Gdevops

全球敏捷运维峰会

海量运维日志异常挖掘

知道创宇高级安全研究员
邓金城

目录

1 概述

2 异常类型

3 异常挖掘方法

4 我们的工作

1 什么是海量运维日志异常挖掘？

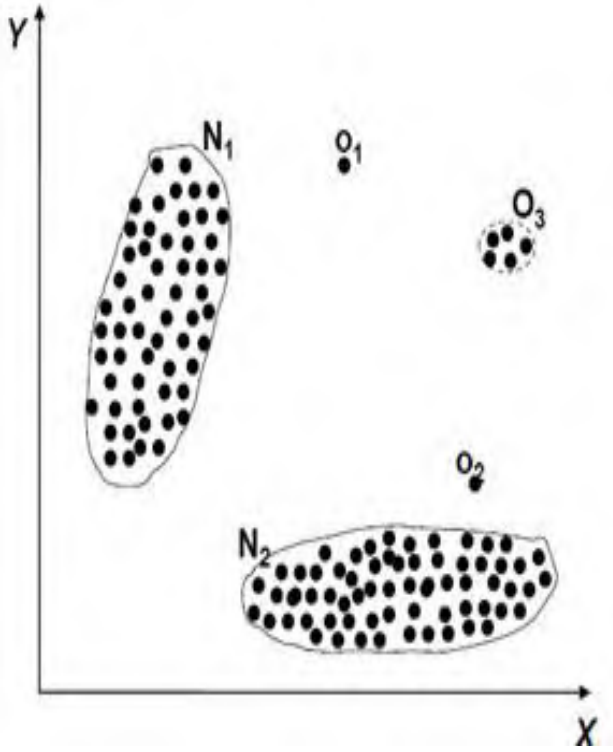
什么是海量运维web日志异常挖掘

2 常见的异常类型有哪些？

异常类型1-单点异常

单点异常

单独的数据实例是异常的

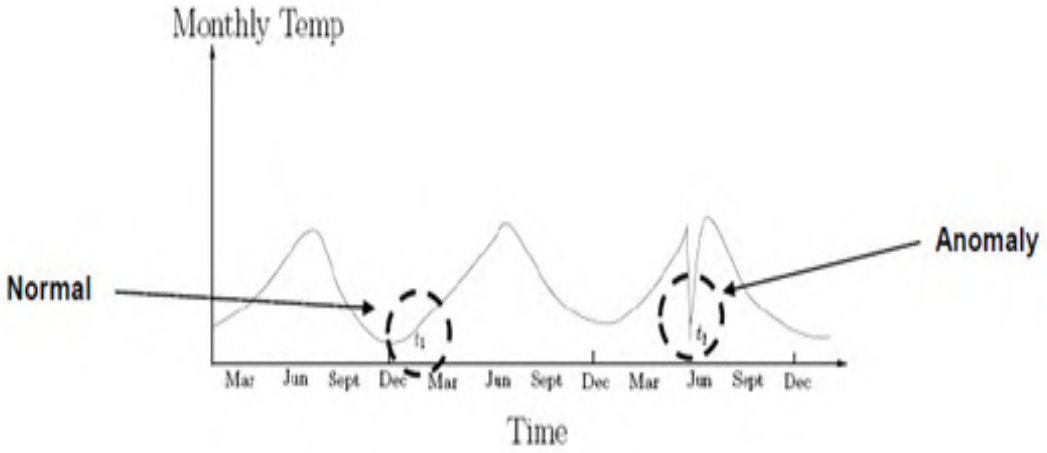


一个部署在内网的kibana，出现一条外网访问日志

异常类型2-上下文异常

上下文异常

- 在一个上下文中单独的或连续几个数据实例是异常的
- 需要一个上下文的概念



账号系统里面的平权访问

异常类型3-集体异常

集体异常

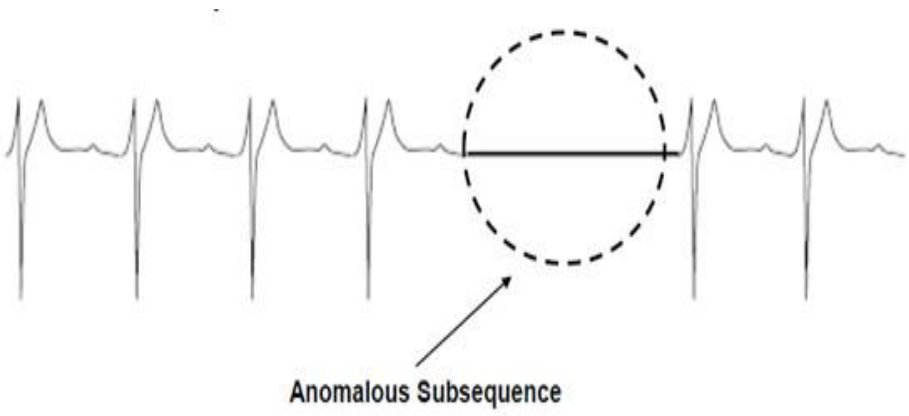
相关数据实例的集体是异常的
在数据实例间需要一个关系
常见的：
-有序数据

-空间数据

-图数据

在一个集体异常中单独的实例，从它们自己看来并不是异常的

金融网站/电商网站的撸羊毛事件



3 异常挖掘方法

异常挖掘方法1-基于经验特征挖掘

基于经验特征挖掘

- 基于分析人员自身经验，使用特定的与日志相关的特征进行挖掘
- 比如特定的字符串，特定的Cookies、特殊的系统命令等等

- 此方法在试探性分析的时候常用

老司机专用

异常挖掘方法2-基于数据统计挖掘

基于数据统计挖掘

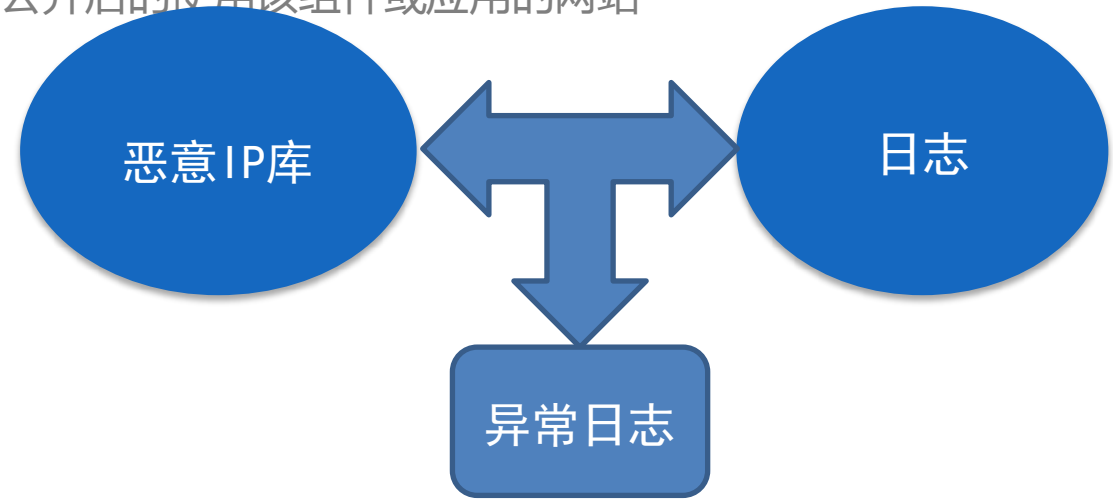
- 通过统计多维度的数据，根据其频次，分散度等信息，挖掘有价值的异常
- 比如统计单个IP24小时内，使用的手机号码数量；短时间内大量登录请求等

- 单IP一天之内使用100000+的电话对某接口进行查询

异常挖掘方法3-基于外部数据关联挖掘

基于外部数据关联性挖掘

- 基于已有的外部数据与日志的关联性进行挖掘
- 例如高危IP库的IP的正常访问日志，代理IP库的IP的正常访问日志，某个通用组件或应用漏洞公开后的使用该组件或应用的网站日志等等

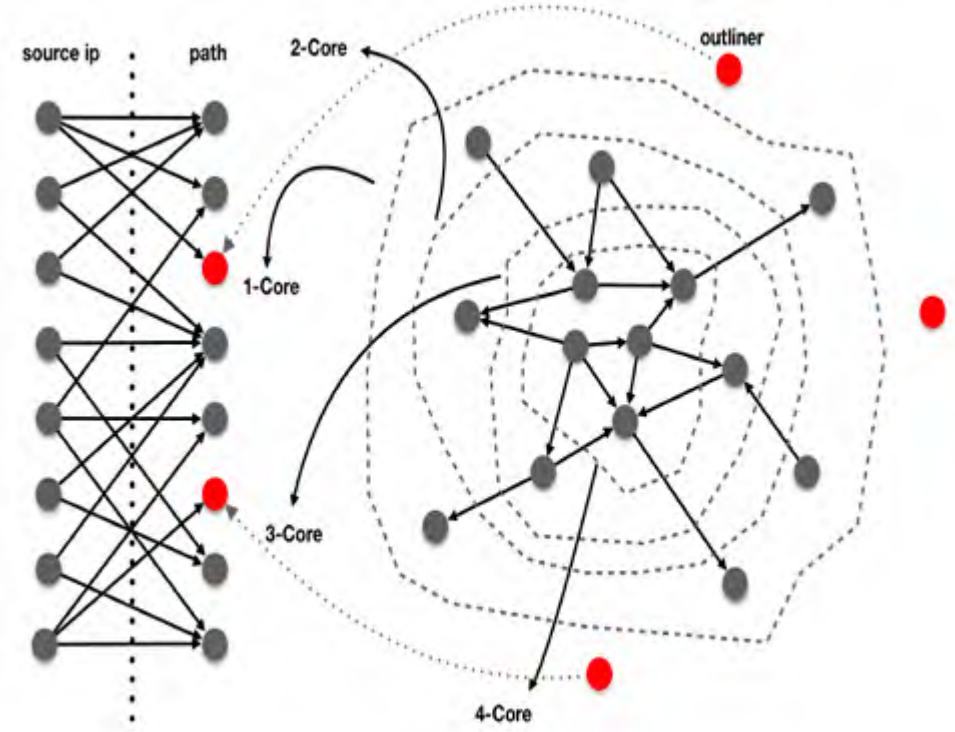


异常挖掘方法4-基于内部数据关联挖掘

基于内部数据关联性挖掘

- 通过分析运维日志的内部关联性进行异常挖掘

- 通过分析Referer与URL二元关系组以及IP与URL二元关系组，挖掘低频访问且是孤立节点的页面



异常挖掘方法5-基于防火墙规则的异常挖掘

基于防火墙规则的异常挖掘

- 将每条防火墙规则转换成多个语句的逻辑条件组合，并给予每个子条件一定的rank，然后对每条日志数据进行评分，根据最后评分以及设定阈值来判定日志是否异常
- 常用于防火墙 Bypass 挖掘、未知漏洞挖掘

异常挖掘方法6-基于网站画像的异常挖掘

基于网站画像的异常挖掘

- 根据网站的目录结构、动静态页面分布及页面参数类型等信息而抽象出的一个多维度的标签化的描述网站合法访问范围的画像模型。通过判断单条日志数据是否在网站画像中，来判定日志是否为异常

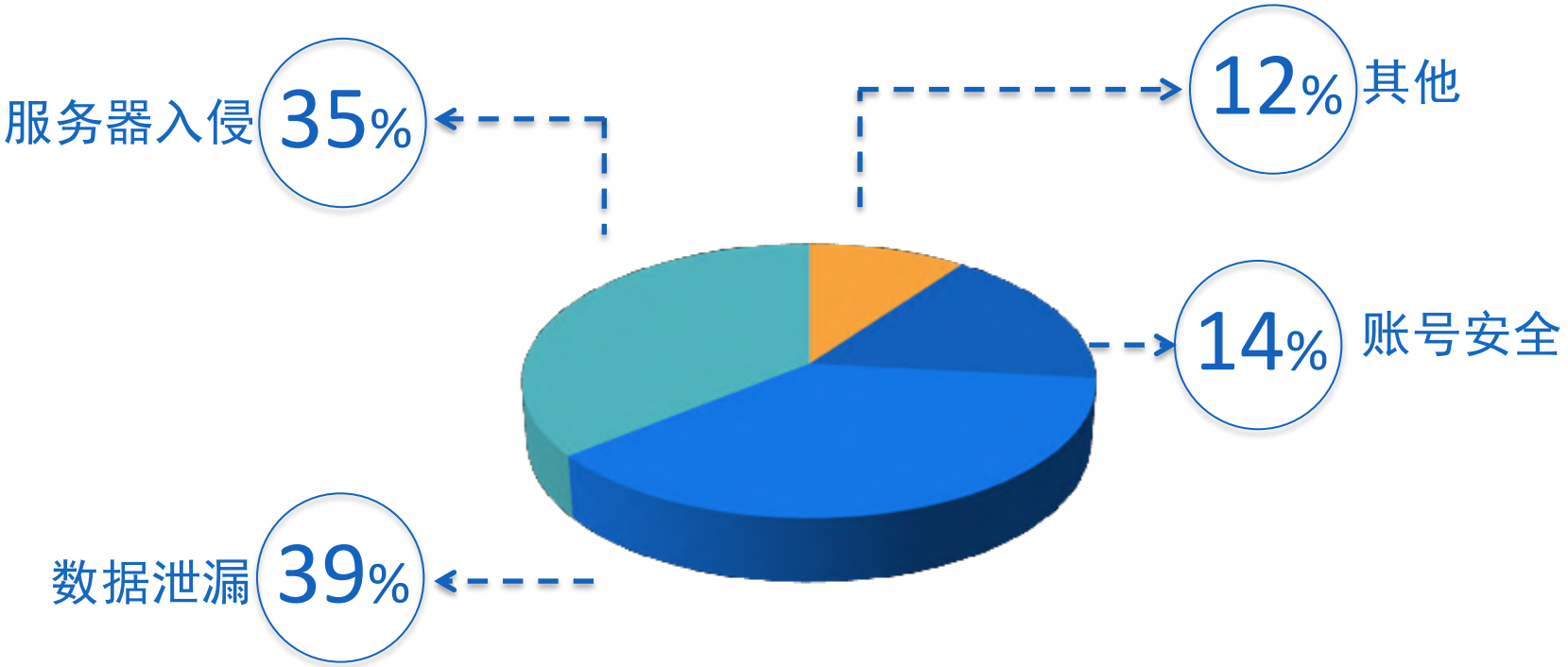
- 适用于单个网站的日常持续分析
- 每个网站可以根据其应用特性和业务特征构建个性化的网站画像

机器学习与异常挖掘

- HMM（隐马尔科夫模型）
- Isolation Forest（孤立森林）

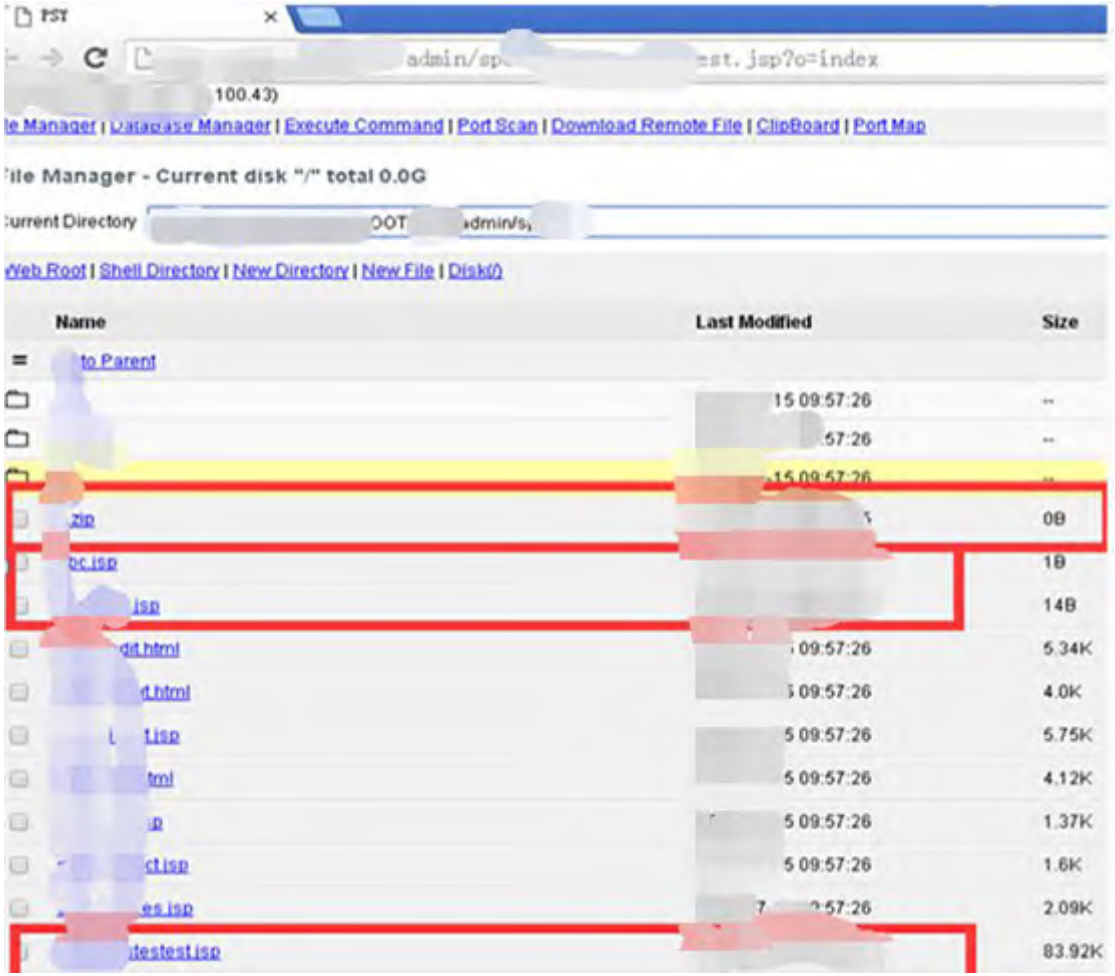
4 我们的工作

我们发现了什么



案例1-从异常日志到系统0Day

- 通过网站画像模型发现我们发现某网站存在一个webshell。而该网站使用某商业信息系统
- 通过IP和时间的上下文关联分析分析我们找到了上传点。
- 通过分析上传点的源码，我们发现其使用了某个通用模板，而此通用模板在未配置好的时候可能导致系统存在命令执行的漏洞。
- 经过验证，此商业信息系统所有版本都存在命令执行漏洞



案例2-一个异常访问引发的连锁反应

- 一个孤立访问页面，重放后发现是一个诈骗页面
- 使用内部大数据平台进行溯源分析

案例2-一个异常访问引发的连锁反应

- 发现网站存在后门
- 获取用户授权，进行深度分析

知道创宇云安全
为企业安全保驾护航

欢迎大家深入交流

Gdevops

全球敏捷运维峰会

THANK YOU！