阿里云
奥运会全球指定云服务商

第三届阿里云安全算法挑战赛

wj3235@126.com

云安全算法挑战赛

阿里云 奥运会全球指定云服务商

检测算法的原理

1 检测算法的原理

2 计算效率的分析

3 攻击

4 聚类

5 其他方法

云安全算法挑战赛

阿里云 奥运会全球指定云服务商

检测算法的原理

云安全算法挑战赛

检测算法的原理-预处理

同一个文件下的记录存成json格式

api_call 里面需要根据调用顺序升序排列

{"lable":0,"threads":{"2828":{"api_calls":[["2202","GetSystemTimeAsFileTime","0"],
[“XXXX"," LdrGetDllHandle ","0"]]}},"file_id":12}

云安全算法挑战赛

阿里云 奥运会全球指定云服务商

检测算法的原理

不同恶意文件其Api的调用顺序，
调用频次，返回值，函数类型
能够很好地反映他们的特征的

例如：勒索软件会高频次调用文件的读写，加密函数

软件的线程数
不同线程API调用的平均次数，最小和最高次数
可以总体上反映软件是不是多线程
发现某些benign软件的api调用次数超过5000次，恶意软件少见

TFIDF分析
如果某个函数（模式）在某个文件频繁出现
而在其他文件少出现那么该函数具有很好区分度

可以发现某些API调用的高阶模式

云安全算法挑战赛

阿里云 奥运会全球指定云服务商

检测算法的原理-特征工程:调用日志的序列化

Example（列表顺序:[index，apiname，返回值]

```
[0",",SetUnhandledExceptionFilter",",0"],
["1",",LdrGetDllHandle",",3221225781"],
["2",",LdrGetDllHandle",",3221225781"],
["3",",LdrGetDllHandle",",0"],
["4",",LdrGetProcedureAddress",",0"],
["5",",LdrGetProcedureAddress",",3221225785"],
["6",",SetErrorMode",",32775"]]
```

按照函数名类别序列化

不同的函数会划分成更加抽象的类别

比如： writeconsolea ， writeconsolew 都是 IO write

ntopenkey， ntopenkeyex ， regopenkeyexw 都属于 reg read

具体划分请查看附件

类别序列化结果：module_handle_get module_handle_get module_handle_get module_address_get module_address_get

按照函数名序列化

结果：SetUnhandledExceptionFilter LdrGetDllHandle LdrGetDllHandle LdrGetDllHandle LdrGetProcedureAddress LdrGetProcedureAddress SetErrorMode

按照函数返回值序列

返回值包括了许多return code

结果：0 3221225781 3221225781 0 0 3221225785 32775

按照函数名结合返回值序列化

结果：SetUnhandledExceptionFilter_0 LdrGetDllHandle_3221225781 LdrGetDllHandle_3221225781 LdrGetDllHandle_0 LdrGetProcedureAddress_0 LdrGetProcedureAddress_3221225785 SetErrorMode_32775

云安全算法挑战赛

阿里云 奥运会全球指定云服务商

检测算法的原理-特征工程(feature number)

5

基本特征

Threadnum
Totalapicall
Maxapicall
Minapicall
meanapicallperthread

1000

Api类型特征

5-gram
Tfid向量化

5300

Api序列特征

1-gram
2-gram
3-gram
通过tfidf向量化

3000

返回值序列特征

1-gram
2-gram
3-gram
Tfidf向量化

1500

API结合返回值序列特征

1-gram
Tfidf向量化

云安全算法挑战赛

检测算法的原理-模型

工具：lightgbm
尝试了几种集成模型
最后选择模型：dart

预测测试数据

最后模型直接使用了所有特征
因为做过特征选择后模型loss稍微有变大
虽然速度快很多

交叉验证选择loss比较小的参数
贪婪法调参

云安全算法挑战赛

阿里云 奥运会全球指定云服务商

检测算法的原理-总结

通过对软件API调用的序列分析

- 提取了API类别的5-gram TFIDF特征
- 提取了API的1-3-gram TFIDF特征
- 提取了API返回值的1-3-gram TFIDF特征
- 提取了API结合返回值的1-gram TFIDF特征
- 以及一些基本特征
- 这些特征通过监督分类算法 DART能够反映不同病毒之间模式和差异
- 并能用来进行预测

云安全算法挑战赛

阿里云 奥运会全球指定云服务商

计算效率的分析

1 检测算法的原理

2 计算效率的分析

3 攻击

4 聚类

5 其他方法

云安全算法挑战赛

计算效率的分析

Sandbox启动会花费几秒级别的时间
Sandbox环境耗费内存较大

文件必须上传到云端服务器的sandbox引擎才能被扫描

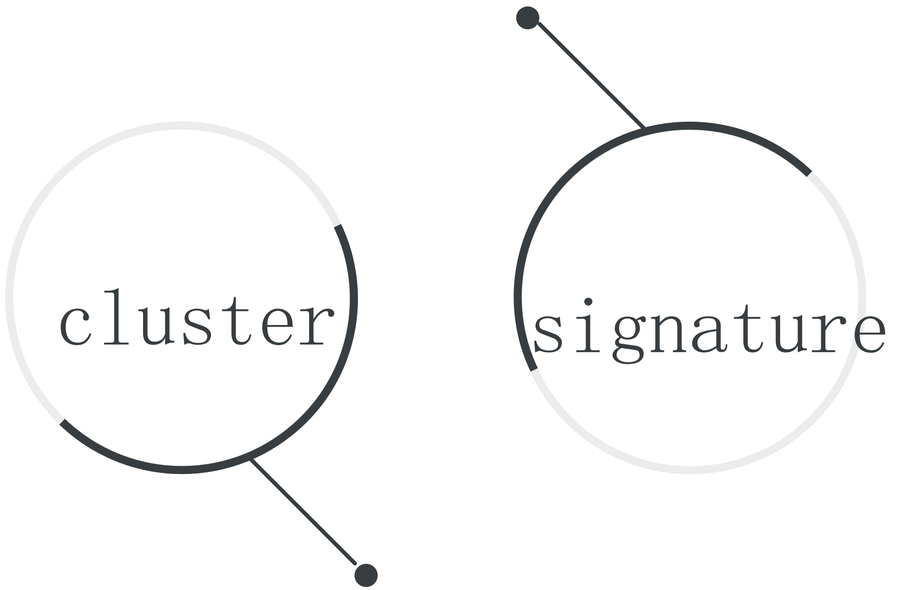
得到API log提取特征调用模型预测
一般都会迅速返回，在整个过程中占用时间相对较少

很多程序的执行要花费几十秒的时间
有的还会超时

云安全算法挑战赛

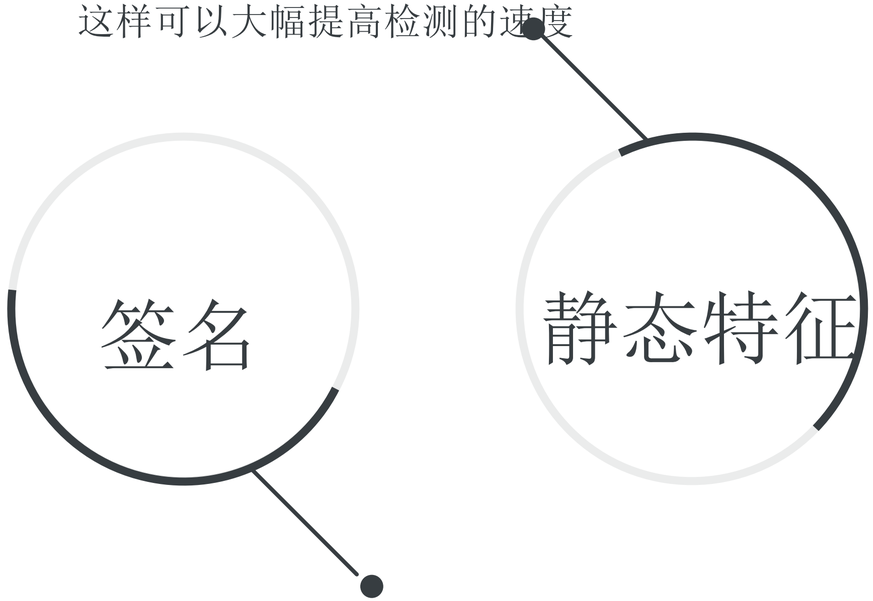
计算效率的分析——提高计算效率

利用已经有的signature（比如yara）扫描文件作为过滤



通过分布式集群并行的提高计算速度

除了可以利用动态特征构建模型
也可以通过静态特征，比如entropy，import API，常量字符串 等来做机器学习
这样可以大幅提高检测的速度



如果软件有签名，通过签名认定

云安全算法挑战赛

攻击

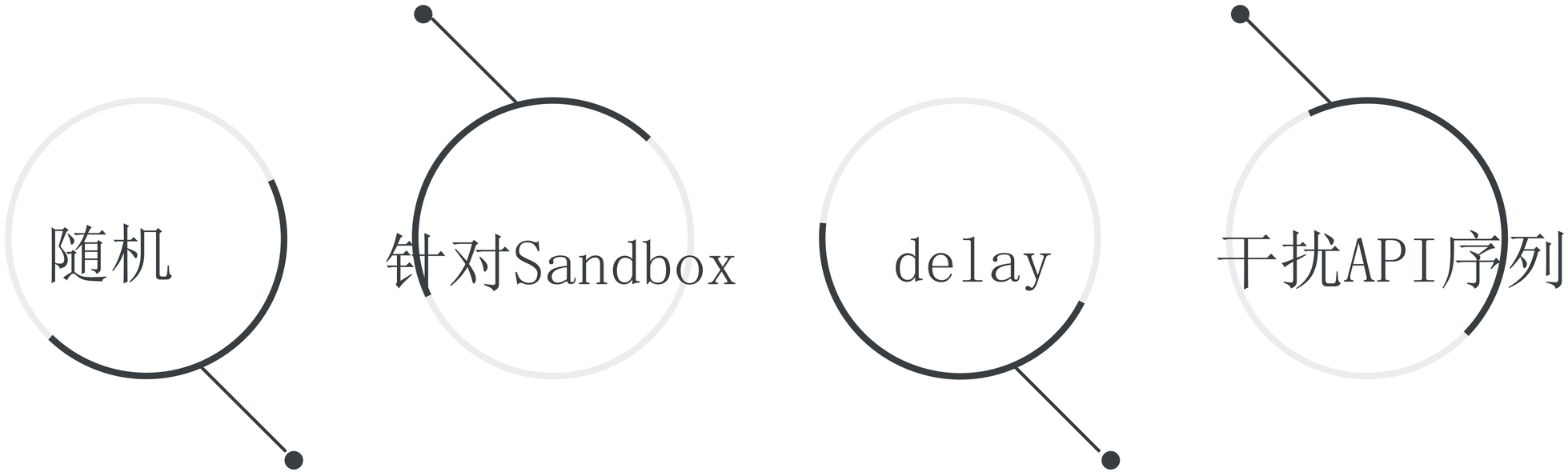
云安全算法挑战赛

阿里云 奥运会全球指定云服务商

攻击-如何逃逸

恶意软件通过某些特征检测自己是在真实环境还是sandbox环境
一旦发现是sandbox隐藏自己的恶意特征
利用Sandbox缺陷：针对hook的攻击
Sandbox可能并不支持所有的文件类型：android，IOS，linux等

在恶意特征的API序列里面调用无关的API视淆视听



通过一些随机数，让软件调用的API序列呈现不同的特征
只有在某些概率情况下才表现恶意特征
当随机调用表现的很正常时会造成引擎漏检

并不立即执行恶意，而是延时表现出来
毕竟sandbox不会长时间一直运行

云安全算法挑战赛

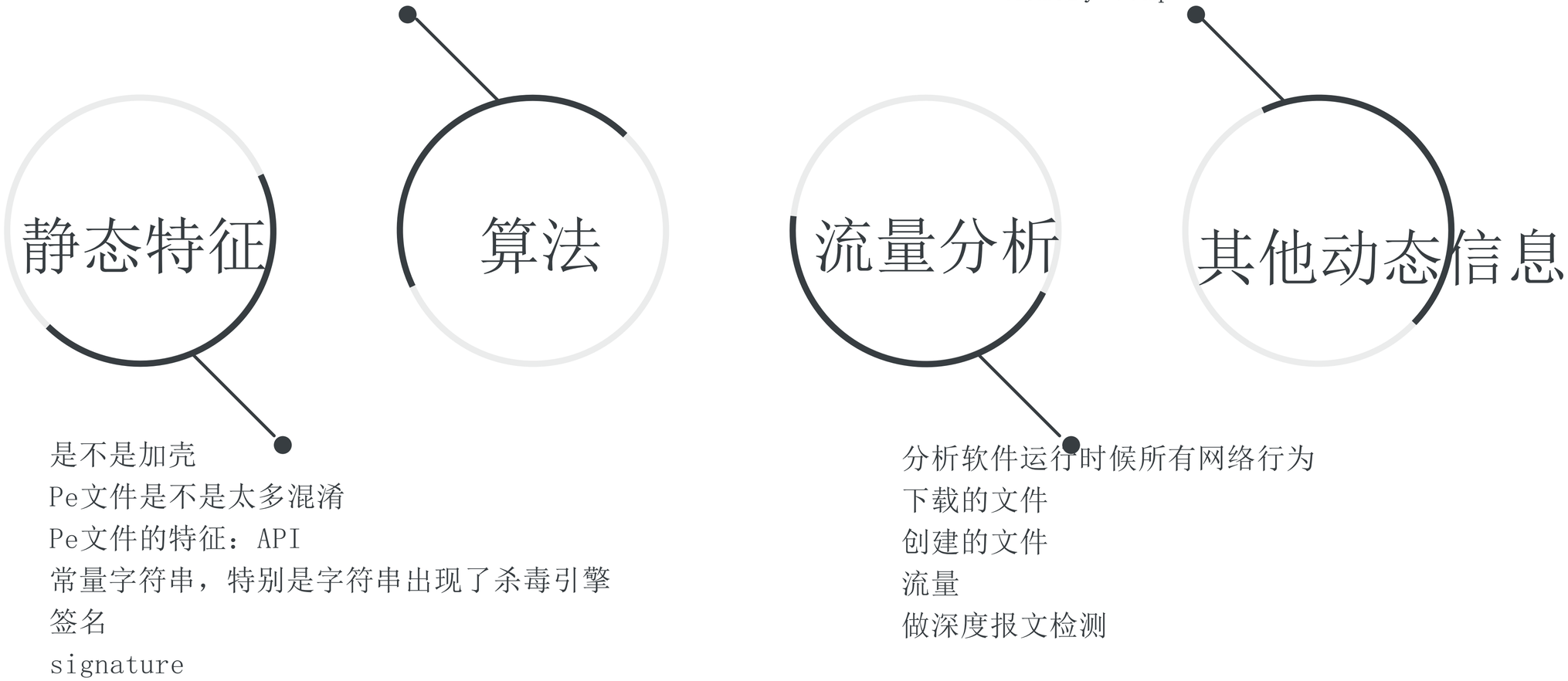
阿里云 奥运会全球指定云服务商

攻击-如何逃逸

只有系统重启才表现恶意
通过鼠标等系统特征判断是不是在虚拟机
PE文件或者其他文件类型有时候会攻击特定版本操作系统或者软件版本

云安全算法挑战赛

如何防御黑客绕过导致算法失效



云安全算法挑战赛

阿里云 奥运会全球指定云服务商

非监督分类-聚类

1 检测算法的原理

2 计算效率的分析

3 防攻击

4 聚类

5 其他方法

云安全算法挑战赛

阿里云 奥运会全球指定云服务商

聚类

对提取的特征做SVD降维，提高计算效率

聚类算法用k-means，K的选择可以参考已经有的病毒类型

考虑到聚类的实时性要求，可以先利用lsh（Locality-sensitive hashing）做初次聚类分析，再用K-means做聚类

考虑到聚类并不清楚聚类以后的类别

也可以尝试利用DBSCAN(Density-Based Spatial Clustering of Applications with Noise)算法做聚类

特征提取依然采用监督分类所描述的方法

云安全算法挑战赛

阿里云 奥运会全球指定云服务商

恶意软件检测方法比较

传统signarute方法

该方法通过提取恶意软件的signature来做检测
速度快
但是对新出现的病毒效果差
用户需要及时更新病毒库

基于静态特征的机器学习

静态特征是指不运行程序直接通过PE文件提取
该方法检测速度快，比signature方法的好
有时候对新出现的病毒会有作用
对于一些加壳的软件会有问题

基于动态特征的机器学习

可以有效防止0day攻击
但是该种检测方法速度较慢
如果文件太多，完全通过动态执行检测会非常耗时
也会有逃逸问题

阿里云 | 为了无法计算的价值