save 杀毒功能

测试指导

深信服科技股份有限公司

目录

第1章 文档说明

随着 AF8.0.5 及以上的版本发布，SAVE 引擎做为 AF 在市场同类产品中的强势功能之一，同时结合当时日益频发的勒索病毒及其变种。针对病毒的防护在用户端也变得尤为重要。

所以此文档，主要指导一线技术人员，后续在用户侧交付项目时，无论测试/实施/售后，遇到 AF8.0.5 及以上版本，如无特殊情况，均要求开启杀毒功能，且尽量给用户现场模拟展示杀毒功能的效果。

第2章 SAVE 说明

2.1 命名说明

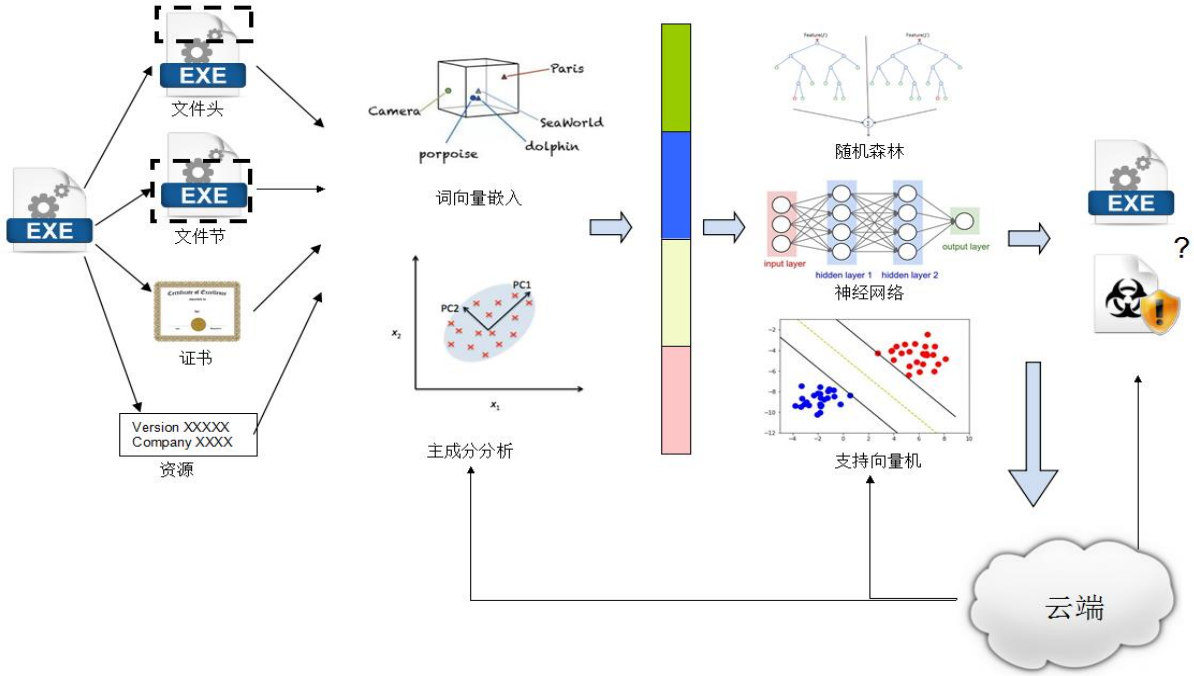
SAVE 的全称是 Sangfor AI-based Vanguard Engine，强调了基于 AI 技术，而 Vanguard 有“先锋、领导者”之意，中文名是“**SAVE 安全智能检测引擎**”

2.2 核心优势

SAVE（Sangfor AI-based Vanguard Engine）是由深信服创新研究院的博士团队联合 EDR 产品的安全专家，以及安全云脑的大数据运营专家，共同打造的人工智能恶意文件检测引擎。该引擎利用深度学习技术对数亿维的原始特征进行分析和综合，结合安全专家的领域知识，最终挑选了数千维最有效的高维特征进行恶意文件的鉴定。相比基于病毒特征库的传统检测引擎，SAVE 的核心优势有：

- ✓ 基于人工智能技术，**拥有强大的泛化能力**，能够识别未知病毒或者已知病毒的新变种；
- ✓ 对**勒索病毒检测效果达到业界领先**，包括影响广泛的 WannaCry、BadRabbit 等病毒，而对非勒索病毒也有较好的检出效果；
- ✓ **云+设备+端联动**，依托于深信服安全云脑海量的安全数据，SAVE 能够持续进化，不断更新模型并提升检测能力，从而形成本地传统引擎、人工智能检测引擎和云端查杀引擎的完美结合。

2.3 检测原理



✓ **AI 在 SAVE 中的应用**：SAVE 引擎基于机器学习算法从海量样本集合中筛选出可用特征，并评估各个备选特征对各类恶意文件判别能力，筛选出最有效的特征集合。对于所选特征，我们选择合理的向量表示方法。通过统计学习中的多种降维方法，在保证信息损耗最小的同时，将所选特征集合的向量表示从数十万维压缩到数千维。最后我们通过集成学习算法组合多个分类器，构建模型和评分系统，以实现对已知和未知恶意文件的检测。 通过云端上亿级的黑白样本对检测模型进行持续的训练、改进、优化形成强大的模型，实现对未知威胁的精确识别。

✓ **基于 AI 的检测流程描述**：首先，SAVE 会从文件的头、节、资源和签名等多个部分提取信息，作为判别输入。对于提取出来的原始信息，SAVE 通过多种向量降维方法（词向量嵌入，主成分分析等）提取出信息量丰富、类间界限明显的向量化特征。基于特征向量，SAVE 使用多种分类算法（随机森林，神经网络，支持向量机等）进行鉴定。最后，综合评分系统整合各模型检测结果，综合判断文件是黑文件、白文件或者灰文件。对于无法在本地检测的文件，在符合相关法律法规的前提下，SAVE 引擎将上报信息到云端进行云查，由于云端部署了检测能力更强同时需要更多计算资源的检测引擎，可以对文件进行

更深入的分析，做出准确判断。同时，SAVE 引擎在云端（安全云脑）也会持续演进，包括特征提取、检测模型更新等，然后通过升级服务下发最新检测模型到终端或者设备上，以不断提升本地检测的能力。

第3章 SAVE 最佳配置

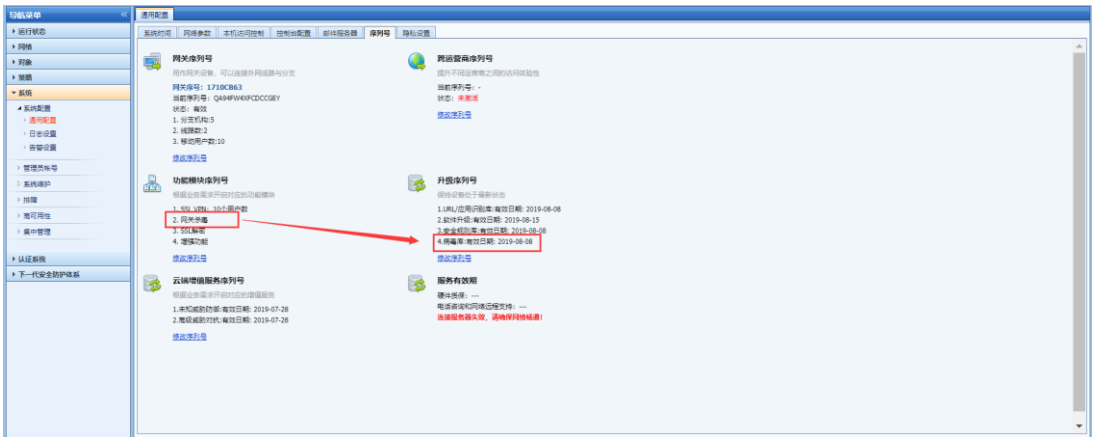
提供常规环境下，SAVE 杀毒功能的最佳实践配置，包括针对上网终端的病毒防护策略以及针对服务器主动上网的病毒防护策略。

3.1 环境准备

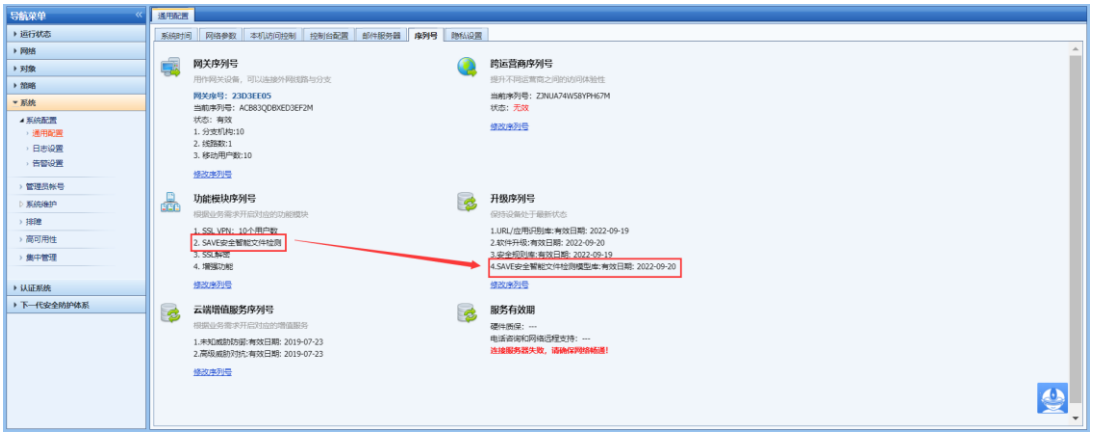
✓ AF 版本要求为 AF8.0.5 及以上；

✓ AF 要求开通杀毒模块及升级序列号（AF8.0.6 及以后版本，更名为“SAVE 安全智能文件检测”和“SAVE 安全智能文件检测模型库”），如下图：

AF8.0.5

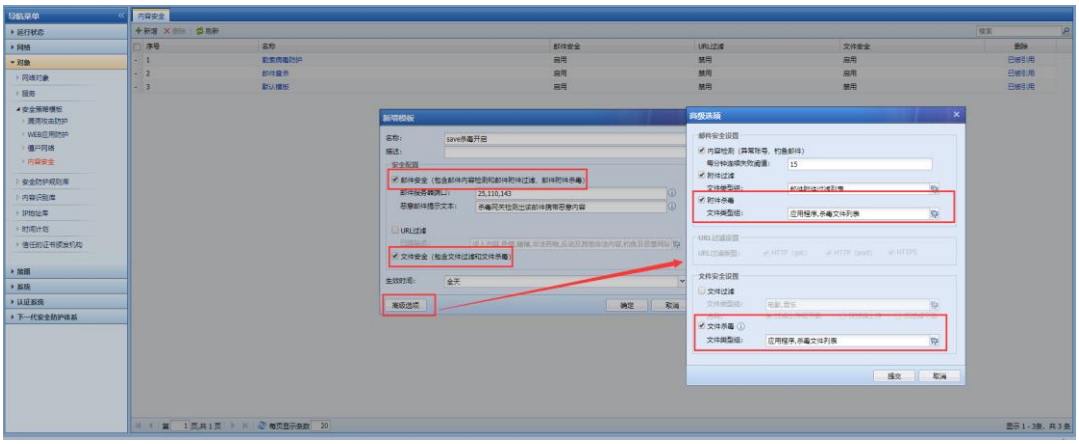


AF8.0.6

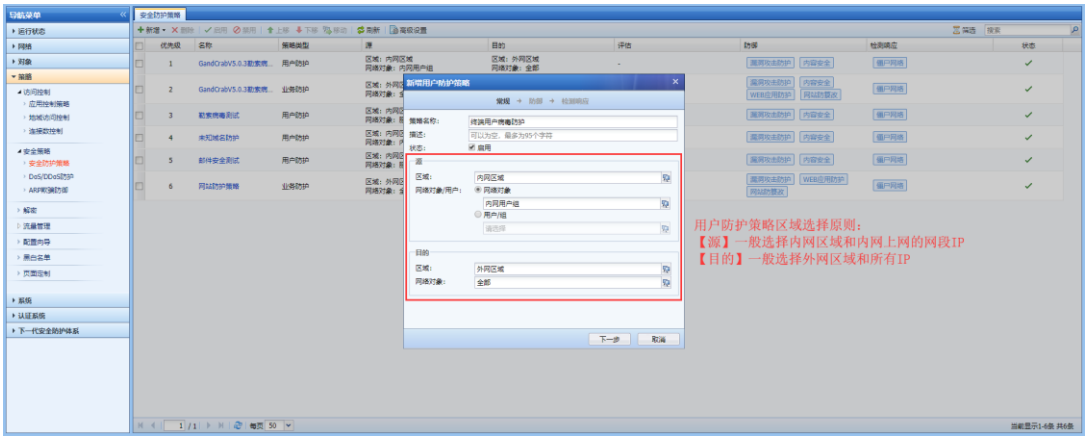


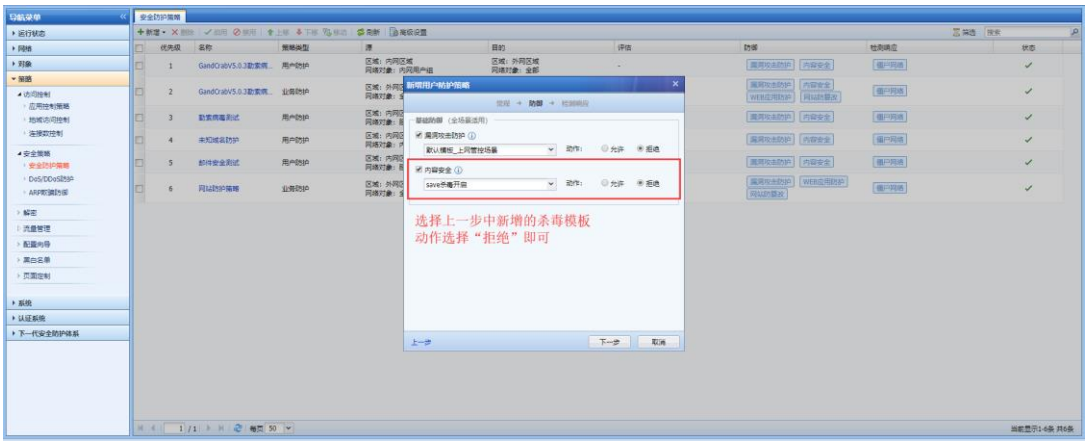
3.2 设备配置

✓ 完成【内容安全】模板定义，【对象】——【安全策略模板】——【内容安全】，要求针对“邮件安全”与“文件安全”均开启杀毒功能，“文件类型组”保持默认即可，其他功能项根据用户需求选择性开启，具体配置如下图所示：

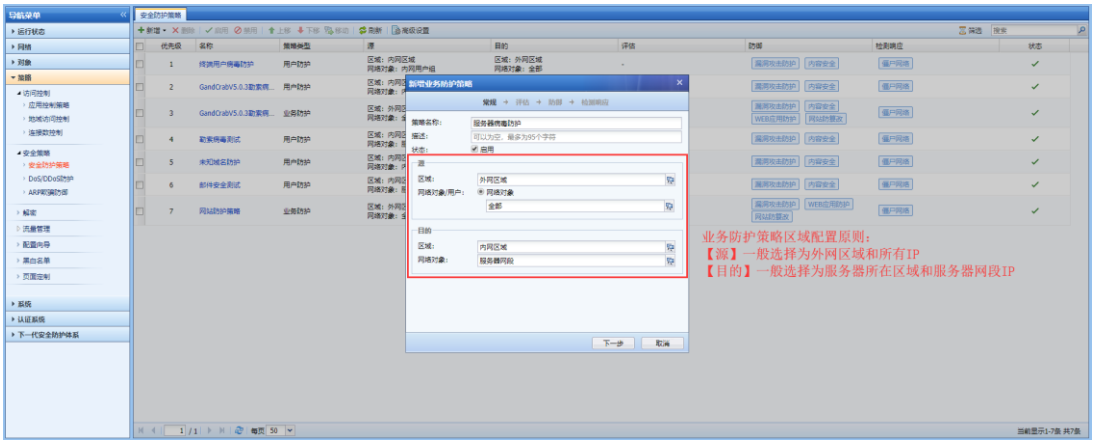


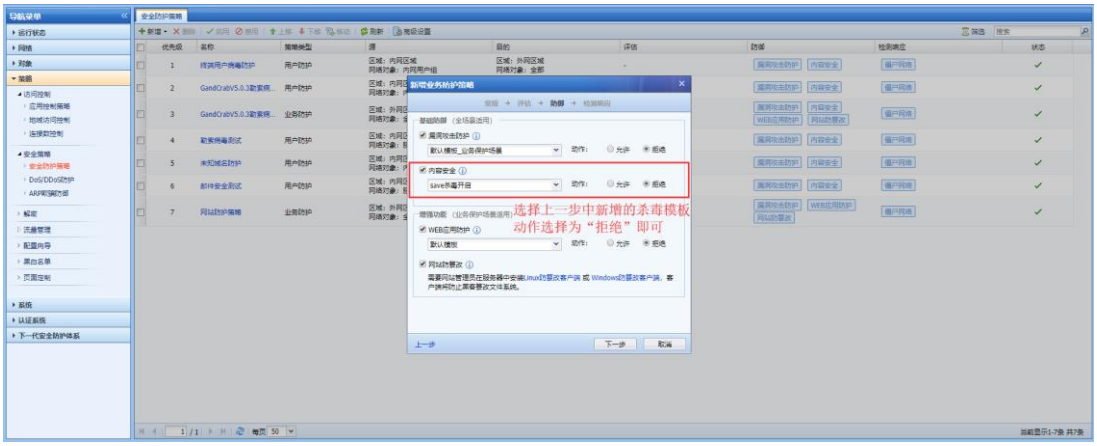
✓ 新增【用户防护策略】的防护策略，【策略】——【安全策略】——【安全防护策略】——新增【用户防护策略】。要求在“内容安全”功能中，选择上一步中新增的内容安全模板，动作选择“拒绝”。其它配置根据用户需求选择性开启，配置截图如下；





✓ 新增【业务防护策略】的防护策略，【策略】——【安全策略】——【安全防护策略】——新增【业务防护策略】。要求在“内容安全”功能中，选择上一步中新增的内容安全模板，动作选择“拒绝”。其它配置根据用户需求选择性开启，配置截图如下；



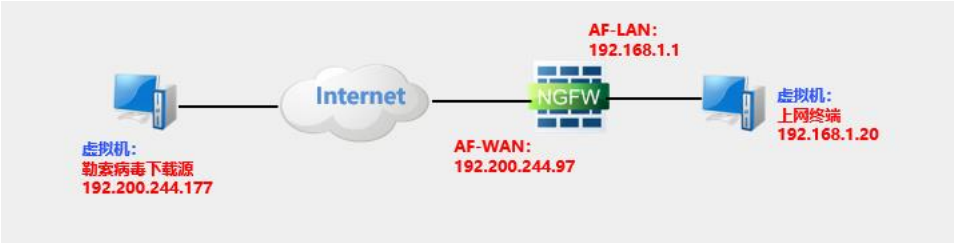


第4章 SAVE 效果展示

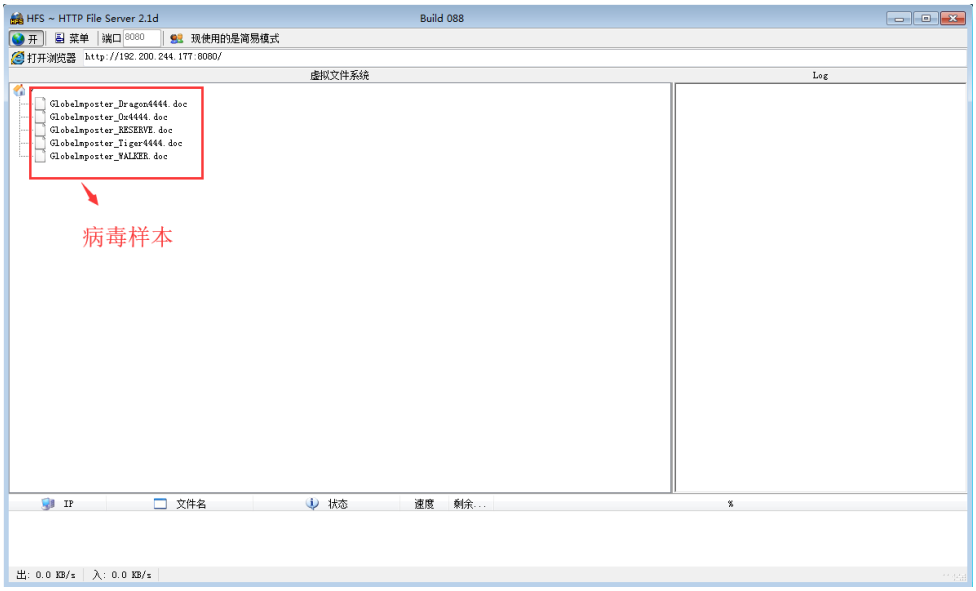
4.1 环境准备

该测试会通过真实的**最新的一批 globelmposter2.0/3.0** 勒索病毒样本（病毒样本已包含在工具包里）进行。所以**需要谨慎操作，一旦样本被执行，将会导致电脑文件加密，目前无法恢复**。目前把样本的.exe 后缀改为.doc，防止误执行。而后续的操作，有条件下，都需要在虚拟机里进行。

1）搭建拓扑，具体用户现场的拓扑可以根据条件自行搭建。本例拓扑如下：

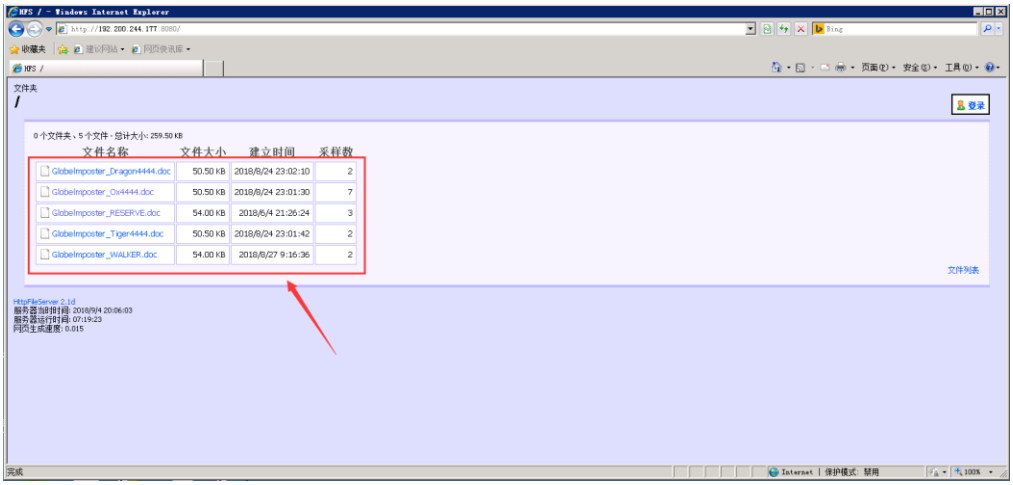


2）勒索病毒下载源的主机上，搭建一个 hfs（工具已包含在工具包里）的服务器，同时把病毒样本加载到界面，如下图：

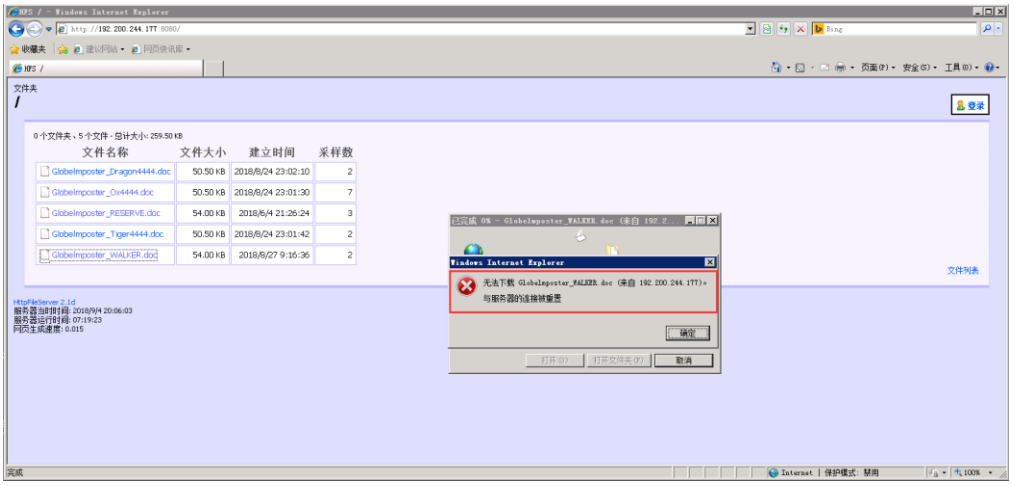


4.2 效果展示

1）上网终端访问带勒索病毒样本的网页，如下图：

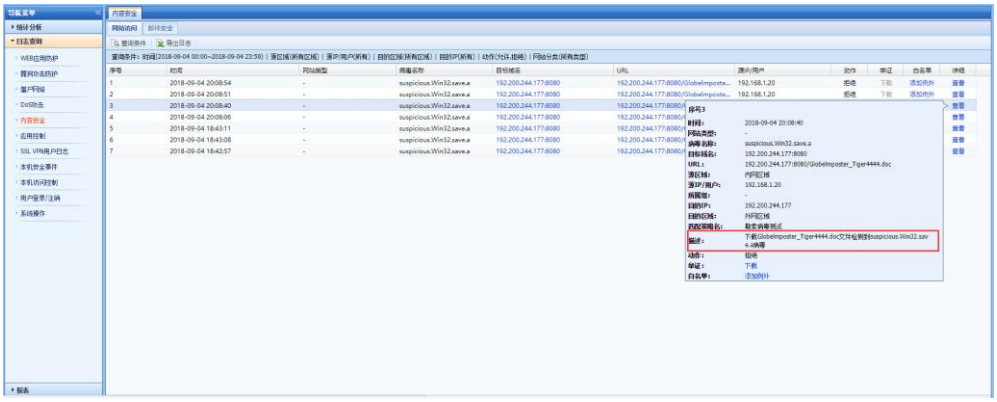


2）任意点击一个样本进行下载，返回如下效果，如下图：

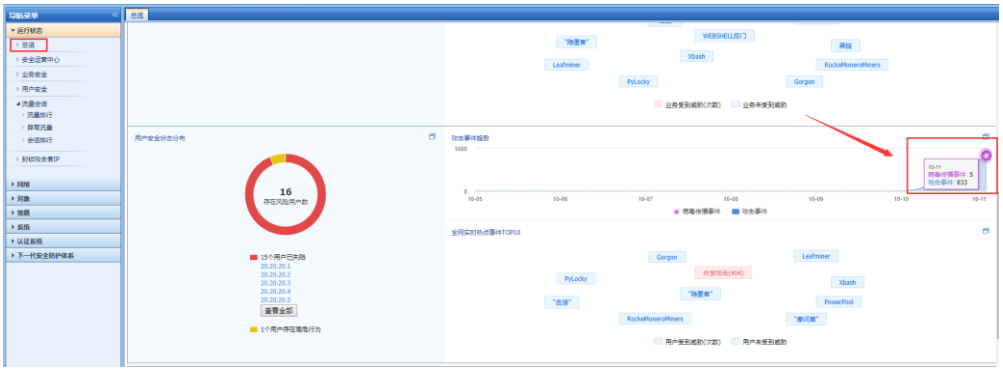


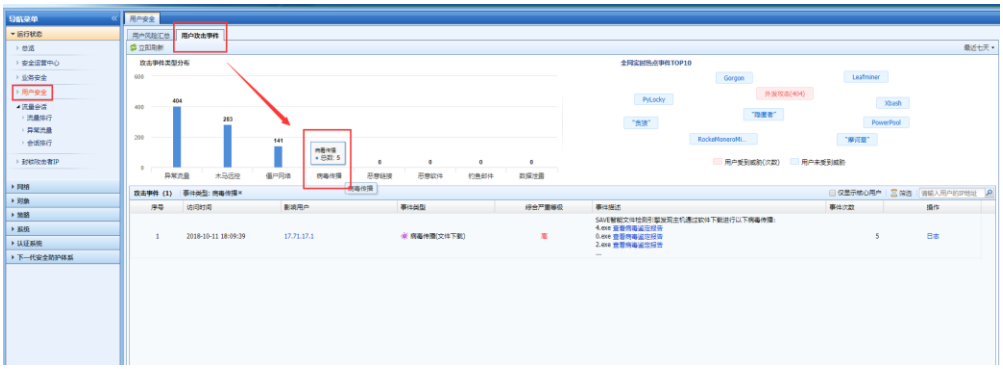
3）右击任意样本，在新页面打开，会重定向到一个检测告警页面，如下图：

4）查看 AF 的日志，【内置数据中心】——【内容安全】——【网站访问】，如下图：



5）AF8.0.6 及以后的版本，在控制台如下界面也可以看到 save 杀毒的效果图：





第5章　注意事项

1）目前版本的杀毒功能，AF8.0.5 不支持压缩文件，AF8.0.6 已支持部分压缩文件，包括.7z .rar .zip .gz .tar 这几类

2）目前版本的杀毒功能，默认不支持大于 2M 的文件（后台可调整，但建议不要超过 10M，一方面很少有病毒是大文件传播，另一方面文件太大会对设备内存损耗过高）；

3）目前版本的杀毒功能，不支持 http/ftp 上传方向，只支持下载方向；

4）测试时**注意安全、注意安全、注意安全**，不允许把病毒样本扩展名改为.exe。有条件的环境下，一律在虚拟机里进行操作

第6章　常见 Q&A

◆ **Q1：SAVE 和传统杀毒软件的比较**

传统技术有：MD5，脚本，虚拟执行，沙箱。从前到后在理论上的检测能力依次变好，但是性能依次下降，资源开销也依次变大。SAVE 的优势是有着接近 MD5 匹配的近线性的扫描速度。

同时，得益于机器学习的泛化能力，SAVE 能够识别未知病毒或者已知病毒的新变种。更进一步地，在资源开销方面，SAVE 只占用 200MB 以下的内存，比已知所有传统引擎都要小。

◆ Q2：为什么脚本引擎会比较慢，以及其相对于 MD5 检测技术的好处

MD5 的方案最大的问题在于一个 MD5 只能表征一个恶意文件，只要稍微进行一点点的修改就不能够检出，即泛化能力为零。为了解决这一问题，研究人员设计了一些脚本语言，通过允许恶意软件分析人员写一些识别脚本来检测病毒。这些脚本可能只是抓病毒中关键的某一部分的内容，因此如果恶意软件的变种没有修改这一关键部分的话就可以抓住这些变种。因此提升了一些的泛化能力。

具体来说，脚本引擎可能会使用的特征包括：1）某一个特定节的 MD5；2）某一小段字符串；3）信息熵值的计算等等。总的来说就是这一类引擎仅仅是一个类似正则匹配引擎的东西，具体特征还是得病毒分析师去写。如果某一类病毒具有某种特别固定的特征的话就可以一条规则匹配一类病毒，但是完全取决于规则写的好坏。

性能方面，MD5 只需要计算一次就可以了。相对的，规则引擎在计算 MD5 的基础上还会需要根据分析师写的规则去做一些正则匹配，熵值计算，每节各自的 MD5 之类的计算，因此效率较差。同时，和 MD5 只要计算一次然后在特征库里匹配不同，脚本引擎每多一条规则就会增加一些计算量，因此规则库越大性能越差。

◆ Q3：虚拟执行和沙箱的区别

虚拟执行一般是通过指令虚拟机来实现的，它是通过软件来模拟常用 CPU 指令、操作系统仿真环境来实现的，是一种轻量级的解决方案，被广泛用于杀毒软件的脱壳、病毒行为的提取等。目前沙箱主要是采用成熟虚拟化技术（Intel、AMD 提供的 CPU、IO 虚拟化驱动）承载的虚拟机来实现的，技术路线主要是虚拟机加上应用层、内核层的各种 hook 技术，从这个角度看，沙箱具有完整的 CPU 特性、可以无差别的运行各种操作系统，和物理硬件几乎没有差别。

综上来看，虚拟执行相对沙箱主要有如下缺点和不足：

✓ 能够模拟的 CPU 指令有限，MMX、3DNow 等指令无法很好的模拟；

- ✓ 多线程、多 CPU 支持不好；
- ✓ 操作系统兼容性差，不能模拟全部操作系统功能；
- ✓ 虽然虚拟执行有上述缺点和不足，但相比基于虚拟机技术的沙箱，具有轻量级、资源占用少、易于部署、速度快等优点。

◆ Q4：SAVE 和其它 AI 引擎的比较

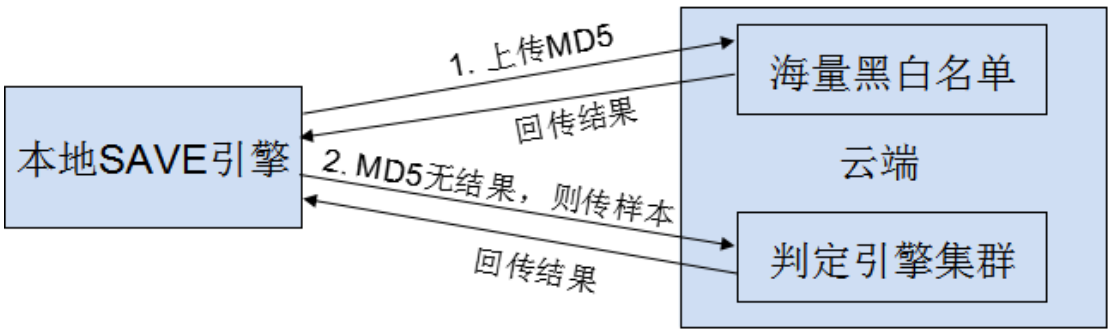
SAVE 引擎主要有几大优势。第一，优质的数据来源。通过安全云脑、EDR 和 AF 等产品持续汇聚分析热门数据，SAVE 引擎能够及时演进，从而提升检测能力，并覆盖最新的病毒。第二，创新的算法设计。SAVE 引擎结合深度学习等多种机器学习算法，使用集成学习充分利用各个算法的检测优势。第三，精细的特征工程。通过自研的特征提取算法和特征向量设计，我们的引擎能快速、准确捕捉文件的有效信息，因此我们可以用轻量化引擎快速识别病毒。第四，高出同行的勒索病毒检出率。基于对勒索病毒的深刻理解和深厚的算法积累，我们的引擎对于勒索病毒有很好的检测效果，实际测试结果表明，SAVE 引擎对勒索病毒的检出率达到业界领先水平。

◆ Q5：基于 AI 的杀毒引擎 vs 传统杀毒引擎对比

AI 引擎的主要优势有以下几点，特别是前两点。

传统引擎	AI 引擎
病毒专家人工提取病毒指纹、特征码，不仅成本高，而且有很大的滞后性，可能导致病毒出现很久了，传统杀毒厂商才能够更新病毒库	模型自动学习提取特征，无需人工参与；模型在云端不断进化，检测能力和自动化程度大幅提高
频繁的应急响应	AI 自带泛化能力，查杀未知病毒或新变种
不断膨胀的规则库	AI 强大表征能力，使用小巧的模型文件实现检测功能

◆ Q6、 SAVE 如何与云端（安全云脑）联动



◆ **Q7: 病毒类别与不同的传播途径和方式说明**

深信服公司对计算机病毒分类采用如下规则。

- ✓ Trojan：木马，是一种秘密潜伏的能够通过远程网络进行控制的恶意程序
- ✓ Backdoor: 后门，是指绕过安全控制而获取对程序或系统访问权的程序方法
- ✓ Worm：蠕虫，是一种通过网络进行自我传播的的恶意程序
- ✓ Virus: 感染型病毒，是具有自我复制能力感染其他正常文件的病毒
- ✓ Rootkit：内核劫持恶意程序，通过隐藏进程、文件等方式避免被使用者感知
- ✓ Bootkit：通过感染 MBR 的方式，实现绕过内核检查和启动隐藏自身
- ✓ Exploit：利用漏洞来进行攻击的恶意代码
- ✓ Hacktool：黑客工具，也许本身并不破坏你的计算机，但会被别人加以利用作为替身去破坏其它计算机
- ✓ PUP：Potentially unwanted Program，是指尽管使用者可能同意下载但其实并不想要的程序，包括间谍软件、广告软件和拨号器等
- ✓ RiskWare：风险软件，运行此类程序可能导致不确定风险
- ✓ AdWare：广告软件
- ✓ Suspicious：可疑软件
- ✓ Ransom: 勒索病毒