Netis天旦

www.netis.com

运维排障案例解析

案例 1

- **事件：**

某保险一次大型营销活动中，出现中间件宕机。管理人员尝试重启进行恢复，但中间件重启后再次宕机。
运维人员排查发现，中间件本身一切正常。经过几个轮回的排查，发现最终原因竟然是：

- **排障解析：**

运维人员查看日志后，发现web服务器接受了很多请求、互联网带宽也被占得很满。第一反应是怀疑发生了攻击，于是排查了防火墙、IPS等安全设备，发现没有问题；没有头绪之下，只好又重新把日志全部看了一遍，发现在所有访问中，提交环节的操作占比远超出正常占比水平；抽查部分用户访问记录后，发现存在一份保单提交多次的现象。
最终找到问题原因：**网络慢**。

由于网络较慢，保险经纪人提交订单后不会立即显示成功，此时用户往往会重复点击提交，导致一份保单重复提交多次，由于活动期间用户量巨大，当大量用户提交多次，服务器接收的访问量呈指数级增长，最终拖垮了服务器。

而活动期间之所以出现网络慢：
一方面是因为网络没有规划好，带宽相对不足；
另一方面，正值运营商业务高峰期，导致网络较平时要慢。

案例 2

- **事件：**

保险公司某系统平时一切运行正常，但与某大型支付平台合作进行营销活动时，却发现从合作平台端发出的部分用户请求无响应，数据中心反馈没有收到相关请求。你觉得是哪个环节出了问题呢？

- **排障解析：**

运维人员发现，营销活动开始后，当交易量上升到一定水平后，建立成功率开始下降。
经排查发现问题环节在**安全设备**。

营销活动中，系统访问量短时间内急剧上升，远超出平时访问量水平。这导致所有的安全设备认为这是攻击行为，从而对访问进行了阻拦，导致请求全部失败。

案例 3

- **事件：**

某公司的支付系统过去一直都很正常，某日大约有半天时间，交易成功率明显降低，经对问题交易类型进行分类，逐步排查后，发现是因为某地区过来的所有支付请求都失败了。这是什么原因导致的呢？

- **排障解析：**

运维人员通过BPC进一步分析发现，该地区过来的交易，都被认为是非法请求而被拒绝。再进一步通过报文分析发现，该地区发出的交易请求不知为何触发了运营商HTTP广告注入机制。

由于广告注入，导致整个请求的完整性被破坏。所以，从该地区过来的所有请求都被认为是非法请求而被拒绝了。

案例 4

- **事件：**

某日上午，某保险公司核心服务器出现宕机，经排查发现是网厅触发核心系统高并访问所致。应用部通过日志和某APM工具看到，从应用服务器发出一笔请求，经F5，到核心服务器端接受的请求变成了2-5笔。
这些多出来的请求是从哪里发出的？

- **排障解析：**

运维人员排查发现该2-5笔请求都为无响应请求，且每一笔间隔时间都是固定300秒。经查证，F5中TCP超时限制时间为300秒。同时，问题交易在核心服务器实际所需执行时间为12分钟。故障形成过程如下：

1、应用服务器发出交易请求，由于交易所需执行时间为12分钟，大于F5超时时间设置300秒，交易被中断
2、**应用服务器底层的JAVA HTTP CLIENT小程序**重新发出交易请求（JAVA HTTP CLIENT小程序负责将APP封装好的请求通过网络发出去，该程序的默认配置是只要发出的请求被异常中断就会retry重试）。
3、交易超时，被异常中断
4、JAVA HTTP CLIENT再次retry
……

- 案例番外：

当时该保险功公司的应用部门配备了日志分析和某APM工具，但由于日志分析和APM关注的是应用本身，无法了解应用底层发生了什么。只能看到应用端发送了1笔请求A，核心收到了2-5笔请求A。由此判断问题出在F5，即网络部。

当时恰巧天旦BPC在该公司做POC测试，网络部坚信F5没有问题，却无法证明，便向天旦BPC技术人员求助。基于网络镜像技术的BPC关注的是应用与其他组件、应用与应用之间的数据交互，所以能够看到从应用服务器端发送的请求确实是2-5笔，而非日志和某APM看到的1笔。进而根据请求无响应及间隔时间逐步排查，找到了问题原因。


Netis天旦
中国创造，走向全球