AI增强勒索病毒：工作机理与防御方法

李业深[1,2]，董鹏[3]，朱贺[3]，郭孝天[1,2]，尹晨旭[1,2]，熊轲[1,2]
（1. 北京交通大学高速铁路网络管理教育部工程研究中心，北京 100044；
2. 北京交通大学计算机科学与技术学院，北京 100044；3. 中铁信（北京）网络技术研究院有限公司，北京 100044）

摘 要：随着数字经济的迅猛发展，网络安全风险日益加剧。据相关报道，勒索病毒已成为网络空间最具破坏性的威胁之一。值得警惕的是，网络黑客正在不断尝试利用先进的人工智能（AI, artificial intelligence）技术培育新型勒索病毒，使得病毒更智能、更具隐蔽性和破坏力。因此，如何全面审视 AI 对网络安全带来的新影响，深入揭示其工作原理并研究和构建有效的防御方法迫在眉睫。目前，尚未有文献系统全面地总结并分析 AI 增强勒索病毒危害的原理和影响。为此，首先，对勒索病毒进行了分类；接着，剖析了勒索病毒的攻击流程；然后，结合最新研究进展，深入阐述了 AI 增强勒索病毒的工作机理；最后，从预防、预测、检测、识别及缓解 5 个方面，系统归纳了基于 AI 的勒索病毒应对措施，并分析了 AI 增强勒索病毒的发展趋势与未来可能研究方向，旨在为网络安全领域的从业者提供有价值的参考与启示。

关键词：勒索病毒；网络安全；人工智能；防御体系
中图分类号：TP18；TP393.08
文献标志码：A
doi：10.11959/j.issn.2096−3750.2026.00511

AI-assisted ransomware: operating principles and defense methods

Li Yeshen[1,2], Dong Peng[3], Zhu He[3], Guo Xiaotian[1,2], Yin Chenxu[1,2], Xiong Ke[1,2]
1. Engineering Research Center of Network Management Technology for High-Speed Railway of Ministry of Education, Beijing Jiaotong University, Beijing 100044, China
2. School of Computer Science and Technology, Beijing Jiaotong University, Beijing 100044, China
3. Network Security Research Office, China Railway Information (Beijing) Network Technology Research Institute Co., Ltd., Beijing 100044, China

Abstract: With the rapid development of the digital economy, cybersecurity risks have become increasingly severe. According to relevant reports, ransomware has emerged as one of the most destructive threats in cyberspace. Alarmingly, cybercriminals are continuously leveraging advanced artificial intelligence (AI) technologies to develop next-generation ransomware, making these attacks more intelligent, covert, and damaging. Consequently, it is imperative to comprehensively examine the new impact of AI on cybersecurity, deeply reveal the operating principles of AI-assisted ransomware, and build effective defense strategies. At present, there is a lack of systematic and comprehensive literature analyzing the operating principles and impacts of AI-assisted ransomware. To address this gap, firstly, ransomware was categorized. Subsequently, the attack process of ransomware was analyzed. And then, combined with the latest research progress, the operating principles of AI-assisted ransomware were elaborated in depth. Finally, response measures to operating principles ransomware were systematically summarized from five key perspectives: prevention, prediction, detection, identification and

收稿日期：2025−03−18；**修回日期：**2025−07−01
通信作者：熊轲，kxiong@bjtu.edu.cn
基金项目：国家自然科学基金资助项目（No. 62071033）；北京市自然科学基金昌平创新联合基金资助项目（No. L234084）；中国国家铁路集团有限公司科研专项（No. L2023W001）
Foundation Items: The National Natural Science Foundation of China (No. 62071033), The Changping Innovation Joint Fund of Beijing Natural Science Foundation (No. L234084), The Project of China Railway Corporation (No. L2023W001)

mitigation. Additionally, the development trends and potential future research directions of AI-assisted ransomware were analyzed, aiming to provide valuable insights and guidance for practitioners in the field of cybersecurity.

Key words: ransomware, cybersecurity, artificial intelligence, defense system

0　引言

随着数字经济的快速发展，网络安全风险日益加剧。在众多网络安全威胁中，勒索病毒已成为最具破坏力的危害之一，被视为人类已知的最具传染性的“数据疾病”[1]。加之，提供勒索病毒定制服务的新型商业化运作模式“勒索软件即服务（RaaS, ransomware as a service）”大大降低了勒索攻击的技术门槛[2]，使得勒索病毒的传播和危害影响越来越大。据报道，2024年瑞星“星核”平台共截获病毒样本总量6 848万个，病毒感染次数8 474万次，其中，勒索软件样本总量4 243万个[3]。此外，Resecurity报告[4]指出，到2031年，勒索软件攻击造成的全球经济损失预计达到2 650亿美元，对全球企业造成的潜在总损失或达到10.5万亿美元。全球范围内，美国NCR公司、加拿大航空公司Air Canada等多个知名企业与机构曾遭受LockBit等勒索组织的攻击，受害企业涉及广泛，涵盖金融服务、科技、能源、医疗、运输等多个产业，严重影响着全球各国的经济和民生。

勒索病毒是一种新型的恶意程序，具有破坏性强、隐蔽性高、易传播等特点[5]。攻击者常利用钓鱼邮件、系统漏洞等方式传播病毒，通过加密用户文件、锁定用户设备等方式阻止用户访问，以文件解密密钥、设备解锁密钥或公开重要数据为筹码要挟受害者进行钱财勒索，或者将所窃取网络算力挖出的加密货币售出进行获利。不同于其他恶意程序，勒索病毒更注重以非法索取受害者财物为目的，巨大的经济利益直接推动了整个勒索病毒行业的快速发展。勒索病毒起源于1989年，由美国生物学博士Joseph Popp制造的勒索程序aids-trojan[6]，后经过不断升级迭代，由计算机程序演变为计算机病毒，具备了更强的传播与破坏能力。

近年来，勒索病毒发展迅速，衍生出多种类型。尤其是，随着人工智能（AI, artificial intelligence）技术的不断发展与普及，网络黑客正在不断尝试利用AI技术衍生出各类新型勒索病毒，使其具备智能和动态调整攻击策略的能力，具有更强的隐蔽性与破坏力。AI增强后的勒索病毒将具有新型传播手段与自动化勒索攻击的能力，因此会给个人、企业、政府等各类网络用户造成严重的经济损失和社会影响。鉴于此，网络防御者更应充分利用AI技术，构建更加智能、灵活的防御机制，为网络安全提供更加有效的增强手段。此外，在AI时代，如何全面审视AI对网络安全带来的新影响，深入揭示其工作原理并研究和构建有效的防御方法迫在眉睫。

为此，首先，对勒索病毒进行了分类；接着，剖析了勒索病毒的攻击流程；然后，结合最新研究进展，深入阐述了AI增强勒索病毒的工作机理；最后，从预防、预测、检测、识别及缓解5个方面，系统归纳了基于AI的勒索病毒应对措施，并分析了AI增强勒索病毒的发展趋势与未来可能研究方向，旨在为网络安全领域的从业者提供有价值的参考与启示。

1　勒索病毒种类

勒索病毒种类及特征如图1所示，根据勒索手段，可分为加密型、锁屏型、挖矿型、恐吓型、泄露型5种类型。

(1) 加密型：通过加密用户重要数据以勒索赎金[7-8]，是作案范围最广、造成经济损失最严重的勒索病毒，典型的有TeslaCrypt[9]、WannaCry[10]、CryptoLocker[11]、CryptoWall[12]等。

(2) 锁屏型：创建新桌面并使它持续覆盖在最上层，锁定用户设备，禁止用户操作，受害者通常只被允许查看被锁定并带有赎金支付指示的屏幕，典型的有LockScreen[13]、Winlocker[14]等。

(3) 挖矿型：从感染设备或网络中窃取计算资源以挖掘加密货币，然后通过售出所窃取的加密货币以获利，是一种与加密货币相关的勒索病毒，典型的有Coinhive[15]、Cryptoloot[16]等。

(4) 恐吓型：使用弹出式广告显示令人恐慌的信息来恐吓用户，使其在慌乱中下载或购买带有勒索病毒的恶意软件，典型的有FakeAV[17]、Reveton[18]等。

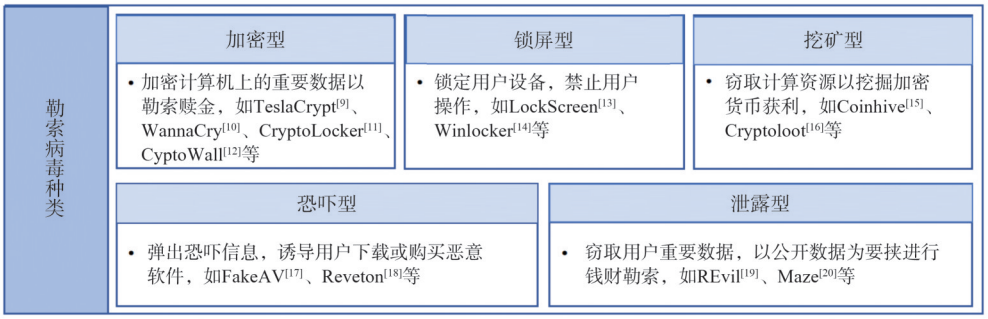

图1　勒索病毒种类及特征

(5) 泄露型：以处理机密或敏感信息的组织为主要攻击目标，以公开所加密或窃取的重要数据为要挟进行钱财勒索，典型的有REvil[19]、Maze[20]等。

2　勒索病毒攻击流程

虽然勒索病毒的种类不同，但其具有相似的攻击流程[21]。勒索病毒攻击流程如图2所示，通常由病毒传播、感染系统、通信、文件搜索、加密、勒索、解密、赎金转移8个步骤构成。

步骤1　病毒传播：勒索病毒具有多样化的传播方式，勒索病毒传播方式见表1，传统传播方式是“广撒网”式的被动型传播，包括社会工程学[7]、钓鱼邮件[22]、恶意广告[6]、恶意下载[6]等方式。而如今，更倾向于针对特定攻击目标的主动型传播[23]，包括系统漏洞[10]、协议漏洞[22]和僵尸网络[24]等方式。

图2　勒索病毒攻击流程

表1　　**勒索病毒传播方式**

传播类型	传播方式	特点	代表性工作
被动型传播	社会工程学	伪装成合法组织以获取对个人信息或密码的访问权限	文献[7]
	钓鱼邮件	发送伪装成合法来源的邮件或短信	文献[22]
	恶意广告	将恶意代码嵌入网页广告中	文献[6]
	恶意下载	将恶意代码植入合法网站中	文献[6]
主动型传播	系统漏洞	攻击系统漏洞获得受害系统的控制权	文献[10]
	协议漏洞	攻击系统协议漏洞获得受害系统的控制权	文献[22]
	僵尸网络	在终端执行僵尸程序，将其感染成为僵尸主机，执行恶意任务	文献[24]

步骤 2　感染系统：当勒索病毒感染系统后会启动感染程序进行一系列操作，包括收集用户信息、生成唯一的设备标识符来标识感染系统、频繁访问注册表等。

步骤 3　通信：勒索病毒在感染受害系统后，会与 C&C 服务器建立双向连接以实现加强控制、发送指令、传递数据、获取密钥等目的。

步骤 4　文件搜索：勒索病毒与 C&C 服务器完成通信后，病毒通常会对感染设备的文件进行扫描寻找重要数据，如格式为 .doc、.pdf 等类型的用户常用文件。

步骤 5　加密：寻找到目标文件后勒索病毒正式开展加密，加密方式主要有以下两种，一是使用 C&C 服务器获得的加密密钥来进行加密，二是在本地调用随机数应用程序接口（API, application program interface）生成加密密钥进行加密，被加密的文件通常带有勒索标识的后缀名。

步骤 6　勒索：勒索病毒会向用户弹出勒索弹窗，告知用户赎金数额以及支付方式。为使受害者尽快妥协并支付赎金，勒索病毒通常会通过设定支付时限、公开部分数据等方式向受害者施加压力。

步骤 7　解密：受害者根据勒索信息支付赎金后，攻击者会通过匿名网络发送解密密钥。

步骤 8　赎金转移：攻击者获取比特币、门罗币等形式的赎金后，会立即将其转入一个由交易所控制的钱包以换取法定货币。为防止交易被追踪溯源，有的攻击者会使用混币进行赎金转移，极大提高了追踪难度。

3　AI 增强勒索病毒

随着 AI 技术的飞速发展，勒索病毒的演进进入了一个全新的阶段。攻击者开始将机器学习（ML, machine learning）、深度学习（DL, deep learning）、强化学习（RL, reinforcement learning）等先进技术融入勒索病毒的各个攻击环节，一方面提升了勒索攻击的成功率与破坏力，另一方面使勒索病毒具备了更强的逃避传统安全检测和防御机制的能力，极大地提升了勒索攻击的隐蔽性。因此，融合 AI 技术的勒索病毒，进一步加剧了对网络安全的威胁。

3.1　AI 增强勒索病毒特点

AI 增强勒索病毒主要呈现智能决策、高效执行、精准攻击和跨平台运行 4 个特点。

(1) 智能决策：AI 增强勒索病毒可以智能执行初始逻辑流程以自动导航目标系统，智能选择特定的目标类型，并主动推送数据到勒索攻击者。它能够自主进行决策，选择最佳感染路径、加密策略等。因此，攻击将变得更具针对性和威胁性。

(2) 高效执行：人工智能可以以机器速度执行类似于人类的分析功能。因此，将 AI 技术应用于勒索病毒，可以利用大规模的软件漏洞，高效执行大范围攻击，如每小时数千台机器的自动攻击[25]。

(3) 精准攻击：AI 增强勒索病毒能够智能化地综合考量具体环境和受感染的机器自动选择攻击对象以最大化利润。例如，恶意软件可以根据受害者的通信来判断是否感染了公司重要人员的计算机。在受害者设备上，窃取敏感信息或锁定勒索文件都将获得更多的利润。

(4) 跨平台运行：跨平台 AI 增强勒索病毒携带了可以在不同环境中运行的各种有效负载工具，根据其对目标环境（包括平台信息）的评估，病毒将智能选择、组装并执行对目标的攻击。可以触发跨多个平台的传染，从而使检测和解决更加困难。

3.2　AI 增强勒索病毒研究现状

目前，已经有一些研究工作探索了 AI 技术在勒索病毒攻击领域的应用。AI 增强的勒索病毒代表性工作见表 2。

文献[26-30]中，研究者采用 RL 方法来规避检测系统。其中文献[26-28]针对的是静态检测。静态检测是指基于 ML，通过分析勒索病毒样本的静态结构来判断软件是否具有恶意行为。研究者通过对抗性技术来规避基于 ML 的检测系统，旨在执行勒索攻击前找到构建勒索病毒样本的最佳方法，进而规避静态检测系统。文献[29-30]针对的是动态检测。与静态检测不同，动态检测是指通过分析软件在运行时的行为特征来识别恶意软件，特别适用于勒索病毒的检测。它通过在虚拟隔离的环境中执行勒索病毒样本，并监视其行为来识别其是否具有恶意行为。文献[29]研究了针对资源受限设备的 AI 增强勒索病毒，借助 RL 框架使勒索病毒具备规避动态检测机制的能力，并加剧其对目标设备的影响。文献[30]提出了一个基于 RL 的框架 RansomAI，该

表2　　AI增强的勒索病毒代表性工作

技术	检测方式	攻击类型	混淆技术	执行方式	评估方式	代表性工作
RL	静态	对抗性	是	离线	模拟环境	文献[26-27]
		对抗性	是	离线	真实环境	文献[28]
	动态	对抗性	是	在线	真实环境	文献[29]
		恶意软件	否	在线	真实环境	文献[30]
GA	静态	恶意软件	是	离线	模拟环境	文献[31]
ML	-	恶意软件	否	在线	模拟环境	文献[32]
	-	对抗性	否	在线	模拟环境	文献[33]
DL	静态	对抗性	是	离线	模拟环境	文献[34-35]
	混合	恶意软件	是	离线	真实环境	文献[36]

框架可以集成到现有勒索软件样本中以动态调整勒索软件的加密行为，并规避动态检测系统。

除了使用RL方法之外，文献[31]提出了利用遗传算法（GA, genetic algorithm）进行字节级修改的方法，以规避恶意软件检测。文献[32]则提出了一种ML模型，用于在网络物理系统中注入战略性的系统故障。该模型并未用于规避检测，而是用于优化故障发生的时间和位置，因此没有采用对抗性技术。文献[33]提出了能够评估勒索攻击影响的RoboTack模型。它设计了具有3层隐藏层的神经网络，能够提供关于攻击目标、攻击时间的最佳策略。文献[34-36]将DL应用于勒索攻击中。文献[34]利用生成对抗网络（GAN, generative adversarial network）构建了一个基于DL的域名生成算法（DGA, domain generation algorithm），并利用该DGA绕过基于DL的检测器，在一系列对抗中生成越来越难以检测的域名。文献[35]提出了一种基于GAN的MalGAN算法来生成对抗性恶意软件示例，以此绕过基于黑盒ML的检测模型。以上两种检测模型都是采用静态规避策略。文献[36]提出了首个不仅依赖静态混淆实现规避的通用恶意软件DeepLocker。它通过引入动态混淆机制，在攻击发生前对任意来源进行实时混淆，从而增强了规避能力。DeepLocker采用深度神经网络（DNN, deep neural network）对攻击载荷进行加密，并巧妙地将自身嵌入目标系统。尤为独特的是，它能够通过学习目标系统的特征构造有针对性地加密密钥。

3.3　AI增强勒索病毒的攻击流程

如前所述，传统勒索病毒的攻击流程由8个环节构成。网络黑客们不断尝试将AI技术融入勒索攻击的这8个环节，用以提升勒索攻击的成功率、增强勒索病毒对目标系统的破坏能力。AI增强勒索病毒攻击流程如图3所示，AI技术在不同攻击环节的增强手段和所达到效果不同，步骤和内容如下。

步骤1　病毒传播

在病毒传播阶段，勒索病毒凭借AI技术的助力，实现了传播效率的大幅提升与破坏范围的显著扩大，同时其传播手段更加隐蔽、智能化。AI增强勒索病毒传播与感染方式见表3，AI增强勒索病毒传播手段主要有AI语音合成、AI钓鱼邮件生成和AI社交机器人等。

(1) AI语音合成：AI语音合成技术会引发针对生物识别安全流程的新型欺诈行为。通过模仿合法用户的语音模式包括音色、语气等特征来合成攻击性的语音命令，然后通过智能环境监测，在不被用户注意的情况下，以最佳时间和最佳音量播放“攻击性声音”[37]。

(2) AI钓鱼邮件生成：AI技术可以帮助勒索攻击者起草钓鱼邮件，根据受害者的个人信息和兴趣爱好，生成个性化邮件内容，增加受害者的信任度，诱骗用户点击恶意链接或下载附件[38]。

(3) AI社交机器人：AI社交机器人能够利用先进的AI技术创建自动化程序，在社交媒体平台上模拟人类行为，执行如数据收集、分析、信息传播、互动等一系列复杂任务。同时具备生成高度仿真的人类文本、图像乃至视频内容的能力[39-41]。然而，此类技术若被勒索攻击者所利用，则可能演化为执行恶意活动的工具，策划并实施钓鱼攻击及欺诈行为，对社会网络安全构成潜在威胁。

步骤2　感染系统

当勒索病毒传播至目标系统后，便通过AI密码攻击、智能僵尸网络等方式感染系统。AI增强勒索病毒具备智能化分析目标系统环境特征能力，精确解析目标系统操作系统类型、安全配置状况等

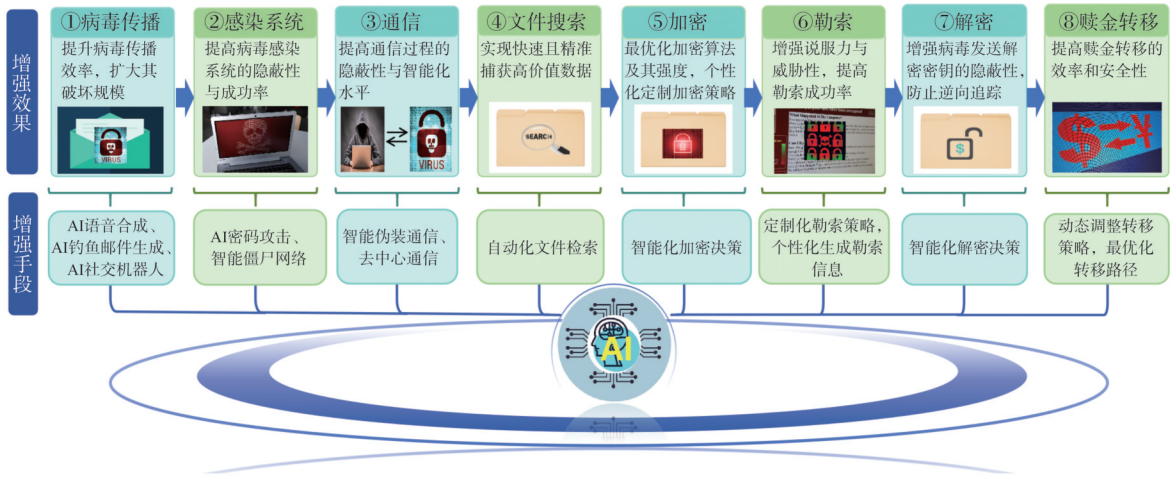

图3　AI增强勒索病毒攻击流程

表3　　AI增强勒索病毒传播与感染方式

传播与感染方式	特点	代表性工作
AI语音合成	通过模仿某人的语音模式包括音色、语气等特征进行欺诈	文献[37]
AI钓鱼邮件生成	借助AI技术生成高度仿真的邮件内容，诱骗用户点击恶意链接或下载附件	文献[38]
AI社交机器人	利用AI技术创建自动化程序，在社交媒体平台上模拟人类行为，执行恶意活动	文献[39-41]
AI密码攻击	借助AI技术学习先前密码模式，快速构建攻击字典来生成新的潜在密码	文献[42]
智能僵尸网络	利用AI技术构建智能僵尸网络，智能分析网络环境变化，持续地自主学习与进化，动态调整病毒攻击策略	文献[43]

关键环境信息，进而利用系统漏洞或伪装成系统内部可信元素以有效规避安全检测，实现隐蔽入侵。更进一步，借助AI技术动态生成智能攻击策略，选择最佳感染路径，显著提升感染的隐蔽性与成功率。如表3所示，AI增强勒索病毒的感染方式主要有AI密码攻击和智能僵尸网络等。

(1) AI密码攻击：传统密码攻击技术是基于遍历的，仅能捕获特定的密码空间子集。其破解成功率不仅深受解密者经验与人工生成密码规则的影响，而且与所使用字典的质量和更新频率密切相关。相比之下，借助AI技术，勒索病毒能够分析并学习先前密码特征，快速构建攻击字典，并据此生成新的潜在密码组合，极大地提高了破解密码的成功率[42]。

(2) 智能僵尸网络：网络攻击者可以利用AI技术构建由自主智能机器人组成的智能僵尸网络，它具备实时监测防御系统状态与深入分析网络环境变化的能力，并能够动态调整其行为模式与攻击策略，以有效规避系统检测与响应措施[43]。除此之外，通过采用分布式的指挥和控制架构，智能僵尸网络降低了对单一控制服务器的依赖性，显著增强了网络的生存能力。

步骤3　通信

在通信阶段，传统勒索病毒通常具有一个中心控制节点，该节点作为勒索病毒与受感染设备间信息交互的枢纽。为提高隐蔽性，AI增强勒索病毒能够智能伪装通信隐藏C&C节点，甚至通过采用去中心通信模式以躲避安全系统的检测与识别。

(1) 智能伪装通信：AI增强勒索病毒能够学习受感染设备的网络通信模式，通过模拟其通信数据包格式以伪装成合法通信流量，从而隐匿其真实意图。甚至在面临潜在的暴露风险时，勒索病毒可以智能地暂停通信并进入潜伏状态，以规避检测，直至出现更为有利的攻击时机。

(2) 去中心通信：借助群体智能算法，如蚁群优化算法、粒子群优化算法等AI技术，勒索病毒能够实现去中心化的通信机制。在该通信机制中并无中央控制节点，所有节点均可通过不同方式与其他节点进行通信。该机制可避免因C&C服务器被意外关闭而导致通信链路中断，进而增强勒索病毒通信系统的鲁棒性。

步骤4　文件搜索

勒索病毒与C&C节点建立通信连接后，会迅速遍历受感染设备的文件系统，识别并定位关键数

据。AI增强勒索病毒能够深度剖析感染系统中的文件特征及其存储架构，迅速辨识并精确锁定系统中的高价值数据资源（如敏感文件、财务信息等）。这一过程不仅实现了对高价值数据的智能分类与精准捕获，还显著缩短了搜索时间，有效降低了病毒被安全系统检测的风险，提升了其隐蔽性。

步骤5　加密

一旦搜到目标数据，勒索病毒便进行加密操作。AI增强勒索病毒能够智能决策，选择最优的加密算法及加密强度。具体而言，病毒能够依据受感染设备中防御系统的实际检测效能，针对性地选择最难以被检测到的加密手段；同时根据目标数据的类别和重要性差异，灵活制定最佳加密策略。此外，凭借AI技术，勒索病毒还可根据目标系统的资源配置状况（如CPU性能、内存容量等），动态调控加密过程的资源占用，避免因系统性能波动而引发检测系统察觉。AI增强勒索病毒的智能加密策略与AI技术的工作原理、适用场景等因素密切相关，以文献[30]提出的基于RL的RansomAI模型为例，该模型能够动态调整加密行为，进而智能规避防御机制的检测。RansomAI模型框架如图4所示，RansomAI指纹识别智能体能够自主学习最优加密算法组合、速率与持续时间，在奖励机制的驱动下持续优化加密策略，最终实现在最小化被检测概率的同时最大化破坏效果。

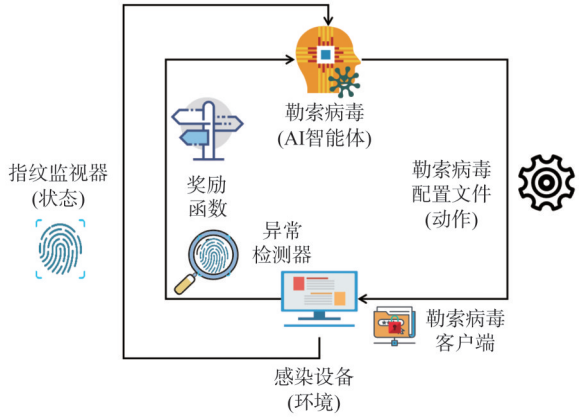

图4　RansomAI模型框架

步骤6　勒索

加密操作完成后，勒索病毒便启动勒索环节。AI增强勒索病毒能够通过对目标系统用户行为模式的深度分析和数据内容的细致解读，精准掌握受害者的语言偏好和文化背景，进而构建详尽的目标人物画像。据此定制化勒索策略，运用自然语言处理技术生成贴合受害者特征的勒索信息，显著增强其说服力与恐吓效果。

步骤7　解密

受害者交付赎金并满足黑客的要求后，便可步入解密阶段。传统勒索病毒的密钥通常是通过C&C通信连接传送，而运用AI技术可显著增强勒索病毒发送密钥过程的隐蔽性。例如，在去中心通信机制中，可以通过随机节点发送密钥以防止逆向追踪。

步骤8　赎金转移

网络黑客获得赎金后，立即转移并消除赎金账户，降低被公安系统追踪风险。借助AI技术，网络黑客能够深入分析全球金融交易网络的运作模式与潜在安全漏洞，进而选择最为安全且高效的赎金转移路径。此外，还可根据赎金转移过程中的实际情况（如交易受阻、账户被冻结等）动态调整赎金转移策略。一旦发现某一赎金转移渠道被阻断，则可即刻自动切换至备用路径，或探索并采用其他可行的方式完成赎金转移。

4　AI增强勒索病毒防御

对于网络安全而言，AI犹如一把双刃剑，其影响具有双重性。一方面，网络黑客可将AI技术深度融入勒索病毒攻击的各个环节之中，显著提升攻击的成功概率，对网络安全构成更为严峻的威胁。另一方面，网络防御者亦可充分利用AI技术，构建基于“AI+”的勒索病毒防御体系，以强化网络的总体安全防护能力，有效抵御黑客的入侵[44]。

4.1　AI增强勒索病毒防御体系

AI增强勒索病毒防御体系可分为事前防御、事中防御和事后防御3个层级。AI增强勒索病毒防御体系如图5所示，事前防御层级包括基于AI的勒索病毒预测和基于AI的勒索病毒预防两个核心环节。预测环节是指利用AI技术深度分析历史数据和威胁情报，识别潜在风险并预警未来勒索攻击；预防环节是指借助AI技术提前部署防御措施，主动阻止勒索攻击发生。事中防御层级包括基于AI的勒索病毒检测和基于AI的勒索病毒识别两个核心环节。检测环节是指通过AI技术实时监控系统行为和网络流量，发现异常攻击行为；识别环节是指借助AI技术进一步准确识别并归类威胁类型，

以制定针对性的防御策略。事后防御层级包括基于AI的勒索病毒缓解环节，它是指运用AI技术迅速缓解勒索攻击影响，最小化损失，通过快速隔离受感染系统、终止恶意进程，防止勒索攻击进一步蔓延，最终实现系统的快速恢复与正常运行。AI技术的全面渗透与融合，极大地提升了勒索病毒防御体系的智能化水平与高效性能。

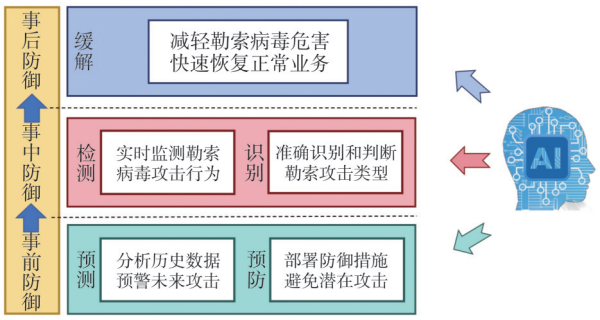

图5　AI增强勒索病毒防御体系

4.2　AI增强勒索病毒防御研究现状

AI增强勒索病毒防御已引起学术界的关注，主要围绕基于AI的勒索病毒预防、预测、检测、识别和缓解5个方面展开研究。勒索病毒防御方法文献总结见表4。

(1) 基于AI的勒索病毒预防

勒索病毒预防是指在勒索病毒攻击之前提前部署有效防御措施以保护网络、系统和数据免受勒索病毒的侵害。传统预防手段包括文件访问控制（生物识别、CAPTCHA机制等）、安全备份加密密钥、移动目标防御（MTD, moving target defense）机制等。Lee等[45]应用了MTD的概念，通过持续随机变换文件扩展名降低用户文件被勒索病毒锁定的概率。Ami等[46]设计了一个Antibiotics系统，通过实施严格的文件访问控制策略预防勒索病毒攻击，使用生物识别认证和全自动公共图灵测试（CAPTCHA）方案以验证用户身份的合法性以及人类属性。尽管当前AI在勒索病毒预防的应用相对较少，但其在自适应防御、动态响应、智能备份策略以及智能安全策略优化（如智能更新防火墙规则、入侵检测系

表4　勒索病毒防御方法文献总结

目标	方法	特征	分析类型	精确度	数据集			平台	实验环境	文献
					来源	勒索家族	样本			
预防	MTD	API调用	动态分析	98.6%	-	-	-	Windows	-	[45]
	文件访问机制	I/O请求、系统调用	动态分析	-	-	-	-	Windows	-	[46]
预测	LogR，DT，RF，Boosting	比特币交易特征	静态分析	97%	-	-	-	-	-	[47]
	KNN	IP地址	静态分析	99%	-	-	-	-	-	[48]
	支持向量机	IP地址、API接口等	动态分析	-	-	-	-	Windows	Windows	[49]
检测	CNN，LSTM，RNN	BGP更新信息参数	-	84.3%	-	1	-	-	-	[51]
	CNN	操作码序列	-	89.5%	VirusTotal	8	100 B	-	-	[52]
	D-CNN	权限、API	静态分析	96.8%	Drebin、Google Play Store data sets	179	5 560 R 5 560 B	Android	Android	[53]
识别	CNN，LSTM	系统调用	-	97.2%		3	660 R 219 B	-	-	[55]
	DNN	API调用	-	95.96%	VirusTotal	14	483 R 754 B	-	-	[56]
	CNN、多层感知机	图像的局部纹理特征、全局统计特征	动态分析	98%	-	11	5 472 R	-	-	[57]
				99.1%、94.3%	Dumpware10	10	4 294 R	Windows/Android	Windows	[58]
缓解	基于签名、异常分析	API调用等	混合分析	-	-	-	-	Android	Android 8 emulator	[59]
	ML、SDN	流量持续时间、源端口等	动态分析	-	-	-	-	ICE	OpenICE	[60]
	ML	文件访问模式、网络流量等	-	-	系统日志、网络流量等	-	-	Ubuntu	Ubuntu	[61]

统签名库以及安全策略配置参数）等方面的潜力巨大。

(2) 基于AI的勒索病毒预测

勒索病毒预测是指深入剖析勒索病毒的活动规律和进化路径，实时评估潜在风险并预警未来威胁，是一种前瞻性的安全防护方法。AI技术能够从丰富的历史数据中挖掘勒索病毒特征与行为模式，从而提前识别并防范潜在威胁。典型的预测方法包括时间序列预测技术，通过分析时间维度上的数据变动以预测攻击趋势；行为分析技术，通过监测系统和网络行为识别异常模式；以及比特币交易分析技术，追踪与勒索软件相关的加密货币交易活动。此外，卷积神经网络（CNN, convolutional neural network）、循环神经网络（RNN, recurrent neural network）、长短期记忆神经（LSTM, long short term memory）网络等ML算法也广泛应用于勒索病毒预测领域。Xu等[47]利用比特币交易中的大量特征，进行了描述性统计分析，并构建基于ML的勒索病毒攻击预测模型，实现对勒索攻击的精准预测。Chang等[48]设计了一个基于K最近邻（KNN, k-nearest neighbor）算法的网络异常流量监测与预测系统，实时监测网络异常流量并基于历史数据精准预测勒索攻击。

为了深入阐释AI技术应用于勒索病毒预测领域的工作机理，本文以文献[49]提出的基于上下文感知的勒索病毒攻击预测算法为例进行解析。基于AI的勒索病毒预测模型如图6所示，首先，该模型借助网络流量捕获工具采集物联网设备通过网络栈传输的数据包，包括正常流量和勒索软件攻击的恶意流量，并以结构化JSON格式存储；继而针对物联网场景设计上下文本体模型，动态关联攻击者行为模式、目标系统脆弱性及网络事件等多维特征；随后，基于活动上下文信息构建攻击文本过滤器，筛选关键判别性特征；最后，将特征向量输入预测模型中，实现对勒索软件攻击的精准预测。

(3) 基于AI的勒索病毒检测

勒索病毒检测是指在一个系统或网络环境中检测勒索病毒是否存在的过程。利用DL、RL、对抗性ML等AI技术学习历史数据，深度分析勒索病毒的行为和特征，实时监控并分析网络流量，一旦发现异常访问模式或行为特征便自动调整防御策略，有效阻断勒索病毒的入侵路径，极大地提高了检测的准确性和响应速度[50]。此外，AI技术还赋予了检测系统识别新型及未知勒索病毒变体的能力，以应对不断衍生出的各类勒索病毒。Li等[51]提出了结合CNN、LSTM等多种ML模型的病毒检测程序BGP-Guard，监视网络用户的恶意行为并精准识别勒索病毒和其他类型的攻击。Zhang等[52]提出了基于CNN的ML检测算法，并结合静态分析框架实现对指纹勒索病毒的有效检测。

不同AI算法应用于勒索病毒检测的工作机理基本相同，本文以文献[53]提出的深度卷积神经网络（D-CNN, deep convolutional neural network）模型为例，深入剖析AI技术应用于勒索病毒检测的工作机理。基于D-CNN的勒索病毒检测模型如图7所示。首先，该模型获取Android应用程序数据集，包括恶意软件样本与良性样本；接着，通过APK-Tool等工具反编译APK文件，解析AndroidMani-

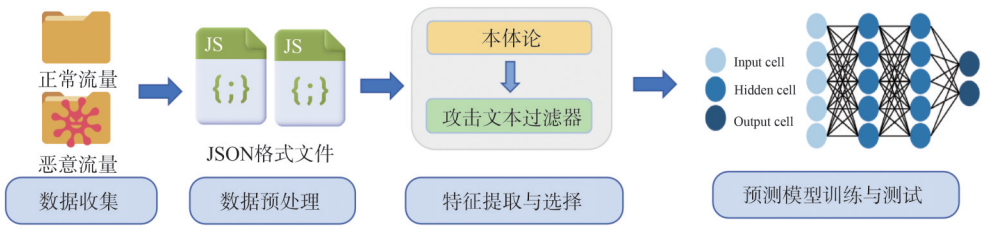

图6　基于AI的勒索病毒预测模型

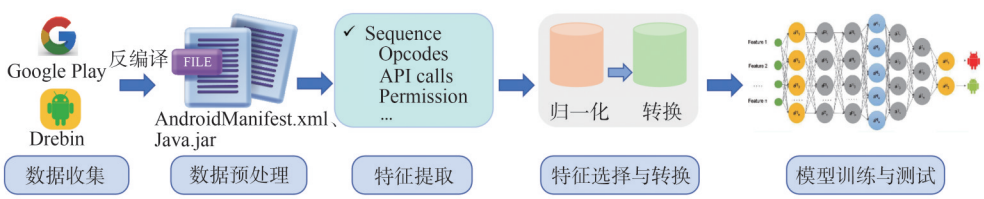

图7　基于D-CNN的勒索病毒检测模型

fest.xml 中的权限声明与 smali 代码中的 API 调用序列；然后，对提取的原始特征进行归一化、降维等操作；最后，将处理后的特征输入到 D-CNN 模型中进行勒索病毒检测，准确率高达 96.8%。

(4) 基于 AI 的勒索病毒识别

勒索病毒识别是指通过特征抽取、DL 与集成学习等先进技术，实现对勒索病毒的精细化识别与归类[54]。利用 CNN、RNN 等 AI 算法，结合静态分析或动态分析手段对恶意代码进行深入剖析，提取代码结构、行为模式、通信特点等多维度特征，并将其与已知的恶意软件家族进行匹配，从而确定相似性并进行分类。经过不断地模型训练与优化，最终构建能够有效识别不同勒索病毒家族的智能分类器。AI 技术的应用显著提升了对新型病毒变种的高效识别能力。Homayoun 等[55]提出了一种基于使用 CNN 和 LSTM 的方法 DRTHIS（deep ransomware threat hunting and intelligence system），展示出高准确率的识别能力。Sharmeen 等[56]提出了一种基于 ML 算法的勒索软件识别方法，通过动态分析收集的数据，实现了高准确率和低误报率。Hamad Naeem 等[57-58]提出了一种基于深度堆叠集成模型的恶意软件分类框架，通过动态分析方法获取进程内存转储，并将其转换为灰度图像。随后设计混合局部与全局特征描述符对图像进行结构与纹理的联合分析，实现了跨平台恶意软件的高精度检测与识别。

基于 AI 的勒索病毒识别和图 7 所示的勒索病毒检测模型，其工作机理基本相同，但因目标差异，在数据收集来源、特征提取与选择、模型架构设计等方面有所不同。

(5) 基于 AI 的勒索病毒缓解

勒索病毒缓解是指勒索病毒攻击事件发生后，有效减轻其对系统造成的损害并迅速恢复业务运营。传统缓解策略包括加密密钥托管、软件定义网络、取证分析等。近年来，AI 技术的融入为勒索病毒缓解提供了新的视角与手段。具体而言，借助 AI 技术实现智能动态漏洞管理，持续扫描系统，识别并优先修补高危漏洞，从而显著降低潜在攻击面。此外，基于 AI 的情境感知响应机制，能够自动化执行应急剧本，针对情境快速做出反应，提升响应精准度与效率。Faghihi 等[59]提出了一种以数据为中心的缓解和检测技术，以防御针对智能手机的加密勒索病毒攻击。Maimó 等[60]提出了基于 SDN 框架和网络功能虚拟化的方案，实现对受感染系统的迅速隔离，以防勒索病毒进一步传播。Alexander Panaras 等[61]提出了一个协作聚类框架，利用 ML 技术实现检测已知的和新兴的勒索病毒变体攻击并最大限度地减少勒索攻击对受害系统的整体影响。

5　AI 增强勒索病毒发展趋势与未来可能研究方向

近年来，勒索病毒的发展态势愈发严峻。勒索病毒具有变异迭代速度快、传播扩散范围广、勒索赎金金额高等典型特征，正因如此，它已然成为数据安全领域备受瞩目的核心威胁之一。尤其近年来，随着 AI 技术的融入，新一代勒索病毒攻击变得更加难以预防和抵御，在传播速度、攻击精准度和逃避安全检测的能力上都实现了质的飞跃[62]。因此本节深入探究 AI 增强勒索病毒发展趋势与未来可能研究方向，旨在为网络安全领域研究人员与行业人员提供有价值的启示。

5.1　AI 增强勒索病毒发展趋势

(1) 大模型增强勒索病毒

随着大模型在深度学习、自然语言处理等领域的广泛应用，网络攻击者开始探索如何将大模型技术融入勒索病毒的制作与传播中，旨在提升勒索攻击的成功率和破坏力。首先，借助大模型强大的数据处理能力和自我优化机制，勒索病毒能够更快速地识别并加密受害者的关键数据，同时更加隐蔽地潜藏于受害系统之中，加剧了检测与清除的难度。其次，大模型强大的数据分析能力能够提高勒索病毒的定向攻击能力，使攻击者能够更精准地锁定潜在的高价值目标，如大型企业、金融机构或政府机构，进而实施更为精确且高效的勒索攻击。

(2) 生成式 AI 增强勒索病毒

随着生成式 AI 技术的日益成熟，其强大的内容生成能力为勒索病毒的进化提供了新的可能。首先，网络攻击者利用生成式 AI 的深度学习机制和高级文本生成能力，自动生成高度逼真的虚假警告信息、加密通知以及伪装成合法来源的恶意链接或附件，创建虚假的支付页面和客户服务系统，显著增强勒索病毒的隐蔽性和欺骗性。其次，生成式 AI 技术还可用于用户行为模式分析和数据解析，

通过深入分析用户的在线活动习惯与数据特征，勒索病毒能够更准确地锁定更具备勒索价值的潜在受害者。这种个性化的攻击方式将大幅提高勒索攻击的成功几率。

5.2　AI增强勒索病毒未来可能研究方向

(1) 大模型技术在增强勒索病毒攻击的应用

大模型在增强勒索病毒方面的应用具有重要的研究价值。如前所述，大模型具备强大的多模态数据分析与深度关联挖掘能力，能够有效解决当前AI技术在增强勒索病毒攻击应用中面临的样本分析效率低、漏洞定位精准度不够等问题。攻击者可借助大模型整合全球漏洞库和目标公开数据，结合时序预测模型，精准预判目标系统未来可能存在的薄弱点，进而制定长期渗透攻击策略。此外，大模型的强化学习机制能够模拟防御系统的响应模式，针对特定AI防御模型生成对抗样本，形成“攻击–反馈–优化”的闭环迭代模式，显著提升勒索病毒攻击的持续性和隐蔽性。

(2) 生成式AI技术在增强勒索病毒攻击的应用

生成式AI技术在增强勒索病毒方面的应用值得进一步探索。如前所述，生成式AI具备强大的内容生成能力和数据分析能力，能够显著提升勒索攻击的隐蔽性与欺骗性。攻击者可借助变分自编码器、扩散模型所具有的深度生成能力，加速勒索病毒变种的更新迭代，快速合成具备新型传播逻辑和加密特征的攻击样本。同时，在社会工程学攻击方面，生成式AI卓越的自然语言处理能力能够解决钓鱼邮件内容同质化、易被识别的问题。它可通过情感分析、风格迁移技术，针对不同行业、人群的语言习惯和心理弱点，生成极具个性化的钓鱼邮件。此外，攻击者还能借助其跨模态生成技术，将文本攻击指令转化为语音、图像形式，拓宽攻击载体的维度，大幅提升攻击成功率。

(3) 大模型技术、生成式AI技术在增强勒索病毒防御的应用

随着大模型与生成式AI技术在勒索病毒领域的持续渗透，不仅要警觉其在增强勒索病毒攻击方面的潜力，更须深刻认识到这些技术在勒索病毒防御中的关键作用。未来，构建“AI+”动态智能防御架构，强化AI驱动的入侵检测与响应机制，深入研究基于大模型与生成式AI技术的勒索病毒防御框架，将成为网络安全领域新的研究方向。

针对传统检测技术存在的滞后性与高误报率问题，生成式AI能够借助GAN高度仿真勒索病毒新型变种的传播、加密等行为模式，为大模型训练提供充足的攻击场景数据，有效解决数据匮乏与样本单一的问题，助力大模型实现对新型攻击的精准预测。同时，大模型与生成式AI协同工作，基于模拟攻击场景开展多维度交叉验证。通过对网络流量、系统日志等多源数据进行深度分析，精准识别恶意行为，从而显著降低误报率。此外，针对大模型训练周期长、计算资源需求巨大的难题，生成式AI可预生成攻击数据样本，大幅缩减大模型训练时长，降低资源消耗，推动智能化防御体系加速落地应用。

6　结束语

AI技术的快速发展不仅为网络安全捍卫者提供了强有力的支撑，也为网络黑客带来了新的前景，他们正尝试将AI技术应用于勒索攻击中，不断培育新型勒索病毒，使得病毒更智能、更具隐蔽性和破坏力。因此，为全面审视AI对网络安全带来的新影响，深入揭示其工作原理，并研究和构建有效的防御方法。首先，对勒索病毒进行了分类；接着，剖析了勒索病毒的攻击流程；然后，结合最新研究进展，深入阐述了AI增强勒索病毒的工作机理；最后，从预防、预测、检测、识别及缓解5个方面，系统地归纳了基于AI的勒索病毒应对措施，并分析了AI增强勒索病毒的发展趋势与未来可能研究方向，为构建多层次“AI+”生态体系提供有价值的参考，助力网络空间安全建设。

参考文献：

[1] 曾敏, 戴卫龙. 勒索病毒原理分析与企业有效防范勒索病毒研究[J]. 现代信息科技, 2019, 3(18): 124-125, 128.
Zeng M, Dai W L. Principle analysis of blackmail virus and research on effective prevention of blackmail virus in enterprises[J]. Modern Information Technology, 2019, 3(18): 124-125, 128.

[2] 李白咏. 网络勒索赎金创新纪录，“勒索软件即服务”成新趋势[J]. 中国电信业, 2022(4): 62-63.
Li B Y. Online ransom set a new record, and “ransomware as a service” became a new trend[J]. China Telecommunications Trade, 2022(4): 62-63.

[3] 北京瑞星网安技术有限公司. 瑞星2024年中国网络安全报告[R]. 2024.
Beijing Rising Cybersecurity Technology Co., Ltd.. Rising 2024

China cybersecurity report[R]. 2024.

[4] Sarkar S, Sharma P. Preventing ransomware attacks: countermeasures and best practices[M]. Hershey: IGI Global, 2022: 32-45.

[5] 郑啸宇, 杨莹, 汪龙 . 基于 ATT&CK 模型的勒索软件组织攻击方法研究[J]. 信息安全研究, 2023, 9(11): 1054-1060.
Zheng X Y, Yang Y, Wang L. Analysis of attack methods of ransomware organizations based on ATT&CK[J]. Journal of Information Security Research, 2023, 9(11): 1054-1060.

[6] 董昱宏, 宋广佳 . 勒索病毒技术发展研究综述[J]. 计算机应用与软件, 2023, 40(1): 331-343.
Dong Y H, Song G J. Review on the technology development of ransomware[J]. Computer Applications and Software, 2023, 40(1): 331-343.

[7] O'Kane P, Sezer S, Carlin D. Evolution of ransomware[J]. IET Networks, 2018, 7(5): 321-327.

[8] Kok S H, Abdullah A, Jhanjhi N, et al. Prevention of crypto-ransomware using a pre-encryption detection algorithm[J]. Computers, 2019, 8(4): 79.

[9] Lemmou Y, Souidi E M. Infection, self-reproduction and overinfection in ransomware: the case of TeslaCrypt[C]//Proceedings of the 2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security). Piscataway: IEEE Press, 2018: 1-8.

[10] Chen Q, Bridges R A. Automated behavioral analysis of malware: a case study of WannaCry ransomware[C]//Proceedings of the 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). Piscataway: IEEE Press, 2018: 454-460.

[11] Liao K, Zhao Z M, Doupe A, et al. Behind closed doors: measurement and analysis of CryptoLocker ransoms in Bitcoin[C]//Proceedings of the 2016 APWG Symposium on Electronic Crime Research (eCrime). Piscataway: IEEE Press, 2016: 1-13.

[12] Thakkar S. Ransomware: exploring the electronic form of extortion[J]. Department of Computer Applications, 2014, 2.

[13] Ameer M, Murtaza S, Aleem M. A study of android-based ransomware: discovery, methods, and impacts[J]. Journal of Information Assurance & Security, 2018, 13(3).

[14] Verma M, Kumarguru D P, Deb S B, et al. Analysing indicator of compromises for ransomware: leveraging IOCs with machine learning techniques[C]//Proceedings of the 2018 IEEE International Conference on Intelligence and Security Informatics (ISI). Piscataway: IEEE Press, 2018: 154-159.

[15] Varlioglu S, Gonen B, Ozer M, et al. Is cryptojacking dead after coinhive shutdown?[C]//Proceedings of the 2020 3rd International Conference on Information and Computer Technologies (ICICT). Piscataway: IEEE Press, 2020: 385-389.

[16] Bijmans H L J, Booij T M, Doerr C. Inadvertently making cyber criminals rich: a comprehensive study of cryptojacking campaigns at Internet scale[C]//Proceedings of the 28th USENIX Security Symposium (USENIX Security 19). Berkeley: USENIX Association, 2019: 1627-1644.

[17] McCormack C. Five stages of a web malware attack[EB]., 2016.

[18] Mcknight J. The evolution of ransomware and breadth of its economic impact[D]. Utica: Utica College, 2017.

[19] Hacquebord F, Hilt S, Sancho D. The near and far future of ransomware business models[J]. Trend Micro Research, 2022.

[20] Chimmanee K, Jantavongso S. Digital forensic of Maze ransomware: a case of electricity distributor enterprise in ASEAN[J]. Expert Systems with Applications, 2024, 249: 123652.

[21] Oz H, Aris A, Levi A, et al. A survey on ransomware: evolution, taxonomy, and defense solutions[J]. ACM Computing Surveys, 2022, 54(11s): 1-37.

[22] 刘国宏 . 勒索病毒研究与企业应对实例[J]. 网络安全技术与应用, 2017(11): 113-114, 131.
Liu G H. Research on blackmail virus and examples of enterprise's response[J]. Network Security Technology & Application, 2017(11): 113-114, 131.

[23] Lindorfer M, Neumayr M, Caballero J, et al. POSTER: cross-platform malware: write once, infect everywhere[C]//Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security-CCS'13. New York: ACM Press, 2013: 1425-1428.

[24] 诸葛建伟, 韩心慧, 周勇林, 等 . 僵尸网络研究[J]. 软件学报, 2008, 19(3): 702-715.
Zhuge J W, Han X H, Zhou Y L, et al. Research and development of botnets[J]. Journal of Software, 2008, 19(3): 702-715.

[25] Brundage M, Avin S, Clark J, et al. The malicious use of artificial intelligence: forecasting, prevention, and mitigation[PP]. V2. arXiv (2024-12-01)[2025-02-10]. arXiv.1802.07228.

[26] Anderson H S, Kharkar A, Filar B, et al. Learning to evade static PE machine learning malware models via reinforcement learning[PP]. V2. arXiv (2018-01-30)[2025-02-10]. arXiv: 1801.08917.

[27] Labaca-Castro R, Franz S, Rodosek G D. AIMED-RL: exploring adversarial malware examples with reinforcement learning[C]// Proceedings of the Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track. Cham: Springer International Publishing, 2021: 37-52.

[28] Song W, Li X, Afroz S, et al. MAB-malware: a reinforcement learning framework for blackbox generation of adversarial malware[C]//Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security. New York: ACM Press, 2022: 990-1003.

[29] Luchinger J. AI-powered ransomware to optimize its impact on IoT spectrum sensors[D]. Zurich: University of Zurich, 2023.

[30] Von D A J, Celdrán A H, Luechinger J, et al. RansomAI: AI-powered ransomware for stealthy encryption[C]//Proceedings of the GLOBECOM 2023 IEEE Global Communications Conference. Piscataway: IEEE Press, 2024: 2578-2583.

[31] Castro R L, Schmitt C, Dreo G. AIMED: evolving malware with

genetic programming to evade detection[C]//Proceedings of the 2019 18th IEEE International Conference on Trust, Security and Privacy In Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE). Piscataway: IEEE Press, 2019: 240-247.

[32] Chung K, Kalbarczyk Z T, Iyer R K. Availability attacks on computing systems through alteration of environmental control: smart malware approach[C]//Proceedings of the 10th ACM/IEEE International Conference on Cyber-Physical Systems. New York: ACM Press, 2019: 1-12.

[33] Jha S, Cui S K, Banerjee S, et al. ML-driven malware that targets AV safety[C]//Proceedings of the 2020 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). Piscataway: IEEE Press, 2020: 113-124.

[34] Anderson H S, Woodbridge J, Filar B. DeepDGA: adversarially-tuned domain generation and detection[C]//Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security. New York: ACM Press, 2016: 13-21.

[35] Hu W W, Tan Y. Generating adversarial malware examples for black-box attacks based on GAN[C]//Proceedings of the Data Mining and Big Data. Singapore: Springer, 2022: 409-423.

[36] Stoecklin M P, Jang J, Kirat D. Deeplocker: how AI can power a stealthy new breed of malware[J]. Security Intelligence, 2018, 8(2018): 2018.

[37] Bendel O. The synthetization of human voices[J]. AI & Society, 2019, 34(1): 83-89.

[38] Teichmann F. Ransomware attacks in the context of generative artificial intelligence: an experimental study[J]. International Cybersecurity Law Review, 2023, 4(4): 399-414.

[39] Adams T. AI-powered social bots[PP]. arXiv (2017-06-16)[2025-02-11]. arXiv: 1706.05143.

[40] Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners[J]. OpenAI Blog, 2019, 1(8): 9.

[41] Danziger M, Henriques M A A. Attacking and defending with intelligent botnets[C]//Proceedings of XXXV Simpósio Brasileiro de Telecomunicaç oes e Processamento de Sinais-SBrT, 2017: 457-461.

[42] Trieuk K, Yang Y. Artificial intelligence-based password brute force attacks[C]//Proceedings of the 13th Annual Conference of the Midwest AIS (MWAIS'18), 2018: 1-7.

[43] Kudo T, Kimura T, Inoue Y, et al. Stochastic modeling of self-evolving botnets with vulnerability discovery[J]. Computer Communications, 2018, 124: 101-110.

[44] Razaulla S, Fachkha C, Markarian C, et al. The age of ransomware: a survey on the evolution, taxonomy, and research directions[J]. IEEE Access, 2023, 11: 40698-40723.

[45] Lee S, Kim H K, Kim K. Ransomware protection using the moving target defense perspective[J]. Computers & Electrical Engineering, 2019, 78: 288-299.

[46] Ami O, Elovici Y, Hendler D. Ransomware prevention using application authentication-based file access control[C]//Proceedings of the 33rd Annual ACM Symposium on Applied Computing. New York: ACM Press, 2018: 1610-1619.

[47] Xu S Y. The application of machine learning in Bitcoin ransomware family prediction[C]//Proceedings of the 2021 the 5th International Conference on Information System and Data Mining. New York: ACM Press, 2021: 21-27.

[48] Chang H Y, Lin T L, Hsu T F, et al. Implementation of ransomware prediction system based on weighted-KNN and real-time isolation architecture on SDN networks[C]//Proceedings of the 2019 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW). Piscataway: IEEE Press, 2020: 1-2.

[49] Mathane V, Lakshmi P V. Predictive analysis of ransomware attacks using context-aware AI in IoT systems[J]. International Journal of Advanced Computer Science and Applications, 2021, 12(4).

[50] Shu L H, Dong S, Su H D, et al. Android malware detection methods based on convolutional neural network: a survey[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2023, 7(5): 1330-1350.

[51] Li Z D, Rios A L G, Trajković L. Machine learning for detecting the WestRock ransomware attack using BGP routing records[J]. IEEE Communications Magazine, 2023, 61(3): 20-26.

[52] Zhang B, Xiao W T, Xiao X, et al. Ransomware classification using patch-based CNN and self-attention network on embedded N-grams of opcodes[J]. Future Generation Computer Systems, 2020, 110: 708-720.

[53] Dong S, Shu L H, Nie S. Android malware detection method based on CNN and DNN hybrid mechanism[J]. IEEE Transactions on Industrial Informatics, 2024, 20(5): 7744-7753.

[54] 乐任. 基于机器学习的医院网络勒索软件攻击识别与防御策略研究[J]. 大数据与人工智能, 2024, 5(1): 16-18.
Le R. Research on hospital network ransomware attack identification and defense strategies based on ML[J]. Big Data and Artificial Intelligence, 2024, 5(1): 16-18.

[55] Homayoun S, Dehghantanha A, Ahmadzadeh M, et al. DRTHIS: deep ransomware threat hunting and intelligence system at the fog layer[J]. Future Generation Computer Systems, 2019, 90: 94-104.

[56] Sharmeen S, Ahmed Y A, Huda S, et al. Avoiding future digital extortion through robust protection against ransomware threats using deep learning based adaptive approaches[J]. IEEE Access, 2020, 8: 24522-24534.

[57] Naeem H, Cheng X C, Ullah F, et al. A deep convolutional neural network stacked ensemble for malware threat classification in Internet of things[J]. Journal of Circuits, Systems and Computers, 2022, 31(17): 2250302.

[58] Naeem H, Dong S, Falana O J, et al. Development of a deep stacked ensemble with process based volatile memory forensics for platform independent malware detection and classification[J].

Expert Systems with Applications, 2023, 223: 119952.
[59] Faghihi F, Zulkernine M. RansomCare: data-centric detection and mitigation against smartphone crypto-ransomware[J]. Computer Networks, 2021, 191: 108011.
[60] Fernández M L, Huertas C A, Perales G A L, et al. Intelligent and dynamic ransomware spread detection and mitigation in integrated clinical environments[J]. Sensors, 2019, 19(5): 1114.
[61] Paanaras A, Silverstein B, Edwards S. Automated cooperative clustering for proactive ransomware detection and mitigation using machine learning[J]. Authorea Preprints, 2024.
[62] Ferdous J, Islam R, Mahboubi A, et al. AI-based ransomware detection: a comprehensive review[J]. IEEE Access, 2024, 12: 136666-136695.

[作者简介]

李业深（2000– ），男，北京交通大学计算机科学与技术学院博士生，主要研究方向为人工智能、高铁无线通信、网络安全。

董鹏（1977– ），男，中铁信（北京）网络技术研究院有限公司高级工程师，主要研究方向为网络安全。

朱贺（1994– ），女，中铁信（北京）网络技术研究院有限公司工程师，主要研究方向为网络安全。

郭孝天（2002– ），男，北京交通大学计算机科学与技术学院硕士生，主要研究方向为机器学习、目标检测。

尹晨旭（2001– ），男，北京交通大学计算机科学与技术学院硕士生，主要研究方向为扩散模型、语义通信。

熊轲（1981– ），男，博士，北京交通大学计算机科学与技术学院教授、副院长，主要研究方向为人工智能+5G/6G网络、无线大数据分析与处理、AI赋能的移动网络优化设计、绿色智慧物联网、网络大数据分析、雾计算/边缘计算、室内定位、基于无线大数据的人体姿态识别。