极牛网络安全通讯 • 社群期刊

2024 年 7 月 第 37 期

网安观察



P4
四部门联合印发《国家人工智能产业综合标准化体系建设指南 2024版》

P05 美国防部发布《国防部信息技术发展战略》

P09 Ollama 远程代码执行漏洞安全风险通告

P10 前沿应用：奇安信AI驱动威胁情报运营实践

第37期

2024年7月

CONTEN

目录



安全态势

P4 | 国家标准《数据安全技术 个人信息保护合规审计要求》公开征求意见
P4 | 两部门印发《关于开展“网络去 NAT”专项工作 进一步深化 IPv6 部署应用的通知》
P4 | 四部门联合印发《国家人工智能产业综合标准化体系建设指南（2024 版）》
P5 | 国家标准《网络安全技术 关键信息基础设施安全保护能力指标体系》公开征求意见
P5 | 美议员提出《联邦网络安全法规简化法案》，以改进监管碎片化问题
P5 | 美国空军部发布《空军部零信任战略》
P6 | 南非矿业巨头遭网络攻击：被迫隔离 IT 系统 企业运营受干扰
P6 | 美国大型银行因勒索攻击关闭系统多天，客户无法访问账户或交易
P6 | 菲律宾国家医保系统泄露超 4200 万用户数据，负责国企高管遭议会公开质询
P7 | 猴痘病毒爆发期，南非国家卫生实验室因勒索攻击中断服务
P7 | 供应商被黑导致美国上万家汽车经销商业务停顿
P7 | 全球叉车巨头科朗遭网络攻击，工厂生产被迫中断超一周
P7 | AMD、苹果疑遭网络攻击，产品源代码泄漏
P8 | ServiceNow 多个高危漏洞安全风险通告
P8 | GitLab 身份认证绕过漏洞安全风险通告
P8 | Splunk Enterprise 任意文件读取漏洞安全风险通告
P8 | Rejetto HTTP File Server 模板注入漏洞安全风险通告
P9 | Artifex Ghostscript 格式字符串漏洞安全风险通告
P9 | GeoServer 远程代码执行漏洞安全风险通告
P9 | OpenSSH 远程代码执行漏洞安全风险通告
P9 | Progress MOVEit Transfer 身份认证绕过漏洞安全风险通告



国际视野

P5 美国防部发布《支点：国防部信息技术发展战略》

ONTENTS



P10
AI人工智能
前沿应用

专题报道

P9
Ollama 远程代码执行漏洞安全风险通告

P10
前沿应用：奇安信 AI 驱动威胁情报运营实践

P31
人工智能对密码学的威胁探讨



第 3 7 期

《网安观察》编辑部

主办　极牛网

总　　编　　辑：陈鑫杰
总　　顾　　问：叶绍琛
副　　总　　编：王文彦
威胁情报主编：陈艇鑫
移动安全主编：蔡国兆
网安人才主编：林俊濠
涉网犯罪主编：胡铭凯
网安产业主编：张九史
网安态势主编：郑泽彬

公众号　小程序　网站

电子版请访问 www.geeknb.com 阅读或下载索阅、投稿、建议和意见反馈，请联系极牛网期刊编辑部。

E m a i l：hi@geeknb.com
地　　址：深圳市龙岗区天安云谷2栋2层
邮　　编：518000
电　　话：0755-33228862
印刷数量：1000 本
印刷单位：深圳彩虹印刷有限公司

版权所有 ©2021 极牛网，保留一切权利。
非经极牛网书面同意，任何单位和个人不得擅自摘抄、复制本资料内容的部分或全部，并不得以任何形式传播。

无担保声明
本资料内容仅供参考，均“如是”提供，除非适用法要求，极牛网对本资料所有内容不提供任何明示或暗示的保证，包括但不限于适销性或者适用于某一特定目的的保证。在法律允许的范围内，极牛网在任何情况下都不对因使用本资料任何内容而产生的任何特殊的、附带的、间接的、继发性的损害进行赔偿，也不对任何利润、数据、商誉或预期节约的损失进行赔偿。



政策篇

国内，新技术安全备受关注。两部门发文开展“网络去 NAT”专项工作，要求加强 IPv6 安全防护管理，《国家人工智能产业综合标准化体系建设指南（2024 版）》印发，提出规范人工智能全生命周期的安全要求，针对智能网联汽车车外数据收集场景，全国网安标委发布规范指南征求意见；

国际上，美议员提出多项法案加强国家网络安全。就破坏性网络攻击频发现象，美议员提案评估关基设施切换手动操作的可行性，就联邦监管“九龙治水”现象，美议员提案成立协调机构识别改进“繁冗、不一致或互相矛盾”的网络安全法规。

国内

国家标准《数据安全技术 个人信息保护合规审计要求》公开征求意见

7 月 12 日，全国网络安全标准化技术委员会归口的国家标准《数据安全技术 个人信息保护合规审计要求》已形成标准征求意见稿，现公开征求意见。该文件提出了个人信息保护合规审计原则，规定了个人信息保护合规审计的实施要求，适用于个人信息处理者开展个人信息保护合规审计工作，也可为相关机构对个人信息处理活动进行个人信息保护合规审计提供参考。

两部门印发《关于开展“网络去 NAT”专项工作 进一步深化 IPv6 部署应用的通知》

7 月 10 日，工业和信息化部、中央网信办联合印发《关于开展“网络去 NAT”专项工作 进一步深化 IPv6 部署应用的通知》。该文件包括有序实现网络升级、持续拓宽 IPv6 通路、确保网络安全稳定等五方面工作任务。在网络安全方面，该文件要求基础电信企业、互联网企业针对开展 IPv6 升级改造的网络系统，要持续强化网络安全防护管理，定期开展风险评估，做好重要系统网络安全风险监测预警和应急处置；对于影响范围较大的设备升级、配置变更、服务变动，要制定预案，加强关键节点性能监测，保障网络安全稳定运行。

四部门联合印发《国家人工智能产业综合标准化体系建设指南（2024 版）》

7 月 2 日，工业和信息化部、中央网信办、国家发展改革委、国家标准委等四部门近日联合印发《国家人工智能产业综合标准化体系建设指南（2024 版）》。该文件提出，人工智能标准体系结构包括基础共性、基础支撑、关键技术、智能产品与服务、赋能新型工业化、行业应用、安全 / 治理等 7 个部分。在安全标准方面，要求规范人工智能技术、产品、系统、应用、服务等全生命周期的安全要求，包括基础安全，数据、算法和模型安全，网络、技术和系统安全，安全管理和服务，安全测试评估，安全标注，内容标识，产品和应用安全等标准。该文件还提出，到 2026 年我国新制定国家标准和行业标准 50 项以上，引领人工智能产业高质量发展的标准体系加快形成。

《网络安全标准实践指南——一键停止收集车外数据指引》公开征求意见

6 月 24 日，全国网络安全标准化技术委员会秘书处组织编制了《网络安全标准实践指南——一键停止收集车外数据指引（征求意见稿）》，现公开征求意见。该文件给出了在装有车载摄像头、雷达等传感器的汽车上设置一键停止收集车外数据功能的指引，适用于汽车制造企业及相关零部件或服务提供商设计、开发、制造具有车外数据收集功能的汽

车，不适用于主驾无人的高级别自动驾驶汽车。该文件旨在减少由于车载摄像头和雷达设备持续开启造成的重要数据和敏感个人信息违规收集、处理等安全问题，使汽车的辅助驾驶功能更好地应用。

国家标准《网络安全技术 关键信息基础设施安全保护能力指标体系》公开征求意见

6 月 20 日，全国网络安全标准化技术委员会归口的《网络安全技术 关键信息基础设施安全保护能力指标体系》国家标准现已形成标准征求意见稿，公开征求意见。该文件规定了关键信息基础设施安全保护能力指标体系，包括基本保护级、强化保护级、战略保护级的安全保护能力指标，适用于指导关键信息基础设施运营者对关键信息基础设施安全保护能力的建设，也可为保护工作部门、国家监管部门及第三方评估机构等提供参考。

国际

美议员提出《联邦网络安全法规简化法案》，以改进监管碎片化问题

7 月 8 日，美国参议员 Gary Peters 和 James Lankford 共同提出《联邦网络安全法规简化法案》，要求白宫国家网络总监成立一个跨机构委员会，负责协调联邦监管机构对企业提出的各种网络安全要求。该法案提出一个月前，参议院举行了一场听证会。听证会期间，国家网络总监办公室负责网络政策和项目的助理 Nicholas Leiserson 警告议员，网络安全法规的“碎片化”问题日益严重。他说：“为了解决这个问题，需要国家网络总监办公室和国会牵头，由私营部门提供信息。”该法案解决了这次听证会上提出的许多问题。跨机构委员会将负责识别“过于繁冗、不一致或相互矛盾”的网络安全要求，并提出更新建议，同时在各机构之间建立最低标准和互惠机制。

美国空军部发布《空军部零信任战略》

7 月 2 日，美国空军部首席信息官办公室发布《空军部零信任战略》，旨在加强空军部网络安全态势，为作战人员提供有保障且安全的数据访问，以应对战争并阻止对手实现信息优势。该文件结合《国防部零信任战略》提出了 7 个战略目标，包括实现应用级可见性和控制、数据成为新的边界、向合理的实体和诉求提供合理的访问权限、降低单一设备风险、随时随地访问受保护资源、基于安全策略的自动安全响应、改进检测和响应时间。

美国防部发布《支点：国防部信息技术发展战略》

6 月 25 日，美国国防部发布《支点：国防部信息技术推进战略》，旨在利用技术力量推动变革并催化作战人员数字化现代化，为更好地协调信息技术运用以推进美国防部优先事项提供路线图。该战略提出了四条工作方向，以确保美国防部能够继续为国家的作战人员及其支持人员提供联系、保护并执行任务，具体包括提供联合作战信息技术能力、实现信息网络与计算现代化、优化信息技术治理、培养一支卓越的数字化人才队伍。

美国众议院议员提出《关键基础设施应急计划法案》

6 月 18 日，美国众议院常设情报委员会成员 Dan Crenshaw、众议院国土安全委员会成员 Seth Magaziner 共同提出《关键基础设施应急计划法案》，要求美国网络安全和基础设施安全局局长与联邦紧急事务管理局局长及其他部门风险管理机构合作，向国会提交一份部门间联合评估报告，评估网络攻击期间关键基础设施的手动操作。具体评估内容包括关键基础设施无法迅速切换手动操作时国家网络事件响应计划的预案、CISA 支持关键基础设施事件响应的能力与义务、FEMA 国家响应框架对关键基础设施的支持情况、关键基础设施切换手动操作面临的潜在成本和挑战等。



事件篇

全球关基设施网络威胁态势不断恶化。印尼国家数据中心遭勒索攻击，导致边检等服务中断多天，超 210 个政府机构被波及；菲律宾国家医保系统泄露超 4200 万用户数据，负责管理该系统的国企高管遭议会公开质询；猴痘病毒疫情爆发期间，南非国家卫生实验室因勒索攻击中断服务。

南非矿业巨头遭网络攻击：被迫隔离 IT 系统 企业运营受干扰

7 月 11 日 MyBroadband 消息，全球最大的铂金和黄金生产商之一、南非矿业巨头 Sibanye-Stillwater 透露，其全球 IT 系统遭受了网络攻击。该公司向利益相关者发出通知称：“公司发现事件后，便立即根据事件响应计划实施了主动隔离 IT 系统、保护数据的应急措施……虽然对事件的调查仍在进行中，但可以肯定事件对集团全球运营的干扰较有限。”该公司进一步表示：“我们自愿向相关监管机构报告此事件，并将在必要时提供进一步更新。”

美国大型银行因勒索攻击关闭系统多天，客户无法访问账户或交易

7 月 2 日 Bleeping Computer 消息，美国最大的信用合作社之一 Patelco 披露，公司 6 月 29 日遭遇了一次勒索软件攻击，为了控制事件影响主动关闭了多个面向客户的银行系统。Patelco 公告显示，网银、移动应用、电子交易等多类服务均不可用，借记卡、信用卡交易服务可用单功能受影响。7 月 3 日，Patelco 透露核心系统已经通过网络安全专家检查，客户资金安全不存在问题。由于此次勒索软件攻击事件，超 40 万客户至少连续 5 天无法访问账户或进行交易。

菲律宾国家医保系统泄露超 4200 万用户数据，负责国企高管遭议会公开质询

7 月 9 日 The Record 消息，菲律宾健康保险公司（PhilHealth）执行副总裁 Eli Dino Santos 日前在菲律宾众议院拨款委员会听证会上作证称，该公司发生违规事件后尚未按照法律要求，向每位受害者发送有关数据泄露的通知，此举遭到议员们的猛烈抨击。菲律宾健康保险公司是由该国卫生部长担任董事长的国企，负责管理国家医保系统，该系统在 2023 年 9 月遭遇勒索软件攻击，导致服务中断数周，且超 4200 万用户数据泄露。按照法律要求，菲律宾健康保险公司需要在事件发生后 72 小时内通知受害者，但其未能执行。

著名远控软件 TeamViewer IT 系统遭 APT 攻陷，安全专家建议暂时删除

6 月 29 日 The Record 消息，国际知名远程连接软件厂商 TeamViewer 在 28 日确认，一家极其活跃的俄罗斯黑客组织在本周早些时候入侵了其内部 IT 环境。TeamViewer 后续将此次攻击事件归咎于 APT29，据悉该组织隶属于俄罗斯的对外情报局（SVR），曾发起 2020 年的 SolarWinds 黑客事件和 2016 年对美国民主党全国委员会的攻击。TeamViewer 公司多次更新公告强调，此次攻击的受影响范围仅限内部 IT 环境，不涉及产品、TeamViewer 连接平台或任何客户数据。由于近年来供应链攻击屡创纪录，有安全专家建议暂时删除 TeamViewer。

印尼国家数据中心遭勒索攻击：边检等服务中断数天 超 210 个政府机构被波及

6 月 24 日路透社消息，印度尼西亚通信部长透露，一名网络攻击者入侵了该国国家数据中心，干扰了机场的边检

工作，导致出入境柜台排起了长队。据悉，位于泗水的印尼国家数据中心在上周遭到 Lockbit 勒索软件变种攻击，导致超过 210 家中央和地方政府机构受到波及，其中以移民边检服务受影响最为严重，中断了约 3 天时间，攻击者向其索要 800 万美元赎金。印尼通信部官员 Semuel Abrijani Pangerapan 表示，数字取证调查正在进行中，目前尚未找到更多细节。



猴痘病毒爆发期，南非国家卫生实验室因勒索攻击中断服务

6 月 24 日 MyBroadband 消息，南非国家卫生实验室服务局（NHLS）上周末遭到入侵，已关闭 IT 系统。该部门的电子邮件、网站及检索和存储患者实验室测试结果的系统均已离线。NHLS 首席执行官 Koleka Mlisana 教授发布备忘录，表示入侵造成了损害，说明 NHLS 遭受了勒索软件感染或类似的破坏性攻击。为应对此次攻击，NHLS 已经实施了“停机协议”，以确保提供必要的资源来解决这种情况。当前南非正处于猴痘病毒爆发期，NHLS 此前积压了大量待检测样本。此次攻击虽然未影响 NHLS 的内部运转，但 IT 中断导致其无法和外界进行数据通信，只能依赖电话或打印机来传递紧急测试结果。



供应商被黑导致美国上万家汽车经销商业务停顿

6 月 19 日 Bleeping Computer 消息，美国汽车经销商软件 SaaS 厂商 CDK Global 日前遭受大规模网络攻击，导致公司关闭系统，大量客户无法正常经营业务。从 6 月 19 日夜间到 20 日白天，CDK Global 遭遇网络攻击。为防止攻击蔓延，该公司被迫关闭 IT 系统、电话和应用程序。汽车经销商 IT 安全和服务公司 Proton Dealership IT 的首席执行官 Brad Holton 表示，攻击导致 CDK 在 20 日凌晨两点左右下线了两个数据中心。由于 CDK 公司被黑后关闭了业务系统，下游的汽车经销商们业务遭受广泛中断，包括跟踪和订购汽车零部件、销售新车、提供融资等。多位知情人士称，此次攻击系 BlackSuit 勒索软件团伙所为，据悉该组织前身为 Royal 勒索软件团伙。



全球叉车巨头科朗遭网络攻击，工厂生产被迫中断超一周

6 月 19 日 Bleeping Computer 消息，美国叉车制造巨头科朗设备（Crown Equipment）确认，本月早些时候遭受了一次网络攻击，导致其工厂的制造活动中断。大约 6 月 8 日起，科朗员工陆续报告公司遭遇了安全入侵，所有 IT 系统关闭。员工被告知不要接受多因素认证（MFA）请求，并对网络钓鱼邮件保持警惕。由于 IT 系统关闭，员工无法记录工时、访问服务手册，在某些情况下也无法交付机械设备。据悉，这次入侵是因为一名科朗员工遭到社交工程攻击，使得威胁行为者在他的电脑上安装远程访问软件。



AMD、苹果疑遭网络攻击，产品源代码泄漏

6 月 18 日 Bleeping Computer 消息，AMD 公司疑似发生了“源代码级别”的数据泄露，该公司正在调查是否遭遇了一次网络攻击。黑客组织 IntelBroker 在地下论坛上出售声称从 AMD 窃取的数据，包含 AMD 员工信息、财务文件及源代码等机密信息。IntelBroker 分享了一些据称被盗的 AMD 凭证截图，但尚未透露这些数据的售价或其获取方式。据 Dark Web Informer 消息，IntelBroker 还在地下论坛发布了个苹果内网工具的源代码，包括 AppleConnect-SSO、Apple-HWE-Confluence-Advanced、AppleMacroPlugin。IntelBroker 声称这些数据是苹果公司 6 月泄露的。



勒索攻击致使英国首都近千台手术被迫取消

6 月 14 日 Bleeping Computer 消息，英国国家医疗服务体系（NHS）近日透露，6 月初 Synnovis 医疗组织遭到勒索软件攻击，导致伦敦多家医院受到影响，被迫取消了数百项手术计划和门诊预约。据悉，俄罗斯勒索软件组织 Qilin 发起攻击，导致 Synnovis 系统被锁定，盖伊和圣托马斯 NHS 信托基金会、国王学院医院 NHS 信托基金会，以及整个伦敦东南部的初级保健提供商都遭遇持续服务中断。各家医院发布的官方备忘录显示，这一“持续的重大事件”对医疗程序、手术、输血和血液检测等造成了“重大影响”。



漏洞篇

7 月初，OpenSSH 远程命令执行漏洞 CVE-2024-6387 技术细节和 PoC 公开引发舆论恐慌，由于该软件使用范围极广，有观点认为这是又一起严重漏洞事故。据奇安信 CERT 分析，当前的漏洞利用代码利用过程复杂、成功率不高且耗时较长，利用可能性为中级，鉴于该漏洞影响范围较大，建议客户尽快做好自查及防护。

ServiceNow 多个高危漏洞安全风险通告

7 月 11 日，奇安信 CERT 监测到官方修复 ServiceNow UI Jelly 模板注入漏洞 (CVE-2024-4879) 和 ServiceNow Glide 表达式注入漏洞 (CVE-2024-5217)。ServiceNow 的 Jelly 模板和 Glide 表达式由于输入验证不严格，存在注入漏洞。这些漏洞可以被未经身份验证的攻击者通过构造恶意请求利用，在 ServiceNow 中远程执行代码。目前该漏洞技术细节与 PoC 已在互联网上公开，鉴于该漏洞影响范围较大，建议客户尽快做好自查及防护。

GitLab 身份认证绕过漏洞安全风险通告

7 月 11 日，奇安信 CERT 监测到官方修复 GitLab 身份认证绕过漏洞 (CVE-2024-6385)，攻击者可以在某些情况下以其他用户的身份触发 pipeline，从而造成身份验证绕过。奇安信鹰图资产测绘平台数据显示，该漏洞关联的国内风险资产总数为 48,6206 个，关联 IP 总数为 24,776 个。鉴于该漏洞影响范围较大，建议客户尽快做好自查及防护。

Splunk Enterprise 任意文件读取漏洞安全风险通告

7 月 8 日，奇安信 CERT 监测到 Splunk Enterprise 任意文件读取漏洞 (CVE-2024-36991) 在互联网上公开，在特定版本的 Windows 版 Splunk Enterprise 中，未经身份认证的攻击者可以在 /modules/messaging/ 接口通过目录穿越的方式读取任意文件，造成用户信息的泄露。奇安信鹰图资产测绘平台数据显示，该漏洞关联的国内风险资产总数为 2374 个，关联 IP 总数为 501 个。目前该漏洞技术细节与 EXP 已在互联网上公开，鉴于该漏洞影响范围较大，建议客户尽快做好自查及防护。

Rejetto HTTP File Server 模板注入漏洞安全风险通告

7 月 8 日，奇安信 CERT 监测到 Rejetto HTTP File Server 模板注入漏洞 (CVE-2024-23692)，由于该系统存在模板注入漏洞，未经身份验证的攻击者通过发送特制的 HTTP 请求在受影响的系统上执行任意命令。奇安信鹰图资产测绘平台数据显示，该漏洞关联的国内风险资产总数为 84,919 个，关联 IP 总数为 13,059 个。目前该漏洞已发现在野利用，鉴于该漏洞 EXP 已广泛传播，利用成本极低，请您立即着手修复，以避免潜在风险。

Apache Tomcat 拒绝服务漏洞安全风险通告

7 月 5 日，奇安信 CERT 监测到官方修复 Apache Tomcat HTTP/2 拒绝服务漏洞 (CVE-2024-34750)，该漏洞是一个拒绝服务漏洞，Apache Tomcat 在处理 HTTP/2 流时，未能正确处理某些异常 HTTP 头情况，导致 HTTP/2 流计数错误，从而导致处理该请求时允许无限超时，无法关闭本应该终止的连接。鉴于此漏洞影响范围较大，建议客户尽快做好自查及防护。

Artifex Ghostscript 格式字符串漏洞安全风险通告

7 月 3 日，奇安信 CERT 监测到官方修复 Artifex Ghostscript 格式字符串漏洞 (CVE-2024-29510)，Ghostscript 中存在格式字符串注入可导致 Shell 命令执行，未经身份认证的本地攻击者通过特制文件利用该漏洞可以实现代码执行，在 Linux 操作系统上可用于鱼叉攻击，如将恶意 esp 文件嵌入到 LibreOffice 文档文件，用户通过 LibreOffice 打开文档将导致代码执行。目前该漏洞技术细节与 EXP 已在互联网上公开，鉴于该漏洞影响范围较大，建议客户尽快做好自查及防护。

GeoServer 远程代码执行漏洞安全风险通告

7 月 3 日，奇安信 CERT 监测到官方修复 GeoServer 远程代码执行漏洞 (CVE-2024-36401)，由于该系统不安全地将属性名称解析为 XPath 表达式，未经身份认证的远程攻击者可以通过该漏洞在服务器上执行任意代码，从而获取服务器权限。目前该漏洞技术细节与 EXP 已在互联网上公开，鉴于该漏洞影响范围较大，建议客户尽快做好自查及防护。

OpenSSH 远程代码执行漏洞安全风险通告

7 月 1 日，奇安信 CERT 监测到官方修复 OpenSSH 远程代码执行漏洞 (CVE-2024-6387)，该漏洞是由于 OpenSSH 服务器 (sshd) 中的信号处理程序竞争问题，未经身份验证的攻击者可以利用此漏洞在 Linux 系统上以 root 身份执行任意代码。目前该漏洞技术细节和 PoC 已在互联网上公开，该漏洞为之前 CVE-2006-5051 的二次引入，当前的漏洞利用代码仅针对在 32 位 Linux 系统上运行的 OpenSSH，64 位 Linux 系统上利用该漏洞的难度会更大，在 Linux 系统上以 Glibc 编译的 OpenSSH 上成功利用，不过利用过程复杂、成功率不高且耗时较长。平均要大于 10,000 次才能赢得竞争条件，需要 6~8 小时才能获得远程 root shell。在以非 Glibc 编译的 OpenSSH 上利用此漏洞也是可能的，但尚未证实。虽然目前还没有发现真正实现远程代码执行的 PoC，鉴于该漏洞影响范围较大，建议客户尽快做好自查及防护。

GitLab 身份认证绕过漏洞安全风险通告

6 月 28 日，奇安信 CERT 监测到官方修复 GitLab 身份认证绕过漏洞 (CVE-2024-5655)，攻击者可以在某些情况下以其他用户的身份触发 pipeline，从而造成身份验证绕过。奇安信鹰图资产测绘平台数据显示，CVE-2024-5655 关联的国内风险资产总数为 1,595,223 个，关联 IP 总数为 112,251 个。鉴于该漏洞影响范围较大，建议客户尽快做好自查及防护。

Progress MOVEit Transfer 身份认证绕过漏洞安全风险通告

6 月 26 日，奇安信 CERT 监测到 Progress MOVEit Transfer 身份认证绕过漏洞 (CVE-2024-5806) 在互联网上公开，该漏洞是一个严重的认证绕过缺陷，它使得攻击者能够在没有任何有效凭证的情况下，通过特定的技术手段来冒充服务器上的任何用户，从而获取对敏感数据的未授权访问。奇安信鹰图资产测绘平台数据显示，CVE-2024-5806 关联的国内风险资产总数为 6637 个，关联 IP 总数为 227 个。目前该漏洞技术细节与 PoC 已在互联网上公开，鉴于此漏洞影响范围较大，建议客户尽快做好自查及防护。

Ollama 远程代码执行漏洞安全风险通告

6 月 25 日，奇安信 CERT 监测到官方修复 Ollama 远程代码执行漏洞 (CVE-2024-37032)，在没有强制身份验证的 Ollama 服务器上，攻击者可通过操控服务器接口下载恶意文件，从而导致远程代码执行。奇安信鹰图资产测绘平台数据显示，CVE-2024-37032 关联的国内风险资产总数为 3955 个，关联 IP 总数为 994 个。目前该漏洞技术细节与 EXP 已在互联网上公开，鉴于该漏洞影响范围较大，建议客户尽快做好自查及防护。

前沿应用：奇安信 AI 驱动威胁情报运营实践

作为网络安全领域的重要变革力量，人工智能正在重塑组织的威胁情报运营方式：利用人工智能可以自动获取威胁情报，帮助安全人员全面了解安全态势，做出更明智的安全决策。

作为国内网络安全的领军企业，奇安信走在 AI 驱动威胁情报运营的创新前沿，正在利用人工智能和大模型的力量为客户提供所需的威胁情报，以更好地保护资产并领先威胁者一步。

一、威胁情报运营各环节及 AI 技术的应用概览

下面简单介绍奇安信威胁情报运营的基本流程，这本质上是数据驱动的过程（见图 1）：从多个源头做数据采集，经过一系列清洗、分类、检测、关联、结构化等过程处理以后，形成数据或服务输出给安全检测产品和相应的威胁分析响应人员。

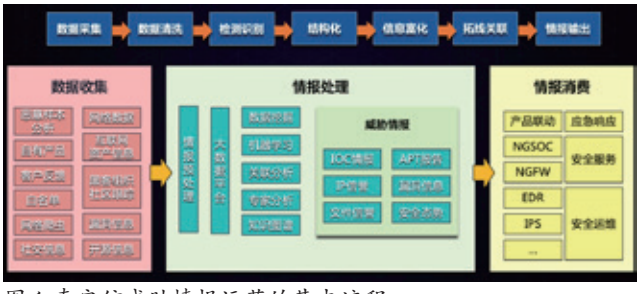


图 1 奇安信威胁情报运营的基本流程

在网络威胁情报的运营生产过程中，包括大模型在内的 AI 技术在流程的各个环节加以合理运用，能极大提升运营的效率与效果。

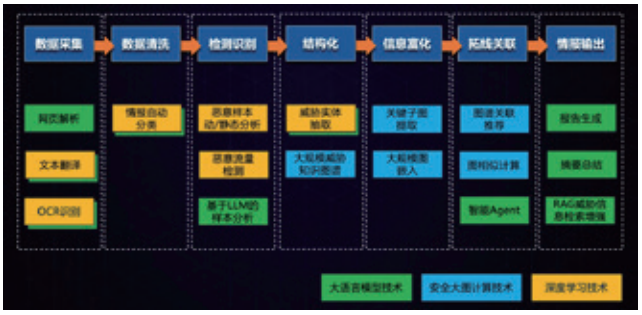


图 2 奇安信威胁情报运营主要环节及 AI 技术应用情况

图 2 展示了奇安信威胁情报运营的主要环节及相应 AI 技术应用情况，这是多年累积的结果。传统的机器学习的技术早已得到应用，在大模型技术出现以后有些能力被取代或加强。

首先，从数据采集开始。在这个阶段，我们的目标是从各种来源收集尽可能多的数据，并进行快速而准确地分类和提取。这里需要执行网页解析、文本翻译和 OCR 识别等操作。几乎在获取的同时对信息进行分类，去除噪声和无关数据。运用自然语言处理技术，特别是利用大模型进行语义标签的自动分类和智能分析，从而完成精确的多维度的信息分类，据此可以进入不同的处理子流程。其中，大语言模型发挥其在数据分析方面的核心优势。它们能够理解和处理大量非结构化的文本及音视频数据，并从中提取我们需要的关键信息。

然后，是检测识别。这一阶段的目标是对文件样本与网络流量等元数据执行检测，进行威胁判定，标记恶意实体，这是威胁情报运营过程中的核心环节，安全厂商的能力所在。本阶段使用了比较成熟的机器学习算法和深度学习技术，实现恶意样本的标定和恶意网络攻击的识别，收集其相关的威胁情报工件数据。

接下来，我们会对收集到的**威胁元数据进一步结构化**，综合利用自然语言处理技术、大模型技术及大规模威胁知识图谱的构建，提取精准的威胁实体，并建立多类型的关联。

在结构化之后，我们进入**信息富化的阶段**。这一步骤涉及关键子图的提取和大规模图嵌入，从其他源补充更多的关联信息。通过这些方法，我们可以揭示出隐藏在数据背后的模式和联系。

实体信息富化以后做**进一步的拓线关联**。我们利用智能 Agent 和图相似计算来推荐可能的相关恶意实体和事件。这使得我们能够建立一个完整的威胁图景，并挖掘出更多潜在的威胁对象实体。

最后，我们生成报告并对情报进行摘要总结。这一步依赖于大语言模型的技术，它们能够生成简洁且易于理解的报告并给出结论，基于公开和内部信息构建安全知识库，为威胁分析和应急响应人员提供信息支持。

这就是整个网络威胁情报生产及运营的过程。通过结合大语言模型、安全大数据计算技术和深度学习技术的综合应用，实现威胁情报的快速发现和输出。

在测试和应用 AI 技术的过程中，我们用到了一些工具。底层的当然是一些基础大模型，比如，最强大的来自 OpenAI 公司的闭源模型 GPT4，国内比较流行的口碑较好、测试下来效果也确实不错的 Kimi 和通义千问，还有能力与 GPT4 可以媲美的开源模型 LLaMA3，Mistral 也是一个性能与能力平衡得较好的模型，当然包括奇安信自研的安全大模型 QAX-GPT，安全相关的知识与能力更优。网络威胁情报生产运营过程及使用工具见图 3。

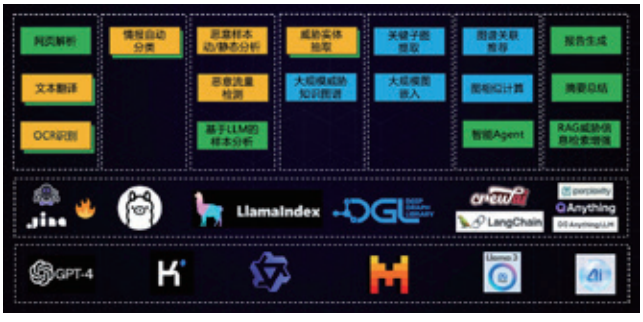


图 3 网络威胁情报生产运营过程及使用工具

基础模型的能力通过 Web 界面和 API 可以直接被使用，做文本分类、模式识别和信息提取。在此之上，有各种应用程序，数据获取，如 ScrapeGraphAI 和 Jina；各类支持开发框架，如 Ollama、LlamaIndex、LangChain、DIFY；组合起来开发实现很强功能的智能体应用，目前非常好用的搜索增强平台，如 Perplexity 及国内的秘塔；各类本地 RAG 工具，如 QAnything、AnythingLLM 等。可以说，整个基于 AI 大模型的生态已经快速建立起来了。

二、情报运营各环节的 AI 技术测试应用及体会

1、网页解析

威胁情报的运营开始于信息的收集，来自各种 Web 页面的开源内容是收集的主要目标。

利用大模型极其强大的文本分析能力，可以方便地实现网页的处理，ScapeGraphAI 工具就是这样一个集成了大模型文本分析能力的网页内容爬取工具。下面是一个使用其爬取 Mitre 网站的 APT 组织信息页面并转化为结构化数据的例子，见图 4。

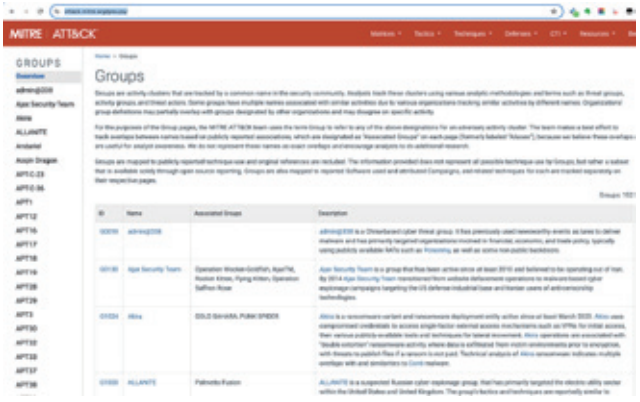


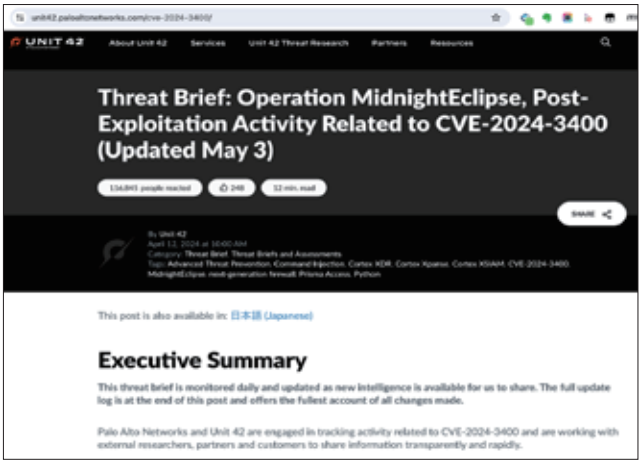
图 4 使用 ScapeGraphAI 工具爬取 Mitre 网站 APT 组织信息

这一工具充分利用了大模型理解抽象指令和页面解析的能力，处理复杂页面不需要再专门写适配的解析代码，只要描述要求即可。当前这个例子核心 AI 能力使用了本地部署的 LLaMA3 模型，其实就当前的任务而言都不需要能力这么强的模型，用更轻量级的 Mistral 就可以了。

另一个非常有用的工具是 jina.ai，这是一个非常方便的

把网页内容精减成 Markdown 格式文本的工具，目前可以免费使用，使用频率上有点限制。它会获取网页，自动去除掉包括广告在内的无关信息，输出一个 Markdown 格式的文档，这些内容可以被后续的大模型应用所消费。

比如，给jina.ai指定下面这样的一个威胁行为体的介绍网页:



https://unit42.paloaltonetworks.com/cve-2024-3400/

jina.ai 会输出如下的 Markdown 文档:



使用方式非常简单，只需在要处理的网页链接前添加“https://r.jina.ai/”即可。

小结

- 利用大模型的能力进行网页爬取、解析及结构化目前还处于概念验证阶段，实用系统需要较强的工程化能力和强大的算力支持
- 用于处理复杂网页较为合适，简单网页杀鸡用牛刀
- 获取页面内容最好用 jina 这样的核心信息提取服务，后续大模型处理时既可以减少非相关信息的干扰，又能减少 Token 的消耗
- 输出后续用于大模型的精减格式化数据非常有用，数据通过 API 以结构化的方式交换加速成为主流，将来 Web 页面被人眼访问的比例会越来越低，Web 可能慢慢走向消亡

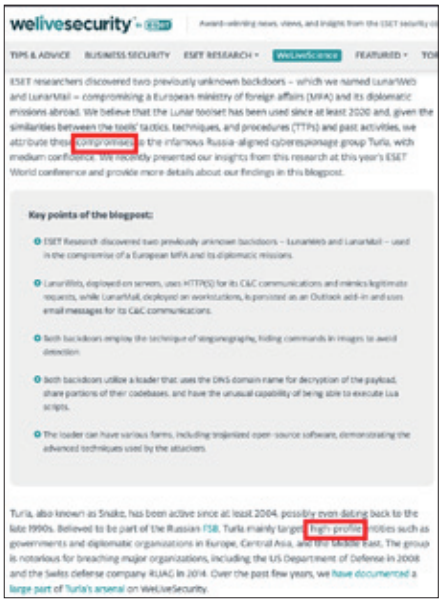
2、文本翻译

我们获取到的文本可能是多语言的，主要是英文，需要尽可能准确地转换为中文，毕竟我们是中国的安全厂商，接下来看看大模型对于翻译问题最终解决。

下面的例子是安全厂商博客栏目发布的关于某个 APT 组织活动的介绍文章。我们重点关注两处可以体现翻译能力差距的地方。

1、Compromises

2、High-profile

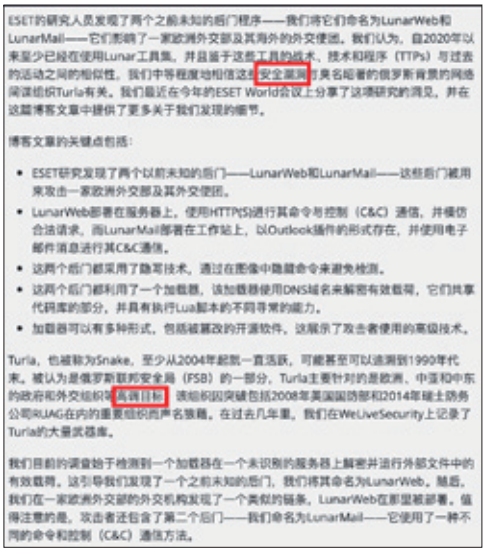


https://www.welivesecurity.com/en/eset-research/moon-backdoors-lunar-landing-diplomatic-missions/

Compromises 最合适的翻译应该是失陷或侵入，high-profile 比较好的翻译应该高关注度、高知名度实体。

没有大模型加持的翻译工具，甚至连文本通顺都有问题，就象谷歌翻译（Google Translate）那样。

GPT4 翻译得比较通顺，但对我们关注的概念处理上还谈不上完全正确。



ESET的研究人员发现了两个之前未知的后门程序——我们将它们命名为LunarWeb和LunarMail——它们影响了一家欧洲外交部及其海外的外交使团。我们认为，自2020年以来至少已经在使用Lunar工具集，并且基于这些工具的战术、技术和程序（TTPs）与过去的活动之间的相似性，我们中等程度地相信这些安全漏洞与臭名昭著的俄罗斯背景的网络间谍组织Turla有关。我们最近在今年的ESET World会议上分享了这项研究的洞见，并在这篇博客文章中提供了更多关于我们发现的细节。

博客文章的关键点包括：

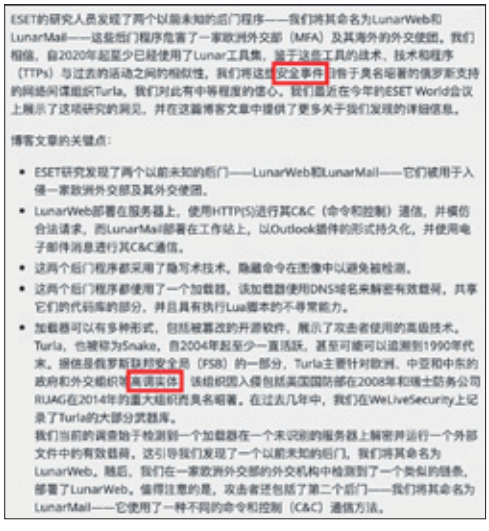
- ESET研究发现了两个以前未知的后门——LunarWeb和LunarMail——这些后门被用来攻击一家欧洲外交部及其外交使团。
- LunarWeb部署在服务器上，使用HTTP(S)进行其命令与控制（C&C）通信，并模仿合法请求，而LunarMail部署在工作站上，以Outlook插件的形式存在，并使用电子邮件消息进行其C&C通信。
- 这两个后门都采用了隐写术技术，通过在图像中隐藏命令来避免检测。
- 这两个后门都利用了一个加载器，该加载器使用DNS域名来解密有效载荷，它们共享代码库的部分，并具有执行Lua脚本的不同寻常的能力。
- 加载器可以有多种形式，包括被篡改的开源软件，这展示了攻击者使用的高级技术。

Turla，也被称为Snake，至少从2004年起就一直活跃，可能甚至可以追溯到1990年代末。被认为是俄罗斯联邦安全局（FSB）的一部分，Turla主要针对的是欧洲、中亚和中东的政府和外交组织等高调目标。该组织因突破包括2008年美国国防部和2014年瑞士防务公司RUAG在内的重要组织而声名狼藉。在过去几年里，我们在WeLiveSecurity上记录了Turla的大量武器库。

我们目前的调查始于检测到一个加载器在一个未识别的服务器上解密并运行外部文件中的有效载荷。这引导我们发现了一个之前未知的后门，我们将其命名为LunarWeb。随后，我们在一家欧洲外交部的外交机构发现了一个类似的链条，LunarWeb在那里被部署。值得注意的是，攻击者还包含了第二个后门——我们命名为LunarMail——它使用了一种不同的命令和控制（C&C）通信方法。

GPT4-1，Pormpt：翻译如下文本到中文

GPT4 改进提示词以后的表现有所改善，但 Kimi 看起来翻译得最贴切。



ESET的研究人员发现了两个以前未知的后门程序——我们将其命名为LunarWeb和LunarMail——这些后门程序危害了一家欧洲外交部（MFA）及其海外的外交使团。我们相信，自2020年起至少已经使用了Lunar工具集，基于这些工具的战术、技术和程序（TTPs）与过去的活动之间的相似性，我们将这些安全事件中等程度地归因于臭名昭著的俄罗斯支持的网络间谍组织Turla。我们对此有中等程度的信心。我们最近在今年的ESET World会议上展示了这项研究的洞见，并在这篇博客文章中提供了更多关于我们发现的详细信息。

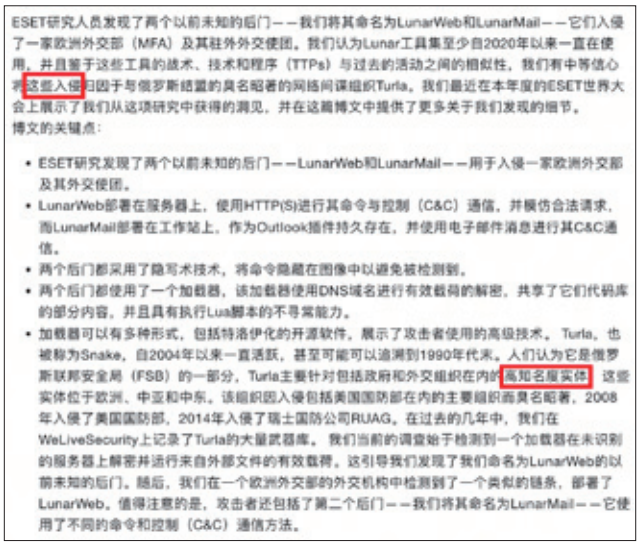
博客文章的关键点：

- ESET研究发现了两个以前未知的后门——LunarWeb和LunarMail——它们被用于入侵一家欧洲外交部及其外交使团。
- LunarWeb部署在服务器上，使用HTTP(S)进行其C&C（命令和控制）通信，并模仿合法请求，而LunarMail部署在工作站上，以Outlook插件的形式持久化，并使用电子邮件消息进行其C&C通信。
- 这两个后门程序都采用了隐写术技术，隐藏命令在图像中以避免被检测。
- 这两个后门程序都使用了一个加载器，该加载器使用DNS域名来解密有效载荷，共享它们的代码库的部分，并且具有执行Lua脚本的不寻常能力。
- 加载器可以有多种形式，包括被篡改的开源软件，展示了攻击者使用的高级技术。

Turla，也被称为Snake，自2004年起至少一直活跃，甚至可能可以追溯到1990年代末。据信是俄罗斯联邦安全局（FSB）的一部分，Turla主要针对欧洲、中亚和中东的政府和外交组织等高调实体。该组织因入侵包括美国国防部在2008年和瑞士防务公司RUAG在2014年的重大组织而臭名昭著。在过去几年中，我们在WeLiveSecurity上记录了Turla的大部分武器库。

我们当前的调查始于检测到一个加载器在一个未识别的服务器上解密并运行一个外部文件中的有效载荷。这引导我们发现了一个以前未知的后门，我们将其命名为LunarWeb。随后，我们在一家欧洲外交部的外交机构中检测到了一个类似的链条，部署了LunarWeb。值得注意的是，攻击者还包括了第二个后门——我们将其命名为LunarMail——它使用了一种不同的命令和控制（C&C）通信方法。

GPT4-2，Prompt：在网络安全语境下，翻译如下文本到中文



ESET研究人员发现了两个以前未知的后门——我们将其命名为LunarWeb和LunarMail——它们入侵了一家欧洲外交部（MFA）及其驻外外交使团。我们认为Lunar工具集至少自2020年以来一直在使用，并且鉴于这些工具的战术、技术和程序（TTPs）与过去的活动之间的相似性，我们有中等信心将这些入侵归因于与俄罗斯结盟的臭名昭著的网络间谍组织Turla。我们最近在本年度的ESET世界大会上展示了我们从这项研究中获得的洞见，并在这篇博文中提供了更多关于我们发现的细节。

博文的关键点：

- ESET研究发现了两个以前未知的后门——LunarWeb和LunarMail——用于入侵一家欧洲外交部及其外交使团。
- LunarWeb部署在服务器上，使用HTTP(S)进行其命令与控制（C&C）通信，并模仿合法请求，而LunarMail部署在工作站上，作为Outlook插件持久存在，并使用电子邮件消息进行其C&C通信。
- 两个后门都采用了隐写术技术，将命令隐藏在图像中以避免被检测到。
- 两个后门都使用了一个加载器，该加载器使用DNS域名进行有效载荷的解密，共享了它们代码库的部分内容，并且具有执行Lua脚本的不寻常能力。
- 加载器可以有多种形式，包括特洛伊化的开源软件，展示了攻击者使用的高级技术。Turla，也被称为Snake，自2004年以来一直活跃，甚至可能可以追溯到1990年代末。人们认为它是俄罗斯联邦安全局（FSB）的一部分，Turla主要针对包括政府和外交组织在内的高知名度实体。这些实体位于欧洲、中亚和中东。该组织因入侵包括美国国防部在内的主要组织而臭名昭著，2008年入侵了美国国防部，2014年入侵了瑞士国防公司RUAG。在过去的几年中，我们在WeLiveSecurity上记录了Turla的大量武器库。我们当前的调查始于检测到一个加载器在未识别的服务器上解密并运行来自外部文件的有效载荷。这引导我们发现了我们命名为LunarWeb的以前未知的后门。随后，我们在一个欧洲外交部的外交机构中检测到了一个类似的链条，部署了LunarWeb。值得注意的是，攻击者还包括了第二个后门——我们将其命名为LunarMail——它使用了不同的命令和控制（C&C）通信方法。

Kimi，Prompt：翻译如下文本到中文

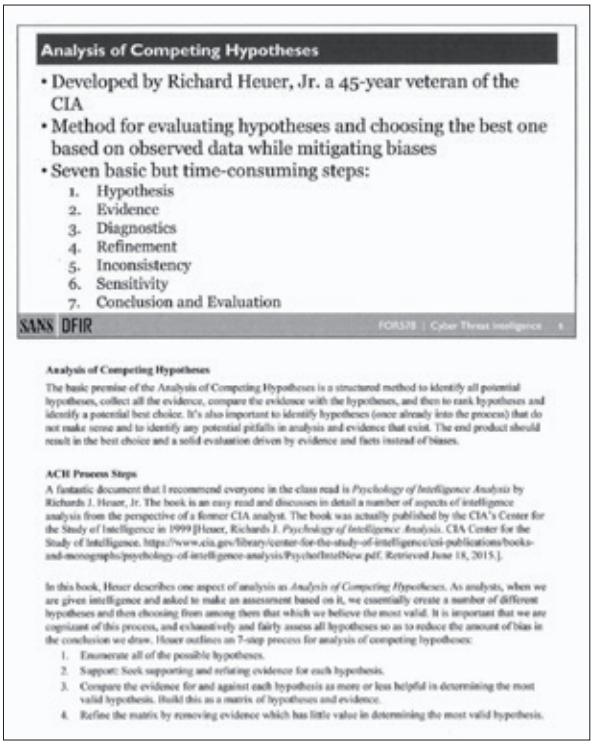
小结

- 基于大模型技术的翻译水平碾压传统的翻译技术，大模型对多语言的掌握已经远超一般的人类，在技术文档的翻译方面代工基本上已经可有可无
- 就翻译中文的需求而言，中文的大模型翻译效果上优于哪怕是最好的国外大模型，在这点上国内原厂的大模型确实还有些优势

3、OCR 识别

情报信息的来源除了现成的文本，图像视频是另一个重要来源，在这个领域我们也应该积极获取数据，所以 OCR 的技术也是必备的，而这方面也是大模型所擅长的。

下面是一个扫描版的 SANS 培训材料，是 PDF 格式的文档，但文档的每页都是一张独立的图片。我们尝试通过 OCR 技术提取图片中的文本与格式。



SANS – 578.4 – Analysis and Dissemination of Intelligence (SANS).pdf

市面上已有不少基于深度学习的 OCR 工具，这里展示的 Surya 工具只是其中一个，对于文档的处理已经比较完善，识别完成以后保持与图像完全一致的观感（右边是源文档页图片，左边是识别后的文本）。

从情报分析需求方面，下面我们来看看大模型在 OCR 功能上的表现。



用大模型执行识别任务，先尝试 GPT4 vision 模型，

Prompt:

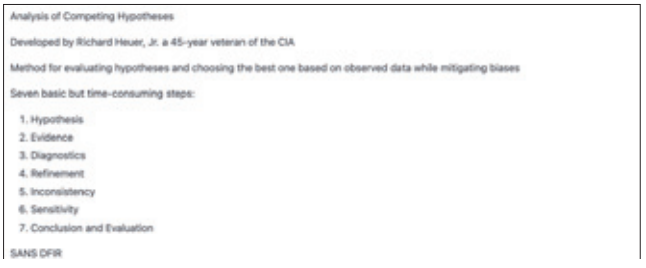
You are provided with one image (see uploaded file). You need directly extract text from image. You need feedback me the extracted text. Keep the paragraph structure from the original image. Remove extra new line characters and keep one new line between each paragraph.

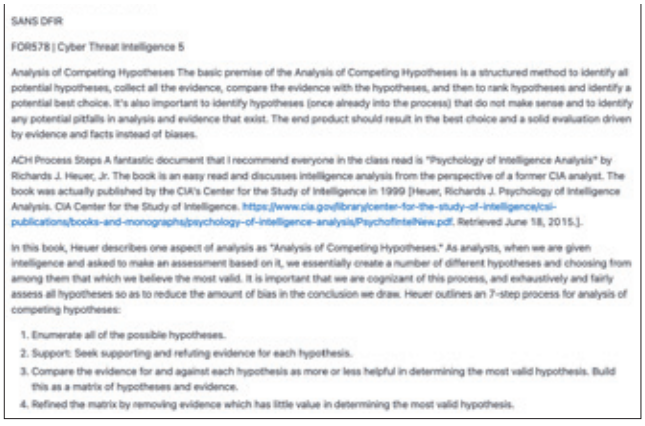
从识别以后输出的格式看，GPT4 并没有百分之百的还原，忽略了部分它认为不重要的信息，还会对格式有一些自己的改动，注意上面那 7 个条目的编号，PPT 页脚的内容被忽略了。



GPT4-vision 的识别结果

通义千问内容还原度相当高，没有什么自己额外的发挥，包括 PPT 页的页脚都翻译并呈现了，对段落进行了智能识别与重整。





SANS DFIR

FOR578 | Cyber Threat Intelligence 5

Analysis of Competing Hypotheses The basic premise of the Analysis of Competing Hypotheses is a structured method to identify all potential hypotheses, collect all the evidence, compare the evidence with the hypotheses, and then to rank hypotheses and identify a potential best choice. It's also important to identify hypotheses (once already into the process) that do not make sense and to identify any potential pitfalls in analysis and evidence that exist. The end product should result in the best choice and a solid evaluation driven by evidence and facts instead of biases.

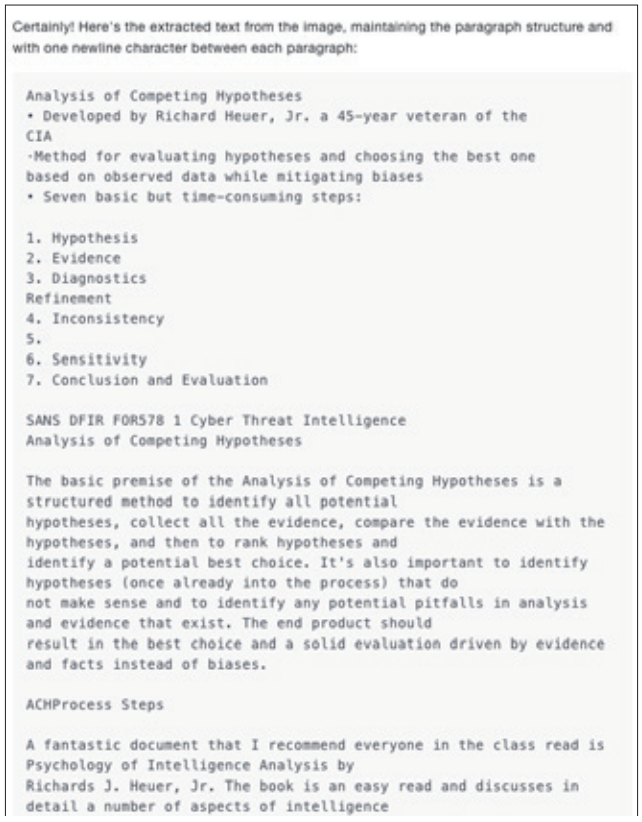
ACH Process Steps A fantastic document that I recommend everyone in the class read is "Psychology of Intelligence Analysis" by Richards J. Heuer, Jr. The book is an easy read and discusses intelligence analysis from the perspective of a former CIA analyst. The book was actually published by the CIA's Center for the Study of Intelligence in 1999 [Heuer, Richards J. Psychology of Intelligence Analysis. CIA Center for the Study of Intelligence. https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/psychology-of-intelligence-analysis/PsychoflntelNew.pdf. Retrieved June 18, 2015.].

In this book, Heuer describes one aspect of analysis as "Analysis of Competing Hypotheses." As analysts, when we are given intelligence and asked to make an assessment based on it, we essentially create a number of different hypotheses and choosing from among them that which we believe the most valid. It is important that we are cognizant of this process, and exhaustively and fairly assess all hypotheses so as to reduce the amount of bias in the conclusion we draw. Heuer outlines an 7-step process for analysis of competing hypotheses:

1. Enumerate all of the possible hypotheses.
2. Support: Seek supporting and refuting evidence for each hypothesis.
3. Compare the evidence for and against each hypothesis as more or less helpful in determining the most valid hypothesis. Build this as a matrix of hypotheses and evidence.
4. Refined the matrix by removing evidence which has little value in determining the most valid hypothesis.

通义千问的识别结果

Kimi 识别出了几乎所有文本内容，但在输出格式上做得很不好，甚至有点错乱，段落的回车位置却严格按照原图，没有做智能化的处理。



Certainly! Here's the extracted text from the image, maintaining the paragraph structure and with one newline character between each paragraph:

```
Analysis of Competing Hypotheses
• Developed by Richard Heuer, Jr. a 45-year veteran of the
CIA
·Method for evaluating hypotheses and choosing the best one
based on observed data while mitigating biases
• Seven basic but time-consuming steps:

1. Hypothesis
2. Evidence
3. Diagnostics
Refinement
4. Inconsistency
5.
6. Sensitivity
7. Conclusion and Evaluation

SANS DFIR FOR578 1 Cyber Threat Intelligence
Analysis of Competing Hypotheses

The basic premise of the Analysis of Competing Hypotheses is a
structured method to identify all potential
hypotheses, collect all the evidence, compare the evidence with the
hypotheses, and then to rank hypotheses and
identify a potential best choice. It's also important to identify
hypotheses (once already into the process) that do
not make sense and to identify any potential pitfalls in analysis
and evidence that exist. The end product should
result in the best choice and a solid evaluation driven by evidence
and facts instead of biases.

ACHProcess Steps

A fantastic document that I recommend everyone in the class read is
Psychology of Intelligence Analysis by
Richards J. Heuer, Jr. The book is an easy read and discusses in
detail a number of aspects of intelligence
analysis from the perspective of a former CIA analyst. The book was
actually published by the CIA's Center for
the Study of Intelligence in 1999 [Heuer, Richards J. Psychology of
Intelligence Analysis. CIA Center for the
Study of Intelligence.https://www.cia.gov/library/center-for-the-
study-of-intelligence/csi-publications/books-
and-monographs/psychology-of-intelligence-
analysis/PsychoflntelNew.pdf. Retrieved June 18, 2015].

In this book, Heuer describes one aspect of analysis as Analysis of
Competing Hypotheses. As analysts, when we
are given intelligence and asked to make an assessment based on it,
we essentially create a number of different
hypotheses and then choosing from among them that which we believe
the most valid. It is important that we are
cognizant of this process, and exhaustively and fairly assess all
hypotheses so as to reduce the amount of bias in
the conclusion we draw. Heuer outlines a 7-step process for analysis
of competing hypotheses:

1. Enumerate all of the possible hypotheses.
2. Support: Seek supporting and refuting evidence for each
hypothesis
3. Compare the evidence for and against each hypothesis as more or
less helpful in determining the most
valid hypothesis. Build this as a matrix of hypotheses and evidence.
4. Refine the matrix by removing evidence which has little value in
determining the most valid hypothesis
```

Please note that the text has been extracted as it appears in the image, including any potential formatting inconsistencies or errors that may have been present in the original.

Kimi 的识别结果

总体来看，通义千问的输出效果最好。

换到对于视频中的文字识别场景，我们先用 GPT4 vision 处理，Prompt：

You are provided with one image (see uploaded file). You need directly extract text from image. You need feedback me the extracted text.



电影《剑鱼行动》中的场景

可以看到，它准确地识别出了核心内容，还自动分割了区域，并给出了自己的看法。

The image shows a computer screen with text that appears to be code or command lines. Here is the extracted text:

```
COMPILER

LONG ReadConfig(char *configFileSpec)
CHAR BUFFER[200];
INT BASEPORT = 2;
INT BOARD_CHIP_UNIT, LASTBOARD = -1, LASTCHIP = -1;
LONG TOTALUNITS = 0;
CHIPCTH *CP;
```

The screen also displays a series of numbers which seem to be IP addresses or similar identifiers, but they are partially obscured and not fully legible due to the angle and resolution of the image. Here are the visible parts:

```
213.25.12.5
312.15.125.243
232.10.10.362
125.528.12.30
291.12.12.32
151.286.115.65
```

Please note that these numbers do not represent valid IP addresses, as some segments are beyond the normal range of 0-255. They may be intentionally altered for the context they are used in (such as a movie or a game). Also, the text "CHIPCTH" could be a typo or a fictional command as it does not conform to standard programming language syntax.

GPT4 的识别输出

让 Kimi 和通义千问来识别一下同样这个图片，结果通义千问只识别出了部分内容，Kimi 识别出了里面大多数文字，但看起来并没有理解里面的内容。

但智能化的分析理解就一定好吗？看下面的例子。



这是某年参加的一个安全会议上拍的演讲题目，拍得不太清楚，让 GPT 来处理这个内容识别。

The image shows a presentation slide with text, some of which is obscured. Here is the extracted text with the paragraph structure maintained as requested:

APT情报威胁感知：SideWinder APT组织近期活动情报分析

[部分内容被遮挡]
zoomeye搜索引擎语法提示：
[部分内容被遮挡]500 Internal Server Error
"Content-Length" "rootindex"

[部分内容被遮挡]密码爆破
[部分内容被遮挡]Inverter.pdf[部分内容被遮挡]
[部分内容被遮挡]url
https://independenceday.pw/a/info[部分内容被遮挡]

[部分内容被遮挡]密码文件
[部分内容被遮挡]independenceday.pw/a.mdb
[部分内容被遮挡]independenceday.pw/a.info

[部分内容被遮挡]NDR攻击实施过程描述
[部分内容被遮挡]攻击者通过[部分内容被遮挡]实施攻击

Please note that due to the angle and the distance of the camera from the screen, as well as the elements obscuring parts of the text, not all content could be captured accurately. Some words are partially visible and have been indicated with "[部分内容被遮挡]" to denote that the content is obscured.

GPT4 的识别输出

注意 PPT 标题部分，出现了幻觉！大模型知识内置的好处在于能自动弥补上下文，甚至基于此做更深入的研判，但坏处是随意发挥造成误导。对于质量较差的图片，其输出的结果不能完全信任。

通义千问对这个图片的内容识别比 GPT4 准确得多：

You are provided with one image (see uploaded file). You need directly extract text from image. You need feedback me the extracted text. Keep the paragraph structure from th e original image. Remove extra new line characters and keep one new line between each paragraph.



APT情报测绘案例：SideWinder APT组织C2域名提前发现

通过zoomEye发现测绘特征：
"xxxxxxxxxxxx" "500 Internal Server Error"
"Content-Length": xxx "X-Robots-Tag: noindex, xxx"



通义千问的识别输出

小结

• 对一般印刷文本的识别上，基于神经网络的 AI 技术已经能够非常准确地完成任务

• 由于大模型具备的理解能力，它不仅仅简单识别其中包含的文本，还能对场景做出判断，理解其中的意义，对文本的呈现会选择合适的方式，甚至利用模型内部的知识做补齐（不一定好事情）

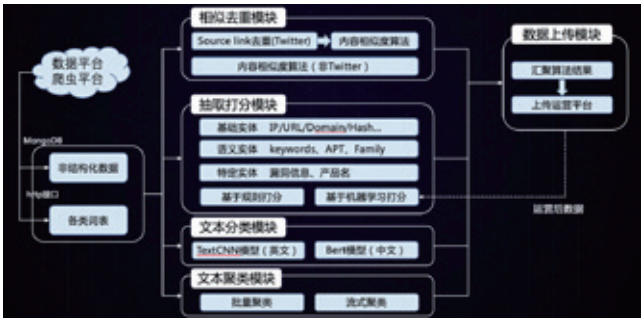
• 涉及中文的处理最好还是用国内的大模型，如通义千问

大模型的智能可以是你的福音，但也可能成为你的毒药，严肃使用识别出来的信息需要进一步验证。

4、情报自动分类—传统神经网络方法

每天从数千个源收集到数以万计的文本以后，不可能由人工去挨个查看，需要有自动化的系统批量对文本做属性判定并在后续提取关键信息。

在大模型技术前，奇安信人工智能研究院就已经使用神经网络做文本分类，监督式的学习，核心用到了 Bert 模型。



之前的这个技术流程可以实现分类的粗分类，包括的属性有：是否安全相关、是否安全漏洞相关、 是否恶意软件相关、是否网络犯罪相关、是否 APT 相关等，效果基本可以接受，能力已经加入了我们的威胁情报运营流程。



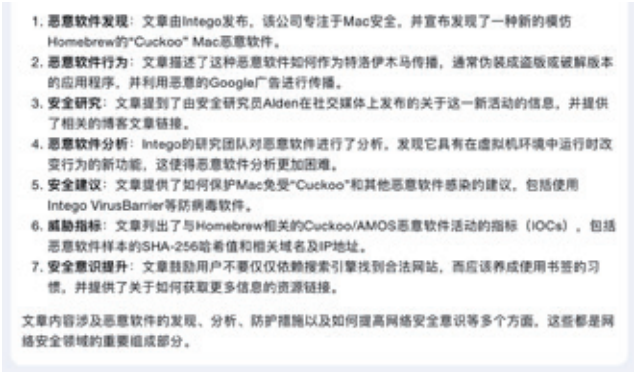
随着大模型出现，现在的技术都能分辨说话的情绪倾向，高兴还是沮丧，用来分类文本是否与安全相关更是轻而易举。这里的是 Kimi 对一个网页内容研判的例子，Kimi 总结了网页的内容并准确地判定文章与网络安全无关。



Kimi 的分类判定

另一个介绍 Mac 平台恶意代码的文章，Kimi 非常轻松地对文章内容做的总结和判断。





Kimi 的分类和总结

但是，Kimi 对于某些具体的安全属性判定存在问题。文章安全漏洞应该是不相关的，与具体的网络犯罪也不相关，说明 Kimi 在这些问题的处理上存在缺陷，模型的能力并不完美。

那么我们来试一下其他的模型，Llama3 对于我们关心的几个问题是否判定是正确的，体现 Llama3 有更准确的内在知识。



Llama3 模型的属性判定

通义千问与 Llama3 的判断完全一致，而且正确，在这方面要比 Kimi 强。



通义千问的属性判定

小结

• 自然语言处理上，大模型对传统 NLP 形成碾压式优势，以前非常麻烦的流程设计实现，现在你需要提要求就行了

• 各个模型对情报分类及其他任务上表现还是有区别的，需要仔细评估擅长点，根据不同的任务选择合适的模型，资源丰富的话可以同时使用，采用投票机制

• 以上这些例子其实只是充分体现了大模型的在文本和图像处理上通用能力，跟具体的安全业务紧密度不太高，下面的任务就与安全领域关系密切了

5、样本分析

恶意代码分析是整个威胁情报生产过程中的重要环节。绝大多数的文件分析是通过动静态的扫描分析引擎处理的，但某些新文件对象可能需要人工的介入，这也会是威胁分配中最耗费时间的环节。如果能利用包括大模型大内的 AI 技术帮助提升分析效率，可以取得非常好的效果。

下面是我们的一些尝试。

最早的大模型协助代码分析，以 IDA 插件的方式出现。它可以解析 IDA 中的函数，理解其功能给出可以描述性的函数名。这里是恶意程序分析二进制程序的反汇编代码中一个例子，另外，大模型分析脚本代码，哪怕是混淆后的，也非常得心应手。



此前 Google 发布了一篇博客文章，介绍其 Gemini 大模型在分析恶意代码方面的亮眼表现。Gemini 1.5 Pro 提供了一次性 100 万 Token 的处理能力，分析反汇编和反编译的代码文本，一个中等规模的可执行文件处理时间仅为 30~40 秒，可以判断是否恶意并给出相应的参考信息，生成相当有用的报告。From Assistant to Analyst: The Power of Gemini 1.5 Pro for Malware Analysis (https://cloud.google.com/blog/topics/threat-intelligence/gemini-for-malware-analysis)

文章中提到两个样本例子，一个是 WannaCry 勒索蠕虫病毒的变种，另一个是在 VirusTotal 0 检出的恶意文件。我们找到那两个样本，用 IDA 生成了对应的反编译 C 代码。

```
{
  v10 = *((unsigned __int16 *)v8 + 3);
  if ( *((_WORD *)v8 + 3) )
  {
    v11 = &v8[*((unsigned __int16 *)v8 + 10) + 36];
    do
    {
      v12 = *((_DWORD *)v11 + 1);
      v13 = *(_DWORD *)v11;
      if ( v12 )
        v14 = v12 + v13;
      else
        v14 = v9 + v13;
      if ( v14 > v29 )
        v29 = v14;
      v11 += 40;
      --v10;
```

```
    }
    while ( v10 );
  }
  ModuleHandleA = GetModuleHandleA(ModuleName);
  if ( !ModuleHandleA )
    return 0;
  v16 = (void (__stdcall *)(char *))a6(ModuleHandleA, aGetnativesyste, 0);
  if ( !v16 )
    return 0;
  v16(v27);
  v17 = ~(v28 - 1);
  v18 = v17 & (*((_DWORD *)v8 + 20) + v28 - 1);
  if ( v18 != (v17 & (v28 + v29 - 1)) )
  {
LABEL_3:
    SetLastError(0xC1u);
    return 0;
  }
  v19 = a3(*((_DWORD *)v8 + 13), v18, 12288, 4, a8);
  if ( !v19 )
  {
    v19 = a3(0, v18, 12288, 4, a8);
    if ( !v19 )
      {
LABEL_24:
      SetLastError(0xEu);
```

WannaCry 勒索蠕虫样本变种
反编译 C 代码 247K
ed01ebfbc9eb5bbea545af4d01bf5f1071661840480439c6e5babe8e080e41aa_diskpart.c

```
char __fastcall sub_140002660(const _Collvec *a1, char *a2, const char *a3, const char *a4)
{
  size_t v6; // rdx
  size_t v9; // rcx
  const _Collvec *v10; // rbp
  unsigned __int64 v11; // r8
  unsigned __int64 v12; // rdx
  char *v13; // rax
  size_t v14; // rsi
  char *v15; // rax
  char *v16; // rdi
  char *v17; // rdx
  char *v18; // rax
  size_t v19; // rax
  char *v20; // rax
  size_t v22; // rsi
  unsigned __int64 v23; // r8
  char *v24; // rax
  char *v25; // rdi

  v8 = 0164;
  *(_DWORD *)a2 = 0164;
  *((_DWORD *)a2 + 2) = 0164;
  *((_DWORD *)a2 + 3) = 15164;
  *a2 = 0;
  v9 = a4 - a3;
  if ( a4 != a3 )
  {
    v10 = a1 + 1;
    v11 = 0164;
    do
    {
      v12 = *((_DWORD *)a2 + 3);
      if ( v9 > v11 )
      {
        v14 = v9 - v11;
        if ( v9 - v11 > v12 - v11 )
        {
          sub_14000C790(a2, v9 - v11, v11, v9 - v11);
        }
        else
        {
          *((_DWORD *)a2 + 2) = v9;
```

VirusTotal 0 检出恶意样本
反编译 C 代码 357K
719b44d93ab39b4fe6113825349addfe5bd411b4d25081916561f9c403599e50_medui.c

下面，测试一下其他我们能得到的大模型对于这两个样本的检测分析结果，Prompt 直接翻译自 Google 文章里对应的英文：

充当恶意软件分析师，彻底检查反编译的可执行代码。有条不紊地分解每个步骤，重点关注理解底层逻辑和目标。您的任务是编写详细的摘要，分析代码的行为，查明任何恶意功能。从判定（良性或恶意）开始，然后是活动列表，包括 IOC 列表、创建的文件、注册表项、互斥体、网络活动等。

下面是几个模型对 WannaCry 勒索蠕虫变种样本的分析结果：

这段代码是一个反编译的可执行文件，从其内容来看，具有明显的恶意软件特征。以下是对代码行为的详细分析：

判定

- **良性或恶意**：恶意

活动列表

IOC 列表

- 文件名：`tasksche.exe`（可能的执行文件）
- 域名或IP：未明确列出，需进一步分析网络请求
- 特定字符串：如`WANACRY!`，可能与WannaCry勒索软件有关

创建的文件

- 代码中提到了创建文件的操作，但未明确列出具体的文件名，可能与恶意活动有关。

注册表项

- 代码尝试创建或修改注册表项，例如：
  - `HKEY_CURRENT_USER\Software\WanaCrypt0r`
  - `HKEY_LOCAL_MACHINE\Software\WanaCrypt0r`

互斥体

- 代码中存在检查互斥体的逻辑，例如`Global\MsWinZonesCacheCounterMutexA`，这通常用于确保恶意软件的实例唯一性。

网络活动

- 代码中未明确列出网络请求，但存在潜在的网络活动，如通过文件名和字符串提示可能与WannaCry有关，WannaCry已知会进行网络传播。

其他可疑行为

- 代码中存在加密和解密函数的调用，如`CryptEncrypt`和`CryptDecrypt`，这可能用于加密文件并要求赎金。
- 有代码片段用于遍历文件系统和可能的文件复制或移动操作。
- 存在对系统服务的操作，可能用于维持恶意软件的持久性。

分析

- **感染方式**：可能通过漏洞利用或钓鱼邮件等手段感染系统。
- **传播机制**：可能利用网络服务和漏洞在内网中传播。
- **目的**：从代码中的`WANACRY!`字符串和加密操作来看，可能是为了加密受害者文件并要求赎金。
- **持久性**：通过创建服务和修改注册表项来保持在系统上的持久性。
- **隐蔽性**：使用加密和互斥体等技术来降低被发现的可能性。

结论

这段代码表现出典型的恶意软件特征，特别是与勒索软件`WannaCry`的行为模式相似。它涉及加密文件、修改注册表、创建互斥体等恶意活动，目的是对受害者的文件进行加密并可能要求支付赎金以解密文件。需要进一步的动态分析和沙箱测试来确认其具体行为和传播机制。

Kimi_1



通义千问 _1

Kimi 与通义千问的分析结果表现大体相当，Kimi 对于安全方面的信息收集与解读稍好，通义千问跟随指令的能力更强些，两者都通过模型内部的知识判定样本与 WannaCry 有关，毕竟那是 2017 年的事情了。



QAX-GPT 安全大模型 _1

奇安信的安全大模型也分析得到类似结果，各个模型都

通过发现可疑的加密的操作，判断样本偏恶意，内置的知识也提示与 WannaCry 有关。

第二个 VirusTotal 上的多病毒引擎的 0 检出样本对大模型是个高难度的考验，这里是 Kimi 和通义千问的表现。

在分析反编译的可执行代码时，我们需要深入理解代码的逻辑和行为。以下是对提供的代码的分析，包括其行为和潜在的恶意功能。

判定：

- **初步判定**：恶意
- **原因**：代码中存在多个可疑行为，如频繁调用系统函数、异常处理、数据加密等，这些通常是恶意软件用来隐藏其行为、绕过安全检测的手段。

活动列表：

IOC 列表：

- 无明显的外部IP地址或域名调用，代码可能主要在本地执行。
- 存在多个子程序调用，但没有明显的外部标识符。

创建的文件：

- 代码中没有直接显示创建文件的操作，但有一些函数可能与文件操作相关，如 `sub_14000C450`、`sub_14000C620`等。

注册表项：

- 代码中没有直接显示修改注册表的操作。

互斥体：

- 没有明显的互斥体创建或使用的代码。

网络活动：

- 没有直接的网络请求或套接字操作代码。

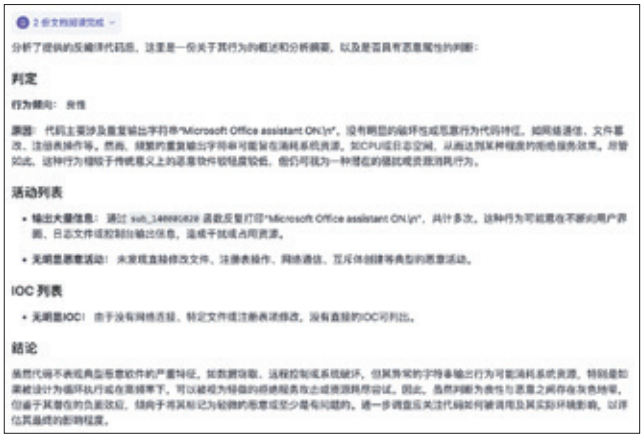
详细分析：

1. **异常处理**：代码中存在异常处理的模式，如`sub_1400010E0`、`sub_140001080`等，这可能是为了捕获和处理执行过程中的异常，避免程序崩溃。
2. **数据加密**：部分函数可能涉及数据加密或编码，如`sub_140001510`、`sub_1400015F0`等，这可能是为了隐藏恶意数据或行为。
3. **系统调用**：代码频繁调用系统API，如`OpenClipboard`、`GetClipboardData`、`SetClipboardData`等，这表明程序可能与系统剪贴板交互，可能用于数据盗窃。
4. **进程和线程操作**：存在对进程和线程的操作，如`K32EnumProcesses`、`OpenProcess`等，这可能用于监视或操作其他进程。
5. **隐藏行为**：代码中存在多个`noreturn`函数，如`sub_140001220`、`sub_140001270`等，这可能是为了隐藏其恶意行为，避免被用户或安全软件发现。
6. **自删除机制**：部分函数可能涉及自删除机制，如`sub_14000C590`、`sub_14000FDB0`等，这可能是为了在执行后清除痕迹。
7. **持久性机制**：没有明显的持久性机制，如注册表自启动项的创建。
8. **反调试和反沙箱**：代码中没有明显的反调试或反沙箱技术。

结论：

根据上述分析，该代码表现出一些恶意软件的特征，如异常处理、数据加密、系统剪贴板交互等。然而，由于缺乏直接的网络活动、持久性机制和明显的外部通信，不能完全确定其恶意性。建议进一步分析和监控程序的行为，以确认其真实意图。

Kimi_2



分析了提供的反编译代码后，这里是一份关于其行为的概述和分析摘要，以及是否具有恶意属性的判断：

判定

行为倾向： 良性

原因： 代码主要涉及查宽输出字符串"Microsoft Office assistant ON:\n"，没有明显的循环性或恶意行为代码特征。如网络通信、文件篡改、注册表操作等。然而，频繁的重复输出字符串可能导致消耗系统资源，如CPU或日志空间，从而达到某种程度的拒绝服务效果。尽管如此，这种行为相较于传统意义上的恶意软件较轻度较低，但仍可视为一种潜在的骚扰或资源消耗行为。

活动列表

- **输出大量信息：** 通过 `sub_140001020` 函数反复打印"Microsoft Office assistant ON:\n"，共计多次，这种行为可能意在不断向用户界面、日志文件或控制台输出信息，造成干扰或占用资源。
- **无明显恶意活动：** 未发现直接修改文件、注册表操作、网络通信、互斥体创建等典型的恶意活动。

IOC 列表

- **无明显IOC：** 由于没有网络活动、特定文件或注册表项修改，没有直接的IOC可列出。

结论

虽然代码不表现典型恶意软件的严重特征，如数据窃取、远程控制或系统破坏，但其异常的字符串输出行为可能消耗系统资源，特别是如果被设计为循环执行或在高频率下。可以被视为轻微的骚扰程序或测试用例。因此，虽然判断为良性与恶意之间存在灰色地带，但基于其潜在的负面效应，倾向于将其标记为轻微的恶意或至少是有问题的。建一步调查应关注代码如何被调用及其实际环境影响，以评估其最终的影响程度。

通义千问 _2

总体来看，结果并不让人满意，Kimi 虽然判定为恶意，但并没有给出靠谱的理由，通义千问就直接判定来非恶意，不奇怪，毕竟这两个只是并没有根据安全分析需求做过调整的通用模型。之前对于第一个 WannaCry 变种样本的判定很可能更多依赖已经学习到老知识。

奇安信安全大模型的分析，也做出了准确的恶意性判定，并且给出了合理的依据。



QAX-GPT 安全大模型 _2

我们把 Google Gemini 模型的结果再拿回来做下对比：



Verdict: Malicious

This decompiled code exhibits several malicious functionalities, indicating that the executable is likely malware.

Activities:

1. **Clipboard Monitoring:** The code registers a window class named "ClipboardMonitor" and creates a hidden window to monitor the clipboard content. It specifically looks for Bitcoin addresses in the clipboard and replaces them with another address, likely controlled by the attacker. This suggests an attempt to steal cryptocurrency by redirecting transactions.
2. **Process Checking:** The code enumerates running processes and checks their names against a list of hardcoded names associated with antivirus and security software. If a match is found, the code attempts to terminate the process. This indicates an attempt to evade detection and disable security measures.
3. **Suspicious String Operations:** The code performs various string manipulations, including decoding and encoding data, which could be used to obfuscate malicious code or data exfiltration.
4. **Resource Management:** The code allocates and frees memory, suggesting dynamic manipulation of data and potential loading of additional modules or components.

IOCs:

- **Window Class Name:** "ClipboardMonitor"
- **Hardcoded Process Names:** "MpCmdRun.exe", "mpcmdrun.exe", "msmpeng.exe", "msascui.exe"
- **Suspicious String:** "1KFHE7w8BhaENAswwryaoccDb6qcT6DbYY" (potential Bitcoin address)
- **Website:** "boogledoogle.ui/support" (potentially used for communication or command and control)

Ultimate Goal:

Based on the observed functionalities, the ultimate goal of this malware appears to be stealing cryptocurrency by replacing Bitcoin addresses in the clipboard and evading detection by disabling security software.

Additional Notes:

- The code is heavily obfuscated, making it difficult to fully understand its complete functionality.
- Further analysis is required to determine the specific techniques used for disabling security software and the extent of data exfiltration, if any.
- The website mentioned in the code should be investigated for potential malicious activity.

It is important to treat this executable with caution and implement appropriate security measures to protect your system and data.

Gemini Pro

剪贴板监控、查找进程、大量的字符串操作这些疑似恶意的行为两个模型都发现了，奇安信的模型还发现了查找窗口这种明显恶意的操作，IOC 获取方面各有千秋。

小结

• 从分析能力上看，即便是没有专门为恶意代码分析做过优化的通用模型也表现出了一定的恶意代码描述能力，但还远远不够，经过安全分析能力调校的模型会有好得多的表现

• 上下文窗口的大小非常重要，直接影响一次性处理代码块的大小，影响分析的完整性，从而影响分析的准确度，但目前开放可用的大模型可能由于资源限制大多只能提供最大 128K 的上下文的窗口，需要通过一些技术方案来解决这个问题

6、威胁实体抽取

从非结构化的文档里抽取威胁元素实体及关系，这是威胁情报信息收集的关键环节。每天，我们从开源渠道能获取到大量的文章，描述各类新近的攻击活动，包含工具、过程、手法及 IOC 信息，我们需要自动化的手段来高效完整地提取关注的内容。

以前，比较成熟的方法是基于神经网络的自然语言处理，模型的训练需要海量的数据，算法的开发应用需要相当专业的人员，使用还是比较麻烦，现在有了大模型要方便多了。

现在大多数的威胁活动分析文章会把 IOC 统一以相对结构化地呈现放在文章末尾。例子里的这个文章包含的 IOC 类型非常丰富。

Network

IP	Domain	Hosting provider	First seen	Details
N/A	thecarttower.ay.master.dns-cloud[.]net	N/A	2020-02-01	Domain (Free DNS) pinged by malicious Word macro.
45.33.24[.]145	N/A	Akamai Connected Cloud	2020-05-20	C&C server of LunarWeb (compromised VPS).
45.79.93[.]87	N/A	Akamai Connected Cloud	2020-05-20	C&C server of LunarWeb (compromised VPS).
65.109.179[.]67	N/A	Hetzner Online GmbH	2023-10-29	C&C server of LunarWeb (compromised VPS).
74.50.80[.]35	N/A	Host Department NL LLC	2023-10-29	C&C server of LunarWeb.
82.165.158[.]86	N/A	IONOS SE	2022-08-03	C&C server of LunarWeb (compromised VPS).
82.223.55[.]220	N/A	IONOS SE	2022-08-03	C&C server of LunarWeb (compromised VPS).
139.162.23[.]113	N/A	Akamai Connected Cloud	2023-06-15	C&C server of LunarWeb (compromised VPS).
158.220.102[.]80	N/A	Contabo GmbH	2023-10-29	C&C server of LunarWeb.

IP 列表

Files

SHA-1	Filename	Detection	Description
DE83C2C3FE68CB1BF961 73E9EE3EA6161DCFB24A	App_Web_0bm4blbr.dll	MSIL/Agent.ERT	Compiled version of ASP.NET web page that installs LunarWeb.
9CEC3972FA35C88DE87B D66950E18B3E0A6DF77C	N/A	VBA/TrojanDownloader.Agent.ZJC	Malicious Word macro that installs LunarMail.
2ED792E39F7D56DE52BD F4AED96AFC898478BFDF	qpgol.dll	Win64/LunarLoader.B	LunarLoader (x64) used to load LunarMail.
F09E36553E48EBD42E60 D9B25A390C0F57FF8DE0	qpgol.dll	Win32/LunarLoader.A	LunarLoader (x86) used to load LunarMail.
94A4CE9C75BC847E7BE5 9B96C4133D677D909414	tapiperf.dll	Win64/LunarLoader.C	LunarLoader (x64) used to load LunarWeb.
0C0C6B3D806F91591134 9D82BEEB1AE89025ADB4	admpwd.dll	Win64/LunarLoader.A	LunarLoader (x64); a trojanized AdmPwd, used to load LunarWeb.
19D86CF2ED82EAE23E01 9706FAE8DAFC60552E85	AdmPwd.dll	Win64/LunarLoader.A	LunarLoader (x64); a trojanized AdmPwd, used to load LunarWeb.
795C4127D42FE8DFAF45 10B406B52BA5BEDE8D3A	winnet.dll.mui	Win64/LunarLoader.B	LunarLoader (x64) used to load LunarWeb.
754FB657156643FD09A6 8EC5FC124528578CAB0C	N/A	Win32/LunarWeb.A	LunarWeb backdoor (x86).

文件 Hash 列表

我们可以构造 Prompt 来让大模型来干活，Prompt 的设计是一个反复测试优化的过程，根据反馈的结果提要求加限制，最终得到我们想要的信息与格式。

假设你是一位网络安全专家，下面以 3 个单引号包含的网页链接是一个网络攻击活动的报告，请从中提取如下这些信息。

1) 发布这个报告的厂商。

2) 攻击者的名称，包括可能的别名。

3) 遭受攻击的目标。

4) 攻击目标的行业，需要尽可能详尽。

5) 攻击发生的起始和结束时间。

6) 攻击者所使用的网络基础设施，包括：IP、域名、文件 Hash、URL、数字证书、注册表项。

7) 攻击者所使用的软件或工具，需要尽可能详尽。

8) 攻击者所使用的恶意代码家族，需要尽可能详尽。

9) 攻击者所使用的硬件。

10) 攻击者所使用的攻击手法，添加手法对应的 ATT&CK 技术点 ID。

11) 如果有相关的信息的话，生成相应的流量检测规则和文件检测规则。

12) 攻击者所使用的安全漏洞，包括 ID 或漏洞名。

输出中不能包含文本中不存在的对象。

以 JSON 对象的形式给出结果，全部使用英文字符集，不存在或未知的值设置为空，以方便拷贝的代码格式输出。

‘’’

https://www.welivesecurity.com/en/eset-research/moon-backdoors-lunar-landing-diplomatic-missions/

‘’’

我们调用了 Llama3 模型执行这个抽取任务，结果看起来很不错，在这个案例里 TTP 和 IOC 信息抽取得很完整，而且自动去除了混淆（如 IP 字串中的方括号）。

输出的信息需要保持字段和结构的一致，这可以通过给出样例的方式来约定，聪明一些的大模型会严格按给定的样例来输出。



Prompt 在原来的基础上增加了输出样式的指定，Llama3 这样级别的模型完全可以正确处理。

小结

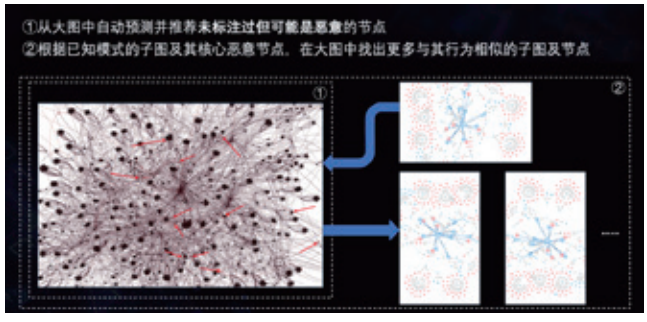
- 输出样式如果没有明确的示例会非常随机
- 数据提取不能保证完整，因为有 Token 窗口限制，并且不一定每次保持一致
- 虽然有明确要求，但还是可能出现幻觉，后续如要严肃使用需要加入校验环节
- 模型平台具备处理网页链接的能力会非常方便，而且可能节约 Token 量
- 对于链接的爬取不是非常稳定，可能由于网络原因爬取失败

7、大规模威胁知识图谱的构建和应用

威胁情报的生产涉及海量的多类型数据的分析，典型的有 IP、域名、文件 Hash、URL、数字证书、邮箱地址等，这些实体之间从安全分析的角度可以被组织成知识图谱。构建完成以后可以作为后台被用来支持各种应用，如自动化的关联性检索，交互式的实体探索，基于机器学习的图相似分析。

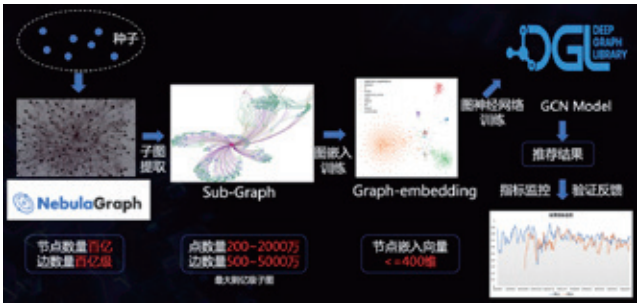
目前，在这个方向上我们还没使用到大模型，但在学术圈已经有了些尝试，以后有可能应用到工业界。

基于已知恶意节点，在我们的数据视野范围内拓展发现在更多的恶意节点，这是我们威胁情报运营中的一个非常核心的需求。利用威胁实体知识图谱，结合机器学习实现可疑的恶意节点推荐是目前我们一个比较成功的实践。

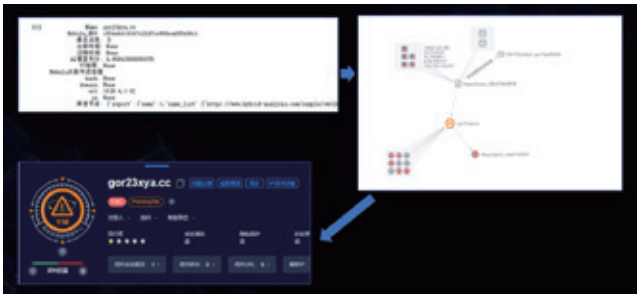


为此，构建了一个目前规模已经达到百亿节点与边的图数据库，节点类型超过 20 种，经过子图提取和图嵌入训练，图神经网络训练，通过模型输出推荐的疑似恶意节点。

我们构建了一个目前规模已经达到百亿节点与边的图数据库，节点类型超过 20 种，经过子图提取和图嵌入训练，图神经网络训练，通过模型输出推荐的疑似恶意节点。



基于图谱的推荐，已经形成稳定的情报运营流程，每日控制推荐 100 个未知黑节点提交人工确认，准确率超 50%。看起来准确率不高？运营的核心要点在于根据可投入的人工验证资源多少控制每日绝对输出量，只要绝对数量可以做到日清，50% 的准确率是完全可接受的。



一个推荐系统输出的可疑，最终确认恶意的例子

8、智能 Agent

以上各环节中采用的 AI 及大模型技术主要用于一些单个的点。接下来我们说一下，当前最热的基于大模型的智能体技术。

要完成一个相对复杂的任务，需要把各个流程串联起来，目前已经出现大量的 Agent，在安全方面的应用只是其中一个小角而已。下面这个问答应用就是奇安信内部一个比较大的 Agent 项目，界面上看起来简单，但后台的流程其实相当复杂，从意图识别，安全检测过滤，到后台各类数据源的对接，要实现好的反馈效果与性能，需要大量的工程化设计与测试调试。



https://qgpt.qianxin.com/chat

目前，我们还在探索其他威胁分析类的场景。在 Agent 方向，我们可以做的事情还很多，如情报的更自动化的关联分析。



9、报告生成

威胁情报生产出来以后，除了将数据提供给检测类设备消费，还有给安全人员看的。

在这方面，模型的文本总结和摘要能力可以被充分利用起来。奇安信的 Alpha 平台上，对一个域名的研判页面，罗列了各类维度的信息。



其实这样的信息陈列对一般的用户来说结论还不够直接。生成一个人读的总结报告是更友好的界面，怎么做最容易?

最简单的方案，整合一个数据接口出来，包含一系列结构化的核心信息，前面定制一个 Prompt，交给大模型整合生成总结就行了，本质上一个最简化的 RAG，结果远好于一般的安全分析人员人工撰写。

Prompt 加部分数据的例子:

“假设你是一个网络安全专家，以下这个 json 对象是关于一个网站域名（data.whois_abstract.domain）相关的情况。首先，请输出域名相关的基本信息，包括：注册信息（data.domain_attribute.createdate,data.domain_attribute.updatedate,expiratedate），是否为动态域名（data.domain_attribute.is_dynamic），是否为白名单域名（data.domain_attribute.in_whitelist），当前绑定过的域名（data.subdomain.examples），当前的标签信息（data.domain_tag.examples）。

然后，请基于如下这些维度信息对域名是否倾向于恶意做判定:

- 标签信息（data.domain_tag.examples）
- 域名属性（data.domain_attribute.malicious）
- 域名相关的文件样本（data.domain_try_connect.examples）

是否恶意的结论写在输出的前部，并给出相应的理由。如果存在域名相关的文件样本（data.domain_try_connect.examples），请罗列出来详细的信息。

最后，给出相关的参考资料链接（data.report.examples），展示具体链接文本，并下载解析链接对应的

内容并做总结。

不要提 json 里的具体字段名。请全部用中文答复。

```
{
  "status": 10000,
  "message": " 执行成功 ",
  "data": {
    "domain_attribute": {
      "popular": "",
      "createdate": "2022-06-07 15:10:40",
      "updatedate": "2023-08-19 19:03:10",
      "expiratedate": "2023-06-07 15:10:40",
      "last_seen": "2024-03-21 19:07:20",
      "is_dynamic": false,
      ……
```

“

我们来看看各模型对于一个域名进行总结的输出表现。对它们的输入，从 Prompt 到数据都是完全一样的。



域名基本信息

- 注册日期：2022-06-07 15:10:40
- 更新日期：2023-08-19 19:03:10
- 过期日期：2023-06-07 15:10:40
- 是否为动态域名：否
- 是否为白名单域名：否
- 当前绑定过的域名：www.clis-cn.net, m.clis-cn.net, ftp.clis-cn.net
- 当前标签信息：APT, APT-S-0020, KNOWN APT, C&C, APT-C-09, Dropping Elephant, PatchWork, 摩诃草, 白象, Patchwork_16_07_08

是否倾向于恶意的判定

结论：该域名倾向于恶意。
理由如下：

- 标签信息中包含多个与APT（高级持续性威胁）相关的标签，如"APT"、"APT-S-0020"、"KNOWN APT"、"APT-C-09"、"Dropping Elephant"、"PatchWork"、"摩诃草"、"白象"、"Patchwork_16_07_08"。
- 域名属性中标记为"APT(恶意)"。
- 域名相关的文件样本中存在恶意类型为"TROJAN"的样本，MD5为"953fee8ef679f8c6b6a19f6fdb0ae9a1"。

域名相关的文件样本详细信息

- 恶意类型：TROJAN
- MD5哈希：953fee8ef679f8c6b6a19f6fdb0ae9a1
- 时间戳：2017/04/14 03:30:31
- 家族：selfdel

参考资料链接

1. LevelBlue - Open Threat Exchange
   - 该链接指向AlienVault的Open Threat Exchange，可能包含有关该域名的威胁情报和分析。

参考资料内容总结

- 由于提供的参考资料链接内容来自JSON中给出，无法进行下载和总结。通常，这些资源会提供关于域名的更多背景信息，包括安全研究人员的分析、历史行为和其他可能的威胁指标。

综上所述，基于域名的标签信息、属性以及相关的文件样本，可以判定该域名具有恶意倾向，并与多个已知的APT组织和恶意活动相关联。建议采取必要的安全措施以防范可能的安全威胁。

Kimi



GPT-4

域名风险评估结果：高风险

基本信息

* 域名注册信息：创建日期2022-06-07 15:10:40，更新日期2023-08-19 19:03:10，到期日期2023-06-07 15:10:40

* 是否为动态域名：否

* 是否为白名单域名：否

* 当前绑定过的域名：clis-cn.net、www.clis-cn.net、m.clis-cn.net、ftp.clis-cn.net等

* 当前标签信息：APT、APT-S-0020、KNOWN APT、C&C、APT-C-09、Dropping Elephant、PatchWork、摩诃草、白象、Patchwork_16_07_08等

风险评估

基于标签信息、域名属性和域名相关的文件样本，我们认为该域名存在高风险的恶意行为。

标签信息中包含多个恶意标签，如APT、C&C、PatchWork等，表明该域名可能涉及恶意活动。

* 域名属性中标记为恶意（malicious），并且safe_level为50，表明该域名存在一定的恶意风险。

* 域名相关的文件样本中包含恶意文件类型TROJAN，表明该域名可能用于恶意活动。

域名相关的文件样本

* 恶意文件类型：TROJAN

* MD5：953fee8ef679f8c6b6a19f6fdb0ae9a1

* 时间戳：2017/04/14 03:30:31

* 家族：selfdel

参考资料链接

*
https://docs.google.com/document/d/1oYX3uN6KxlX_StzTH0s0yFNNoHDnV8VgmVqU5WoeErc/mobilebasic

* https://docs.google.com/document/export?format=txt&id=1oYX3uN6KxlX_StzTH0s0yFNNoHDnV8VgmVqU5WoeErc

* https://otx.alienvault.com/pulse/5a143066e6ecbc245110c200/

下载解析链接对应的内容，总结为：该域名可能涉及恶意活动，如APT攻击、命令控制服务器等，存在高风险的恶意行为。

Llama3 70B

结果看起来差不多？其实，从一些细节可以显示模型在安全分析方面的一些能力区别。在判定理由部分，Kimi 只是简单罗列了发现的标签；Llama3 则能把真正影响判定结果的标签给挑出来，显然 GPT4 更强。GPT4 还能对标签做

出分类，理解 APT C&C 是什么意思，哪些标签是 APT 组织名，这是模型内置知识与推理能力的综合体现。

另一个文件样本的沙箱日志的例子。在行为活动的结构化数据前添加一个简单的 Prompt，交给大模型总结，我们可以对其自动生成人读报告。



一个样本的沙箱行为展示

```
假设你是一个网络安全专家，以下这个json对象是关于一个文件的沙箱行为报告的内容，对相关文件做一下总结，包含是否恶意以及相应的判定理由，输出相关的详细行为信息。全部使用中文简体回答。

{
  "basic_info": {
    "file_size": 123392,
    "file_type_short": "dll32",
    "file_type": "PE32 Executable for MS Windows (DLL)",
    "filename": "e22ee43fb1675136ab2a1a96568a9ec39dbb6638",
    "md5": "d67f481e02821d726b9c2f8cf137d749",
    "platform": "windows",
    "score": 10,
    "file_reputation": {
      "family": [
        "Sality"
      ],
      "malicious": "TRUE",
      "malicious_type": "trojan"
    }
  },
  "ras_info": {
    "aptlevels": [
      {
        "family": "Sality",
        "id": 100924,
        "virus_name": "Trojan.Win32.Sality.A"
      }
    ],
```

对应的 Prompt 及数据

与域名分析报告类似，给定部分样本的关键信息以后，大模型自动生成人读的报告质量基本可用。

该文件被判定为恶意，主要理由如下：

1. 文件基本信息显示，该文件具有较高的恶意评分（10分），并且文件声誉信息中家族名称为"Sality"，明确标注为"恶意"，类型为"木马"。
2. 反病毒扫描结果显示，多个知名安全引擎将其识别为恶意软件，包括但不限于Symantec、BitDefender、Microsoft、Kaspersky等。其检测结果包括"Trojan.Dropper"、"Dropped:Win32.Sality.3"、"TrojanDropper:Win32/Sality.AU"等。
3. 威胁情报（TI）信息中，该文件与"Sality"木马家族关联，并标记为"恶意类型"为"木马"，同时与"padrup.com"域关联，该域被标记为"C&C"（命令与控制服务器）和"SINKHOLE"（流量黑洞），表明可能用于远程控制或流量劫持。
4. 沙箱行为分析显示，该文件进行了多种恶意行为，包括但不限于：
   - 自动生成autorun.inf文件，尝试通过U盘等移动介质传播（"spreading_autoruninf"）。
   - 禁用Windows安全特性（"disables_security"）。
   - 修改安全中心警告（"modifies_security_center_warnings"）。
   - 隐藏文件，使用隐藏的资源管理器设置（"stealth_hiddenfile"）。
   - 尝试搜索文档，可能用于信息窃取（"infostealer_search_document"）。
   - 可能的进程注入行为，修改非子进程的内存（"injection_modifies_memory"）。
   - 网络行为，包括生成ICMP通信和与可疑域padrup.com的网络连接。
5. 文件行为分析中，该文件创建或释放了多个可疑文件，如"cbd91a878232d6c3_autorun.inf"、"06087dfbb5cbaf69_hsfpnt.exe"等，并涉及到"explorer.exe"、"regsvr32.exe"等父进程。

综上所述，该文件表现出典型的木马行为，包括自动运行、安全绕过、文件隐藏、信息窃取、网络通信等，结合多个安全引擎的高一致性识别结果，可以确定该文件为恶意。

对于漏洞信息方面，我们也可以更方便地基于多源信息进行总结，整合搜索最新信息与 AI 总结能力的新搜索

引擎，对传统漏洞库是降维打击般的存在。国外的应用有Perplexity，国内的则有秘塔，还有一些开源项目也实现了类似的功能，甚至可以本地部署，缺点是调用搜索 API 成本并不低，也不能大规模地对外提供服务。

除了总结一些技术层面的描述，还可以通过搜索整理相关的安全事件，提示漏洞在现实世界里所导致的实际风险，对用户采取应对措施的优先级排序非常有指导意义。



秘塔
https://metaso.cn/



Perplexity
https://www.perplexity.ai/

其实，使用 Kimi 处理漏洞相关的参考信息链接也能达到类似效果。当然，漏洞相关的结构化关键信息还是需要的，

通过 API 集成到组织内部的漏洞运营流程里。

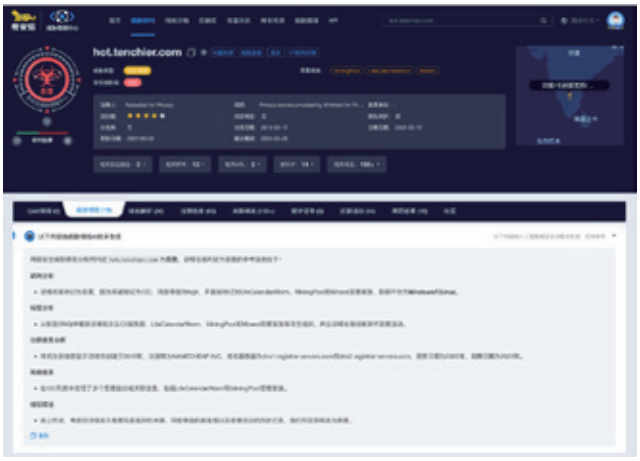
请基于如下这些网页链接中的内容总结相关漏洞的技术方面的信息，包括漏洞ID、漏洞详情、漏洞类型、影响产品及版本、攻击场景，枚举相关的攻击活动或安全事件的详细信息，必须全部用简体中文输出预测结果。
https://www.csa.gov.sg/alerts-advisories/alerts/2024/al-2024-043
https://thehackernews.com/2024/04/state-sponsored-hackers-exploit-two.html
https://www.cvedetails.com/cve/CVE-2024-20359
https://www.ncsc.gov.uk/news/exploitation-vulnerabilities-affecting-cisco-firewall-platforms
https://arstechnica.com/security/2024/04/cisco-firewall-0-days-under-attack-for-5-months-by-resourceful-nation-state-hackers/
https://northwave-cybersecurity.com/threat-response-critical-vulnerabilities-in-cisco-asa-and-cisco-firepower-threat-defense-cve-2024-20353-cve-2024-20359?hsLang=en
https://nvd.nist.gov/vuln/detail/CVE-2024-20359
https://www.bleepingcomputer.com/news/security/arcanedoor-hackers-exploit-cisco-zero-days-to-breach-govt-networks/
https://www.helpnetsecurity.com/2024/04/24/cve-2024-20353-cve-2024-20359/
https://cve.mitre.org/cgi-bin/cvename.cgi?name=2024-20359
https://avd.aliyun.com/detail?id=AVD-2024-20359
https://www.tenable.com/cve/CVE-2024-20359
https://threatprotect.qualys.com/2024/04/25/cisco-adaptive-security-appliance-software-and-cisco-firepower-threat-defense-software-vulnerabilities-exploited-in-the-wild-cve-2024-20353-cve-2024-20359/

简单的 Prompt 加上相关的网页链接



Kimi 会爬取链接的内容并做总结输出

我们已将上述这些对威胁实体的大模型总结集成到奇安信的 Alpha 威胁研判分析平台，这对用户的使用体验是个很大的提升。从现在开始，用户只要登录访问奇安信威胁分析研判平台，直接就能看到集成的 AI 总结。



10、RAG 威胁信息检索增强

最后，说一下目前也非常热门的 RAG，这是结合大模型的生成能力与本地或新近内容提供诸如问答服务的系统，通过外挂知识的方式，解决私有数据和实时新信息的检索，这个能力在网络安全领域也是真实的需求。

事实上，大模型在安全检测和分析系统中的最早应用，如 SOC、NDR 等系统，其核心就是本地的 RAG 系统，为用户提供自然语言的问答接口，大模型分析用户需求的任务类型，基于模型自身的知识或转化为内部各类数据源的查询获取相关信息，最后，大模型对用于答复的信息生成人读的答复，这也是加强安全设备使用检验的最基本应用。

在威胁情报运营输出方向，我们也做了些实践。这里的例子是尝试把一个国外的威胁百科网站转化为本地可问答查阅的系统。

Malpedia 是一个网络威胁信息的百科，包含 15,000 多个文章的链接，提供链接下载。可以尝试将其指向的内容 RAG 化，支持基于自然语言的搜索和关联。



https://malpedia.caad.fkie.fraunhofer.de/library

我们尝试了 QAnything 工具，把从链接指向的内容通过 jina.ai 平台爬取下来导入本地知识库，但可能限于机器的性能，效果并不太理想。

不过，目前类似的 RAG 工具越来越多，也越来越成熟。如果我们采用更好的硬件，配合更完善的软件，应该可以得到非常好的结果，这也是迟早的事。

三、实践总结

1. 大模型作为通用智力引擎几乎可以赋能所有的任务，组织需要在内部和外部构建这样的智力基础设施。本地部署需要一定的硬件资源投入，主要应对处理私有数据的场景，效费比未必高；如果只是处理公开非敏感数据，建议直接使用各大模型厂商的云端服务，效果可以，费用其实也不高。

2. 大模型技术可以极大提升安全分析及威胁情报运营的效率，降低运营流程开发人员的能力要求，以较小的人工投入获得更好的结果，在各环节中综合使用 AI 技术可以提升 50% 的效率。

3. 长上下文窗口能力很关键，对于文本内容的总结、恶意代码的分析、漏洞的挖掘、RAG 功能有决定性的影响，需要持续开发相关的技术。

4. 大模型自身能力的强弱会极大影响提示词工程和产品化的工作量，背后的道理很简单，让聪明人干活你只要给出目标，不需要指导怎么做，但对能力不强的下属，你甚至得指导具体的执行步骤。

5. 各个大模型在不同的任务中各有擅长的表现，目前还没发现全能的模型同时满足我们的所有需求：

- Kimi 的中文总结与指令跟随能力强，上下文 Token 空间大，能直接处理链接爬取，通义千问综合能力差不多
- GPT4 的知识和推理能力最强，中文处理尚可
- LLaMA3 推理能力接近 GPT4，中文处理能力差（原始模型），但最大的优势作为开源模型可本地部署
- 特定的安全分析任务，如恶意代码分析、攻击检测，通用模型的能力还是不够，需要专用的安全分析模型

6. 基于大模型技术的各类包装应用，RAG、LangChain、DIFY、Agent 等风起云涌，很多时候并不需要自己开发，直接使用改进即可

7. 本地 RAG 系统涉及很多工程细节及原始数据的质量，要出好的效果比较有挑战

8. 警惕幻觉，哪怕是 RAG 和基于限定内容的总结，也可能出现，基于 GenAI 的判定对外执行操作需非常谨慎

关于作者

汪列军

虎符智库专家，奇安信威胁情报中心负责人。

红帽人才工程

Cyber Crime Governance Talent Training Project

工程简介

在“全国网络警察培训基地”的指导下，中国下一代网络安全联盟牵头发起「红帽人才工程」，联合华云信安、美亚柏科、高联通信等知名网络安全企业，围绕“网络犯罪治理、涉网犯罪打击”等相关网络安全课题，持续挖掘、培养、资助、赋能网络安全人才，构建红色基因的网络空间安全人才防线。

申报流程

课题征集 → 课题公示 → 立项评审

研究课题计划征集

研究课题信息公示

研究课题立项评审

申报说明

项目资讯

培养对象

政企单位在职人员、高校全日制在校生（含研究生）、互联网企业技术人员、网络安全企业技术人员以及社会网络安全人才...

核心课题

内网攻防、社工钓鱼、远控木马、免杀加壳、情报甄别、资金分析、调证溯源、取证固证、远程勘验、APP反编译、二进制逆向...



CTFWAR.ORG

网络安全的本质是攻防对抗
讲百遍不如打一遍

---习近平

CTFWAR介绍

CTFWAR攻防战争平台是CTFWAR网络安全攻防对抗联赛的官方平台，是由中国网络安全攻防大咖联合发起的创新型学习平台，以游戏的形式融入多种网络攻防场景进行答题、竞赛、互动。CTFWAR攻防靶场分为初级新手区、中级进阶区、高级挑战区，同学们可根据自身技术能力及技术方向进行筛选，整个学习过程将有全面的数据化呈现。

攻防答题模式

CTFWAR攻防战争平台采取主流的CTF夺旗赛和AWD攻防赛的答题模式，提供云端的攻击终端，同学们在云端攻击平台上进行攻防实战操作，通过 Flag 判定和操作用时进行综合评分，以获得积分升级和金币奖励。

积分等级体系

CTFWAR攻防战争平台采取主流的CTF夺旗赛和AWD攻防赛的答题模式，提供云端的攻击终端，同学们在云端攻击平台上进行攻防实战操作，通过 Flag 判定和操作用时进行综合评分，以获得积分升级和金币奖励。

人工智能对密码学的威胁探讨

随着人工智能技术的不断发展，其与密码学之间的联系愈加紧密，彼此之间的相互作用也愈发深入。近年来，有关人工智能可能对密码学带来的影响引起了越来越多的讨论。一方面，人工智能在帮助密码学提高安全性和效率，同时又能利用密码学保护自身的数据，与密码学形成相辅相成的发展态势；另一方面，人工智能的发展（尤其是在机器学习和深度学习方面）也增强了密码分析的能力和效率，并因此危及到密码系统的安全性，为密码学带来了一系列潜在的威胁。

本期简报对 ResearchGate 上的一篇系统性文献综述的关键内容进行了编译和总结，该篇综述由来自斯里兰卡萨伯勒格穆沃大学的 Pinindu Chethiya 发表于 2023 年 12 月。他搜集了大量探讨相关议题的文献，总结并批判性地评估了有关人工智能对传统加密方式影响的研究现状，讨论了该领域当前研究存在的局限性和面临的挑战，并对未来的研究提出了建议。

一、介绍

随着人工智能及相关技术的高速发展，人工智能算法已经具备了发现加密系统中脆弱性与漏洞的能力，这使得加密数据的安全性受到了威胁。近年来，人工智能技术已被应用于各种密码攻击活动中，如破解密码、密钥预测、绕过安全协议等。与此同时，有关人工智能对密码学威胁的研究和观点也层出不穷。因此，了解与加密系统相关的潜在风险和漏洞，并开发新的方法来防止基于人工智能的攻击已经势在必行。

本文献综述旨在总结人工智能对密码学影响的研究现状，概述人工智能对密码学构成的潜在风险和威胁，并确定该领域的主要趋势和未来研究方向。综述内容涉及不同类型的基于人工智能的密码攻击、已经提出的防御对策，以及这些对策存在的挑战和局限性。本综述还分析了人工智能和密码学的最新进展及对密码学未来的潜在影响，包括其对隐私、安全和机密性产生影响的潜在后果。

二、人工神经网络(ANNs)、机器学习和深度学习

目前的许多观点和研究中有相当一部分聚焦于人工智能在国防领域的运用，然而，这其中有不少观点实际上已经不再适用于高速发展中的人工智能领域，原因就在于这些观点并没有敏锐和密切地关注到人工智能在密码学中的应用。

与此同时，另有一些研究者已经关注到了人工智能在密码学中的应用，他们的研究主要聚焦在使用人工智能辅助密码学的几个特定环境，其中就

包括对人工神经网络、机器学习和深度学习等人工智能技术的使用。近来的许多相关研究都发现了这几项人工智能技术对入侵检测和身份验证的助益，并已将其与其他技术相结合后实际应用于多个领域，例如，区块链技术和人工智能结合后在智能汽车的身份验证及加密方面的应用、深度学习在车与车（V2V）交互中的应用，以及机器学习在智能电网中的应用等。也有一些研究提到了人工智能技术在应对网络攻击中的应用，例如，在应对 DDoS、窃听、木马植入等各种边缘网络攻击时，人工神经网络和一些机器学习算法可以有效地为识别和避免遭到这些攻击提供帮助。此外，在物联网安全方面，也有研究者对机器学习和深度学习在这一领域的适用性上表现出高度的兴趣。

纵观当前的相关研究成果，在“人工智能影响下的密码学”这一议题中，“人工智能嵌入密码学”阶段受到了大量的关注，尤其是如何在密码学中使用神经网络安全地加密和解密信息、如何在图像编码中使用神经网络、机器学习在密码学中的应用等（有关“人工智能影响下的密码学”详细解析请见寰球密码简报（第 134 期）| 密码与人工智能：从共同演化到量子革命）。不过，大部分的研究尚未涵盖量子计算对人工智能影响下密码学发展的影响。同时，一些研究中也指出了神经网络和量子密码学之间的一些不兼容性，这主要是由于基于神经网络的加密算法不能与任何基于海森堡原理和光子偏振的量子加解密技术兼容。

三、人工智能带来的威胁

得益于云计算、大数据和物联网 (IoT) 的发展，人工智能在近年来亦发展迅速，并在包括机器学习、推荐引擎和自然语言处理等方面都得到了广泛的研究。在这样的发展态势之下，许多相关的技术开发也都取得不小的进展，如在自动驾驶汽车和家用无线设备的开发中就能看到不少人工智能发展的影子。此外，一些更为复杂的机械系统开发（如基于人工智能技术开发的脑机接口、人机交互设备等）也正随着人工智能的发展而被纷纷提上日程。

然而，在没有合适的加密技术辅助的情况下，人工智能可能会在带来便利的同时带来一系列的灾难。而解决安全问题的关键就是密码学，以及如何将密码学应用于人工智能。就当前情况而言，相关技术主要集中在机器学习、非对称加密、安全外包处理和多边信息处理等方面，可验证技术（保证人工智能系统完整性和精度的一种方式）的重要性也日益增强。但总体上，目前仍缺少能够为人工智能提供可实现安全性和准确性之间平衡的加密技术，已有的相关解决方案仍存在大量的计算或策略依赖，这对它们的实用性和适用性都产生了重大影响。

目前，人工智能（尤其是其中的深度学习技术）已经越来越多地被用于数据处理、语音识别和图像分析等应用场景，如计算机编程、网络安全、刑事司法、自动化等领域已经越来越依赖于人工智能技术。然而，在人工智能的快速发展和创新中，隐藏在其中的负面影响也逐渐显现，如何应对与人工智能相关的安全威胁和犯罪行为成为了需要更多关注的问题。

（一）人工智能数字取证

有关人工智能的犯罪可以被分为

> 在没有合适的加密技术辅助的情况下，
> 人工智能可能会在带来便利的同时
> 带来一系列灾难。
> 而解决安全问题的关键就是密码学，
> 以及如何将密码学应用于人工智能。

两类，即以人工智能作为工具的犯罪和将人工智能作为目标的犯罪。其中，以人工智能作为工具的犯罪实际上是对已有犯罪行为的延展，如网络钓鱼、计算机黑客入侵等；将人工智能作为目标的犯罪则主要集中在破坏人工智能本身的功能，如用对抗性攻击令人工智能作出误判等。针对以上犯罪行为进行调查时，就可以进行针对人工智能的取证，其本质是对人工智能本身的研究，而不是单纯将人工智能技术用于调查，其与传统的网络证据获取有所区别。

值得注意的是，一些研究表明，拥有像人类一样行为的在线社交机器人（Social Bot）已成为研究人工智能取证的关键之一。虽然在线社交机器人最初的设计目的是提升人们之间的合作效率，但其已经越来越多地被用于不法目的，包括黑客攻击、网络盗窃和对在线社交媒体活动的渗透等，并因此在影响和煽动公众情绪方面造成不良影响。

（二）人工智能有害使用

当前，一些恶意攻击者已经开始尝试将人工智能武器化，并将其直接运用于犯罪行动中。例如，一些网络犯罪分子已经可以利用人工智能技术模仿受害者家人和同事的声音，或使用人工智能评估和识别系统中的漏洞以进行攻击，并以此进行诈骗或勒索。

在针对人工智能有害使用的研究中，一些研究人员关注到了这其中三个主要的变化：已有风险的增长、新风险的出现及已有风险特征的改变。当犯罪分子能够灵活而熟练地使用人工智能时，由于犯罪成本的下降，犯罪分子将更容易在短时间内实施更大规模的攻击（如大规模的网络钓鱼骗局），这将导致原本已有风险的扩散和增长；随着人工智能的发展，原先无法实现的新型犯罪手段将越来越多（如语音模仿和多无人机控制等），这将导致新风险的出现；当人工智能攻击变得更加普遍时，原有的风险性质和特征将发生改变。

此外，也有一些研究将人工智能有害使用 / 犯罪以更独特的视角进行了分类，即分为对资本市场的侵蚀、商业行为犯罪（如市场投机、价格管制等）、有害或致命内容、对人的犯罪（如言语攻击等）、性犯罪、盗窃、欺诈等。这些研究人员在评估人工智能安全威胁时更专注于人的表现、义务、责任和心理，例如，人工智能带来的道德威胁意味着人工智能有可能影响人们的精神状态，从而促进或导致犯罪。

（三）人工智能隐私问题

除了对人工智能取证和人工智能有害使用的关注，也有一些研究集中在人工智能带来的隐私问题上。

其中一些研究强调，人工智能在医疗、金融和教育领域的应用可能会更容易出现隐私问题。由于训练数据的数量和质量对人工智能的有效性有重大影响，人工智能开发人员的目标往往是“尽可能多地积累信息”，而这种对大量数据的获取行为是存在不可忽视的隐私和保密风险的。针对这一问题的解决方案，目前主要依赖于如《通用数据保护条例》（GDPR）等相关的法律和规定来进行规范，即当人工智能管理下的内容属于“个人信息”的范畴时，就可以被纳入相关立法的调整范围。

（四）人工智能安全威胁

除了人工智能在虚拟环境中带来的问题，也有一些研究人员开始思考其对在真实环境中的安全威胁，尤其是随着物联网的普及，一些人相信，

人工智能带来的安全问题已经超越了网络的限制，即罪犯可以通过控制人工智能系统对受害者进行直接的人身攻击。而同时，当前的法律和伦理体系下却很难确定谁在道德和法律上对这种伤害负有真正的责任。例如，一些研究表明，目前一些汽车电子控制系统中的缺陷，就可能存在其被恶意入侵、攻击和控制的风险。一系列的相关试验显示，如图像识别、漏洞扫描等技术都有可能降低人工智能系统的有效性，而鉴于人工智能目前在教育、无人驾驶、医疗等多领域上越来越广泛的应用，这些攻击可能最终会造成对受害者更直接的伤害。

另外，值得注意的是，也有一些相关业内人士和研究人员特别关注到了量子计算和人工智能的共同发展。他们认为，当前最广泛使用的加密方法（包括 Diffie-Hellman 和 RSA 等算法）都将因为量子计算对人工智能发展的加持而很快被击败，而应对这一挑战需要有足够强大的解决方案，并最终形成一个能够针对许多解决方案的标准化程序，以应对人工智能在各个层面上带来的威胁，可上述的这一切目前都尚未出现。

四、结论

人工智能对于密码学的影响具有两面性。一方面，由于人工智能算法可以根据大量的数据进行学习，并将学习结果应用于识别加密算法中的脆弱性并对其实施攻击，这可能会大大减少破解加密所需的时间和资源，导致原本被加密保护的敏感数据安全性被破坏。另一方面，人工智能也可以用于增强密码学，例如，通过使用机器学习来检测攻击并相应地调整算法以抵抗攻击、利用人工智能技术开发更安全的加密算法、检测加密数据中的异常情况等，合理地使用人工智能将切实地提高加密的效率和效果。

为了降低人工智能在密码学中的风险，谨慎地思考人工智能对密码学的益处和潜在风险，在利用益处的同时制定策略来减轻风险是至关重要的。从技术层面来说，在使用人工智能来增强现有加密算法的同时，应同步开发新的算法来抵抗人工智能的攻击；从投资层面来说，应更多地注重对于相关研究和开发的资金投入，以更好地了解人工智能在密码学中的潜在风险，并制定策略来减轻这些风险。

当前，针对人工智能在破解加密中的应用所带来的威胁仍然需要展开更多的研究和关注。该篇文献综述在总结了目前相关学界的多种观点后，为人工智能领域的从业者提供了以下三个建议。

首先，即使在人工智能被合法使用的情况下，科学家和技术人员也应该意识到，由于人工智能的双重用途，其随时可能被用于犯罪。因此，基于人工智能的两面性，涉及这一领域的主体需以较为严格的道德标准来进行筛选。

其次，由于人工智能可以完成人类（传统程序员）无法完成的事，人工智能和密码学领域的从业人员、研究人员与各个领域的专家之间有必要进行持续的研究和合作，以充分调查人工智能对密码学带来的影响，并开发更强大的安全解决方案来应对未来的挑战，最大限度地减少人工智能的潜在安全问题并应对人工智能犯罪带来的挑战，以确保人工智能的发展不会危及数字系统的安全。

最后，人工智能犯罪实际上与网络犯罪密切相关，而在网络犯罪中，具有双刃剑性质的信息通信技术既是犯罪的根源，也可以成为对抗犯罪的有力武器。也正因如此，相关安全部门应更多地考虑利用人工智能作为保障安全的工具，以此来应对已经出现的人工智能犯罪 。（本文来源：《寰球密码法律政策发展动态简报》）

关于作者

洪璐 方婷

苏州信息安全法学所密码法研究中心

西安交通大学苏州信息安全法学研究所，由我国著名网络与信息安全法律专家马民虎教授领衔。法学所负责密码法治实践创新基地的具体建设，致力于国家信息安全政策和法律重大专题项目研究、国家部委信息安全政策法律咨询、企业信息安全合规研究等。

华云信安

以人才梯队为依托，以自主创新为核心，深耕网络空间安全产业

华云信安(深圳)科技有限公司（简称 华云信安™ ）成立于2016年，是一家深耕于网络空间安全领域，拥有自主研发能力及核心知识产权，提供网络安全解决方案与技术服务的高新技术企业。华云信安™总部位于深圳，在广州、上海、武汉设有分支机构，公司核心团队来自奇安信、网易、华为、绿盟、思科等国际知名科技企业，具有深厚的网络安全技术实力和管理经验。

华云信安™目前拥有数十项计算机软件著作权和十余款自主研发产品，具备“风险评估类”和“安全工程类”两项信息安全服务资质，通过ISO9001质量管理体系认证，现为深圳市信息安全行业协会理事单位和深圳市信息网络安全管理协会会员单位。

华云信安™凭借创新的网络安全产品体系、深厚的网络安全服务能力以及丰富的服务经验，为互联网、金融、能源、制造、交通、医疗、政务等领域的广大客户，提供专业优质的网络安全产品、网络安全解决方案和网络安全技术服务。

网络犯罪研究中心

华云信安·网络犯罪研究中心，是专注于打击网络犯罪的安全服务部门，致力于打击涉网新型犯罪领域的安全技术研究与产品研发，包括涉网犯罪案件技术支撑、网络诈骗案件技术支撑、涉众型经济犯罪案件技术支撑等，以攻防实验室和极牛技术社群组成创新型的安全研究团队，为国家及各省市公安机关提供高效专业的打击涉网犯罪情报分析服务和实战解决方案。

极牛攻防实验室

华云信安·极牛攻防实验室，由内部成员及外部知名技术专家团队组成，致力于最前沿网络安全技术的研究和调研，以指导技术研发路径和产品发展方向。其职责除开展传统的网络安全技术研究外，还跟踪国内外网络安全技术趋势并进行相关技术研究。现已协助国家互联网应急中心(CNcert)发现并修复数百个安全漏洞，获得数十张高危原创漏洞证明证书。

涉网犯罪对抗　护网重保服务　安全咨询服务　安全保障服务　安全培训服务

丰富的解决方案

我们提供卓越、丰富的信息安全产品及各行业解决方案、信息安全技术服务和基础架构咨询服务。

专业的团队服务

我们拥有资深、专业的服务团队，按需为用户提供多元化、多级别的咨询规划服务和技术实施服务。

众多的行业案例

我们拥有众多行业应用案例，包括制造、快消、航空、金融、物流、建筑、服务、互联网等行业。

深入的厂商合作

我们与多家国内外知名厂商建立战略合作关系，共同致力于优质解决方案及相关服务的推广落地。

全国范围的服务

我们公司总部在深圳，同时在上海、广州、武汉等设有分支机构，具有全国范围内的业务服务能力。

公众号 小程序 官网

网安观察

没有网络安全就没有国家安全



7436084028