DigDog 产品说明书

目录

1. 产品概述

DigDog 是一款基于神经网络和内存取证的恶意软件检测系统。本产品可以在短时间内实现对目标机器内存文件的多角度、多层次的扫描，以获取恶意软件的运行痕迹，扫描结果以网页版报告的形式呈现给用户。该报告包括内存文件的基本信息、恶意进程的详细信息、动态链接库加载信息、注册表自启动项信息、网络活动信息、可疑域名信息、可能的恶意软件家族等信息模块，旨在帮助安全人员快速处理后应急响应环节中的前期工作，降低恶意行为分析的难度，从而达到快速评估的目的。

本产品的云平台地址为 http://digdog-report.cn/，平台主页包括产品的完整执行流程视频及相应描述，以供使用者参考。同时，使用者可以点击所有报告以查看历史报告。

2. 产品环境与结构目录

本产品在 Linux 系统环境下运行，并且支持所有常见的 Linux 发行版，其中对于 Ubuntu 16.04 TLS 版本具有最好的兼容性。同时本产品支持对 Windows7 和 Windows10 两大主流微软操作系统上恶意软件行为痕迹的分析与自动检测，对 x86 系统和 x64 系统都具有良好的扫描支持性。产品正常运行所需的编程环境为 Python2。

产品目录如下：

├──codes

├──csv_file

├──myModel

├──DigDog

├──goodware

├──malware

├──yara

├──MalProcessResult

├──install.sh

└──requirements.txt

DigDog 目录为本产品的主目录，用户可以在终端中运行其下的/APP/Controller/digdog.py 脚本以命令行的形式运行产品；也可以将配套的 DigDog 可执行文件拖动至主目录下的/APP/Controller 点击运行本产品。

codes 目录下为产品源代码，包括所有功能性的实现组件；goodware 和 malware 目录分别用于存放良性样本与恶性样本；csv_file 目录用于存放构建模型所需的 CSV 格式文件；myModel 目录用于存放训练成功的模型文件；yara 目录用于存放样本扫描所需的所有 yara 文件；MalProcessResult 目录用于存放导出的恶意进程文件；install.sh 脚本用于一键安装所需依赖。

3. 依赖安装

为确保本产品的正常运行，请手动安装如下依赖：

- VirtualBox 5.1+版本及其 SDK

除此之外，本产品所需的其他软件包及 Python 库依赖请用户以 root 权限运行产品目录下的 install.sh 脚本进行安装。

4. 用户模式

4.1 使用说明

分析人员可以使用本产品的用户模式来检测内存转储文件中的恶意软件行

为信息。具体操作流程可概括为以下三步：

1）进入产品主目录并运行其下的 install.sh 脚本以自动配置所需的软件包及库依赖；

2）将配套的 DigDog 可执行文件拖至主目录下的/DigDog/APP/Controller/目录；

3）双击可执行文件即可运行用户模式。具体操作参考详细功能说明中的用户模式部分。

4.2 配置说明

为保证用户的正常使用，开发团队已注册相关查询网站的账号并内置于产品配置中，如果用户希望将其修改为自己的账号，需要手动修改 codes 目录下的 DigDogConfig.py 配置文件。用户也可以点击产品主界面中的帮助->设置菜单项以打开配置文件并进行相关编辑活动。

- VIRUSTOTAL_KEY：该部分用于配置用户 VirusTotal 网站的账号信息。为了用户的正常使用，产品已内置开发者团队的 VirusTotal API Key。如果用户希望应用自己的 VirusTotal 账号进行相关查询，可以将该部分修改为自己账号对应的 API Key。

- DGA_ARCHIVE_USER/PASS：该部分用于配置用户的 DGArchive 网站的账号信息。为了用户的正常使用，产品已内置开发者团队的 DGArchive 网站的账号信息。如果用户希望应用自己的 DGArchive 账号进行相关查询，可以将该部分修改为自己账号对应的信息。

4.3 详细功能说明

用户模式主界面如下：

用户模式中，用户可以点击**扫描与报告**按钮以进入扫描报告模块。当用户完成其内部内存镜像导入、扫描模型选择及其他相关配置工作后，本产品对内存转储文件进行相关检测分析，最终结果以网页报告的形式展示给用户。该部分的核心为扫描与报告模块。

用户模式的扫描与报告模块引入训练成功的学习模型，对使用者提供的内存转储文件进行深度扫描。扫描结束后，构造恶意软件检测报告，供使用者阅读分析。模块主页面如下：

用户完成对**内存文件操作系统版本、内存文件路径、模型说明文件路径、是否启用 malfind 或 hollowfind 功能**四大参数的选择与键入后**点击运行按钮**，本产品将使用模型说明对应的模型文件，对用户提供的内存转储文件进行扫描。一次扫描时间在 10-15 分钟不等，取决于用户机器的处理能力及网络状况，扫描结果以网页报告的形式呈现。用户可以在下方的 Logging 框中看到程序的提示性输出，并可以在浏览器中输入 https://digdog-report.cn/archives 来查看历史报告结果。

5. 开发者模式

5.1 使用说明

除用户模式外，本产品还提供开发者模式供使用者完成模型扩展。运行产品主目录下的 install.sh 脚本并完成依赖安装后，需要用户手动下载 VirtualBox 并配置 Windows7 或 Windows10 虚拟环境。为适配使用者机器，用户需按照下述配置说明完成相关参数及路径的选择与键入。完成配置后根据详细功能说明中的提示点击相关按钮即可运行开发者模式。

5.2 配置说明

为保证开发者模式的正常运行，用户需根据机器情况修改 codes 目录下的 DigDogConfig.py 文件中的以下配置。

- characteristics：该部分用于配置特征向量模块。本产品提供二十二个检测特征，用户可以通过注释符进行筛选。
- profiles：该部分用于配置所用虚拟机的操作系统版本。
- dbconfig，hostname，port：该部分用于配置 MongoDB 数据库的环境信息，包括数据库名称、本地主机名、监听端口号等。

- vm：该部分用于配置虚拟机的名称以及用户名、密码，请注意相关参数需与用户 VirtualBox 中的参数保持一致，区分大小写。
- SAVE_PATH：该部分用于选择临时文件存储位置，建议用户选择家目录。
- HOME_PATH：该部分用于配置产品文件夹路径，请使用绝对路径。
- CLASSIFIERS：该部分用于深度学习分类器相关参数的设置。

5.3 详细功能说明

开发者模式中，使用者需要顺次点击**数据提取**及**模型创建**两个按钮，并完成其内部各阶段的操作。模型创建部分结束后，使用者可得到扫描模型，该模型随后用于用户模式中。

5.3.1 数据提取模块

开发者模式中的数据提取模块实现将恶意样本存储到数据库中、挂载至虚拟机运行、GroundTruth 相关操作、提取特征向量数据、导出数据为 CSV 格式等功能，其主页面如下：

1）样本导入

该部分主页面如下图所示。

用户完成对**恶性（良性）样本路径、数据库名称、恶性（良性）标记**这三大参数的选择与键入后**点击运行按钮**，即可导入样本至数据库中。用户可在下方 Logging 框中看到产品的提示性输出。

2）内存转储

该部分主页面如下图所示。

用户完成对**内存转储文件路径和数据库名称**两大参数的选择与键入后**点击运行按钮**，即可将存储在数据库中的样本文件装载到虚拟机中运行，需要注意的是这里的数据库名称应与第一步输入保持一致。等待一定时间后，本产品将运行后的内存文件导出到储存路径下。用户可以在下方的 Logging 框中看到产品的提示性输出。

3） 创建 GroundTruth

该部分主页面如下图所示。

用户完成对 yara 文件夹路径和数据库名称两大参数的选择与键入后点击运行按钮，即可根据每个样本的特定 yara 匹配规则对导出的内存转储文件进行扫描，并得出包含 GroundTruth 数据的 json 文件。用户可以在下方的 Logging 框中看到产品的提示性输出。

4）添加 GroundTruth

该部分主页面如下图所示。

用户完成对 GroundTruth 文件路径和数据库名称两大参数的选择与键入后点击运行按钮，即可将第三步中获取的 GroundTruth 文件保存至 MongoDB 数据库中。用户可以在下方的 Logging 框中看到产品的提示性输出。

5）特征提取

该部分主页面如下图所示。

用户完成对**数据库名称**参数的键入后**点击运行按钮**，本产品将根据前四步中的内存转储文件、GroundTruth 文件以及配置文件中的特征选择情况来提取特征向量，并将提取结果保存至数据库中。用户可以在下方的 Logging 框中看到产品的提示性输出。

6）数据导出

该部分主页面如下图所示。

用户完成对 **CSV 导出路径和数据库名称**这两大参数的选择与键入后**点击运**

行按钮，本产品将参照上一步中特征向量的提取结果将特征数据保存在 CSV 导出路径下的 result.csv 文件中，供后续模型构建使用。用户可以在下方的 Logging 框中看到产品的提示性输出。

5.3.2 模型创建模块

开发者模式中的模型创建模块利用深度学习算法，将数据提取功能得出的 CSV 文件构造为模型文件包。其主页面如下：

用户完成对**模型构建算法、CSV 数据文件路径、模型输出路径、模型输出名称**四大参数的选择与键入后**点击运行按钮**，本产品将结合 CSV 文件中的特征数据，依据用户指定的学习算法生成对应的模型文件，输出 json 格式的模型配置文件以及 model 格式的模型文件。用户可以在下方的 Logging 框中看到产品的提示性输出。

6. 注意事项

1. 若在虚拟机环境中运行本产品（典型如 VMWare），请开启虚拟机的 Intel VT-x/EPT 或 AMD-V/RVI 选项以及虚拟化 CPU 性能计数器选项，以此

来保证虚拟机中 VirtualBox 对 64 位操作系统的良好支持。此外请为虚拟机分配不小于 4G 的内存以获得更好的性能。

2. 配置文件 DigDogConfig.py 中的 SAVE_PATH 路径建议选择家目录或其他空间不小于 20G 的目录（也可以使用外接设备），以此来得更好的性能。

3. 本产品中不包含额外的虚拟机文件，请用户根据需求自行安装 Windows 系列虚拟机作为 VirtualBox 中的沙箱虚拟机环境。