OSCP样本计算科研协作平台及样本数据集介绍

安天技术委员会

执行体治理赋能与大模型辅助

北向守望

01 初心

02 方向与目标

03 建设与规划

01 初心

产业界与高校、研究机构在恶意代码样本方面未能充分合作

• 执法监管方

- 严格管控恶意代码样本

• 产业界

- 反病毒业务的核心资产
  - 病毒库更新、引擎演进
- 样本也是安全风险
  - 污染IT环境、损害公司声誉

• 高校、科研机构

- 科研能力强
- 缺少样本工程化基础

产业界与高校、研究机构如何能更充分地合作?

我们尝试推出样本计算科研协作平台-OSCP

OSCP设计原则

• 开放

- 高校
- 科研机构
- 独立研究者

• 协作

- 研究者提供自己程序
- OSCP提供其它需要的资源
  - 大基数分类样本集
  - 算力
    - CPU/GPU
  - 算法

OSCP设计原则

• 安全合规

- 《计算机病毒防治管理办法》
- 防范样本泄漏、传播
  - 相对封闭的无网络环境
- 防范意外运行恶意代码，污染计算环境
  - 封装、去活、异构
- 不同研究者的环境之间相互隔离

• 规范易用

- 通用计算环境，研究者可以自建开发环境。
- 快速重构
- 护容、迁移。

OSCP样本集构成

• 恶意代码样本

- 大基数
  - 百万级
- 详细分类信息
  - 文件格式、操作系统、是否加壳、恶意代码类型、家族
  - 关联病毒知识百科
- 高质量
  - 人工挑选、核对

• 干扰样本

- 易被误报的样本
- 常见应用
- 干扰集大于样本集

• 多子集

- 功能目标
- 研究/对比测试

OSCP恶意代码样本概览

• 文件类型分布

BinExecute	5337816	87.6594%
Script	418313	6.8697%
Text	143478	2.3562%
Archive	142842	2.3458%
Document	42109	0.6915%
SoftData	2654	0.0436%
Other	1890	0.0310%
Media	119	0.0020%
Picture	46	0.0008%

• 操作系统分布

Windows	5257329	86.5878%
其他(跨平台)	737547	12.1473%
Linux	49109	0.8088%
Android	26950	0.4439%
Apple	737	0.0121%

• 是否带壳

无壳	4462594	72.63%
加壳	1681407	27.37%

02 方向与目标

OSCP预期的研究方向

• 代码分析、识别

- ➢ 恶意代码分析、识别
- ➢ 同源分析
- ➢ 执行体信誉度评估
- ➢ 工具类, 比如脱壳

• AI

- ➢ 分类
- ➢ 聚类/同源
- ➢ 信誉度评估
- ➢ 行为、意图识别
- ➢ 分析报告

OSCP预期目标

• 基准化

- ➢ 标准样本集、数据集
- ➢ 接口
- ➢ 算法
- ➢ AI模型

• 社区化

- ➢ 相互借鉴，相互碰撞，相互促进
- ➢ 共同建设

03 建设与规划

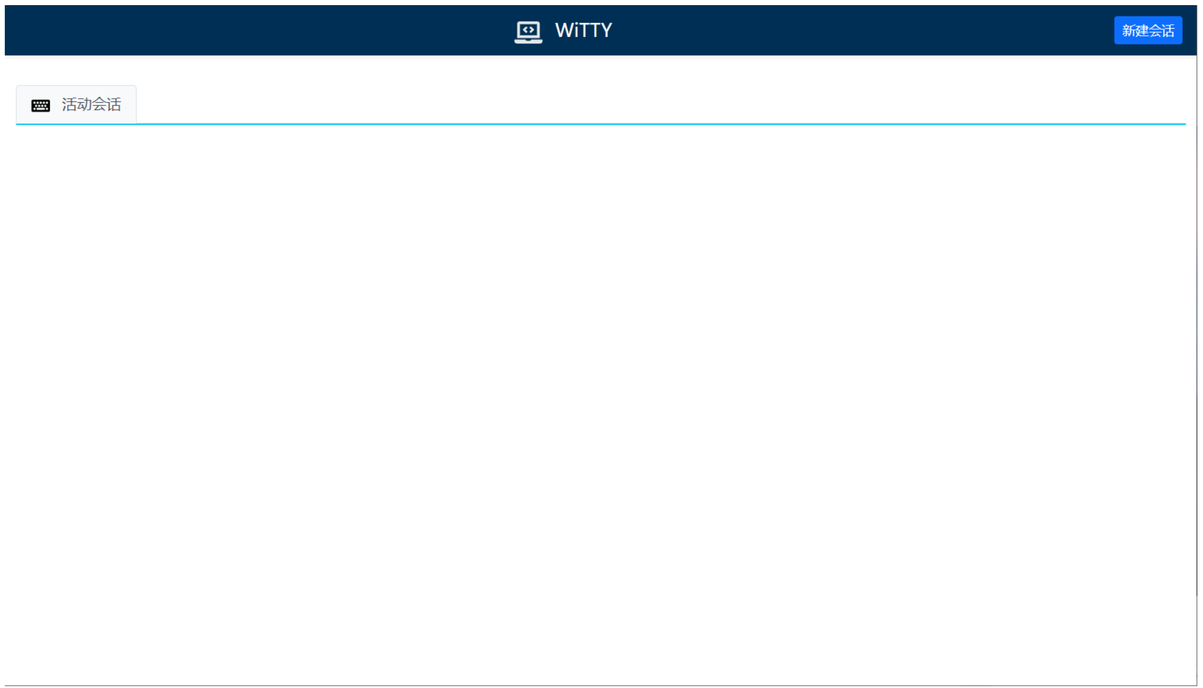
OSCP建设现状

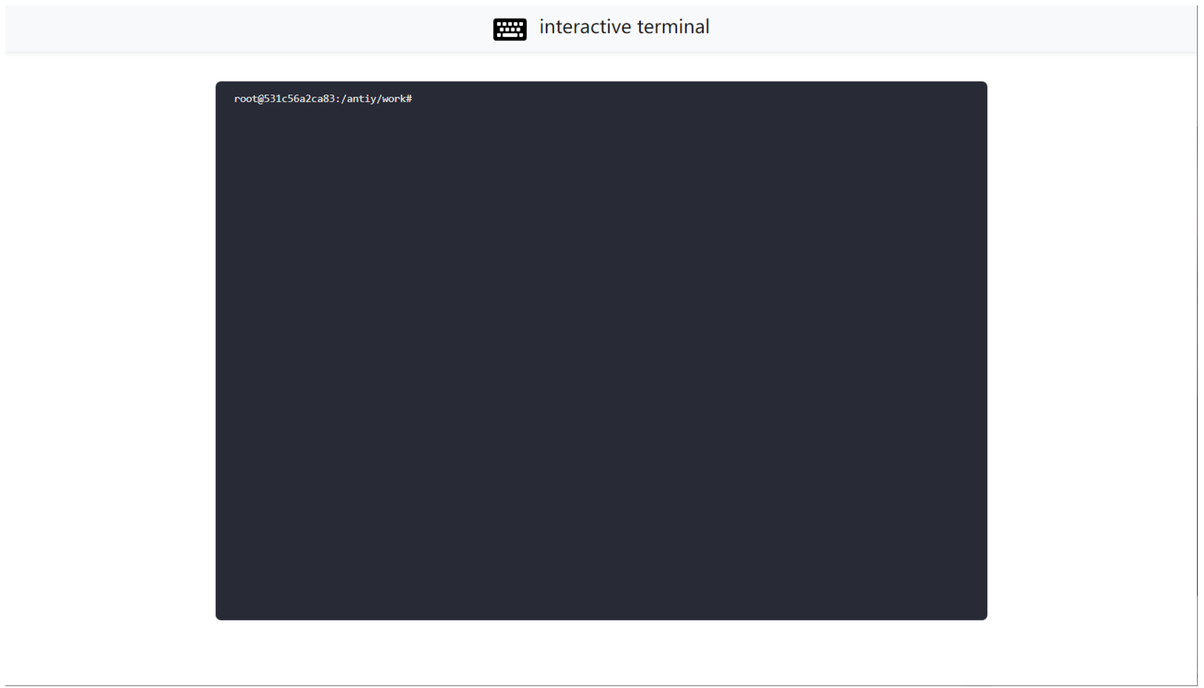
安天开放样本计算环境
OSCP
登录
邮件地址 *
密码 *
记住我
登录
忘记密码?
还没帐号? 申请一个
Copyright © 安天科技集团 2023 .

OSCP建设现状

安天开放样本计算环境
主页
环境管理
交流活动
文档
技术支持
计算环境
transformers-pytorch-cpu:latest
running/Up 2 days
镜像、状态
启动
停止
删除
刷新
终端窗口
用户文件
上传文件
创建目录
用户程序管理
/

OSCP建设现状


WiTTY
新建会话
活动会话


interactive terminal
root@531c56a2ca83:/antiy/work#

OSCP建设规划

• 持续样本集建设

- ➢ 增量迭代
- ➢ 千万级样本集
- ➢ 丰富子集

• 形成基础算法

- ➢ 脱壳
- ➢ 格式解析
- ➢ 反汇编
- ➢ 函数等分析、识别
- ➢ 大模型预数据处理

• 完善计算环境

- ➢ 多样化
- ➢ 基础算法服务

• 社区化

- ➢ 博客
- ➢ 开源代码库
- ➢ 活动

?

意见建议

让我们一起共建OSCP

安天冬训营 wtc.antiy.cn