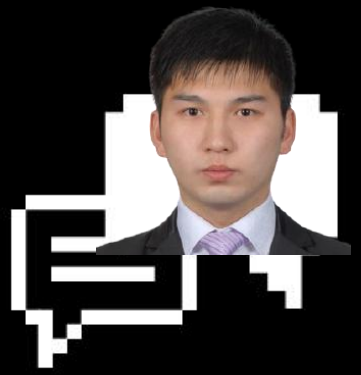


加密流量恶意软件分析在金融场景的实践

张志鹏　　斗象科技 高级安全顾问

CIS 网络安全创新大会
Cyber Security Innovation Summit

目录

网络安全创新大会
Cyber Security Innovation Summit

- 安全概述
- 加密流量的异常检测
- 多模型融合的恶意软件分析
- PRS-NTA
- 斗象&F5联合解决方案

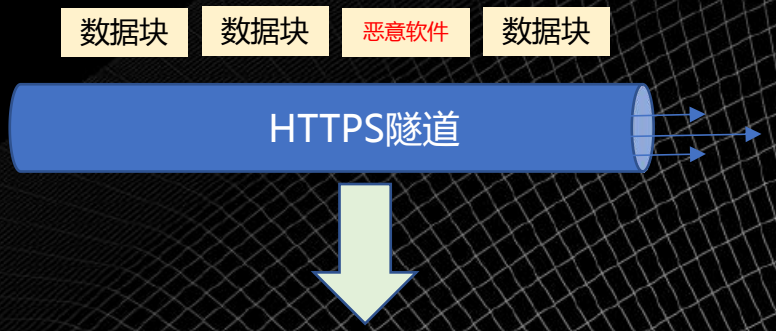
加密技术成为通信主流，超越65%的互联网流量采用https通信。

--加密通信的作用--

- **加密隐私数据**:防止您访客的隐私信息(账号、地址、手机号等)被劫持或窃取。
- **提高页面加载速度**:提高用户体验,防止客户流失。
- **安全身份认证**:验证网站的真实性,防止钓鱼网站。
- **防止网页篡改**:防止数据在传输过程中被篡改,保护用户体验。
- **提升信任度**:地址栏头部的“锁”型图标使您的访客放心浏览网页,提高用户信任度。

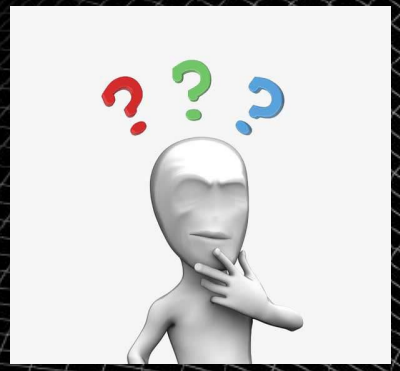
加密的隧道成为攻击者利用的工具

攻击者



感染主机

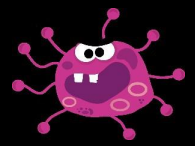
什么鬼？
被搞了？
what？



分析员

```
0000  44 85 00 d3 fc 2a 60 da  83 36 bb e7 08 00 45 20   D····*`· ·6····E
0010  00 f4 d0 3c 40 00 30 06  0f 68 6a 78 9f 7e 0a 00   ···<@·0· ·hjx·~··
0020  56 49 01 bb e1 71 18 9a  e5 10 a6 68 ee 3e 50 18   VI···q·· ···h·>P·
0030  02 48 f3 ca 00 00 17 03  03 00 c7 00 00 00 00 00   ·H······ ········
0040  00 00 01 eb fb eb 03 58  46 41 0f 51 f4 ee a9 4d   ·······X FA·Q···M
0050  6c 9d f8 61 0d 1d 84 2b  88 58 0b 5a bc e1 f1 b9   l··a···+ ·X·Z····
0060  da 3a b5 99 6c 52 7c 3e  82 ef 02 a6 0f d0 76 fe   ·:··lR|> ······v·
0070  35 64 f2 f8 52 aa f2 77  11 aa 19 6b a4 a6 c7 cd   5d··R··w ···k····
0080  8f 66 4e 52 2a 12 aa 96  81 97 2e 05 e5 fd e5 a7   ·fNR*··· ··.·····
0090  40 17 f1 ec f8 66 59 d5  d8 d0 a5 89 22 ea ba 48   @····fY· ····"··H
00a0  68 6e 87 f8 7e 6e cd 4f  26 08 17 34 e2 f8 6c 2c   hn··~n·O &··4··l,
00b0  ad c0 53 90 61 df 91 e7  15 b9 b9 0a 49 ac 43 45   ··S·a··· ····I·CE
00c0  86 65 c8 15 60 9b f4 5f  02 98 af ae 47 74 91 fd   ·e··`··_ ····Gt··
00d0  8f 74 ea 16 6e fd 1a 88  d6 f7 bd 9c b0 e1 56 9a   ·t··n··· ······V·
00e0  33 1a b2 38 a5 36 30 dd  2d a2 0d f9 f2 57 25 d0   3··8·60· -····W%·
00f0  e5 8c 9e 16 9a ee c5 e9  6a 4e 32 ea 9f ea 33 10   ········ jN2···3·
0100  e7 d3                                             ··
```

概述-恶意软件



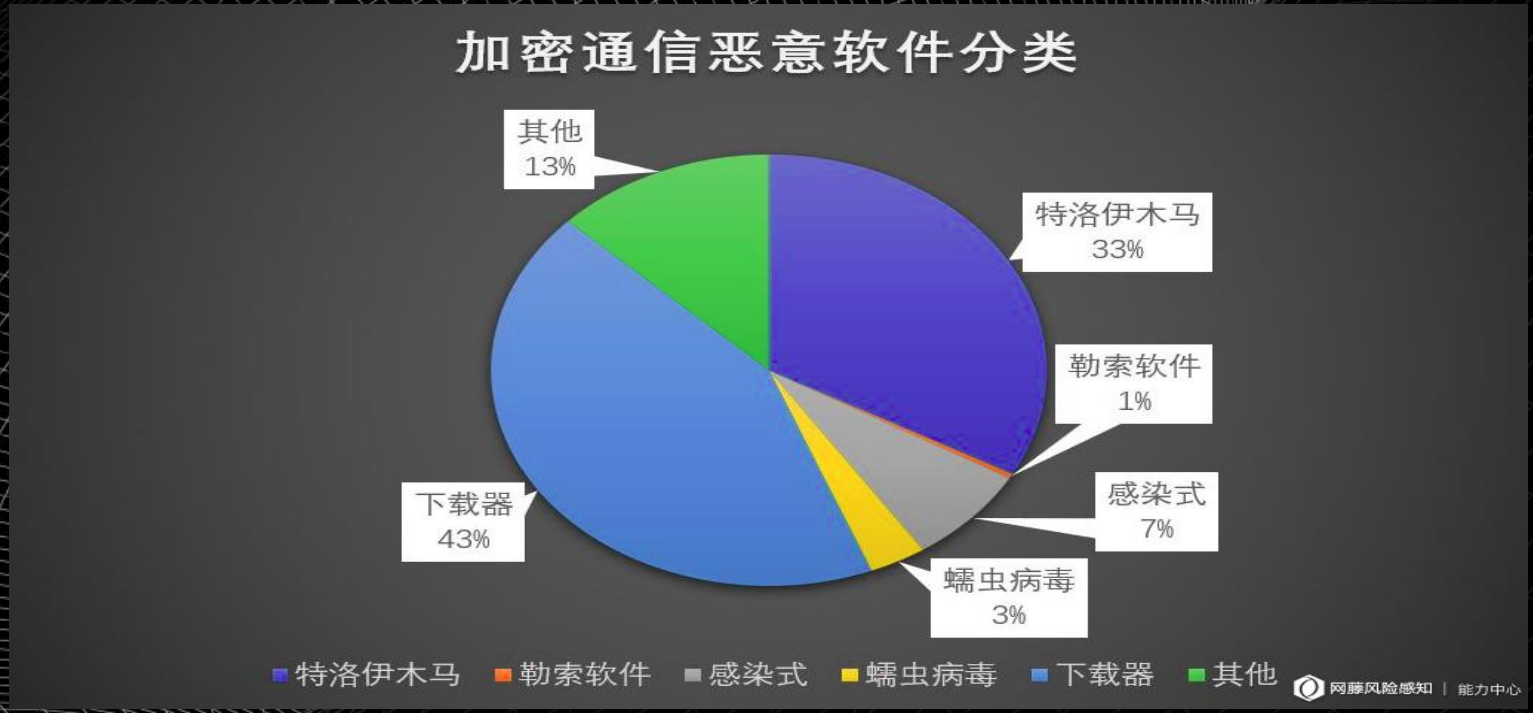
- Malware这个单词来自于Malicious和Software两个单词的合成，是**恶意软件**的专业术语。
- 是植入你电脑中的恶意代码，它可以完全控制、破坏你的PC、网络以及所有数据。

--来自百度百科

Malware包含了以下几个种类：

* Computer Viruses >>>计算机病毒
* Computer Worms >>>计算机蠕虫
* Trojan Horses >>>特洛伊木马
* Logic Bombs >>>逻辑炸弹
* Spyware >>>间谍软件
* Adware >>>广告软件
* Spam >>>垃圾邮件
* Popups >>>弹出

目前使用加密通信的恶意软件家族超过200种，使用加密通信的恶意软件占比超过40%，使用加密通信的恶意软件几乎覆盖了所有常见类型，如：特洛伊木马、勒索软件、感染式、蠕虫病毒、下载器等，其中特洛伊木马和下载器类的恶意软件家族占比较高。



概述-恶意软件

恶意软件产生的加密流量，根据用途可以分为以下六类：C&C直连、检测主机联网环境、母体正常通信、白站隐蔽中转、蠕虫传播通信、其它。

恶意软件类型	产生加密流量类型
特洛伊木马	C&C直连、白站隐蔽中转、检测主机联网环境、其他
勒索软件	C&C直连
感染式	C&C直连、母体正常流量、其他
蠕虫病毒	C&C直连、蠕虫传播
下载器	白站隐蔽中转、其他
其他	C&C直连、广告流量等

蠕虫传播

蠕虫具有自我复制、自我传播的功能，一般利用漏洞、电子邮件等途径进行传播。

C&C直连

恶意软件在受害主机执行后，通过TLS等加密协议连接C&C（攻击者控制端），这是最常见的直连通讯方式。

白站隐蔽中转

攻击者将控制命令或攻击载荷隐藏在白站中，恶意软件运行后，通过SSL协议访问白站获取相关恶意代码或信息。

检测主机联网环境

部分恶意软件在连接C&C服务器之前，会通过直接访问互联网网站的方式来检测主机联网情况，这些操作也会产生TLS加密流量。

母体正常流量

感染式病毒是将恶意代码嵌入在可执行文件中，恶意代码在运行母体程序时被触发。母体被感染后产生的流量有母体应用本身联网流量和恶意软件产生的流量两类。

加密流量与恶意软件有什么关系？

> 由恶意软件生成的加密流量

> 加密流量中携带了恶意软件

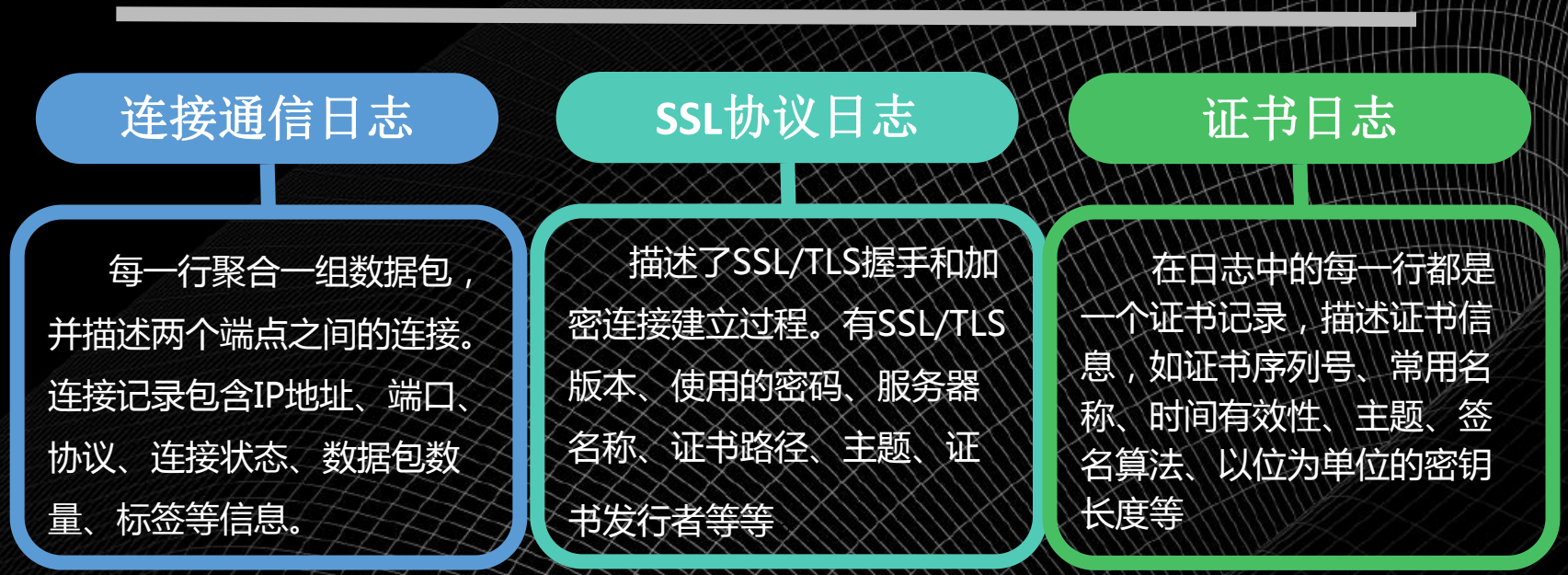
那么斗象科技

用什么的方式去解决

加密流量中恶意软件的检测呢？

课题研究的方法论

通过流量包深度解析方式提取HTTPS流量中的足够信息日志，包括连接通信日志、SSL协议日志、证书日志三部分



基于连接4元组的特征识别

连接4元组

连接4元组是一组SSL聚合和一些单独的连接记录。它们都共享源IP、目标IP、目标端口和协议的4元组。这个4元组是每个连接4元组的唯一键

SSL聚合

SSL聚合是3类数据的串联，SSL聚合仅包含一对连接记录和一个SSL记录，SSL记录是否具有证书路径取决于许多因素，例如，SSL握手是否成功等。

意义？

每个连接4元组总结了恶意软件连接到C&C服务器的行为，或者在大多数情况下普通用户连接到普通网站的行为。

数据集

CTU-13

CTU-13数据集是2011年在捷克共和国CTU大学捕获的。CTU-13数据集由在真实网络环境中捕获的13个不同的恶意软件组成。捕获包括恶意软件，正常和后台流量，但我们只是使用正常和恶意软件。数据集的正常部分是使用运行在虚拟机上的Windows 7捕获的。正常流量使用未知的web浏览器，恶意软件可能使用自己的库与Internet通信。该数据集的作者是Sebastian Garcia, Martin Grill, Honza Stiborek和Alejandro Zunino，他们对僵尸网络检测方法项目进行了实证比较.

MCFP dataset

有来自Stratosphere Malware Capture Facility项目的数据集，由Maria Jose Erquiaga完成。她通过选择使用HTTPS的恶意软件生成这些捕获。该数据集由53个恶意软件捕获和6个正常捕获组成。从2015年到2017年，在运行于虚拟机上的Windows 7上使用ie浏览器，在运行于Linux Debian上的Chrome浏览器上使用谷歌浏览器，都可以捕捉到正常的数据。

My own normal dataset

由于缺乏正常的数据，我们不得不创建更多的正常捕获。方法是浏览普通的网站，如facebook、twitter、gmail等。其中大多数我们都有账户，并进行了一段时间的互动。我们还使用了一个来自Moz.com的网站列表，其中包含了前500个注册域名，以及来自quantcast.com的前700个网站，这些网站是美国人访问最多的网站。数据集的这一部分包含13个普通捕获。2017年4月，运行在虚拟机上的Windows 7上的Mozilla浏览器和运行在Kali Linux上的Iceweasel浏览器分别管理了这次捕获。

特征识别

提取40余个特征，对于这些特征，我们将它们分为3组：**连接特征**、SSL特征、证书特征。

– 连接特征是来自连接记录的特征，描述与证书和加密无关的通信流的常见行为。

– SSL特征是来自SSL记录的特征，描述了SSL握手和加密通信的信息。

– 证书特征是来自证书记录的特性，描述了web服务人员在SSL握手期间提供给我们的证书的信息。

连接通信日志

（1）聚合和连接记录的数量。
（2）持续时间均值。
（3）持续时间标准差。
（4）超出标准差范围的持续时间占比。
（5）总发送包大小。

SSL协议日志

（1）连接记录中ssl连接的占比。
（2）TLS与SSL的比值。
（3）SNI占比。
（4）SNI is IP。

证书日志

（1）公钥均值。
（2）证书有效期的平均值。
（3）证书有效期的标准差。
（4）捕获期间证书周期的有效性。

随机森林

随机森林是一个包含多个决策树的分类器， 并且其输出的类别是由个别树输出的类别的众数而定。

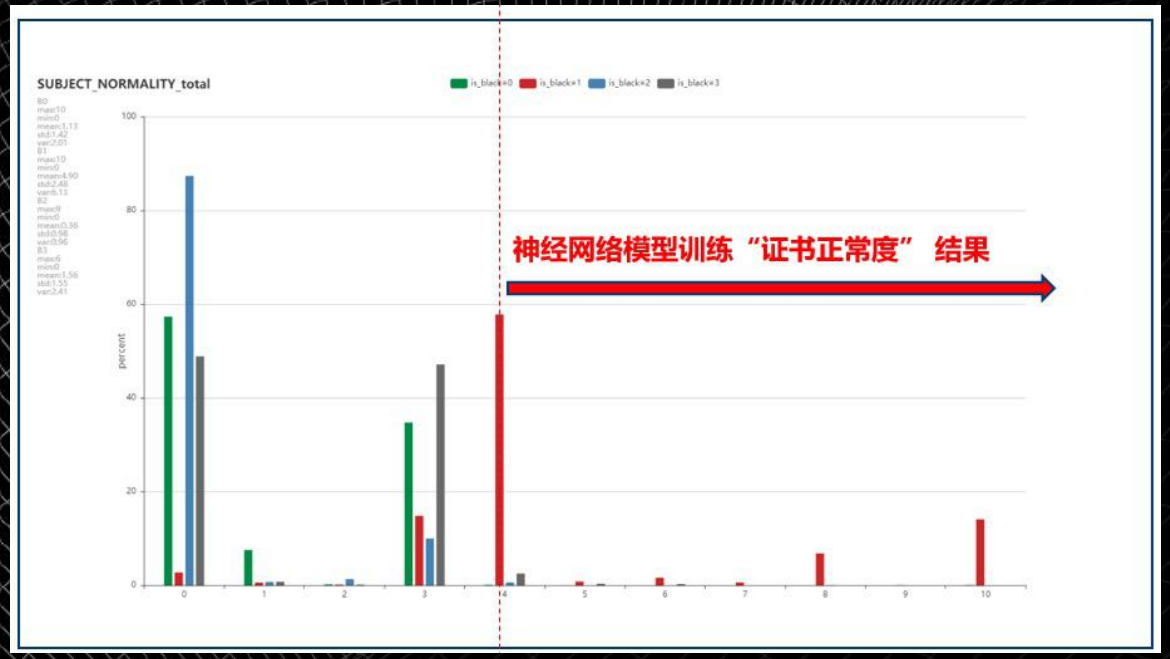
XGboost

XGBoost全名叫（eXtreme Gradient Boosting）极端梯度提升，经常被用在一些比赛中，其效果显著。它是大规模并行boosted tree的工具

经过数据清洗和过滤后，最终得到正样本46949条，负样本45121条。

模型	随机森林	XGBoost
准确率	98.61%	96.54%

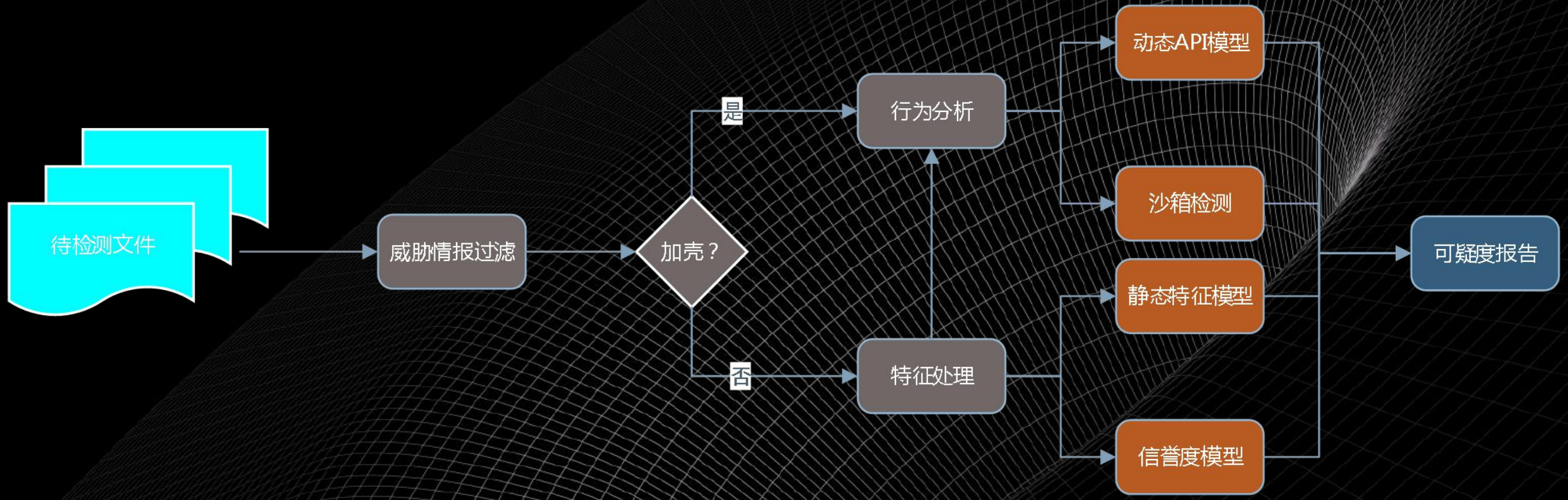
数据绘图



流量中携带的恶意文件怎么检测？

多模型融合检测

- 威胁情报、动态行为分析、静态特征分析、信誉度



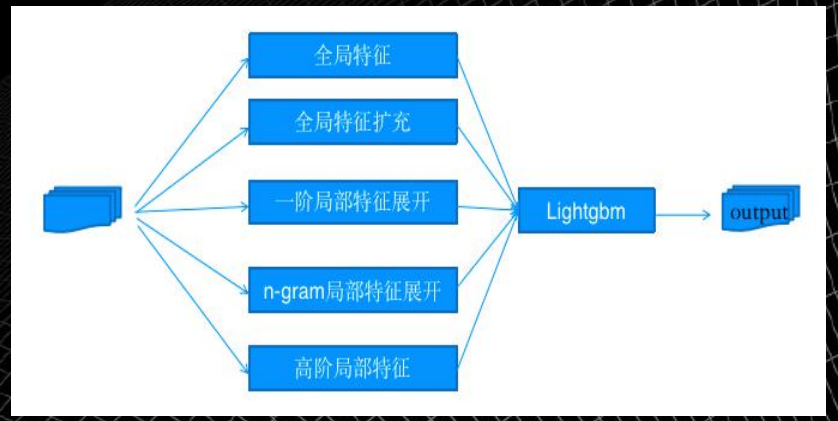
模型案例：基于API序列调用的恶意软件检测

数据源

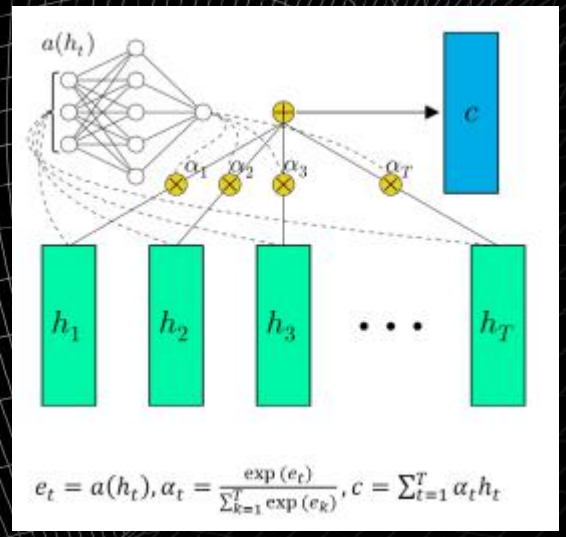
经过沙箱程序模拟运行后的API指令序列

file_id	all_api
1	GetTime GetFileTime NtFreeMemory ...
2	NtFreeMemory NtAllLocal ...
3	...

基于NLP 的文本词袋模型检测



基于统计机器学习模型

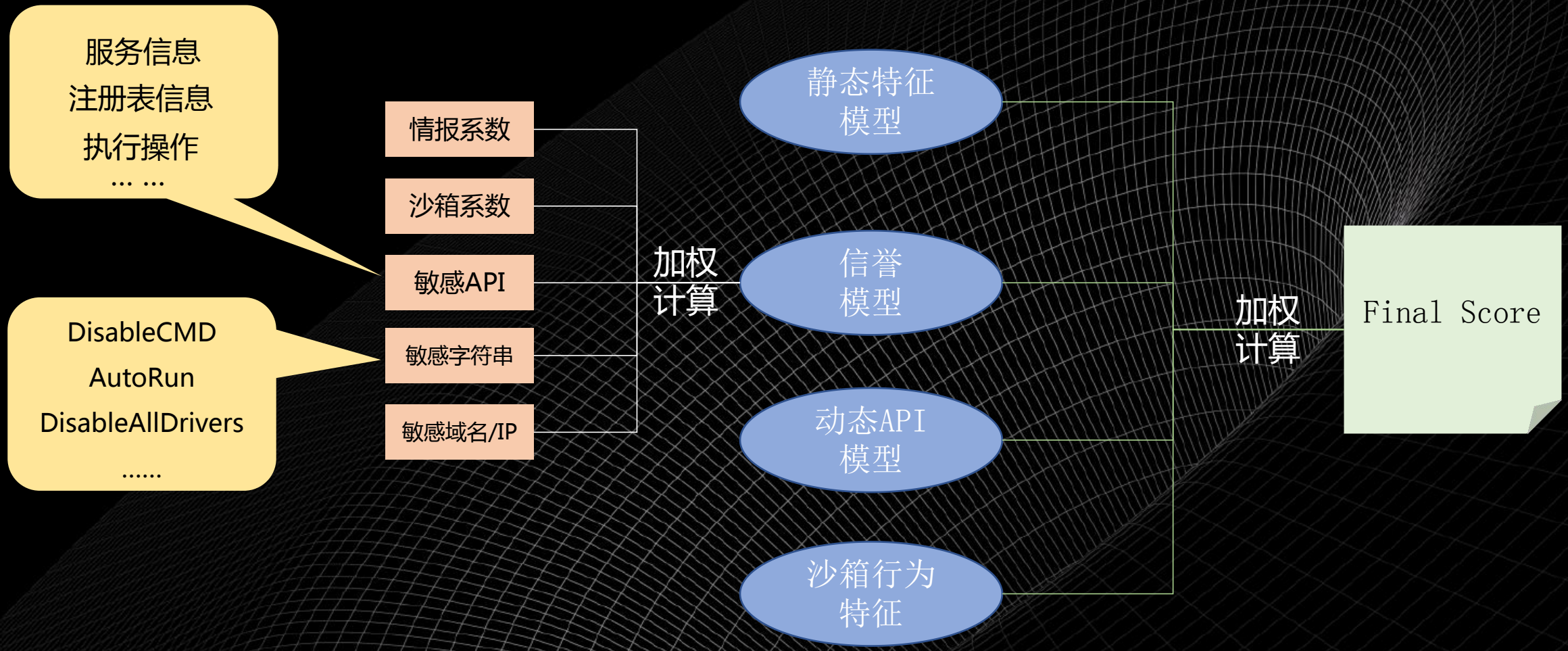


基于文本序列的深度学习检测

实验结果

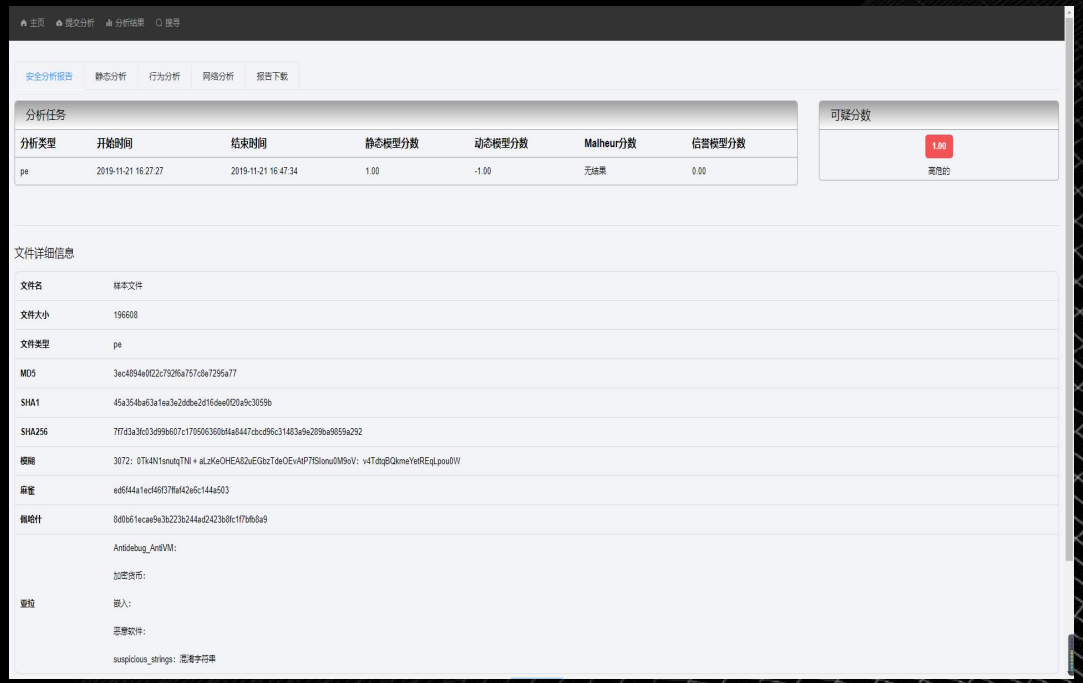
模型	统计机器学习模型	CNN	NLP
准确率	0.98838	0.94562	0.897

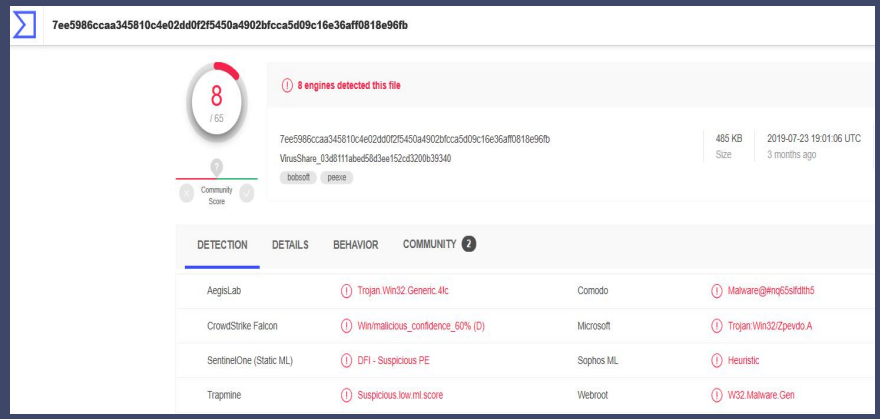
多级融合计解决方案


服务信息
注册表信息
执行操作
… …
DisableCMD
AutoRun
DisableAllDrivers
……
情报系数
沙箱系数
敏感API
敏感字符串
敏感域名/IP
加权
计算
静态特征
模型
信誉
模型
动态API
模型
沙箱行为
特征
加权
计算
Final Score

demo

网络安全创新大会
Cyber Security Innovation Summit





网络安全创新大会
Cyber Security Innovation Summit

THANKS

姓名 ：张志鹏

公司 ： 上海斗象信息科技有限公司

联系方式 ：roc.zhang@tophant.com