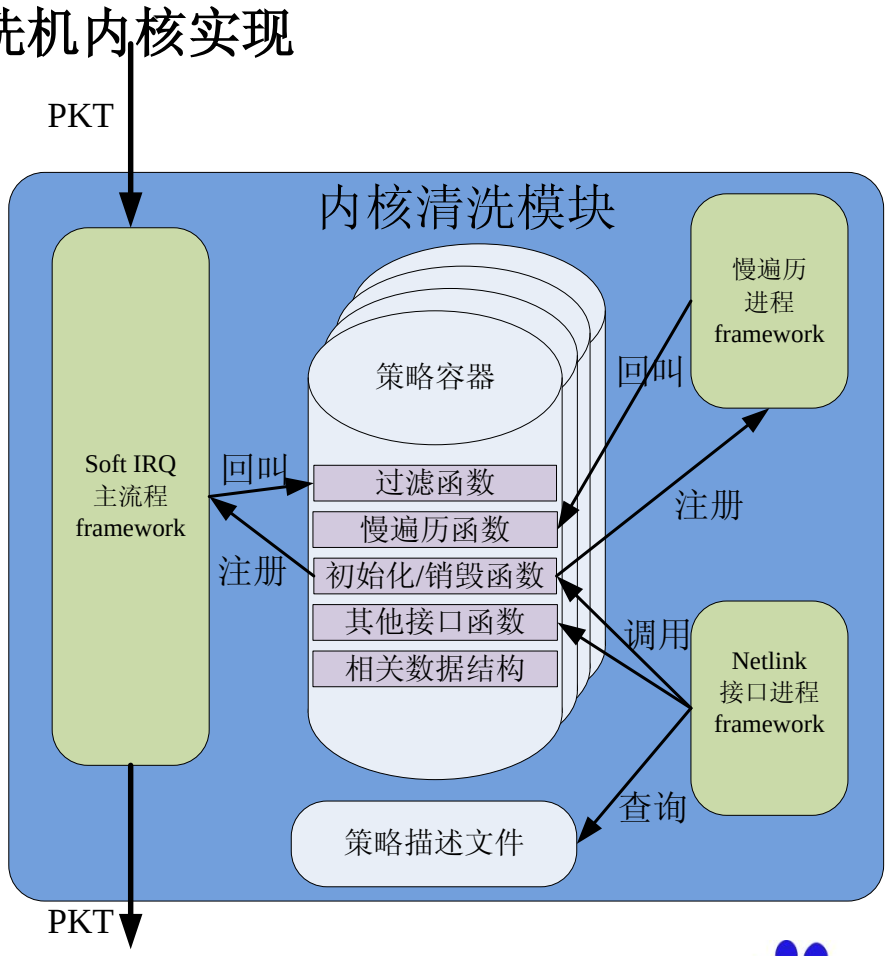
攻击防御系统实现





CASE3

- Netlink接口框架
- 主流程框架
- 慢遍历框架
- 策略容器
- 策略描述
- 过滤算法：令牌算法、
- 优化的AC
- 对内核路由进行优化：per-cpu路由cache表



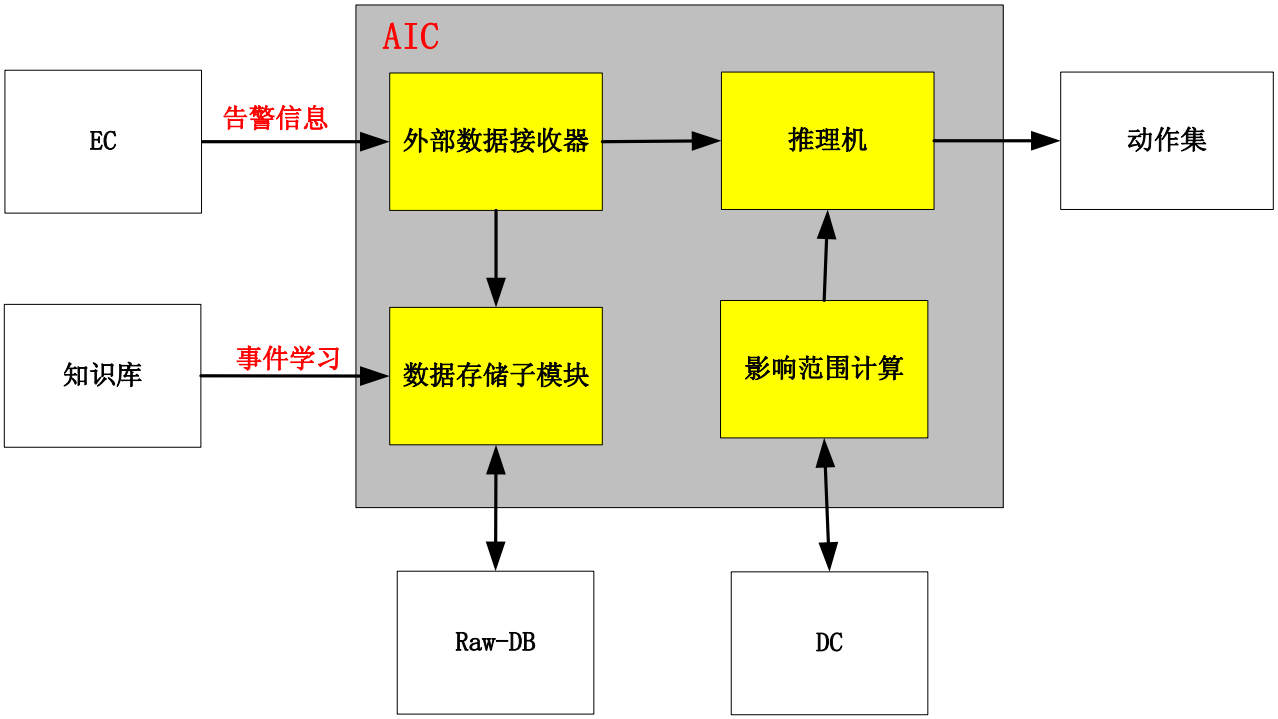
CASE3

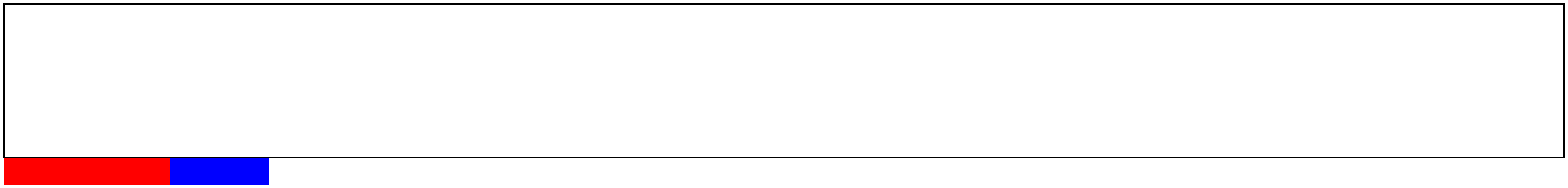
性能评估

Nehalem流量清洗性能测试结果		
测试case	测试结果	备注
数据包丢弃能力	1100wpps	目前BCS DNS防攻击的性能指标
转发能力（单方向）	380wpps	
七层匹配转发	250wpps	全转发
七层匹配并校验baiduID转发	200wpps	全转发

CASE4

专家系统实现





Thanks

Baidu 百度