



Knowledge is all you need

安全知识图谱 & 安全大模型

熊春霖

联通（广东）产业互联网有限公司 AI安全专家

REEBUF

个人介绍

- 浙江大学网络安全安全博士，师从 Yan Chen (IEEE会士 & 美国西北大学终身教授 & 国家千人计划)
- 作为主要成员参与美国国防部高级研究计划局透明计算项目、国家自然科学基金重点项目等国家级课题(APT+AI)
- 曾任杭州奇盾信息技术有限公司（终端、云安全创业公司）联合创始人/CTO
- 深信服XDR 首席算法专家/技术规划专家，负责深信服战略产品部门的技术规划和前沿探索 (AI+)
- 中国科学院深圳先进技术研究院、深信服联合培养 博士后
- 在CCS、Usenix Security、TDSC等世界顶级会议期刊发表论文9篇，专利10+项，主要研究终端安全

01 Knowledge
知识图谱

02 is all you need
大语言模型

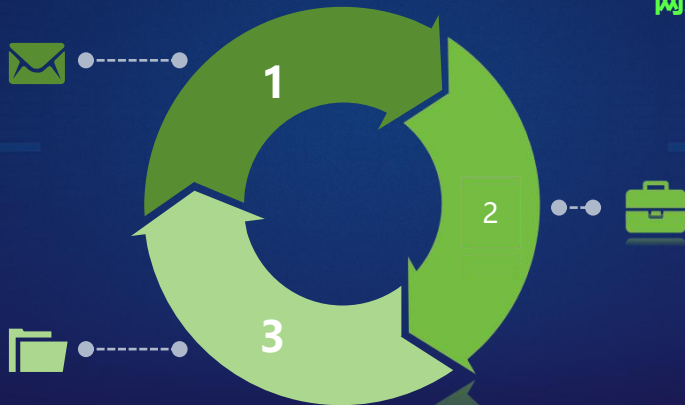
当前企业安全运营的困境

检测效果

检出率、误报率。
提升检测效果，一般会降低告警时效性和增加成本

时效性

攻击越早被发现和处置，造成的损失就越小。
提升时效性，一般会降低检测效果，提升运营成本。



网络安全人才缺口高达327万

运营成本

当前安全运营主要依赖专家，
现有工具很难满足需求。
运营成本降低，一般会同时影响检测效果和时效性。



威胁研判需要什么? 1/3

- 1、echo labkGz&& COMMAND &&echo ILSFYJ: 常见黑客工具特征, 通过随机字符串定位命令返回内容;
- 2、cmd /c "cd /d "C:\"&cd /d "tmp" apache等web server执行命令的特征;
- 3、ipconfig是微软操作系统的电脑上用来控制网络连线的一个命令行工具, 黑客常用它来收集网络信息。

攻击特征

```
sh -c sudo -u#-1  
sudo -u#4294967295 id -u
```

CVE-2019-14287





威胁研判需要什么？ 2/3

- 1、AppData\Local\Temp\7ZipSfx.000\目录下的可执行文件为Sfx（Self-eXtracting，自解压文件）；攻击者常使用自解压文件来隐藏真正的payload和攻击意图；
- 2、该命令行通过WMIC将OInstallLite加入Windows Defender的白名单；攻击者常使用这种方法让恶意文件免杀；

合法应用行为

KMS





威胁研判需要什么？ 3/3

- 1、PowerShell常被攻击者用于下载、执行无文件攻击；
- 2、DownloadFile为PowerShell下载文件的功能；
- 3、temp目录常被用于存储恶意文件。

威胁情报

officecdn.microsoft.com





威胁研判需要什么？

传统方法：

规则引擎

黑白名单

威胁情报

异常分析

AI？

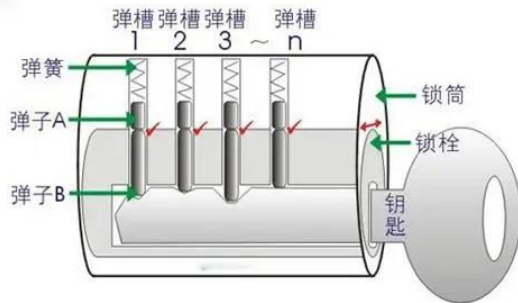
专家需要综合分析，而引擎常常各自为营。
因此自动检测引擎效果不佳





威胁研判需要什么？

安全相关知识十分重要，但碎片且繁多。



弹子锁的组成构件：一般由锁芯、弹子、弹簧、锁舌、钥匙等组成

ATT&CK Matrix for Enterprise

layout: side show sub-techniques hide sub-techniques

Reconnaissance 10 techniques	Resource Development 8 techniques	Initial Access 10 techniques	Execution 14 techniques	Persistence 20 techniques	Privilege Escalation 14 techniques	Defense Evasion 43 techniques	Credential Access 17 techniques	Discovery 32 techniques	Lateral Movement 9 techniques	Collection 17 techniques	Command and Control 17 techniques	Exfiltration 9 techniques
Active Scanning (3) Gather Victim Host Information (4) Gather Victim Identity Information (3) Gather Victim Org Information (4) Phishing for Information (4) Search Open Technical Databases (3) Search Open Websites/Domains (3) Search Victim-Owned Websites	Acquire Access Acquire Infrastructure (6) Compromise Accounts (3) Compromise Infrastructure (7) Develop Capabilities (6) Establish Accounts (3) Obtain Capabilities (6) Stage Capabilities (6)	Content Injection Drive-by Compromise Exploit Public-Facing Application External Remote Services Hardware Additions Phishing (4) Replication Through Removable Media Supply Chain Compromise (2) Trusted Relationship Valid Accounts (4)	Cloud Administration Command Command and Scripting Interpreter (6) Container Administration Command Deploy Container Exploitation for Client Execution Inter-Process Communication (3) Native API Scheduled Task/Job (3) Serverless Execution Shared Modules Software Deployment Tools System Services (2) User Execution (3) Windows Management Instrumentation	Account Manipulation (6) BITS Jobs Boot or Logon Autostart Execution (14) Account Manipulation (6) Boot or Logon Initialization Scripts (3) Browser Extensions Compromise Client Software Binary Create Account (3) Create or Modify System Process (4) Event Triggered Execution (14) External Remote Services Hijack Execution Flow (12) Implant Internal Image Process Injection (12) Scheduled Task/Job (3)	Abuse Elevation Control Mechanism (3) Access Token Manipulation (3) Access Token Manipulation (3) Account Manipulation (6) Boot or Logon Autostart Execution (14) Boot or Logon Initialization Scripts (3) Create or Modify System Process (4) Domain Policy Modification (2) Event Triggered Execution (14) Hijack Execution Flow (12) Process Injection (12) Scheduled Task/Job (3)	Abuse Elevation Control Mechanism (3) Access Token Manipulation (3) BITS Jobs Build Image on Host Debugger Evasion Decfuscate/Decode Files or Information Deploy Container Direct Volume Access Domain Policy Modification (2) Execution Guardrails (1) Exploitation for Defense Evasion File and Directory Permissions Modification (2) Hide Artifacts (11) Impair Defenses (11) Impersonation Indicator Removal (6) Indirect Command Execution	Adversary-in-the-Middle (1) Brute Force (4) Credentials from Password Stores (6) Exploitation for Credential Access Forced Authentication Forge Web Credentials (2) Input Capture (4) Modify Authentication Process (4) Multi-Factor Authentication Interception Multi-Factor Authentication Request Generation Network Sniffing OS Credential Dumping (4) Steal Application Access Token	Account Discovery (4) Application Window Discovery Browser Information Discovery Cloud Infrastructure Discovery Cloud Service Dashboard Cloud Service Discovery Cloud Storage Object Discovery Container and Resource Discovery Debugger Evasion Device Driver Discovery Domain Trust Discovery File and Directory Discovery Group Policy Discovery Log Enumeration Network Service Discovery Network Share Discovery	Exploitation of Remote Services Internal Spearphishing Lateral Tool Transfer Remote Service Session Hijacking (2) Remote Services (4) Replication Through Removable Media Software Deployment Tools Taint Shared Content Use Alternate Authentication Material (4)	Adversary-in-the-Middle (3) Archive Collected Data (3) Audio Capture Automated Collection Browser Session Hijacking Clipboard Data Data from Cloud Storage Data from Configuration Repository (2) Data from Information Repositories (3) Data from Local System Data from Network Shared Drive Data from Removable Media Data Staged (3) Email Collection (3) Proxy (4)	Application Layer Protocol (4) Communication Through Removable Media Content Injection Data Encoding (2) Data Obfuscation (2) Dynamic Resolution (3) Encrypted Channel (2) Fallback Channels Ingress Tool Transfer Multi-Stage Channels Non-Application Layer Protocol Non-Standard Port Protocol Tunneling	Automated Exfiltration (1) Data Transfer Size Limits Exfiltration Over Alternative Protocol (2) Exfiltration Over C2 Channel Exfiltration Over Other Network Medium (1) Exfiltration Over Physical Medium (1) Exfiltration Over Web Service (4) Scheduled Transfer Transfer Data to Cloud Account

威胁研判需要什么？

FREEBUF 知识大陆

广场 帮会 WIKI百科 帮助中心

请输入搜索关键词

注册 / 登录

汇聚专业安全知识 打造行业百科全书

部落拥有超高声望的部落领主、系统的营地知识结构，人人都是内容共建者

热门部落

白帽新手入门

网络安全入门，从入门到精通，汇集通用...

热门营地

Web渗透

内网渗透

防御技术

信息收集

网络对抗

语言框架

领主: AIDb 共建人数: 82 村民: 779 高帮会: 9

渗透测试工具

部落收集网络安全工作、学习中常见的各类工具...

热门营地

外网信息收集

初始访问

权限获取

资源开发

各种cms、OA、...

权限维持

领主: only... 共建人数: 37 村民: 524 高帮会: 3

网安产业全景图

网络安全数字化转型的起步，网络安全领域...

热门营地

网络安全基础

检测

持续改进

防护

响应

业务场景

领主: 流苏 共建人数: 30 村民: 273 高帮会: 14

安全名词百科

汇集网络安全、信息安全、数据安全、CISSP...

热门营地

密码学

数据安全

社会安全

网络安全

领主: 流苏 共建人数: 30 村民: 273 高帮会: 14

CIS网络安全创新大会官方...

CIS 网络安全创新大会(Cyber Security Innov...

热门营地

FIT2019大会合集

CIS2019大会合集

历年CIS大会官网...

CIS2020大会合集

领主: 流苏 共建人数: 30 村民: 273 高帮会: 14

安全标准部落

汇集最全最新的行业法律法规、国家标准、国...

热门营地

个人信息保护

风险管理

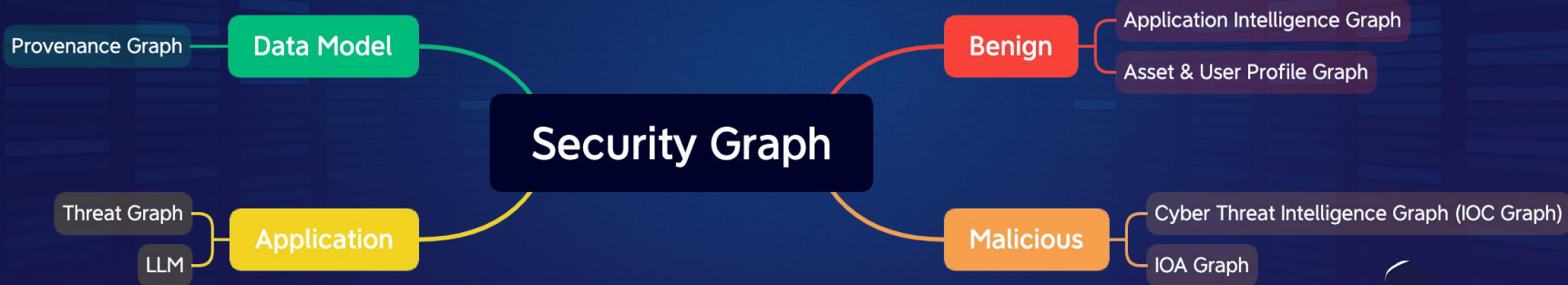
等级保护

发基标准

领主: 流苏 共建人数: 30 村民: 273 高帮会: 14

link

安全图谱



Provenance Graph

Chronicle Security: Unified Data Model

Chronicle Security > Documentation > Reference

该内容对您有帮助吗? 0 0

Unified Data Model field list

Send feedback

This document provides a list of fields available in the Unified Data Model schema. When specifying a field, use the following format: <prefix>.<field_name1>.<field_name2>.<...>.<field_nameN>=<value>

When writing rules for Detect Engine, use the <prefix> pattern "Su" for Event fields and "Se" for Entity fields. For example:

- Su.metadata.event_type
- Su.network.dhcp.opcode
- Su.principal.user.location.city
- Se.graph.entity.hostname
- Se.graph.metadata.product_name

When writing configuration-based normalizer (CBN) parsers, use the <prefix> pattern "event.idm.read_only_udm" for UDM Event fields and "event.idm.graph" for UDM Entity fields. For example:

- event.idm.read_only_udm.metadata.event_type
- event.idm.read_only_udm.network.dhcp.opcode
- event.idm.read_only_udm.principal.user.location.city
- event.idm.graph.entity.user_display_name
- event.idm.graph.entity.asset.hostname

Please Note: Field name and field type values can look similar. This document uses style conventions to help you identify the differences:

- Field type values use CamelCase characters. For example, Platform and EventType.
- Field name values use lowercase characters. For example, platform and event_type.

本页内容

UDM Entity data model

Entity

EntityMetadata

Relation

Entity enumerated types

EntityMetadata.EntityType

EntityMetadata.SourceType

Relation.Directionality

Relation.Relationship

UDM Event data model

Event top level types

Extensions

Metadata

Network

Noun

SecurityResult

Event subtypes

Artifact

Asset

Attribute

Authentication

Certificate

Cloud

Dhcp

Dhcp.Option

Dns

Dns.Question

Dns.ResourceRecord

Domain

Email

File

Ftp

Group

DARPA TC: Common Data Model

```
/*
 * This software delivered to the Government with unlimited rights
 * pursuant to contract FA8750-C-15-7559.
 *
 * ===== TRANSPARENT COMPUTING (TC) COMMON DATA MODEL (CDM) =====
 *
 * The CDM is a property graph (vertices and edges with properties)
 * that has additional typing of the vertices and edges to match the
 * TC domain. All vertex and edge records are atomic and immutable.
 *
 * The schema is defined using Avro's IDL specification language (see
 * http://avro.apache.org/docs/1.8.0/idl.html). The schema is
 * independent of the language bindings used to operate on it. IDL
 * makes it easy and simple to represent the schema. Tools exist to
 * map the IDL to a verbose JSON representation (avsc) as well as to
 * compiled language-specific objects. Optional fields are marked
 * using the notation union {null, <type>} <fieldName> = null
 */
```


Security Ontology Graph (StrikeReady)

SECURITY AUTOMATION & COUNTERMEASURES

NO PLAYBOOK, NO RUNBOOK, AUTOMATION REIMAGINED

CARA assists with real-time reasoning and response using institutional knowledge and practical experiences.

Powered by Security Ontology Graph, CARA helps you skip the management and overhead expense of flowchart-based solutions.

I need help with this URL <http://upgradesrv.890m.com/back/2019/index.php>

I can take the following actions for you:

Check Reputation	99%
Safely Browse URL in Sandbox	98%
Get Linked Files	88%
Generate Snort Signature	65%
Generate Suricata Rule	42%

CARA

CYBER AWARENESS AND RESPONSE

MEET CARA

CARA, an Intelligent System, and the industry's first digital cybersecurity analyst, learns in-real-time from the institutional knowledge and practical experiences of defenders around the world. It then helps analyze, reason, guide and resolve.

Security intelligence Graph (Recorded Furture)

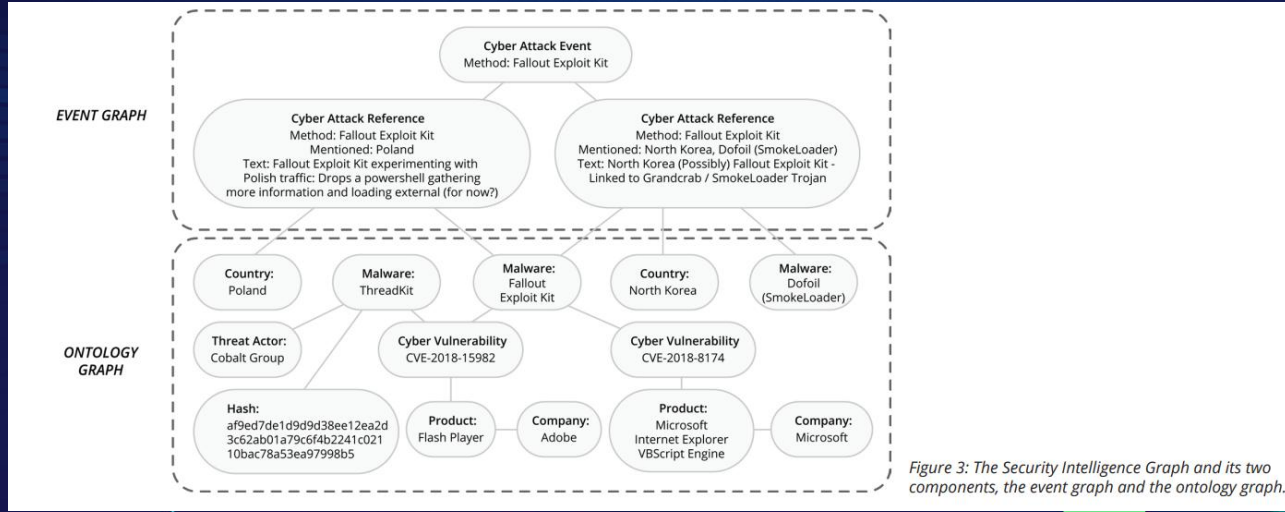


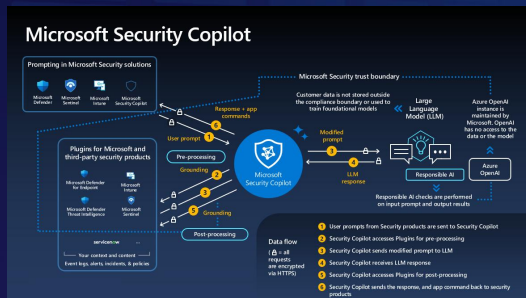
Figure 3: The Security Intelligence Graph and its two components, the event graph and the ontology graph.

01 Knowledge
知识图谱

02 is all you need
大语言模型

垂直领域大模型的一般构造方法

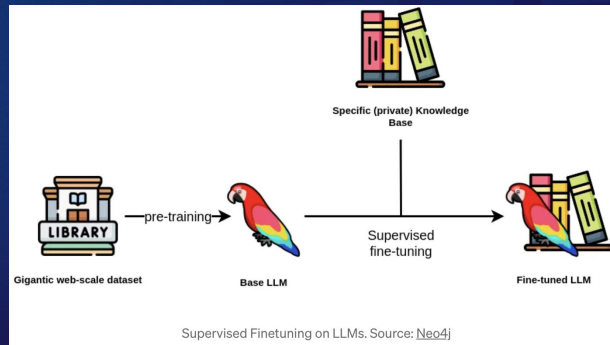
智能体



← 作为智能决策体，负责代替人类作出关键性决策。

通过微调，赋予大模型新的能力；给定一个input，期望获得预期output。

许愿机





垂直领域大模型的一般构造方法

- (1) **端到端训练**：使用通用数据和垂直领域数据混合，从零训练，**训练成本极大**；
- (2) **全参数再训练**：在一个通用模型的基础上做二次预训练，可以调整任何模型参数，**训练成本较大**；
- (3) **有监督微调**（SFT，部分参数微调）：在一个通用模型的基础上做有监督微调，这是开源社区最普遍的做法，方法包括prompt微调、Lora等，可以快速出效果，但可能造成灾难性遗忘，且上限有限，**训练成本较低**；
- (4) **零样本学习**（zero-shot Learning）：以上三种方法也都是遵循传统NLP预训练模型的思路，但大模型最厉害的点在于其对多个下游任务都可以通过一个预训练模型解决，对于新任务可以通过少样本学习（few-shot learning）甚至于零样本学习（zero-shot learning）的方式解决；需要有一个足够好的大模型，prompt工程以及知识库；**成本不高，见效快**；
- (5) **向量数据库**：预置常见问题和对应答案/行为，通过匹配用户问题和预置问题的嵌入向量相似度，返回对应结果。仅使用大模型语句嵌入的能力解决语意相似度问题。**无训练成本，数据库构造成本也不高，但适用范围小。**



大模型用于安全领域的挑战

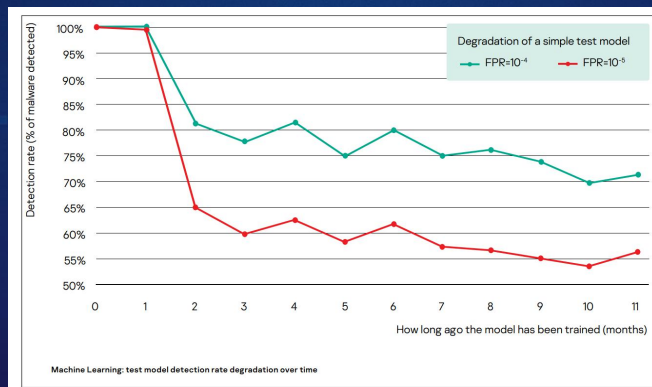
大模型用于安全领域的问题：

- 到底有多少需要用到的安全知识？（碎片化）
- 增量知识怎么解决？
- 高质量标记数据极度匮乏
- 如何保证知识能被训练到模型里面？
- 如何能保证训练进去后，知识能被符合人类理解的方式使用
- 性能/性价比如何平衡

目前来讲：

- 通过重新训练/二次训练的方式
- 通过微调大模型的方式
- 通过向量数据库查询的方式

实现安全领域大模型可能不是最优路径。



数据分布：

在非安全领域，一般假定模型发布后，需要预测的数据分布是保持不变的；然而，在安全领域，恶意程序的作者不停地创造与之前大相径庭的恶意文件、各式的白样本也不停地产生。需要有对样本的运营、周期地训练等。

模型训练流水线比模型本身更重要

安全大模型演进思路

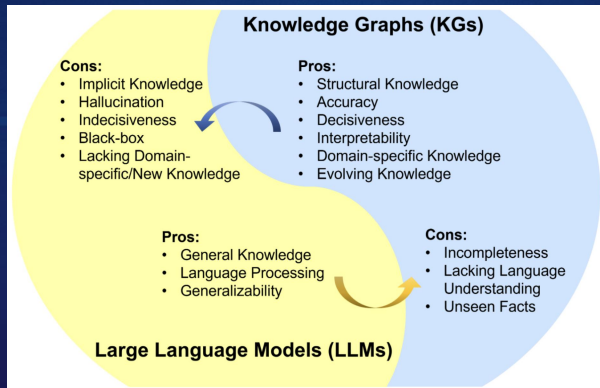
传统安全厂商数据积累：

- 对象：流量包、文件
- 规则：主动防御、IOA规则
- 剧本：响应流程

安全大模型需要的数据：

理解：

- 流量、样本、攻击过程有什么特征，为什么是黑/白的？
- 响应流程为什么要这么设计？



实现上，分为多阶段

- 冷启动：人工标记数据，生产知识，大小模型结合，平衡性能和效果；
- 持续运营：AI辅助人工，提升效率，积攒更多高质量数据；逐步提升自动化程度；
- 统一安全大模型训练：替代绝大部分人工劳动，实现真正的智能化判断。

← “Unifying Large Language Models and Knowledge Graphs: A Roadmap”



解读Demo

命令行	GPT3.5 解读	知识富化后的 GPT3.5
****	<p>这是一条Linux命令，它的作用是从指定的URL下载一个名为"core.png"的文件，并将其作为标准输入传递给bash shell进行执行。具体来说，该命令首先将环境变量"src"设置为"rds"，然后使用curl命令从"https://t.m7n0y.com/ln/core.png?rds*2.0"下载文件，并将其通过管道传递给bash shell进行执行。这个URL中的"?rds*2.0"可能是一个参数，用于指定下载的文件版本或其他信息。需要注意的是，这个命令中使用了"-f"和"-k"选项，它们分别表示"禁止显示进度条"和"允许不安全的SSL连接"。由于这个命令中使用了curl和bash，因此需要确保这两个工具已经安装并且可用。</p>	<p>该命令行通过curl从指定网站下载了一个名为core.png的文件，并使用bash命令执行该文件。这种行为是可疑的，因为它可能会导致恶意软件的执行。为了确定是否存在威胁，需要查询有关该网站的威胁情报。</p> <p>根据恶意威胁情报（Malicious CTI），m7n0y.com是柠檬鸭恶意软件家族使用的服务器。因此，从该网站下载并执行脚本的行为可能会导致系统受到柠檬鸭等恶意软件的感染。</p> <p>综上所述，建议不要执行该命令行，并采取适当的安全措施来保护系统免受潜在的恶意软件威胁。</p>





THANKS

