业务流量数据安全应用实践

蚂蚁金服：刘宇江（羽将）

目录

1. 流量镜像介绍
2. 流量数据的采集
3. 流量数据的处理
4. 应用场景

01 流量镜像介绍

流量镜像介绍

数据对安全来说，是最为重要[illegible]查等，都是以各种数据记录为基础进行，数据的完整性[illegible]的高低。

要保护的系统和资产

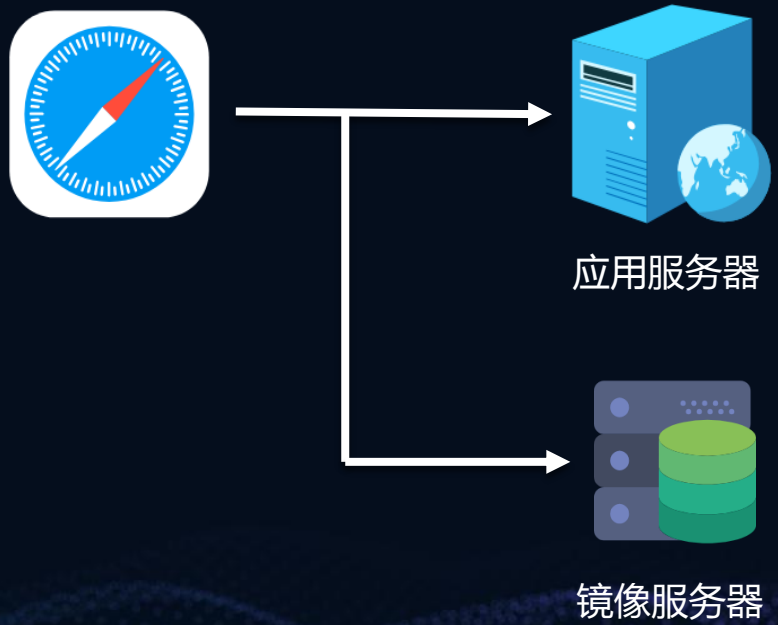
日志数据

写字楼

监控摄像

流量镜像介绍

流量镜像是将网络请求包完整的拷贝一份，
在不阻塞请求的情况下，发送到镜像服务器，
将请求和响应进行存储和解析。



流量镜像介绍

由于是全包采集，因此流量数据信息完整度趋近于100%，比常用的log形式数据，数据量要丰富很多。以HTTP流量镜像为例，对比常用的服务器access log，其信息完整度对比情况如下

数据类型	请求头	请求体	响应	信息完整度
access log	部分内容	可选	无	20% - 40%
HTTP流量镜像	完整	完整	完整	≈ 100%

流量镜像应用介绍

资产梳理

漏洞覆盖

攻击检测

联防联动

业务治理

02 流量数据的采集

通过端口镜像进行采集

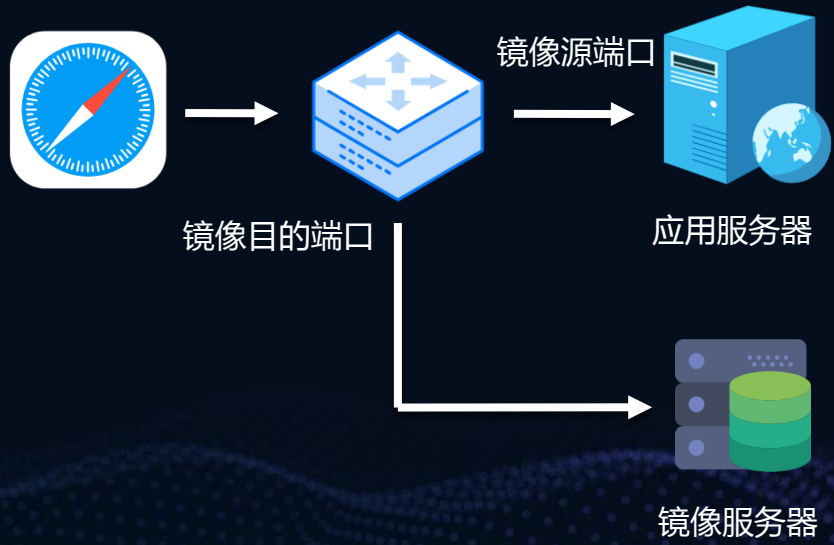
通过交换机或路由器上的端口镜像功能，将一个或多个源端口的数据流量转发到某一个指定端口来实现对网络的监听，如果需要获取响应数据，需要做双向镜像。

优势：

- 旁路镜像
- 网络结构无改变
- 性能影响小
- 故障影响小

劣势：

- 不一定所有设备都有镜像功能
- 占用交换机端口，需要修改交换机配置
- 加密流量无法直接解密



通过分光器进行采集

在宽带接入处，由于核心路由器不支持端口镜像进行流量采集，分光器是一种无源光器件，通过光复制来进行用户访问请求数据的镜像。

优势：

- 旁路镜像
- 网络结构影响小
- 性能影响小
- 故障影响小
- 抓取所有流量

劣势：

- 对原有设备光纤改造时会造成短时间网络中断
- 加密流量无法直接解密

通过软件进行采集

在应用服务器上通过软件直接采集，并将数据发送至服务端进行存储和分析。

优势：

- 纯软件，成本低
- 对网络结构无变更
- HTTPS流量可直接解密
- 可抓取内部网络流量

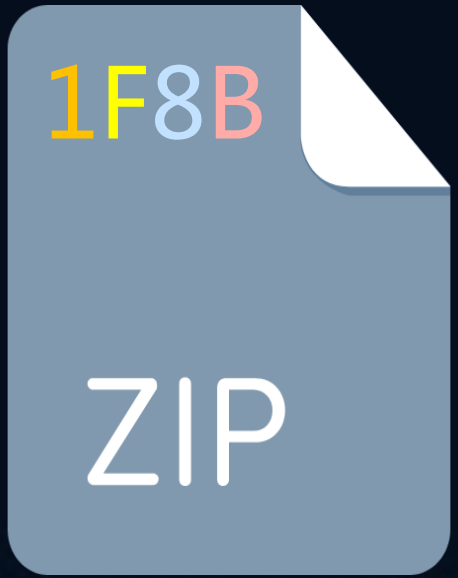
劣势：

- 需要对所有应用服务器安装，推广相对运营麻烦
- 占用应用服务器资源，容易受到业务挑战
- 加密流量无法直接解密

03 流量数据的处理

对压缩数据的处理

数据中如果存在压缩数据，如HTTP中请求或者响应体可能被GZIP压缩等，需要先对数据进行解压。解压过程中，需要准确定位压缩数据块，如HTTP响应体重GZIP压缩数据块由标记 1F8B 开始到数据结尾。由于存储大小是有限的，部分数据可能存在被截断的情况，因此解压时需要考虑数据被截断问题



数据脱敏

需要对请求数据中的敏感信息进行脱敏，并且要保证数据在进入分析端之前已经被脱敏，防止发生数据风险

特征预提取

在对数据做清洗时，可以事先将一些未来可能会需要的特征预先提取出来并打标，为后续的计算或应急需求提高效率，常见的可以用于提取的特征：

- 请求者ID
- 是否可被公网访问
- 是否包含序列化数据、XML、JSON等数据结构体
- 是否含有爬虫、扫描器特征
- 是否是一次有效请求
- 返回的数据量大小
- 参数中是否包含有IP或者URL地址
- 是否具有登陆点

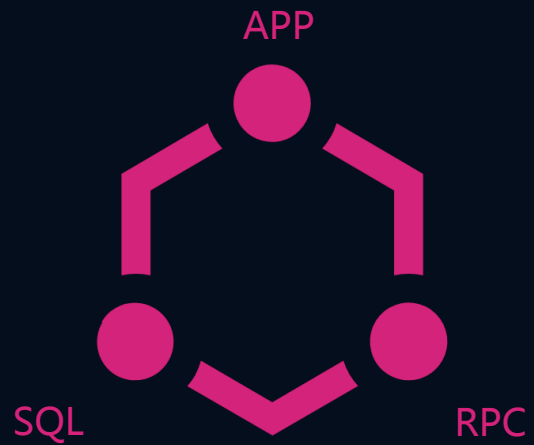
数据质量保障

- 数据缓冲区
- 数据热备份
- 采集端心跳监测
- 主动发送数据探针

数据冷备份

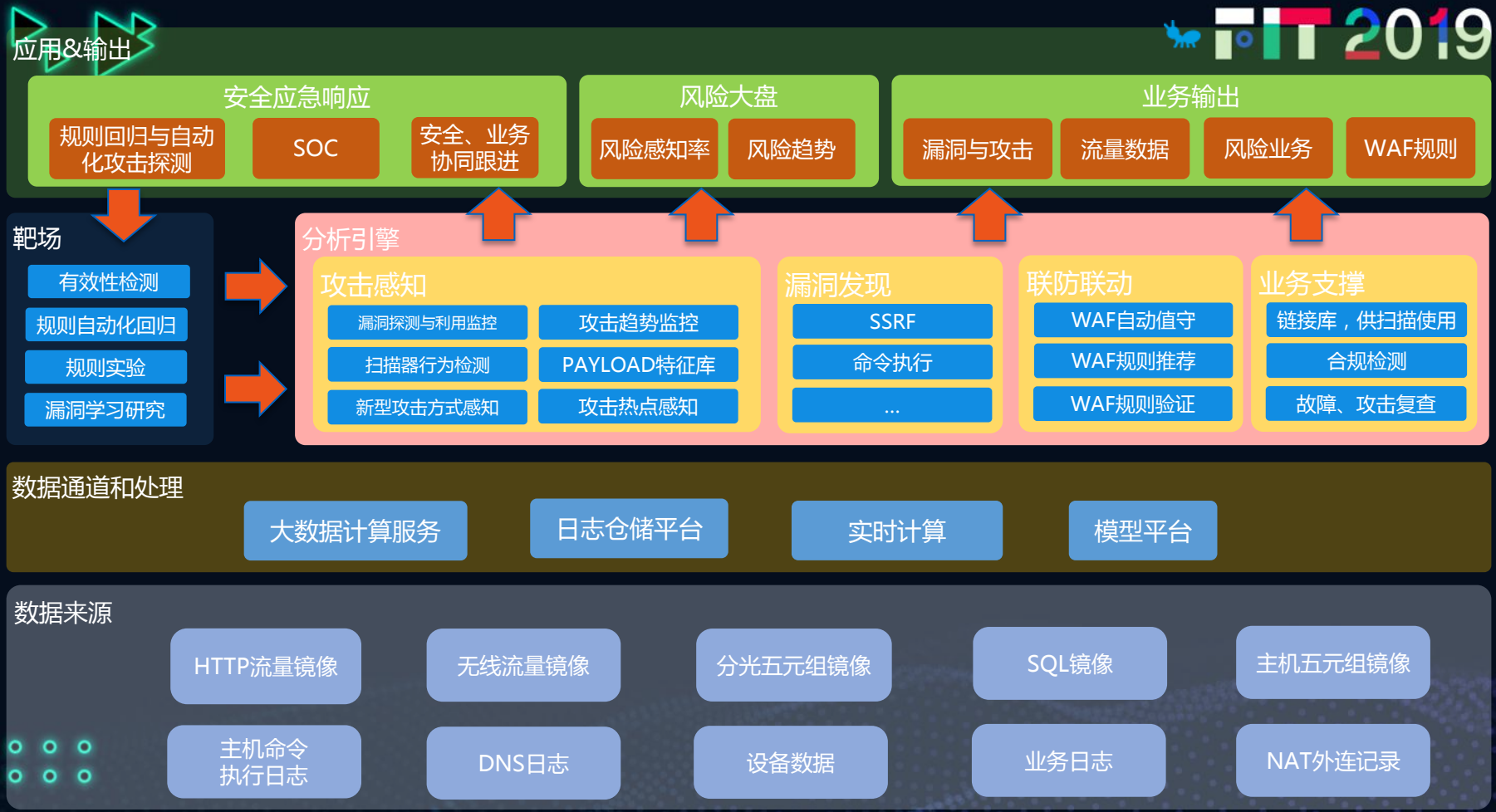
镜像数据的存储会消耗大量的存储空间，大部分数据都属于重要的冷数据，使用频率极低但是又不能删除，且时间越久远，使用的可能性就越低，因此可以对一段时间的历史数据进行高压缩存储，当需要使用时，系统先将数据解压，还原成热数据，再进行使用。

数据关联



通过在数据上加上关联标记，并让标记在各个环节中进行透传，可让不同的数据之间产生关联，从而扩展数据的维度，让之前通过单一数据维度进行安全分析得到的报警之间可进行准确关联，从而相互印证，让真正有价值的报警上浮。

04 应用场景


应用&输出
安全应急响应
规则回归与自动化攻击探测
SOC
安全、业务协同跟进
风险大盘
风险感知率
风险趋势
业务输出
漏洞与攻击
流量数据
风险业务
WAF规则
靶场
有效性检测
规则自动化回归
规则实验
漏洞学习研究
分析引擎
攻击感知
漏洞探测与利用监控
攻击趋势监控
扫描器行为检测
PAYLOAD特征库
新型攻击方式感知
攻击热点感知
漏洞发现
SSRF
命令执行
...
联防联动
WAF自动值守
WAF规则推荐
WAF规则验证
业务支撑
链接库，供扫描使用
合规检测
故障、攻击复查
数据通道和处理
大数据计算服务
日志仓储平台
实时计算
模型平台
数据来源
HTTP流量镜像
无线流量镜像
分光五元组镜像
SQL镜像
主机五元组镜像
主机命令执行日志
DNS日志
设备数据
业务日志
NAT外连记录

应用场景

资产梳理

- 链接库
- 参数模型
- 新业务发现
- 框架梳理
- 管理后台发现

应用场景

漏洞覆盖

漏洞影响面判断：

- FASTJSON漏洞应急
- 水平权限漏洞

漏洞主动发现：

- 代码上传
- 管理后台泄露

应用场景

攻击检测

- SQL注入
- 命令注入
- SSRF
- 任意文件读取
- XSS
- URL跳转
- 管理后台爆破

应用场景

联防联动

- WAF规则回归
- 拦截效果分析
- 一键止血
- 扫描器

应用场景

业务治理

- Cookie治理
- 业务异常排查
- 业务请求回归
- 合规检查

FREEBUF | FIT

THANKS