安全监测告警数据--降噪

原始数据：

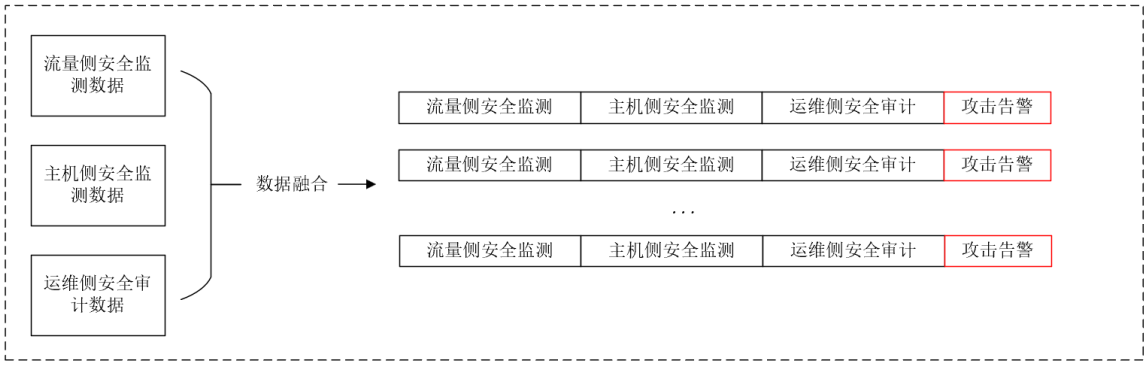
- 业务正常行为数据
- 业务网络流量数据
- 业务日志数据
- 流量侧安全监测数据
- 主机侧安全监测数据
- 运维侧安全审计数据

数据特征：

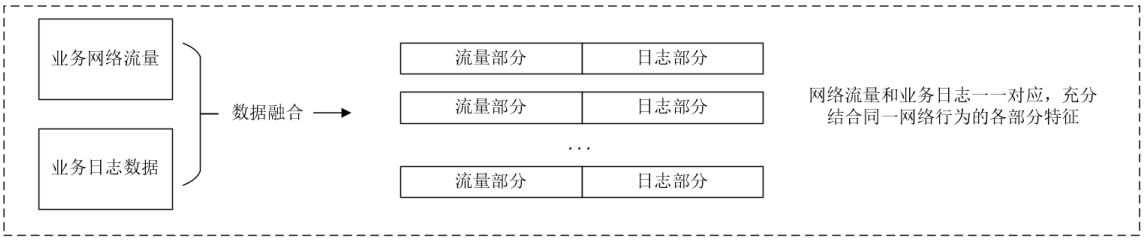
- 业务正常行为数据、业务网络流量数据、业务日志数据：包含大量正常数据，没有必要进行模型训练，通过预处理降低存量
- 流量侧安全监测数据、主机侧安全监测数据、运维侧安全审计数据：数据量递减，且包含的异常特征不完整

方案一

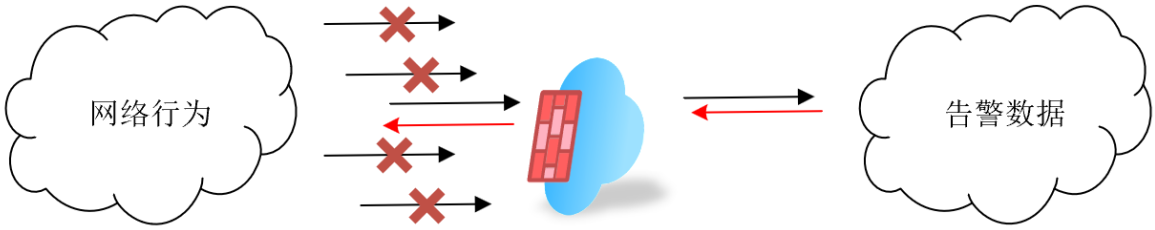
1. 对齐告警数据；



2. 对齐网络行为数据；



3. 依据告警数据向前追溯对应网络行为，做数据融合



要求:

- 网络流量和日志数据的对齐标准是否存在，例如日志和流量能由某个统一字段标识
- 流量侧、主机侧、运维侧监测数据是否具有一致性，即都来自同一原始网络流量

优点:

- 利用大模型弥补本身告警策略的缺陷，学习到通用的误报警的特征，达到降噪效果
- 筛除了大部分对降噪无效的正常流量，降低训练数据量

难点:

- 数据中是否包含误报或者确定攻击行为，可能需要人工标注
- 从监测数据到网络行为数据的追溯可能存在困难，难以一一映射
- 没有结合上下文信息

方案二

1. 对齐告警数据；
2. 对齐网络行为数据；
3. 依据告警数据向前追溯对应**上下文**网络行为，带有时间戳的网络行为

优点:

- 关注了上下文信息，分析网络攻击行为的全貌
- 增加上下文信息，提高降噪效果

难点:

- 如何确定上下文的边界，模型结合的上下文的信息量
- 增加了训练数据
- 数据预处理变复杂