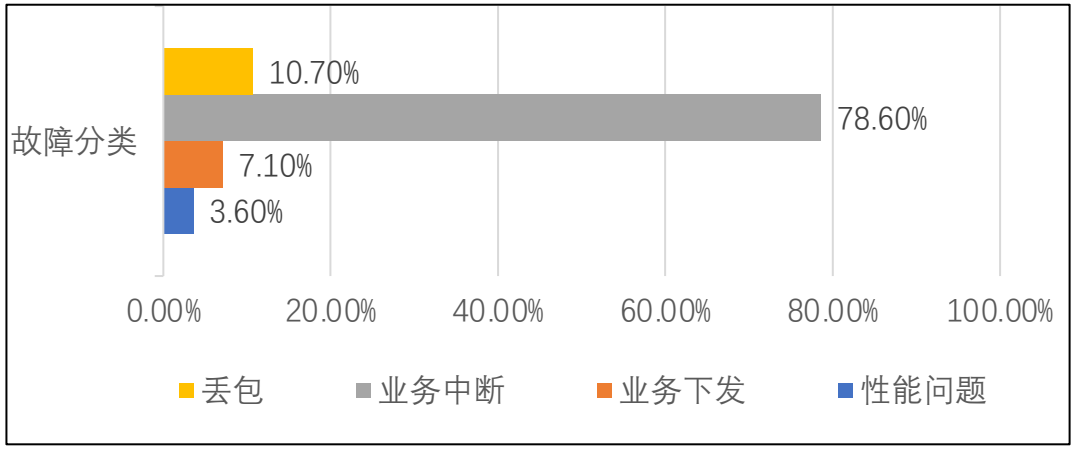
非侵入可观测监控技术eBPF在华为云Stack网络故障诊断中的应用

殷森道 华为云高级软件工程师

开放原子开源基金会 OPENATOM FOUNDATION | OpenEuler

目录

云网络运维难题分析

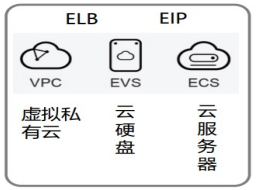


- **大约80%**的问题是**业务中断**类问题

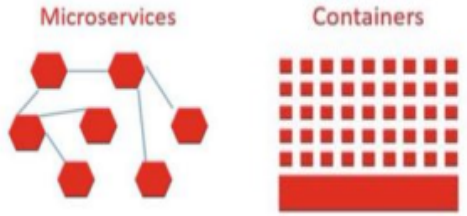
问题总结：

- 部分网络服务的运维能力较弱。
- **云服务+VPC网络**深度调用，流量复杂各异，需要尽快构建**全业务场景中断**的定界定位能力，逐步构建**性能和丢包**问题定界定位能力。
- 云网络路径黑盒，运维难度较大。

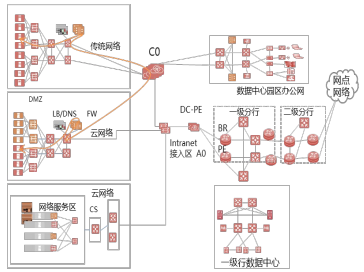
变化1：网络资源云化，从物理设备延伸到服务器内，看不见，摸不着



变化2：微服务化，系统更加复杂，动态性更强，上下游依赖更多



变化3：应用多云多池跨数据中心部署，分布式范围更加分散



痛点-1：应用问题和网络问题定界定位难

应用问题？还是网络问题？无法快速定界定位

痛点-2：虚拟网络和物理网络定界定位难

虚拟网络问题？还是物理网络问题，无法快速定界定位

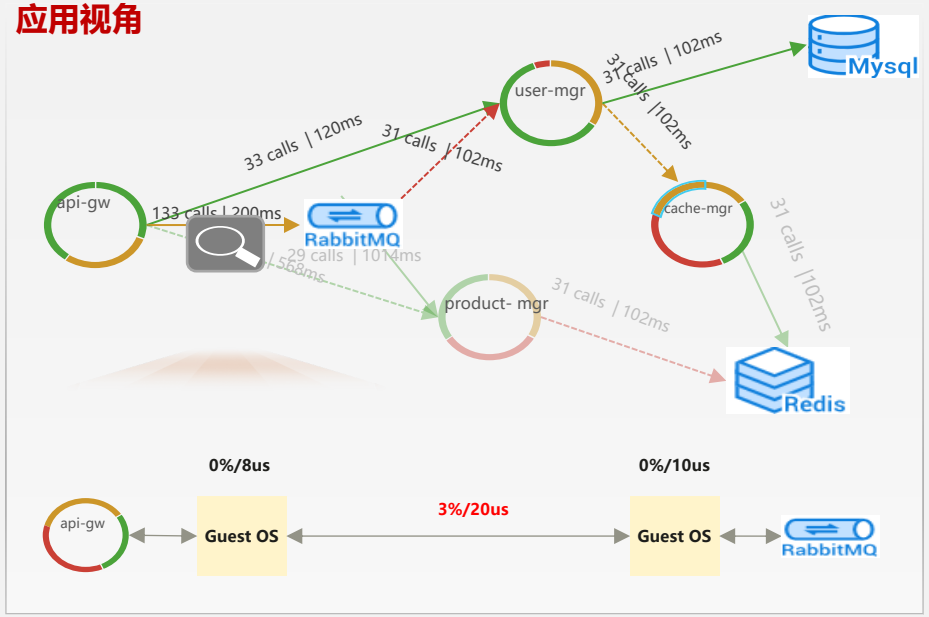
痛点-3：业务调用链路长，关系复杂，定位难

微服务、中间件的拓扑黑盒，缺少自动异常检测机制

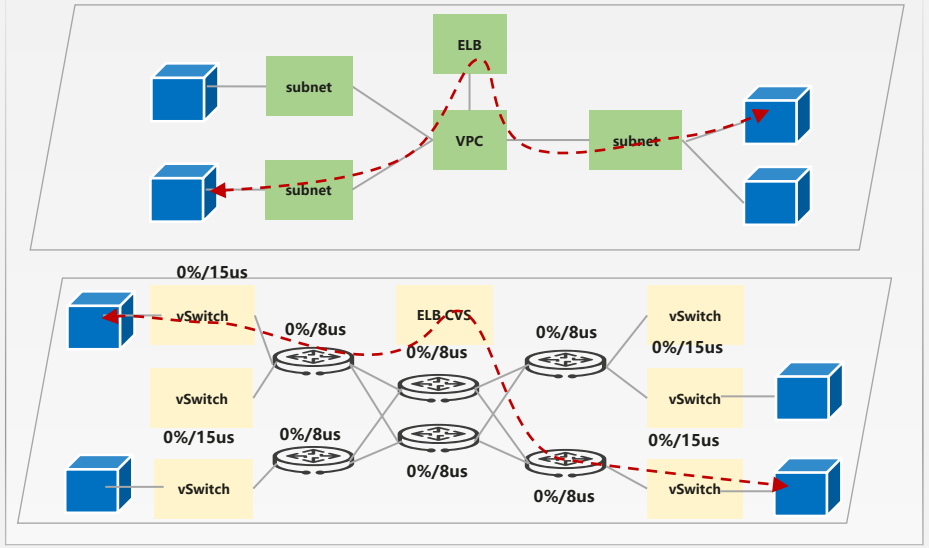
CloudNetDebug：网络运维一体化

全栈一体化监控

应用视角



网络视角



交互拓扑可视 | 网络传输性能 | 主机服务性能

ACK时延 | TCP重传 | TCP握手时间 | 零窗口 | 连接数 | 连接重置

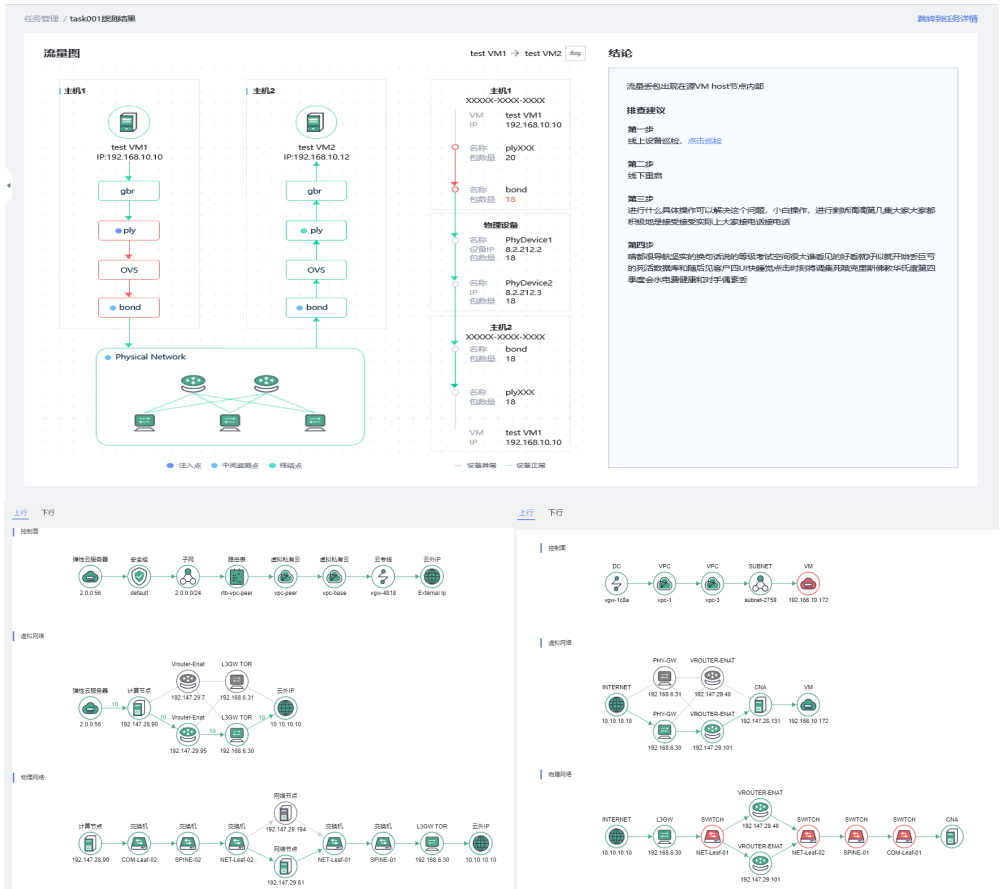
流量路径可视 | 路况质量可视 | 节点负载分析

应用随流检测 | 时延检测 | 丢包检测 | 出入端口可视 | 吞吐量分布 | 指标联动分析

故障诊断

- 流拨测
- 流抓包
- 主动链路监控
- 全链路故障诊断
- 审计/核查
- 网络随流检测

网络视角：全链路故障诊断，根因分钟级定位



业务痛点

- 云服务和VPC网络深度调用，流量复杂且路径是黑盒，运维难度大
- 云上云下互联，涉及软硬联动，问题定界耗时长，分析难度高
- 网络问题中断严重影响业务，对时效性要求高

关键技术

全链路故障诊断服务，网络故障定位分钟级定位

虚拟网络到物理网络可视化呈现

以前：
手工登录页面或者查询API，推断虚拟网络路径，才可获取到物理路径。

现在：
输入五元组，自动生成控制面，虚拟网络，物理网络的三层路径。

控制面配置(路由表，网络ACL，安全组)自动检查

以前：
管理员手工组配置是否正确需要逐跳检查路由表，网络ACL，安全。

现在：
一键式，自动检查来回路径配置，自动提示可疑配置。

硬件网关数据面黑盒检查

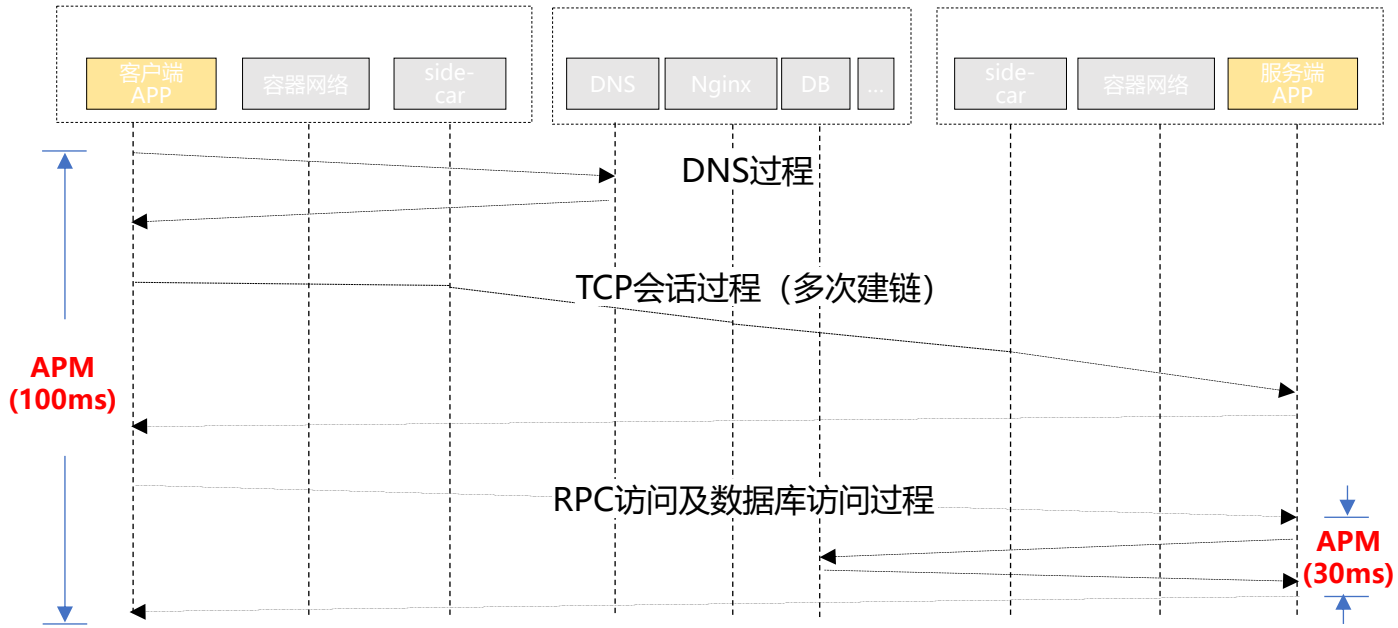
以前：
跨部门协调数通工程师登录交换机进行定位，流程长，耗时多且效果差。

现在：
一键式定位，自动进行流量拨测，专线网关到VM的流量路径可视，故障点直接呈现。

CloudNetDebug是一款专业的网络运维工具，其提供的拨测和全链路故障诊断功能，针对管理员指定的业务流，能够分析业务流的路径，按照控制面，虚拟网络和物理网络三层分别展示路径和故障分析

应用视角：云原生时代全链路故障诊断挑战及解决思路

云原生基础设施之上应用访问全链路



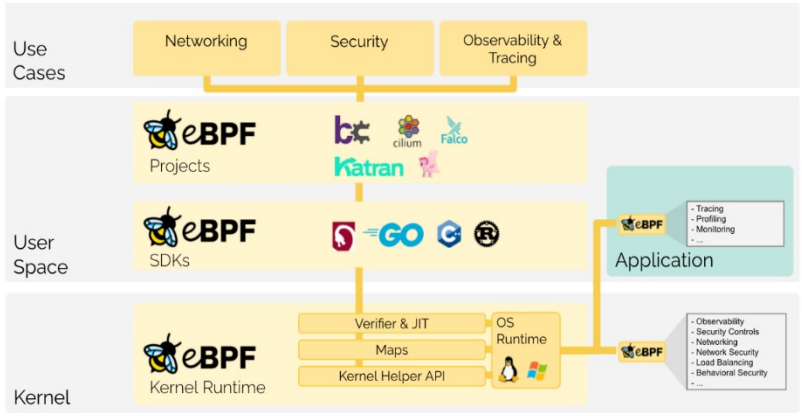
CloudNetDebug运维能力不足：

- **事后运维**：只能在问题发生之后手动触发拨测，非实时流量观测；
- **性能受限**：频繁拨测造成的资源耗费和性能损耗较大，不适合实时观测；

传统全链路故障诊断（APM）方法在云原生场景缺陷：

- **维护困难**：APM SDK侵入式修改在云原生多语言、多版本业务背景下会加剧维护成本，加剧维护责任边界的模糊；
- **盲点多**：Infra node、K8S Node存在大量观测盲区（占比>70%），客户端/服务端APP观测差异大，无法完成定界；

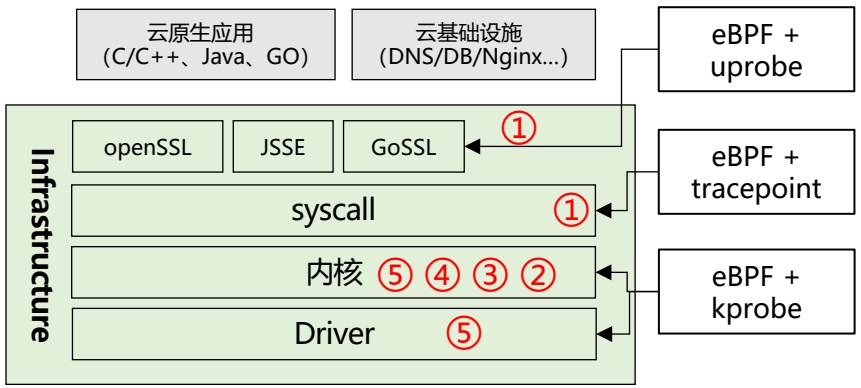
*eBPF 是一个能够在内核运行沙箱程序的技术，提供安全方式的注入代码的机制，通过注入eBPF代码可以**安全、高性能的访问整个系统的运行状态**。*



eBPF技术特征：

- **无侵入**：应用/容器镜像零修改；职责边界清晰；安全&高性能
- **全栈**：通过eBPF + u/kprobe、eBPF + Tracepoint等技术，可以覆盖内核、运行时、基础库等大部分基础软件，轻松应对云原生多语言、多网络协议、厚重软件栈的场景特征。

eBPF覆盖基础设施的观测能力



① 应用性能RED
② 应用访问拓扑
③ 进程网络
④ 进程资源
⑤ 系统资源

网络故障诊断最后一公里 – openEuler社区项目：gala-gopher

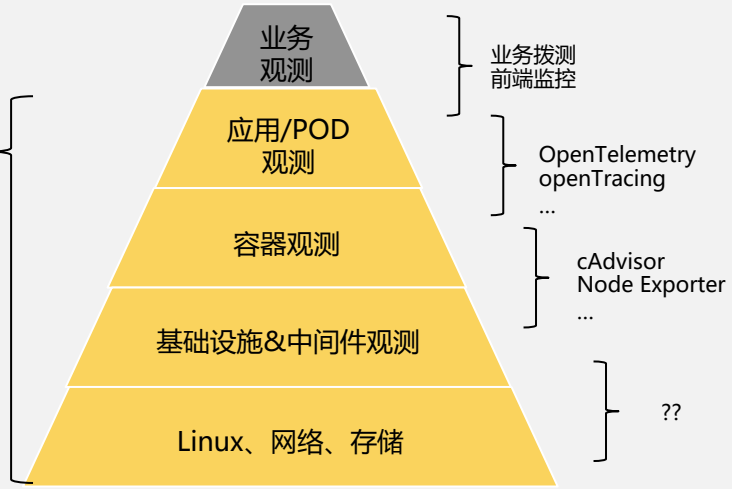
全栈/链路可观测全景图

gala-gopher非侵入观测

业界主流技术

技术特征：
- 应用drill-down观测，避免数据孤岛
- 应用性能可观测（覆盖主流协议*）
- 非侵入式，应用零修改
- 高性能，eBPF JIT执行效率更高
- 插件化架构

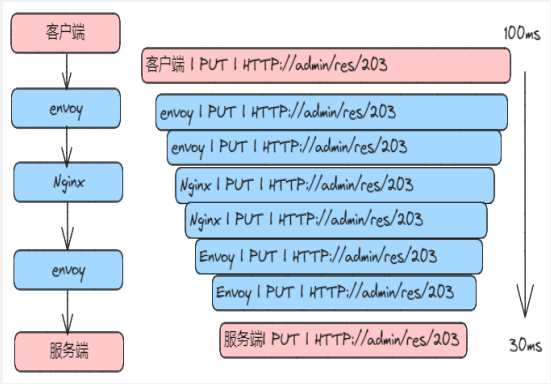
* 网络协议：HTTP 1.X, PG, MYSQL*, Redis*, KAFKA*, HTTP2.0*, DUBBO*



业务效果展示及技术指标

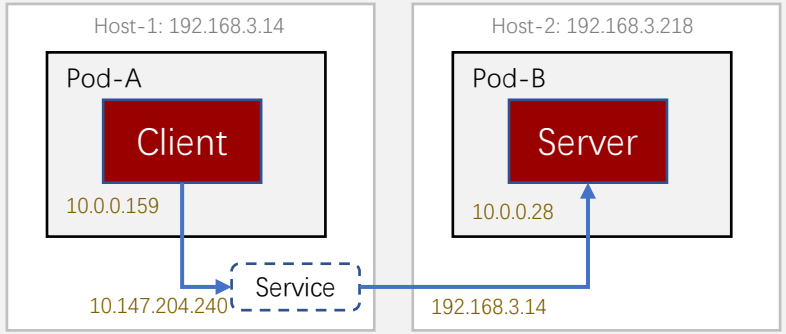
- 业务效果：基础IT系统亚健康诊断时长达到分钟级。
- 软件开销：观测底噪<0.5%，应用性能干扰<1%。

全链路时延观测



- 会话级全链路时延观测，覆盖应用、系统、容器网络、基础中间件（Nginx、DB、DNS等）；

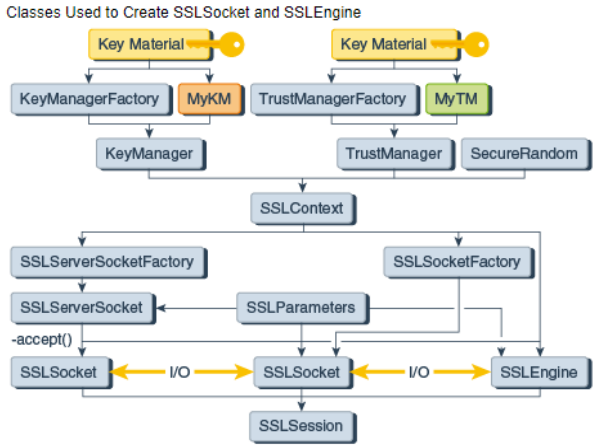
- 基于gala-gopher开放的插件化机制，实现了在云原生流量代理场景下的流量路径还原



gala-gopher架构下eBPF短板及补充方案

Java应用监控场景短板——uprobe能力受限

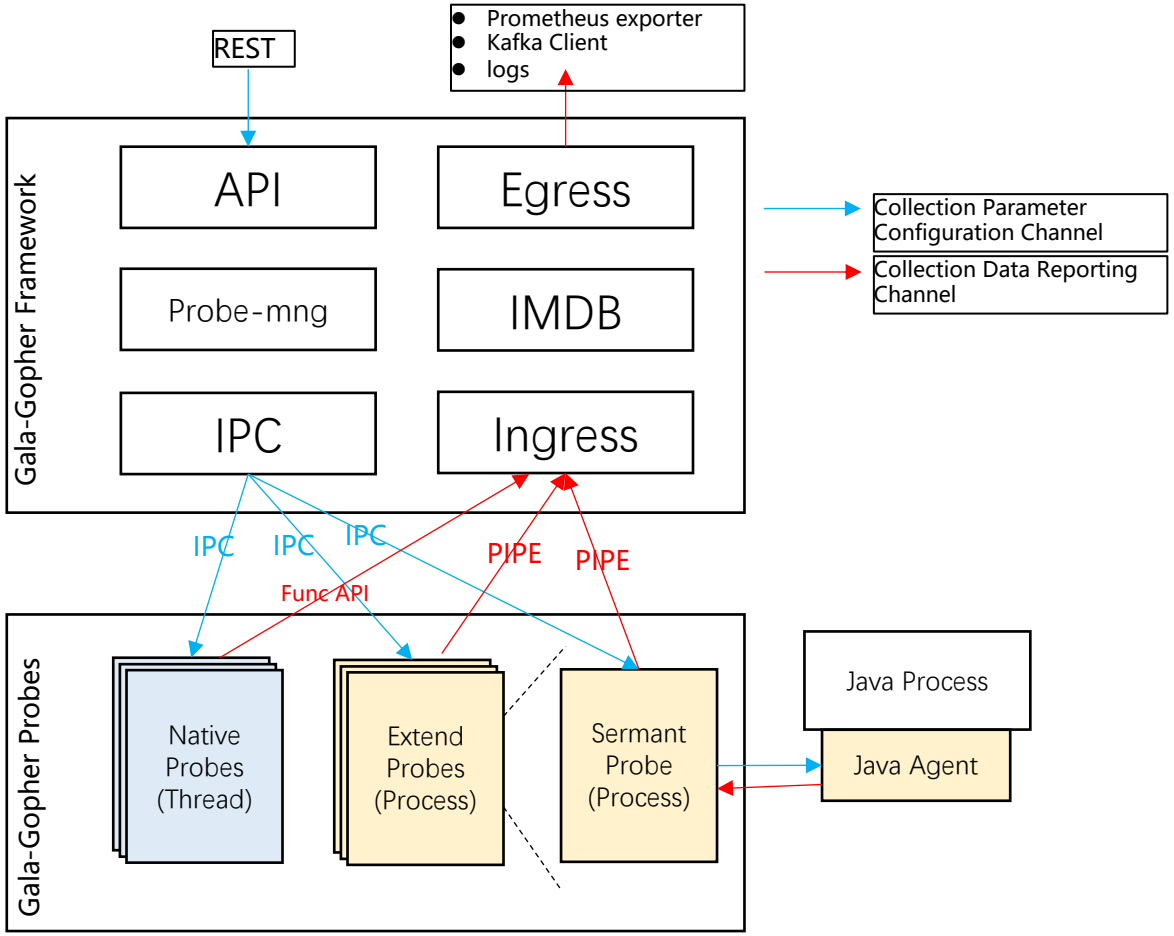
- **场景覆盖难度大**
- **Java场景使用eBPF+uprobe成本较高**
- **不稳定**




```
case NEED_WRAP: {
    socketWriteBuffer.clear();
    SSLEngineResult r = sslEngine.wrap(DUMMY, socketWriteBuffer);
    checkResult(r, true);
    socketWriteBuffer.flip();
    Future<Integer> fWrite = socketChannel.write(socketWriteBuffer);
    fWrite.get();
    break;
}
```

- **案例一**：JSSE提供两种SSL加解密库函数，其中SSLEngine仅提供加解密“工具”，在JSSE中并不维护Socket本身信息
- **案例二**：Java应用场景下gRPC、Dubbo3.0等多种L7应用层指标采集受限
- …

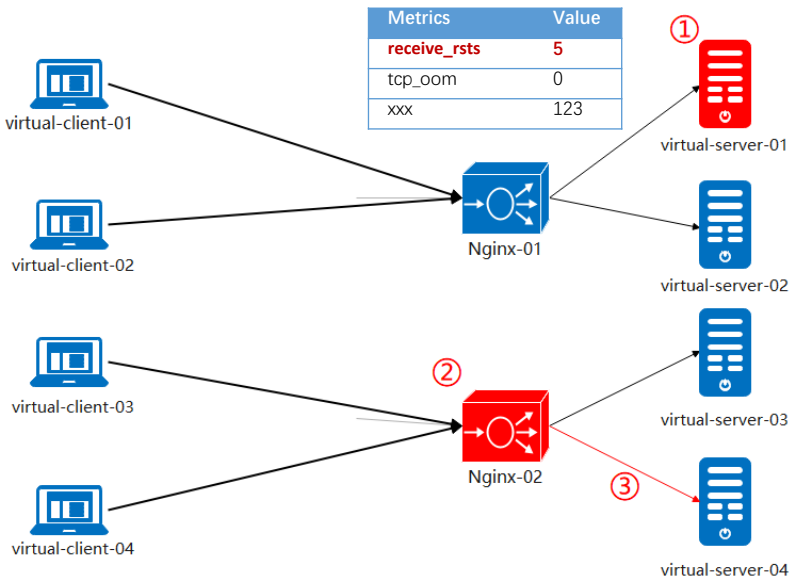
引入Sermant解决Java场景下的能力不足



在Gala-Gopher框架下扩展一个Sermant探针，引入java agent的能力补充Gala-Gopher在java应用监控场景下的短板

eBPF技术在真实场景中解决的问题举例

ELB-Nginx拓扑

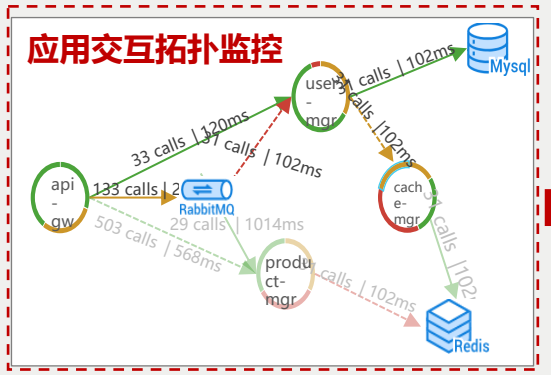


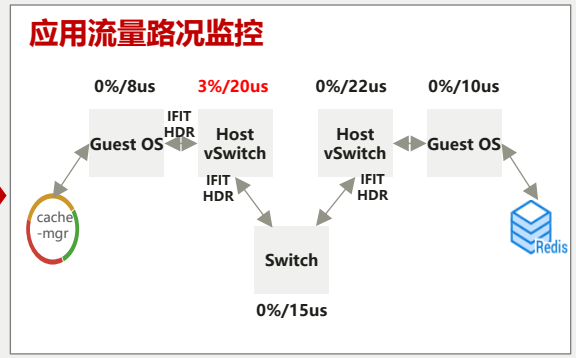
序号	问题	原因	eBPF策略
①	客户通过7层elb压测，三万条有几十条报错	【后端业务问题】后端超时配置错误导致回复reset报文。	**取代抓包，偶现故障记录：** eBPF可以采集到socket数据中的reset报文，拓扑上指标可直接体现后端业务异常，同时可生成系统告警。
②	apic 服务异常，客户反馈影响某实时交易的业务	【ELB数据面问题】Nginx进程单核卡死	**关键指标波动回溯查询：** 1. 采集进程CPU占用率可知nginx进程异常； 2. **Nginx和后端服务的数据量减小，时延增大。**
③	客户某业务经过elb达不到性能要求	【ELB性能问题】后端服务器抓包判断elb负载合理，最终原因是服务经过云外带宽受限	**流量分布快速厘清：** 悬脉拓扑可以直接体现Nginx和后端服务器的连接情况和数据量，判断负载均衡是否合理。

监控策略：使用eBPF监控ELB数据面高频故障组件ELB-Nginx，**采集四层网络通信状态指标数据，并在指标异常时进行特殊标记、告警。**

增强ELB现网监控和问题定位定界的能力，**补齐NGINX网元没有4层相关连接指标监控的缺陷**，同时**补充健康检查离线场景定位定界的能力。**

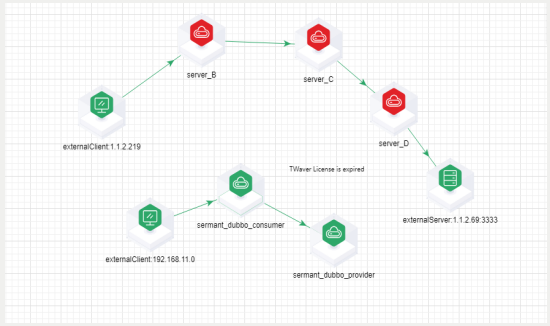
未来演进：应用视角 实时主动监控+网络视角 拨测智能触发

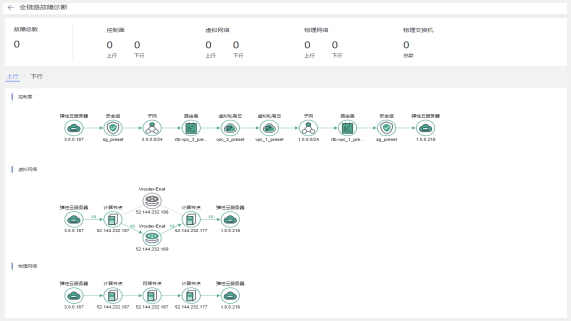




事中：eBPF实时观测网络性能指标

事后：全链路故障诊断流量网络路径是否通畅





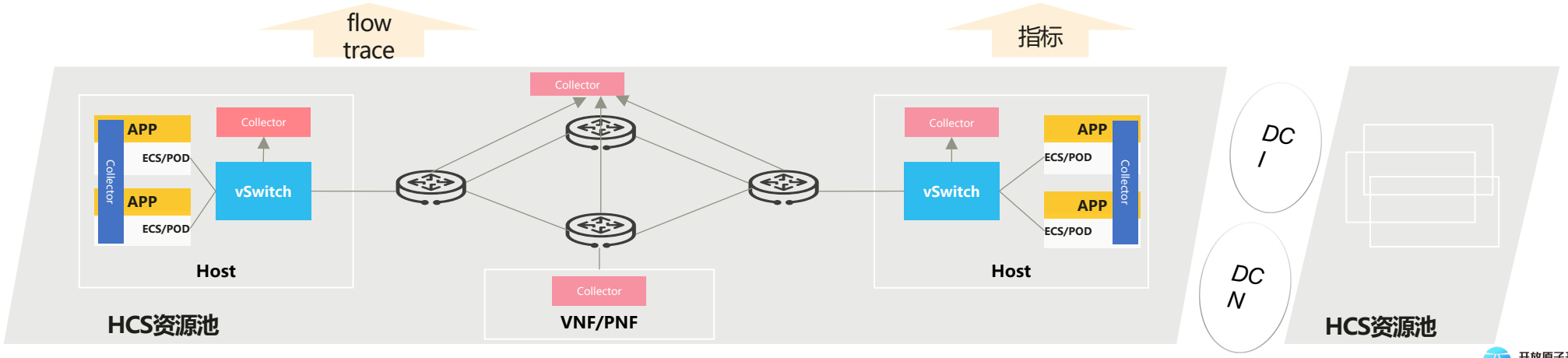
流量路径可视

交互拓扑可视

网络传输性能

路况质量可视

- 基于eBPF实时采集流量指标，实时发现异常流量
- 异常流量自动拨测，及时诊断网络故障
- 实时观测与事后拨测相结合，快速厘清应用/网络问题
- 异常流量实时告警，快速通知运维人员定位排障



THANKS

扫码加入gala-gopher社区微信群

Sermant.io 官网

扫码加入Sermant微信群

开放原子开源基金会 OPENATOM FOUNDATION | OpenEuler

THANKS

扫码加入gala-gopher社区微信群

Sermant.io 官网

扫码加入Sermant微信群

开放原子开源基金会 OPENATOM FOUNDATION | OpenEuler

THANKS

扫码加入gala-gopher社区微信群

Sermant.io 官网

扫码加入Sermant微信群

开放原子开源基金会 OPENATOM FOUNDATION | OpenEuler

THANKS

扫码加入gala-gopher社区微信群

Sermant.io 官网

扫码加入Sermant微信群

开放原子开源基金会 OPENATOM FOUNDATION | OpenEuler