哈勃沙箱

技术白皮书

腾讯哈勃分析系统

HABO.QQ.COM

Tencent 腾讯

目录

第一章　系统简介

1. 概述

哈勃沙箱是哈勃分析系统的核心模块，依靠深度沙箱中自研的动态分析模块、静态分析模块以及稳定高效的任务调度框架，实现自动化、智能化、可定制化的样本分析，对文件具备准确的分析鉴定能力。可以得知样本的基本信息、触发的行为、安全等级等信息，从而更便捷地识别恶意文件。

哈勃通过建设大规模分析集群，沉淀了包括深度学习在内的多个高覆盖率的恶意样本检测模型，能够精准高效地对现网中的恶意样本进行打击。

哈勃解决方案的创新路线

2. 支持的文件格式

支持对常见可执行文件的分析：exe、dll、sys、msi 等

支持对常见脚本类文件的分析：js、vbs、html、bat、ps1 等

支持对常见文档类文件的分析：doc、docx、docm、dotm、xls、xlsx、xlsm、xlsb、xltm、xltx、xlam、ppt、pptx、potx、ppsx、pptm、potm、ppsm、rtf、pdf、swf

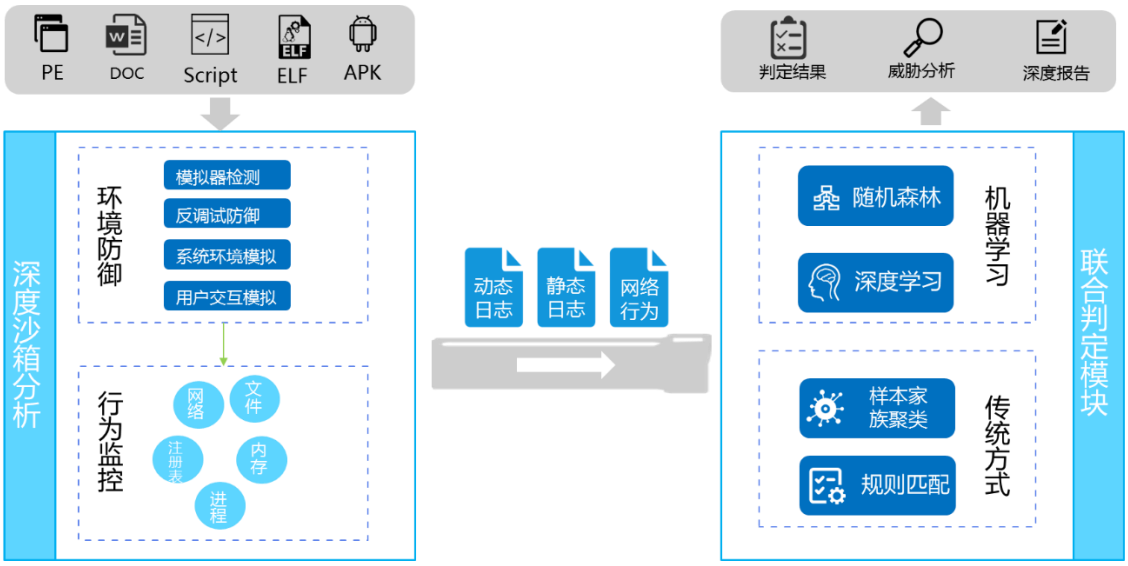
等

支持对常见压缩格式的分析：rar、zip、7z 等

支持对 Linux 可执行文件的分析：elf 等

第二章　系统架构

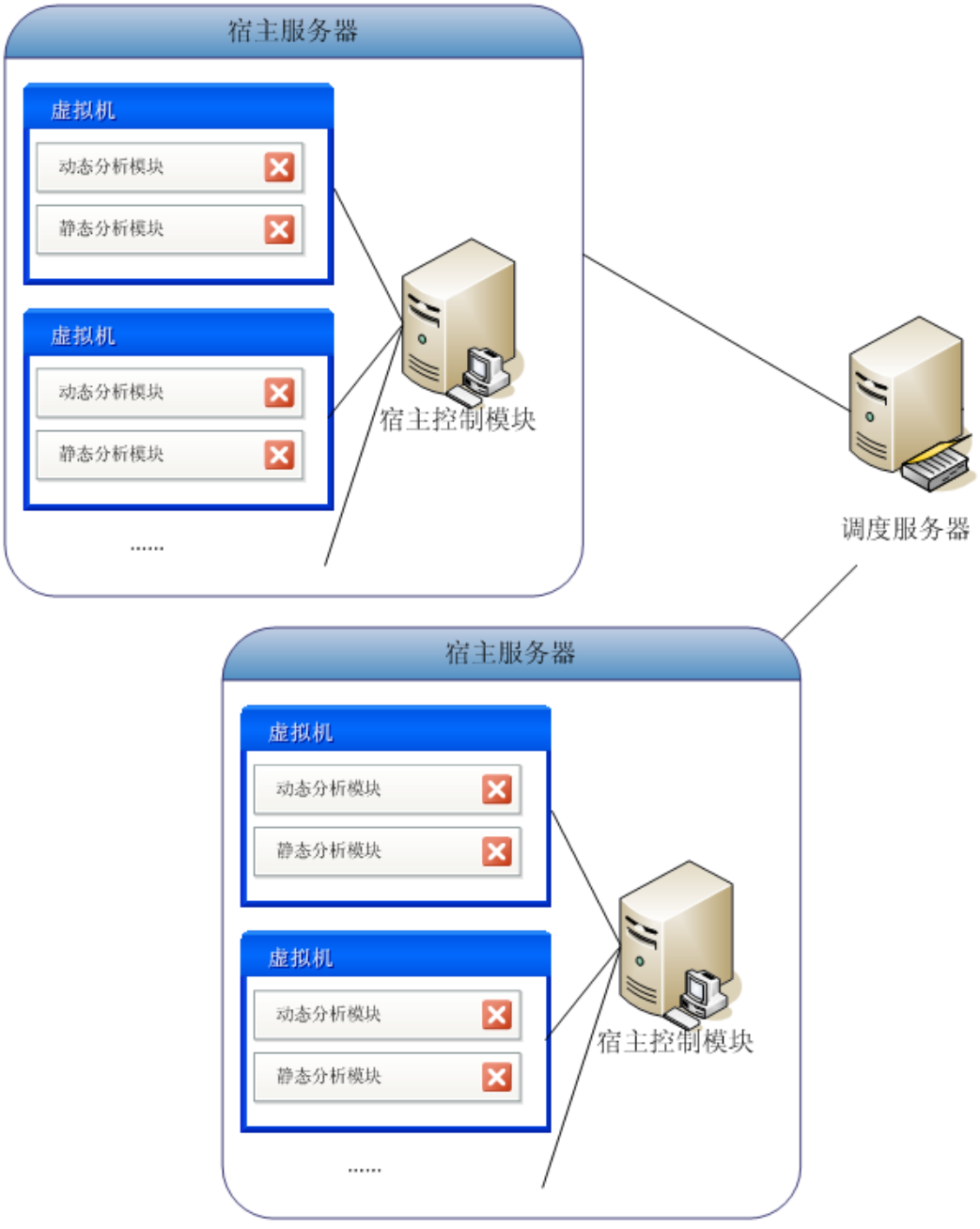
1. 整体流程



哈勃整体流程

哈勃的整体流程是从样本源获得样本，首先经过深度沙箱进行动静态分析，这一环节过后会产生动静态的分析日志，以及恶意行为线索。然后根据不同的检测模型对动静态日志进行处理，将日志中提取出来的特征交给模型鉴定，从而产生鉴定结果。

2. 系统调度和分析集群



系统调度框架

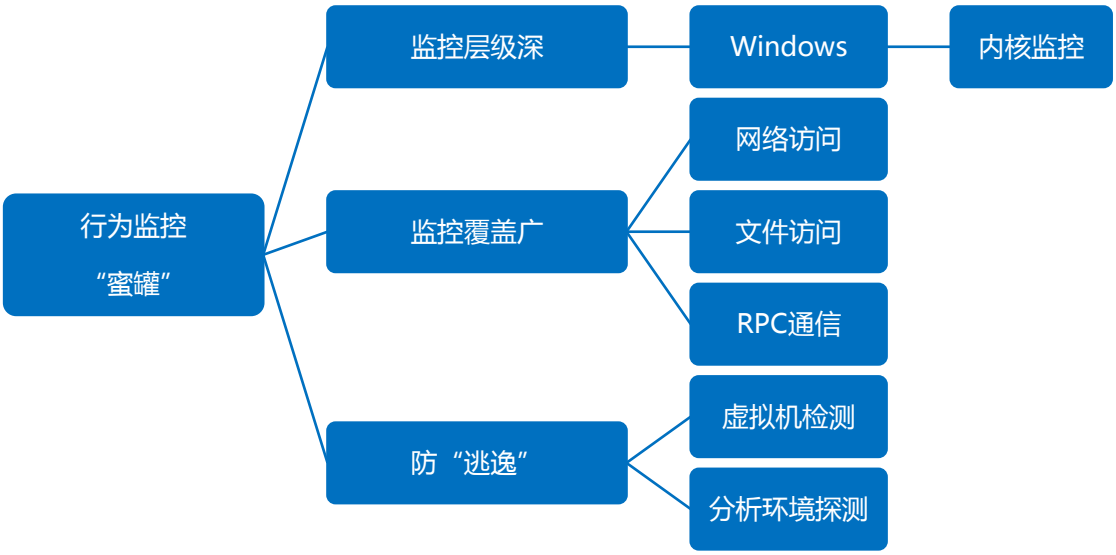
哈勃沙箱以任务为基本单位，对样本进行调度和追踪。系统对大批量样本的并发能力进行了针对性的改进，在流量吞吐、并行计算、耗时控制等方面进行了专项优化，保证了系统能够长时间持续稳定地处理大量样本，提供不间断的支持。

分析集群实现了哈勃沙箱的核心功能。通过主控模块，样本及其对应的任务被分发到合适的宿主服务器中的虚拟机，虚拟机中的分析模块自动运行样本，同时收集样本的各项动态、静态行为，生成行为日志。先进的分析模块保证了样本的各类可疑行为可以被充分的执行和捕捉，同时不会对虚拟机、分析服务器及网络环境产生危害。

第三章　核心功能模块

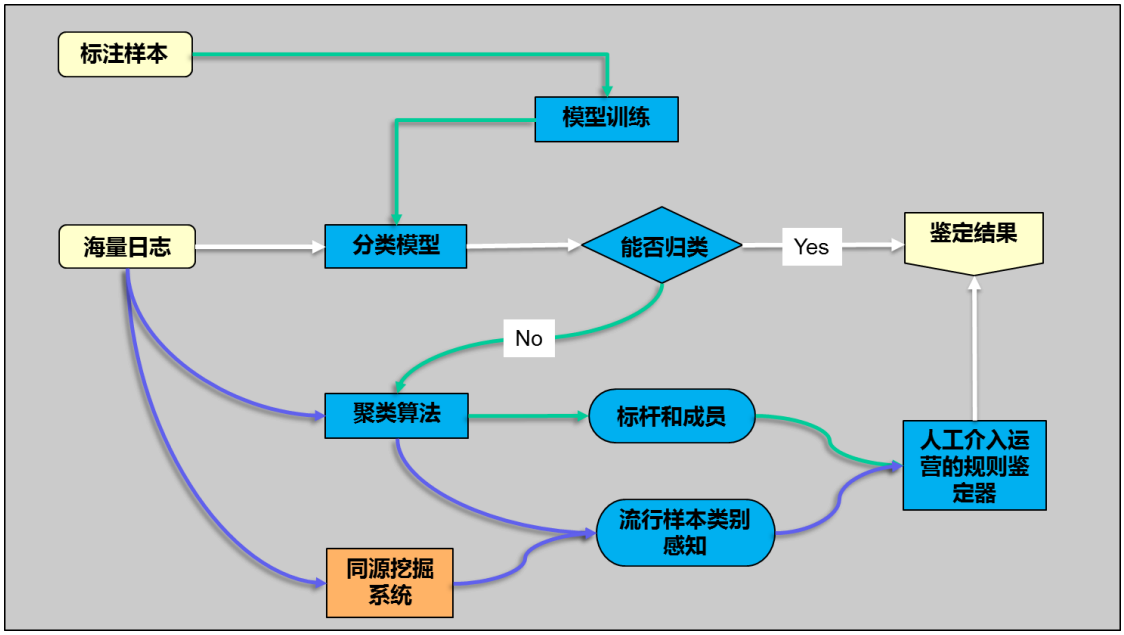
1. 动态分析模块

该模块依靠驱动、服务进程、监控进程三个层次，从多个方位全面监控样本在虚拟机中运行后的动态行为，同时通过独有的手段诱使样本产生通常难以触发的行为，从而将样本的可疑行为完全暴露出来，供后续分析使用。具备高可疑行为监控，网络发包监控，隐私窃取监控等监控能力。



哈勃动态行为监控

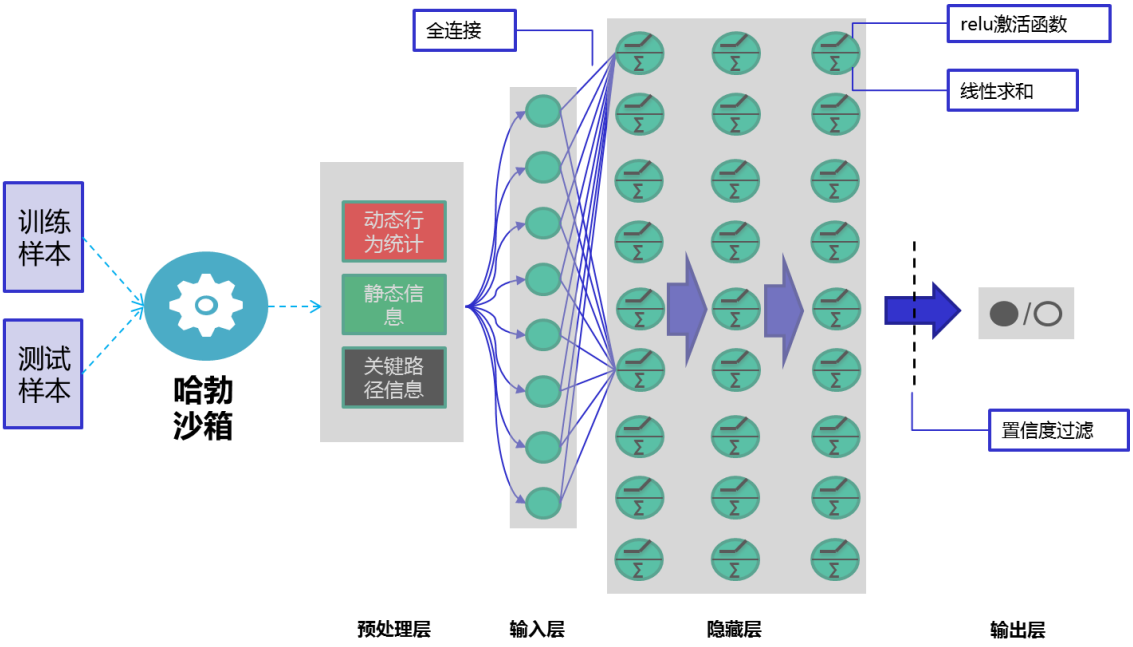
2. 算法模块



算法盒子结构图

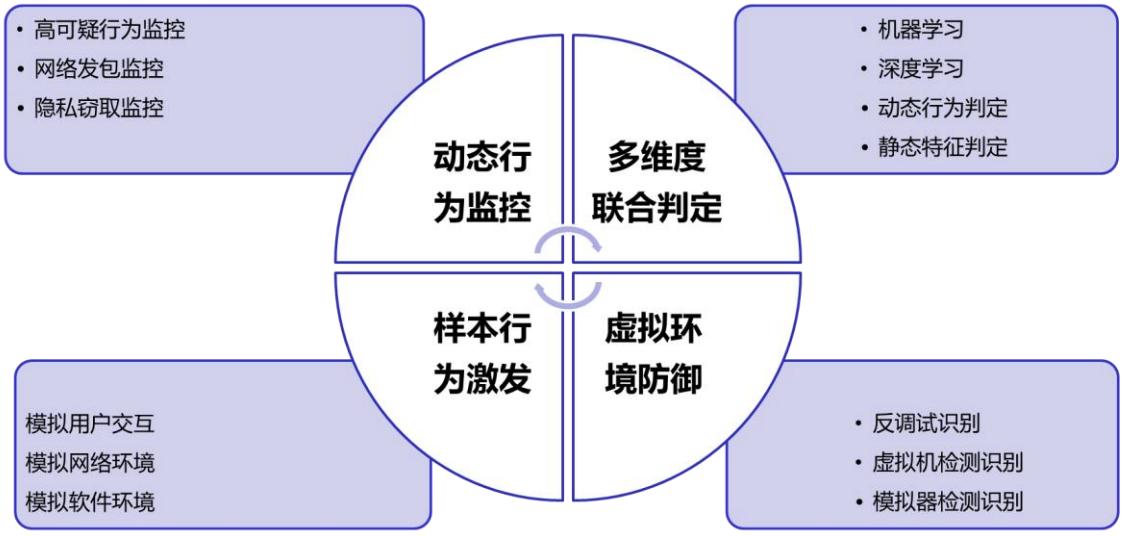
绿色箭头代表的是算法模型学习和使用的流程，白色箭头代表样本鉴定的流程，紫色箭头代表的是同源系统在算法模块中的作用。在处理动静态分析之后的海量日志时，使用了多种机器学习算法（决策树、随机森林、N-Gram 等）组成的算法盒子，对提升识别鉴定能力起到了极大的提升作用。

同时，哈勃还引入了深度学习模型，将沙箱的动、静态分析日志输入到 FNN 等深度网络模型中进行学习和训练，进一步提高了系统的泛化能力。



深度学习模型示意

3. 安全评级模块



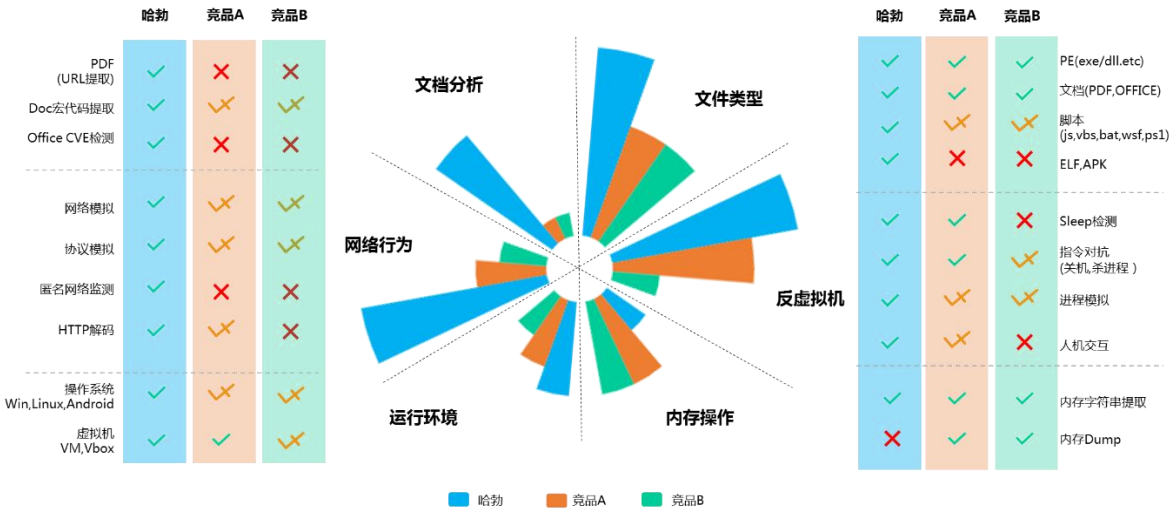
行为安全评级的核心

传统的轻量沙箱仅仅把动态分析作为辅助判定的一种手段。而哈勃沙箱通过多种不同的方式，对经过动静态分析的日志进行自动化的解析和处理，根据日志的内容对样本的安全等级进行判定。在保证极低的误报率前提下，识别率在和传统的杀毒引擎对比中，处于领先水平。

第四章　系统特点

1. 全面的动态分析

自研的监控模块 ,行为覆盖的全面性在和国内外各种沙箱的对比中均处于领先地位。

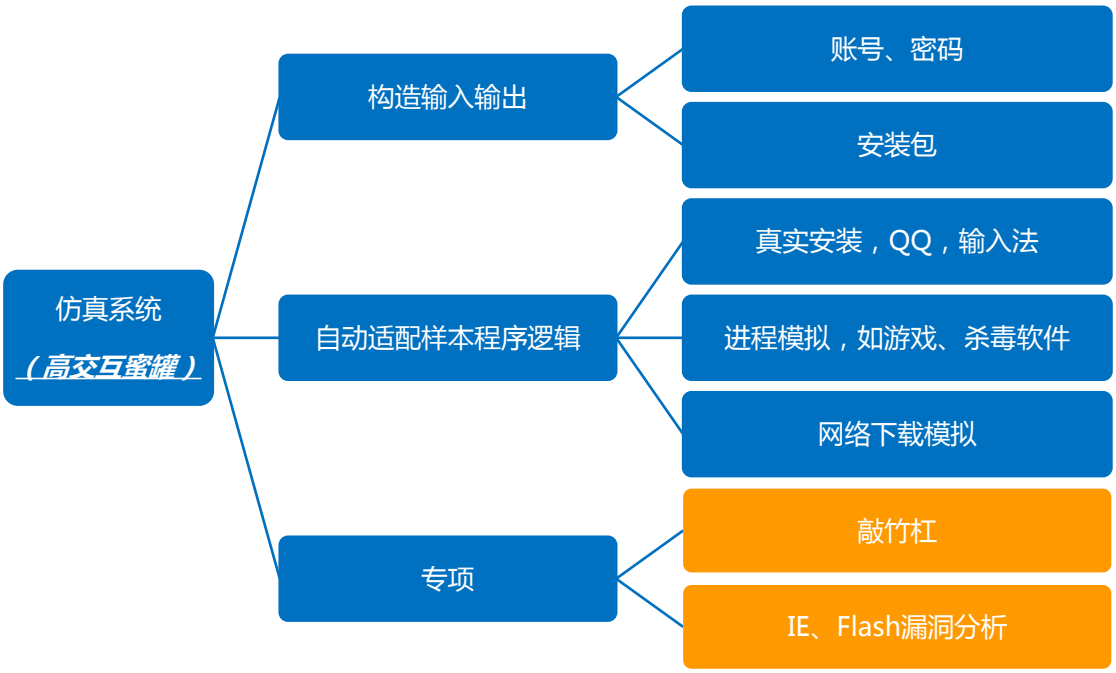


哈勃沙箱的监控能力与竞品的对比

2. 深入性和强对抗性

模拟真实用户环境，模拟用户交互，模拟网络环境，从而激发样本行为。

感知样本检测模拟器和调试器的行为 ,对于样本对抗动态分析环境的行为 ,进行反对抗。



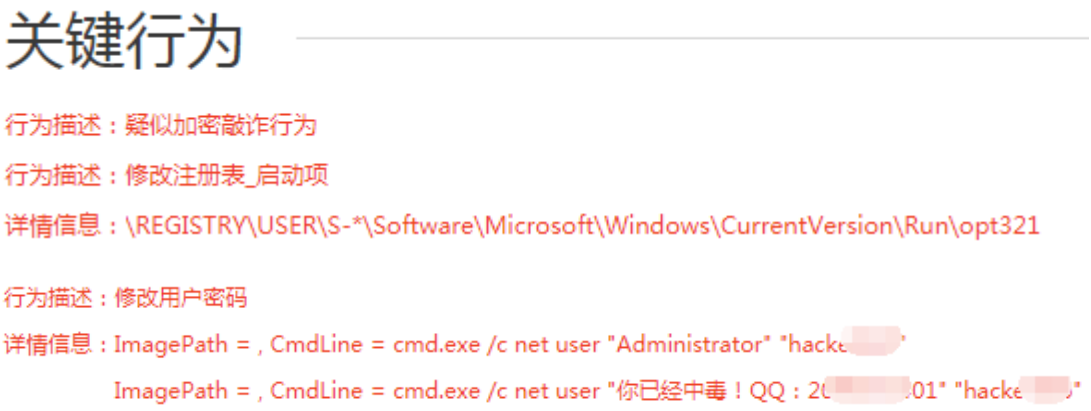
环境仿真

对抗虚拟机检测

3. 漏洞挂马、勒索病毒等高危样本的检出能力

哈勃沙箱通过对网址和脚本的分析，能识别其中的漏洞利用以及挂马攻击。对于漏洞和挂马的攻击，哈勃沙箱部署了针对性适配的分析环境。从漏洞触发，到漏洞利用再到最后的 payload 执行，哈勃在各个环节均进行了监控，能够识别已知和未知的漏洞攻击。

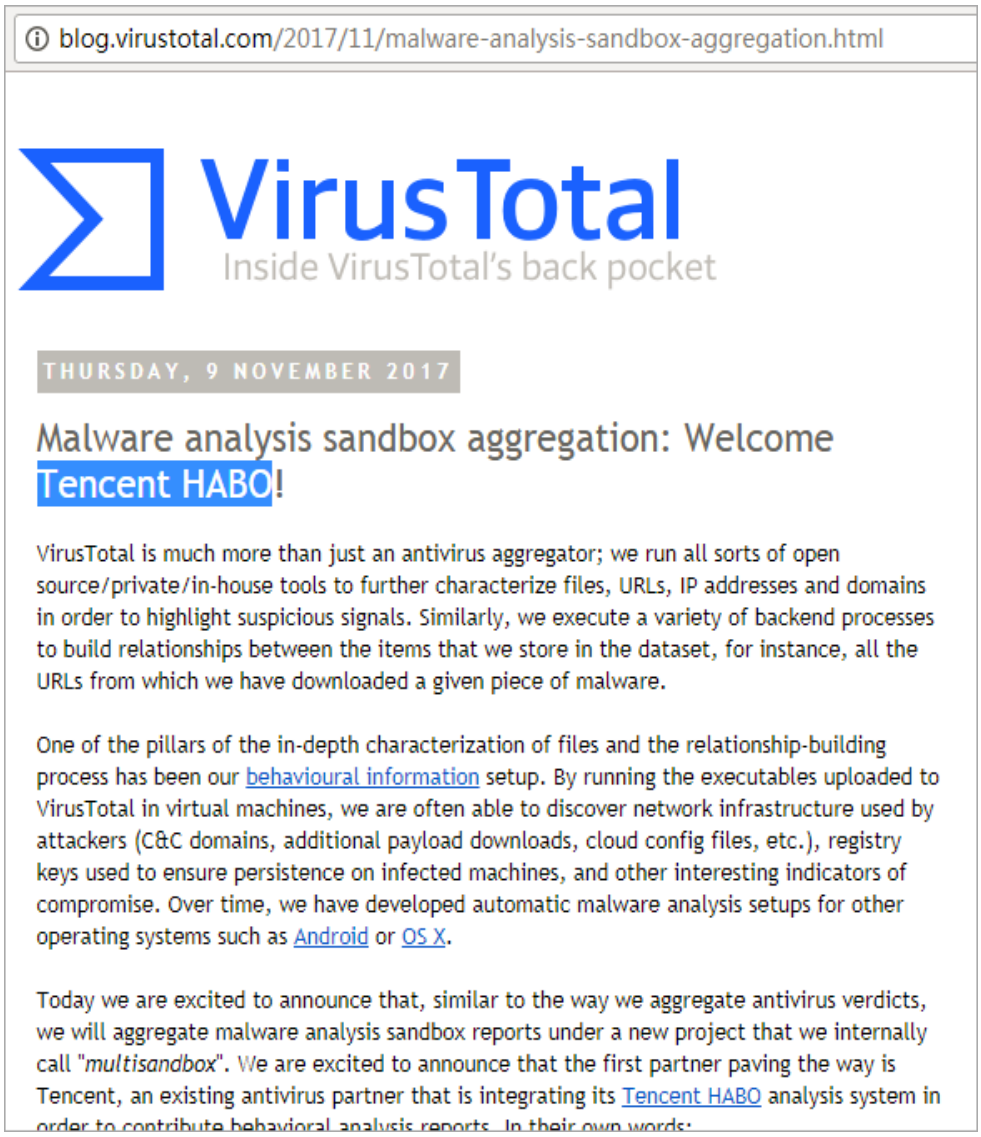
同时，哈勃沙箱还会针对性地诱发和监控勒索病毒行为特征，能够触发和捕获加密文件、匿名网络访问、修改登录密码等典型行为，从而更好地识别已知和未知的勒索病毒威胁。



哈勃沙箱捕获的勒索病毒行为

4. 得到业内高度认可

哈勃系统得到了安全业内的高度认可 ,除了通过来自政府和第三方机构的多项测评之外，还成为了 Google VirusTotal 全球首家沙箱合作伙伴。



blog.virustotal.com/2017/11/malware-analysis-sandbox-aggregation.html

VirusTotal

Inside VirusTotal's back pocket

THURSDAY, 9 NOVEMBER 2017

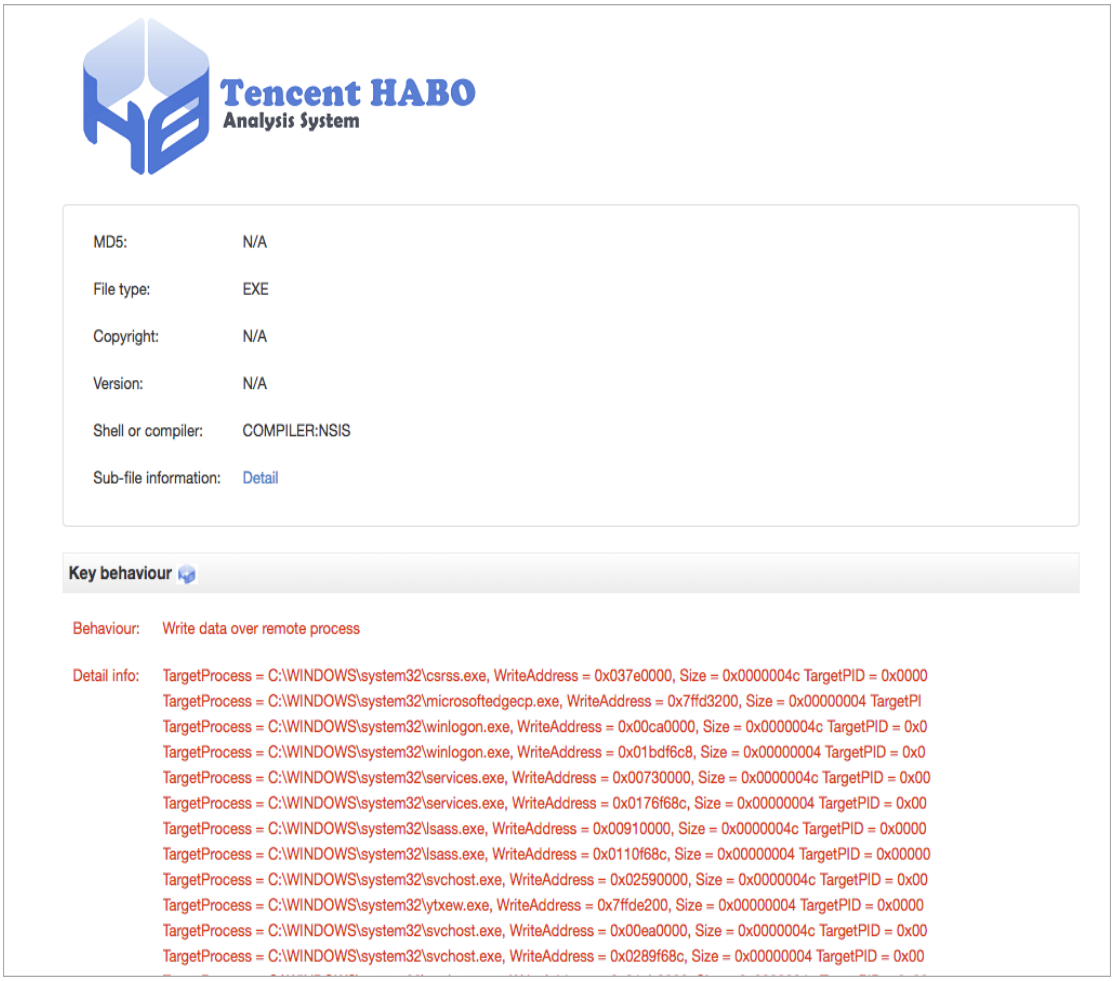
Malware analysis sandbox aggregation: Welcome Tencent HABO!

VirusTotal is much more than just an antivirus aggregator; we run all sorts of open source/private/in-house tools to further characterize files, URLs, IP addresses and domains in order to highlight suspicious signals. Similarly, we execute a variety of backend processes to build relationships between the items that we store in the dataset, for instance, all the URLs from which we have downloaded a given piece of malware.

One of the pillars of the in-depth characterization of files and the relationship-building process has been our behavioural information setup. By running the executables uploaded to VirusTotal in virtual machines, we are often able to discover network infrastructure used by attackers (C&C domains, additional payload downloads, cloud config files, etc.), registry keys used to ensure persistence on infected machines, and other interesting indicators of compromise. Over time, we have developed automatic malware analysis setups for other operating systems such as Android or OS X.

Today we are excited to announce that, similar to the way we aggregate antivirus verdicts, we will aggregate malware analysis sandbox reports under a new project that we internally call "*multisandbox*". We are excited to announce that the first partner paving the way is Tencent, an existing antivirus partner that is integrating its Tencent HABO analysis system in order to contribute behavioral analysis reports. In their own words:

VirusTotal 关于和哈勃达成沙箱合作的官方公告



哈勃沙箱分析结果在 VirusTotal 上的展示形式