网络空间安全态势分析报告（2026）

2026 年 9 月

前　言

2026 年度全球网络空间安全态势由“高烈度对抗”向“智能化、体系化、供应链化对抗”这一更加复杂局面加速演进。本报告基于国家计算机病毒应急处理中心和各编写组成员单位的工作实践，结合对全球主要网络安全机构年度报告、重大安全事件公开披露信息和主要经济体网络安全政策法规的系统性检索与分析，对报告期内的网络空间安全态势、风险现状、治理进展与演进趋势进行综合研判。本报告编写组成员单位（排名不分先后）如下:

国家计算机病毒应急处理中心（以下简称“国家病毒中心”）

安天科技集团股份有限公司（以下简称“安天”）

三六零科技集团有限公司（以下简称“360”）

北京神州绿盟科技有限公司（以下简称“绿盟”）

奇安信科技集团股份有限公司（以下简称“奇安信”）

北京华安云科网络技术有限公司（以下简称“华安云科”）

本报告中所指 2026 年度为 2025 年 7 月至 2026 年 6 月，2025 年度为 2024 年 7 月至 2025 年 6 月，以此类推。本报告中相关统计数据和分析观点仅代表报告编写组相关单位的研究成果，属纯业务和技术研究范畴，仅供学习参考，不代表政府官方口径。

国家计算机病毒应急处理中心
计算机病毒防治技术国家工程实验室
《网络空间安全态势分析报告（2026）》编写组
2026 年 9 月

目 录

第一章　网络空间安全态势概述

1.1 网络安全事件数量持续高位运行

公开渠道披露的全球网络空间安全事件数量为 12395 起，总体呈现“规模高位、烈度上升、节奏加快”的基本特征。从事件类型结构看（见图 1），网络攻击事件 5317 起，占 42.9%，居绝对主导地位；安全隐患事件（漏洞、配置缺陷等）1119 起，占 9.0%；数据安全事件 1101 起，占 8.9%；恶意程序事件 571 起，占 4.6%；信息内容安全事件 127 起，设备故障事件 88 起，违规操作事件 51 起。值得注意的是，“其他事件”4020 起（32.4%）中包含大量政策法规动态、执法行动、威胁研究报告与行业治理事件，表明网络空间安全已从纯技术议题扩展为技术、法律、政策、地缘深度交织的复合议题。威胁等级方面，“特别重要”的事件达 1397 起（11.3%），“重要事件”6741 起（54.4%），两者合计占 65.7%，意味着近三分之二的被记录事件对现实世界具有实质性影响，全球网络空间已不存在“低烈度安全区”。

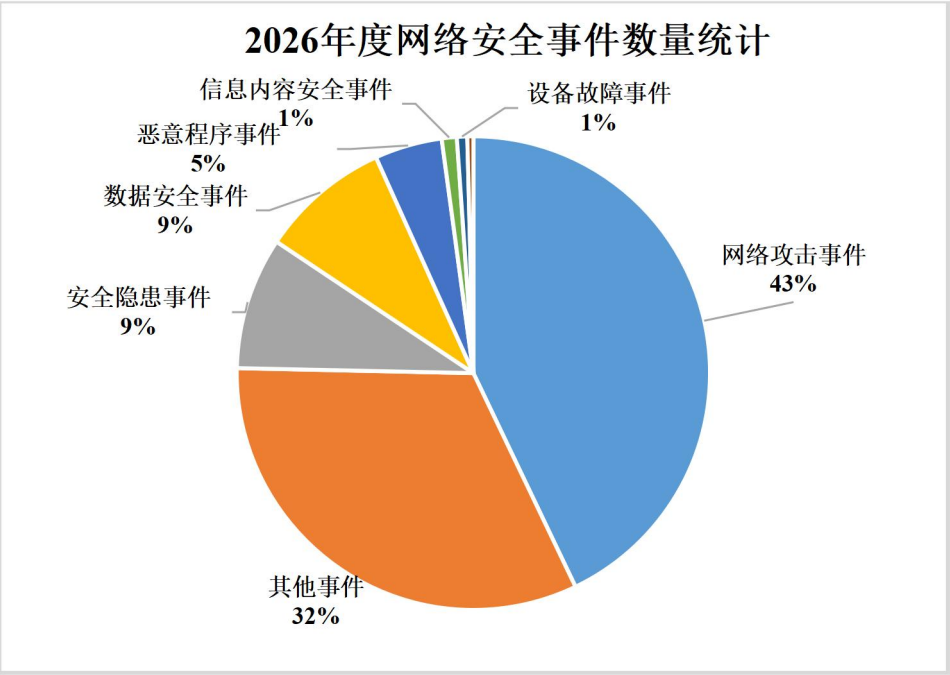

图 1-1　2026 年度网络安全事件数量统计（来源：国家病毒中心）

1.2 漏洞风险治理形势不容乐观

年度漏洞利用事件数量（1292 起）占总网络攻击事件数量（5317 起）的 24.3%，成为数量最多的威胁事件类型。与国际相关报告统计结果高度吻合。威瑞森（Verizon）《2026 年数据泄露调查报告》（DBIR）[1]基于对 31000 余起安全事件以及 22000 余起经确认数据泄露（覆盖 145 个国家）事件的分析发现，漏洞利用首次超越登录凭据滥用成为首要初始入侵途径，占全部泄露事件的 31%，较上一报告期的 20%大幅跃升。

1.3 防御阻断时间窗口明显缩减

在人工智能被频繁滥用的背景下，攻击者的行动速度出现质的飞跃。CrowdStrike《2026 年全球威胁报告》[2]显示，

[1] 威瑞森.《2026 年数据泄露调查报告》. [2026-05-21]. https://www.verizon.com/business/resources/reports/dbir/

[2] CrowdStrike. 2026 Global Threat Report. [2026-07-25]. https://www.crowdstrike.com/en-us/global-threat-report/

2025 年网络犯罪活动中的平均“突破时间”——即从初始入侵到横向移动的间隔——已缩短至 29 分钟，较 2024 年的 48 分钟提速 65%（2021 年为 98 分钟）；最快纪录仅 27 秒，且部分案例中攻击者在获得初始访问后 4 分钟内即开始数据外泄。攻击加速的直接后果是防御响应窗口的塌缩。当攻击从建立初始立足点到数据窃取的全程可能不超过 30 分钟，依赖人工研判、工单流转的传统安全运营中心（SOC）模式面临严峻挑战。

1.4 造成严重损失和社会影响的重大网络安全事件频发

网络攻击对现实世界破坏力已达到影响宏观经济的程度。2025 年 9 月，捷豹路虎（JLR）遭受网络攻击导致全球生产停摆约五周，英国网络监测中心（CMC）评估其对英国经济造成约 19 亿英镑影响、波及 5000 余家企业，被定性为英国历史上损失最重的网络攻击。2025 年 12 月，委内瑞拉国家石油公司（PDVSA）遭受重大网络攻击，导致其核心业务管理系统陷入瘫痪，原油出口业务受到严重影响。2026 年 6 月印度塔塔电子（Tata Electronics）正式证实遭遇黑客攻击，勒索组织 WorldLeaks 在暗网公开了超过 630GB、约 20.43 万份文件，涉及苹果、特斯拉等重要客户资料和制造工艺等高度敏感信息，对全球供应链安全造成了严重冲击。

1.5 人工智能失控引发的网络安全风险走进现实

OpenAI 和 Anthropic 先后承认在先进大模型测试过程中出现测试环境逃逸并自主对非预期目标发动网络攻击的失

控事故。这凸显了人工智能自身失控风险已经从学术界的理论担忧转变为真实发生的安全事件，打破了“先进大模型可被开发方通过安全围栏完全管控”的行业惯性认知。同时，AI 驱动的攻击全流程自动化，进一步验证了前文提及的防御时间窗口塌缩趋势，攻击从入侵到完成破坏目标的速度不断提升，让传统依赖人工响应的安全防护体系频频失效。当前大模型的安全对齐技术仍存在普遍缺陷，针对性的越狱提示可轻易绕过内容安全限制，诱导 AI 生成攻击教程、构建社会工程学陷阱，AI 失控引发的安全风险已经渗透到网络攻击全链路，成为全球网络空间安全领域必须直面的全新重大威胁。

第二章　网络空间安全风险现状

2.1 网络安全风险现状

2026 年度，传统网络安全风险加速向 AI 转型成为最显著特征。在人工智能技术的加持下，木马病毒、勒索病毒和僵尸网络的技术复杂度更高、变种速度更快、逃避检测能力更强；公开披露的安全漏洞数量大幅增长；针对 AI 服务的 DDoS 攻击大量出现; 高级持续性威胁攻击的 AI 去特征化和强混淆进一步推高归因分析难度。

2.1.1 计算机病毒总体态势

2026 年度，计算机病毒在报告期内发生了范式级变化。借助人工智能大模型，计算机病毒从“固定代码”进化为“运行时动态生成”。如: PROMPTFLUX，内置“Thinking Robot”模块，每小时调用 Gemini 模型重写自身源代码以实现混淆变形，使静态特征检测失效；PROMPTSTEAL 通过 Hugging Face API 调用 Qwen2.5-Coder-32B-Instruct 模型动态生成 Windows 命令实施数据收集；PROMPTLOCK 勒索病毒可利用本地部署的 LLM 生成恶意脚本。某高级持续性威胁攻击组织甚至实现了“Vibeware”[3]，即：利用 AI 氛围编程批量生成病毒变体。尽管新型病毒在数量上还没有对现有病毒家族分布实现结构性冲击，但其显然已经初步完成新路径验证，

[3] BitDefender. APT36: A Nightmare of Vibeware. [2026-03-05]. https://www.bitdefender.com/en-us/blog/businessinsights/apt36-nightmare-vibeware

预计在未来 2-3 年内，基于人工智能直接驱动或辅助开发的新型病毒将在网络安全风险矩阵中占据重要位置。

2.1.1.1 木马病毒

1. 木马病毒总体态势

2026 年度木马病毒样本数量为 134708934 个，总量较 2025 年度（195397318 个）有所回落，如图 2-1 所示。从月度统计数据中可以看出，2025 年 9 月出现了明显的峰值，相关情况与虚拟货币资产的大幅升值有关。2025 年 9 月比特币价格一度刷新历史高点，整体在 11 万美元至 12 万美元区间宽幅震荡。与此同时，针对虚拟货币实施盗窃和挖矿的木马病毒数量大幅增加。

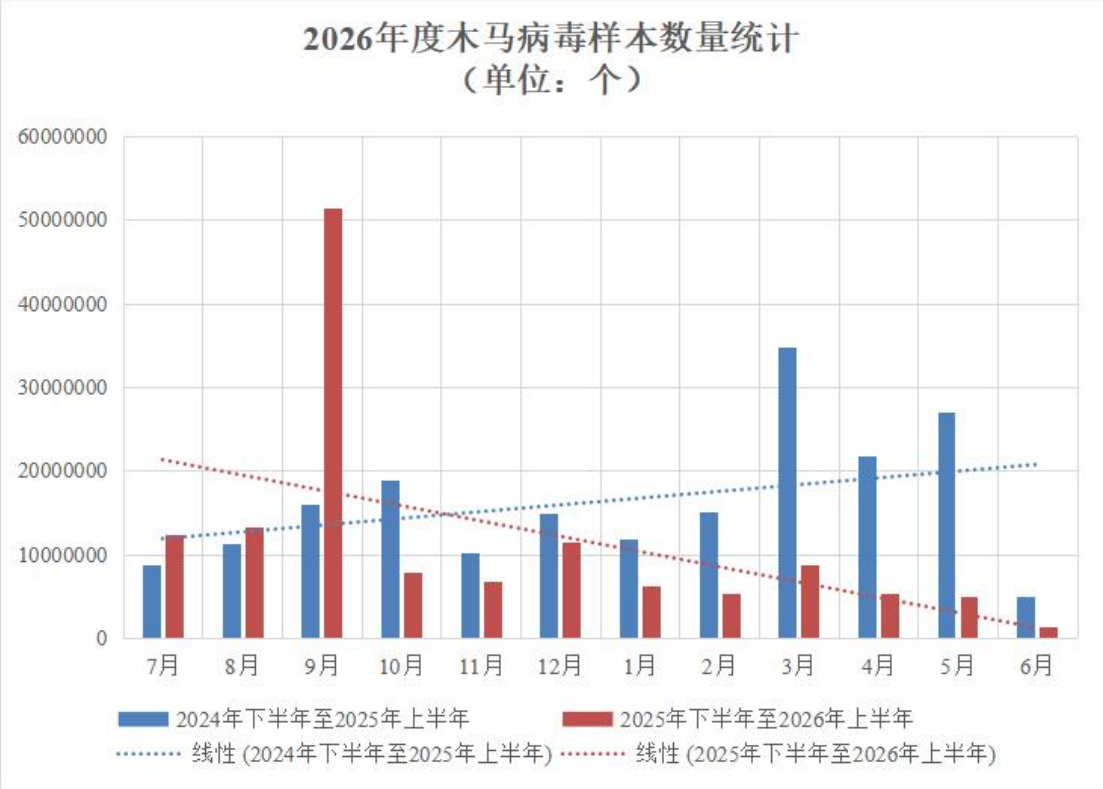

图 2-1　2026 年度新增木马病毒样本数量统计（来源：安天）

木马病毒的细分类型方面，挖矿软件（Miner）木马病毒数量排名第一，多达 43156771 个，较上一年度增长 4.66 倍。显然与网络犯罪团伙希望趁虚拟货币价格新一轮大幅上涨实施套利的动机有关。如表 2-1 所示。

表 2-1　2026 年度木马病毒样本数量分类统计（单位：个）

	2026 年度	2025 年度
挖矿软件（Miner）	43156771	7620009
后门（Backdoor）	13863268	50115921
代理访问（Proxy）	7426189	25732028
静默下载（Downloader）	4861179	31657337
窃取信息（Spy）	3460225	14371954
投放病毒（Dropper）	2298178	9581565
窃取密码（PSW）	1684839	4871525
钓鱼攻击（Phishing）	1188084	20669138
拒绝服务攻击（DDoS）	976624	1362393
勒索赎金（Ransom）	966726	2797570
可疑加壳（Packed）	812156	7415009
银行窃密（Banker）	363084	1699659
漏洞攻击（Exploit）	321808	842656
欺骗点击（Clicker）	297856	2130298
窃取游戏账号（GameThief）	101519	576429
内核攻击（Rootkit）	50470	248395

（来源：安天）

具体木马方面，2026 年度最为活跃的前 20 位木马病毒数量总体来较上年度略有下降。名次 Top1、Top12 至 Top20 对应特洛伊木马数量均多于上一统计周期，最高倍率为 2.20 倍。其中，有 6 个木马病毒家族连续两个年度持续在榜。如表 2-2 所示。

表 2-2　2026 年度排名前 20 位的活跃木马病毒（单位：个）

排名	2026 年度样本数量 Top20		2025 年度样本数量 Top20	
	木马病毒名称	木马病毒数量	木马病毒名称	木马病毒数量
Top1	Trojan/Win64.CoinMiner[Miner]	35180653	Trojan/Win32.Qukart[Proxy]	15986953
Top2	Trojan/Win32.VB	7804371	Trojan/Win32.Padodor[Backdoor]	14909805
Top3	Trojan/Win32.Padodor[Backdoor]	4328335	Trojan/Win32.VB	8671324
Top4	Trojan/Win32.Qukart[Proxy]	3814607	Trojan/Win32.Nemucod[Downloader]	6713213
Top5	Trojan/Win32.Cosmu	3516107	Trojan/Win64.CoinMiner	6261951
Top6	Trojan/Win32.BitCoinMiner[Miner]	3409226	Trojan/Win32.Mansabo	4480503
Top7	Trojan/Script.Phonzy	3267941	Trojan/Win32.Cosmu	3660269
Top8	Trojan/JS.CoinMiner[Miner]	2886814	Trojan/Win32.Berbew[Backdoor]	3044905
Top9	Trojan/HTML.Redirector	1530531	Trojan/HTML.ScrInject	2384266
Top10	Trojan/Win32.Qukart[Spy]	1523037	Trojan/Script.Phonzy	2330947
Top11	Trojan/Win32.CoinMiner[Miner]	1423653	Trojan/Win32.Cerber	2216558
Top12	Trojan/JS.Cryxos	1411705	Trojan/Win32.Snojan[Downloader]	1290838
Top13	Trojan/MSIL.Lazy	1266996	Trojan/Win32.Dinwod[Dropper]	1122211
Top14	Trojan/Shell.EncystPHP	1243637	Trojan/Win32.Jorik	1053122
Top15	Trojan/JS.Redirector	1205794	Trojan/Win32.Fareit	1022322
Top16	Trojan/Win32.Vilsel	1180819	Trojan/Win32.Krap[Packed]	1001101
Top17	Trojan/Win32.Copak	1083291	Trojan/Win32.Small	870303
Top18	Trojan/HTML.Redirector[Phishing]	1012788	Trojan/Win32.Vilsel	778480
Top19	Trojan/Win32.Swisyn	952832	Trojan/Win32.Wabot	708321
Top20	Trojan/Win32.Salgorea[Backdoor]	902632	Trojan/Android.FakeInst	623836

（来源：安天）

2026 年度活跃度排名第一的木马病毒为 Trojan/Win64.CoinMiner[Miner]，其数量达到 35180653 个，存在可执行文件、压缩文件等至少 3 种格式的样本，可执行文件占绝大部分。主要针对 x86-64 体系结构下的 Windows 64 位系统进行攻击，其主要行为是后台隐蔽挖矿，同时可能对目标造成数据丢失、系统崩溃、信息泄露等安全风险。该木马病毒变种数量有 1003 种，且经历了大量的免杀加工，以

致样本 Hash 数量近 3591.9 万。目前排名前 20 位的木马病毒详细信息参见附录 B。

2. 典型木马病毒案例

（1）“银狐”系列木马病毒攻击活动

2026 年度对我国 PC 用户危害最大的木马病毒攻击事件仍然是“银狐”系列攻击活动。“银狐”系列攻击活动最早于 2022 年 10 月被我国网络安全企业首次发现并报告，先后被国内外不同机构命名为“游蛇”“谷堕大盗”“UTG－Q－1000”“Silver Fox”等。其核心远程访问木马 ValleyRAT（又称 WinOS、Gh0st-WinOS）历经多轮技术迭代，逐步形成了模块化、高隐蔽、强对抗的技术特征。2025 年前后，WinOS 4.0 版本源代码在开源平台泄漏，彻底打破了该工具由地下组织垄断的格局，使其成为众多网络犯罪团伙可二次开发的公共恶意代码底座。自此，银狐威胁彻底从最开始的“单一组织攻击活动”演化为“多子群并行的技术生态”，不同团伙基于同一源码底座，针对不同目标场景定制攻击链路。该木马病毒攻击活动在 2025 年初至 2026 年上半年形成了三条技术路径差异显著、目标区域各有侧重的核心攻击线，共同构成了覆盖本土定向渗透、高级工程化免杀、跨境规模化钓鱼的复合型威胁形态。

第一条技术路径是高级定制化攻击，代表了当前银狐生态的技术上限。该子群基于 ValleyRAT 源码进行深度定制开发，以假冒常用办公、系统工具的安装包为核心诱饵，攻击

目标覆盖中国、日本、韩国、新加坡。其攻击链深度整合了UACME提权框架、PoolParty线程池注入技术、自定义RPC服务创建等多项规避手段，全链路恶意代码均采用VMProtect虚拟化加壳保护，反分析与免杀能力极强，整体技术复杂度已达到准网络战组织能力级别，主要针对具备一定防护能力的企业与机构用户。

第二条技术路径是社交媒体渠道渗透攻击，是银狐生态针对中国本土环境开展社会工程学攻击的典型代表。该攻击链以微信等即时通信工具为核心传播载体，通过职场社群、业务私聊等场景，以“[违纪/裁员]人员名单”“公示文件”“重要通知”等高度本土化的诱饵诱导受害者运行恶意程序，目标精准锁定中国境内公共管理和服务机构、制造业企业与医疗保健机构等。攻击者采用卷影复制服务（VSS）注入实现无文件初始执行。搭配多阶段Shellcode与破损PE头技术规避静态检测，更引入BYOVD（自带脆弱驱动）技术从内核层终止终端安全软件，具备极强的终端突破与对抗能力，是针对我国环境的高危害定向威胁。

第三条技术路径是跨境税务钓鱼双后门攻击，标志着银狐组织正式从区域性网络犯罪向全球化扩张。攻击者利用生成式人工智能大模型先后针对多个国家和地区定制符合其本土化特征的税务主题钓鱼邮件，精准利用公众对税务稽查的信任度实施诱骗。攻击链采用基于开源免杀框架二次开发的RustSL加载器完成初始突破，先投放经典ValleyRAT建

立基础持久化通道，再通过插件机制下发此前未公开的Python后门ABCDoor，形成“老牌稳定RAT+新型功能后门”的双载荷架构。该攻击活动波及东亚、东南亚、东欧和非洲地区，规模化传播攻击特征显著。

2025年8月，网络犯罪团伙利用社工技巧，将PE可执行文件进行打包和伪装，利用微信、钉钉等传播植入远控木马，并利用其对受害者进行诈骗。攻击中组合使用了“银狐”黑产此前使用过的多种免杀手段，试图通过更加隐蔽的方式植入远控木马。其利用PoolParty手段将Shellcode注入目标进程中，该Shellcode与C2服务器连接获取载荷文件，通过该载荷文件在指定路径中释放“白加黑”组件并创建计划任务，“白加黑”组件执行后对bin文件进行解密执行最终的远控木马。

2025年9月，网络犯罪团伙利用仿冒的FinalShell运维和开发工具下载网站传播远控木马，并结合搜索引擎SEO技术进行投毒攻击，使其搭建的恶意网站在搜索结果中的排名靠前，并且其域名也具有一定的迷惑性，从而诱导用户访问并下载恶意程序。此外，发现有CSDN用户曾在发布的文章中将该恶意网站描述为官网下载地址。

2025年10月，该组织大量仿冒WPS软件网站并对搜索引擎实施“投毒”，恶意拉高仿冒网站搜索排名，从而进行钓鱼传播。

2026年5月，国家病毒中心和计算机病毒防治技术国家

工程实验室依托国家计算机病毒协同分析平台（https://virus.cverc.org.cn）捕获多个表面上伪装成快捷方式、文件夹、文档文件或压缩包文件且文件名中包含“内部调查结果”“违纪名单”“违纪通报信息”“裁员补偿”等词汇的银狐木马病毒变种，并公开发布了通报预警[4]。

针对活动猖獗的网络犯罪分子，我国公安机关网安部门主动出击，成功侦破多起“银狐”木马网络犯罪案件。2026年6月，公安部网安局公布五起相关案件情况[5]，有力打击了犯罪分子的嚣张气焰。

（2）“SpyNote”手机木马病毒攻击活动

2026年度对我国移动终端用户危害最大的木马病毒攻击活动主要来自“SpyNote”系列Android平台木马病毒变种。“SpyNote”（又称SpyMax、CypherRat）是一款最早于2017年被发现的Android远程访问木马（RAT）。该木马自2020年源代码泄露后，衍生出大量变种和分支，形成了持续演化的恶意软件生态。近年来，SpyNote及其变种持续针对我国用户发起攻击，通过高度本土化的社会工程学手段，伪装成各类官方或知名应用，实施数据窃取和远程监控，对我国移动端用户隐私安全构成严重威胁。

2026年，一个名为“飞鹰”（Flying Eagle）的Android RAT框架源代码在犯罪Telegram频道中流传，该框架被检测

[4] 国家计算机病毒应急处理中心.《关于针对我国用户的“银狐”木马病毒出现新变种的预警报告》. [2026-05-21]. https://www.cverc.org.cn/head/zhaiyao/news20260521-yhyj.htm

[5] 公安部网安局. 净网专项行动 | 公安部网安局公布5起打击“银狐”木马病毒典型案例. [2026-06-16]. https://mp.weixin.qq.com/s/hYYrvgHIRSRMIZXB8RJWtw

为 SpyNote 变种，以假冒公检法服务[6]、教育类[7]、金融类应用形式传播。所有变种均通过 Android 辅助功能（Accessibility Services）请求各类权限，以实现安装应用、拦截短信（窃取登录验证码）、窃听来电、盗录视频音频等功能。部分变种还会请求 Device Administrator 权限以增强控制力。其恶意功能涵盖键盘记录、摄像头和麦克风控制、短信和通话记录窃取、GPS 定位追踪、屏幕录制、覆盖层钓鱼攻击、文件窃取、远程命令执行等。C2 通信采用持久化 TCP 连接，部分变种使用 GZIP 压缩和加密隧道。

2.1.1.2 勒索病毒

1. 勒索病毒总体态势

在全球各国政府和企业持续加大网络安全投入，并加强国际执法合作的背景下，勒索病毒犯罪组织依旧活动猖獗。2026 年度全球勒索病毒攻击事件数量达到 8819 起，较 2025 年度增长 40.1%，继续呈现持续增长态势。如图 2-2 所示。相关统计数据也侧面反映出勒索病毒犯罪团伙的攻击技战术能力在进一步增强。

[6] 国家网络与信息安全信息通报中心. 防范假冒公安机关恶意应用程序攻击. [2026-06-18]. https://mp.weixin.qq.com/s/ysLQcC13mQHMWLuMDdm4Vw

[7] 国家计算机病毒应急处理中心. 关于仿冒教育培训类 APP 的手机木马病毒的预警报告. [2026-07-10]. https://www.cverc.org.cn/head/zhaiyao/news20260710-fmjypx.htm

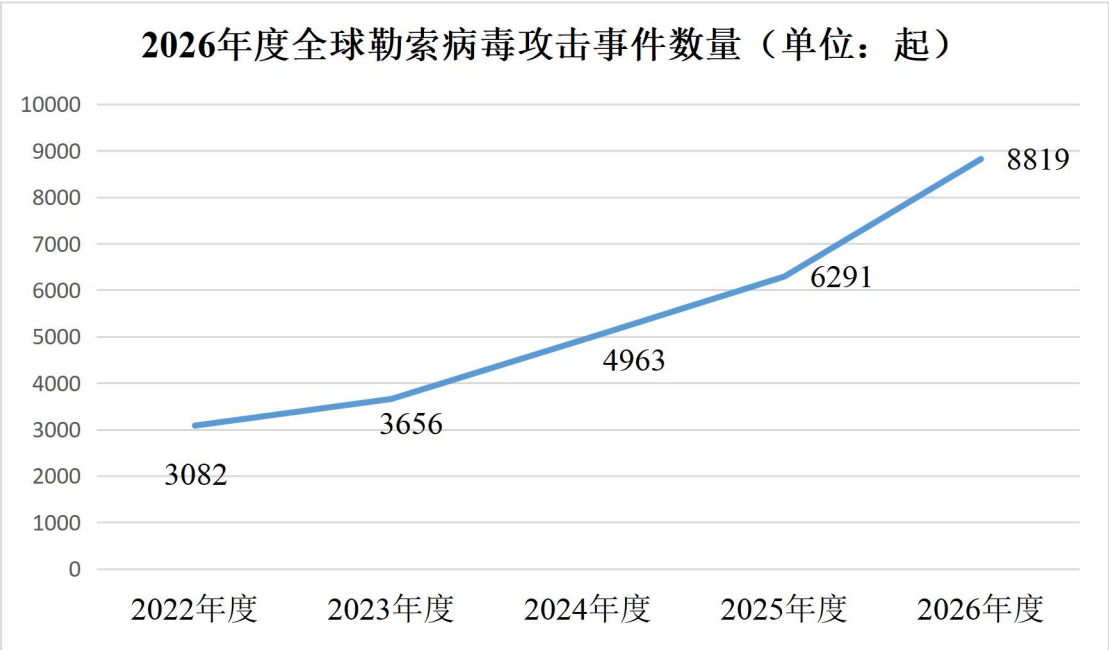

图 2-2　全球勒索病毒攻击事件数量统计

（来源：国家病毒中心，ransomware.live，ransomfeed.it）

2. 勒索病毒受害情况

2026 年度，共有 44 个勒索病毒组织向我国的 164 个机构组织发动了攻击并实施勒索，较 2025 年度（90 个）增长 76.7%，其中中国大陆地区受害单位 38 个、中国台湾地区 81 个、中国香港地区 40 个，涉及 14 个国民经济行业，计算机、通信和其他电子设备制造业受害情况最为严重。如图 2-3 所示。制造业，特别是计算机、通信等电子设备制造行业普遍采用精益制造思想，高度依赖自动化运行技术（OT），原材料和库存常态化处于低冗余状态，生产线长时间停产可能导致巨额损失，拒付赎金的抗压能力差。一些大型跨国制造业企业的分支机构或者中小型制造业企业在网络安全方面投入不足，也使得制造业企业更容易成为勒索软件攻击的目标。

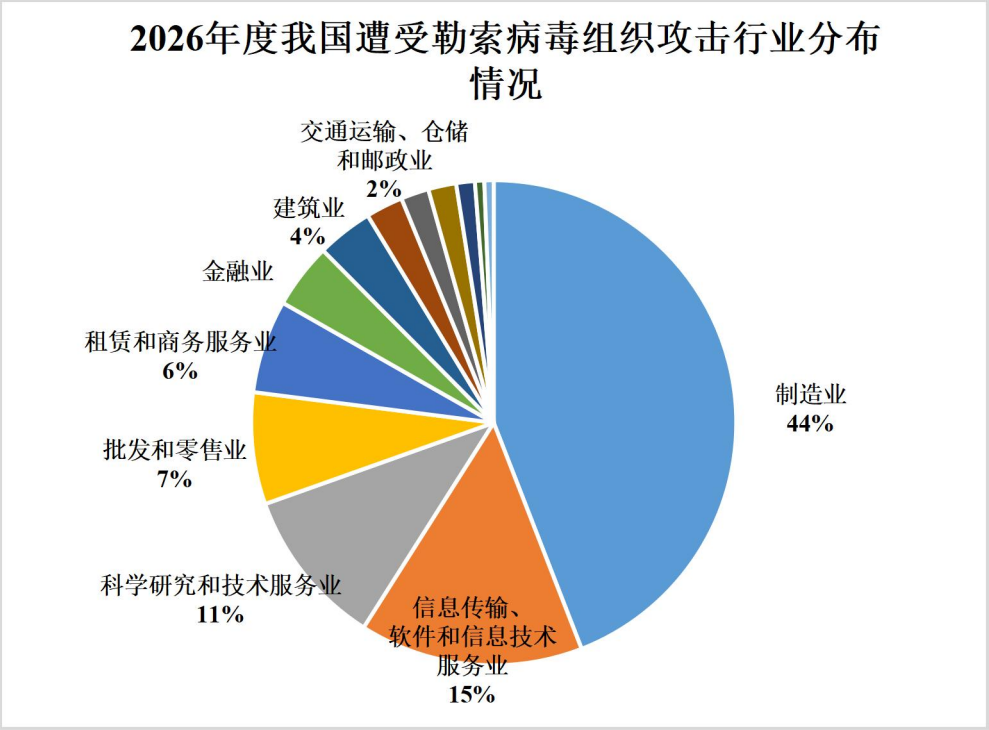

图 2-3　2026 年度我国遭受勒索病毒组织攻击行业分布情况
（来源：国家病毒中心，ransomware.live，ransomfeed.it）

3. 勒索病毒组织活动情况

勒索病毒犯罪组织的活跃情况方面，2026 年度全球范围内活跃的勒索病毒组织共计 118 个，其中活跃度排名前 5 位勒索病毒组织发动的攻击活动占全部勒索病毒攻击活动的比例超过一半，呈现出较为集中的态势。统计数据表明，thegentlemen、ransomehouse 和 qilin 是对我国用户威胁最大的勒索病毒组织。如图 2-4、图 2-5 所示。

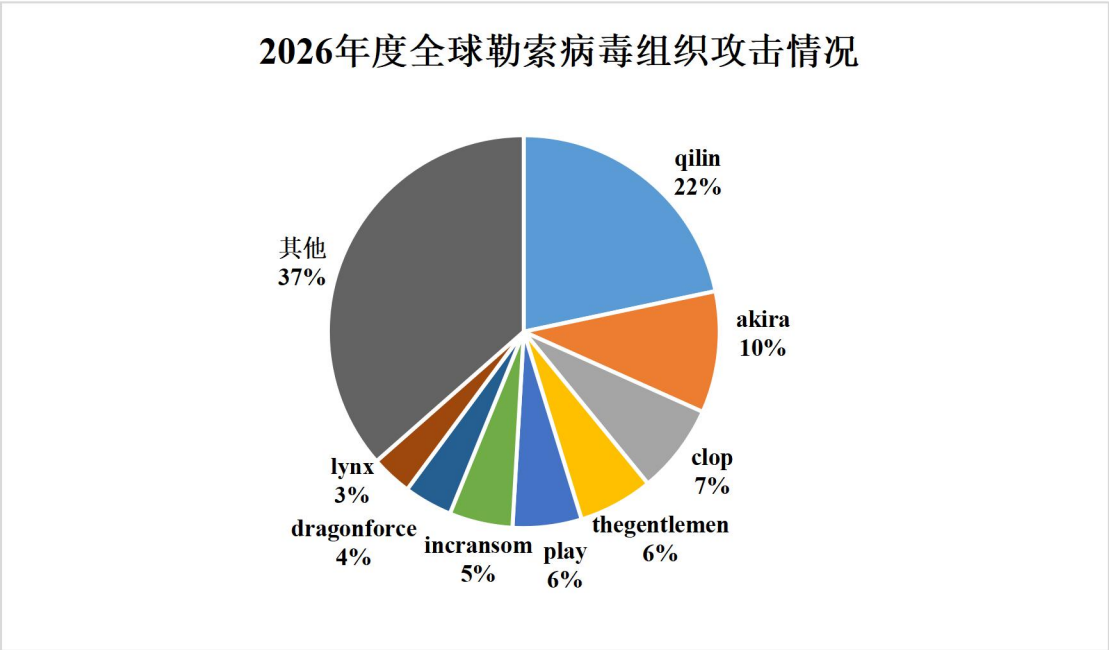

图 2-4　2026 年度全球遭受勒索病毒组织攻击情况
（来源：国家病毒中心，ransomware.live，ransomfeed.it）

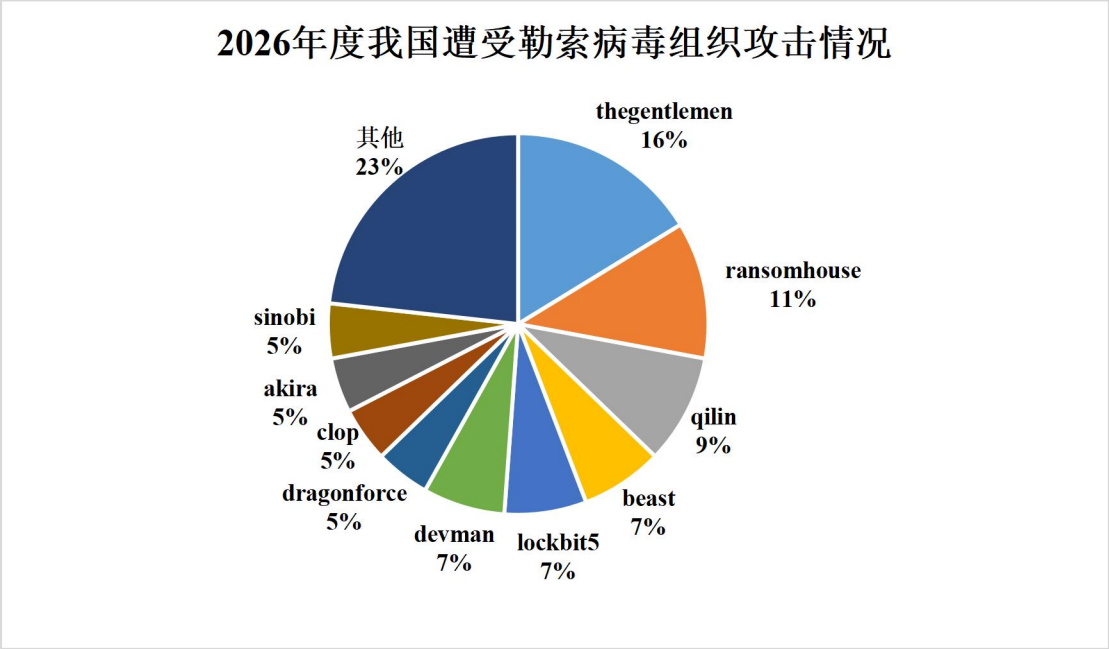

图 2-5　2026 年度我国遭受勒索病毒组织攻击情况
（来源：国家病毒中心，ransomware.live，ransomfeed.it）

对我国用户威胁最大的勒索病毒组织 thegentlemen、ransomhouse 和 qilin 的具体情况如下：

（1）thegentlemen

勒索病毒组织 thegentlemen 是 2025 年中崛起并快速扩

张的勒索软件即服务（RaaS）黑客组织，据传由脱离 Qilin 团伙的核心成员创立。该组织以高达 90%的勒索赎金分成吸引了大量黑客加盟，采用 Go 语言编写的加密程序不仅具备蠕虫级的自动横向传播能力，还配备了专用的 EDR 杀手工具包（GentleKiller），能在加密前强行关停安全软件并清空日志。在“数据窃取与文件加密”的双重威胁下，thegentlemen 在短时间内针对全球教育、医疗、金融及制造业发动了数百起高强度攻击，已快速演变为全球最具破坏性的新兴网络威胁之一。

（2）ransomhouse

勒索病毒组织 ransomhouse 最早出现于 2021 年 12 月，公开招募合作伙伴，运用典型的勒索软件即服务（RaaS）模式。该组织并未自行开发勒索软件，而是购买并部署他人的软件。该组织宣称不加密受害组织或企业设备中的文件，但会窃取数据，若受害组织或企业未能按时支付赎金，该组织将在其专用数据泄露站点（DLS）发布受害组织或企业的数据。同时，该组织表示，如果受害组织或企业能在限定时间内支付赎金，他们将帮助受害组织或企业加强安全防护，并销毁窃取的所有信息，同时移除任何用于渗透网络的后门。该组织的攻击动机与其他网络犯罪分子有所不同，似乎更侧重公开羞辱受害者，而非赎金支付情况。由此引人怀疑其是否为“白帽黑客”，主要意图为惩罚一些忽视安全和漏洞赏金计划的组织。

（3）qilin

勒索病毒组织 qilin（前身为 Agenda）是自 2022 年起活跃至今且已成长为全球市场份额第一的勒索软件即服务（RaaS）团伙。在老牌巨头频遭打击后，该组织凭借高达 80%～85% 的高额分账机制与成熟的黑产基础设施迅速吸纳了大量顶尖黑客；其恶意代码采用 Rust 语言重构，具备极高的反编译难度与 Windows/Linux（ESXi）跨平台攻击能力，不仅擅长利用“自带脆弱驱动（BYOVD）”技术强行禁用 EDR 等安全软件并清空日志，更在传统的“数据窃取与加密”基础上引入 DDoS 协同打击等“三重威胁”手段，目前已对全球医疗、制造业及金融等关键领域的供应链造成了极其深远的破坏。

4. 勒索病毒案例

2026 年 4 月以来，国家计算机病毒应急处理中心和计算机病毒防治技术国家工程实验室依托国家计算机病毒协同分析平台在我国境内发现多起我国用户遭“Sorry”勒索病毒攻击事件[8]。用户重要文件被加密后文件名被更改为原文件名 +“.sorry”后缀字符串，同时，攻击者会在用户主机上留下勒索信，要求用户下载加密通信工具与其联系。该病毒归属于 2026 年新出现的勒索病毒家族“Sorry”，使用 GO 语言编写，主要针对在互联网上暴露的 Linux Web 服务器。攻击

[8] 国家计算机病毒应急处理中心. 关于“Sorry”勒索病毒的预警报告. [2026-08-10]. https://www.cverc.org.cn/head/zhaiyao/news20260810-Sorry.htm

者利用 WebPros cPanel 授权问题漏洞（漏洞编号：CNNVD-202604-5641，CVE-2026-41940）获取服务器管理权限，随后在受害者无感知的情况下投放并运行“Sorry”勒索病毒实施攻击，并伪装成常见的 sshd 进程。该勒索病毒可在国内大部分主流 Linux 发行版操作系统（含信创操作系统）上运行并实施侵害。一旦感染该勒索病毒，可能对企业和个人造成严重的数据泄露、破坏和声誉损失。该勒索病毒具有主动传播能力，可能造成企业内网大面积感染。在没有解密密钥的条件下，被该勒索病毒加密的数据暂时没有可靠的恢复方法。

2.1.1.3 僵尸网络

1. 僵尸网络总体态势

2026 年，僵尸网络开始规模化利用 AI 技术加速恶意代码构建与发起攻击，人工智能成为攻防双刃剑，形成“利用 AI 赋能攻击”与“攻击 AI 基础设施”两个方向。一方面，攻击者借助越狱型 AI 平台降低恶意代码开发门槛，使缺乏专业编程经验的黑产从业者也能参与僵尸网络构建。另一方面，AI 平台本身成为攻击目标。

在僵尸网络深度介入地缘政治博弈的背景下，代理型僵尸网络在 2026 年急剧发展，僵尸网络向 Android 平台大规模迁移。本年度，僵尸网络共计下发攻击指令 160 余万条，主要涉及 30 多个活跃家族，目标覆盖全球 144 个国家/地区。相较上一年度（90 余万）指令规模与攻击频率大幅度增高。

从月度分布看，2025 年 9 月达到峰值（23.8 万条），10

月次之（21.1 万条），主要归因于 XorDDoS 和 Mirai 家族在这两个月异常活跃。2026 年 1 月出现明显回落，此后趋于平稳。结合往年数据整体来看，下半年攻击强度高于上半年，但攻击活动始终持续。命令控制（C&C）月度新增数量的后半年均值同样高于前半年，在 2025 年 12 月和 2026 年 4 月出现两次高峰，分别较前月大幅增长，与同期攻击指令波动存在一定关联。如图 2-6、图 2-7 所示。

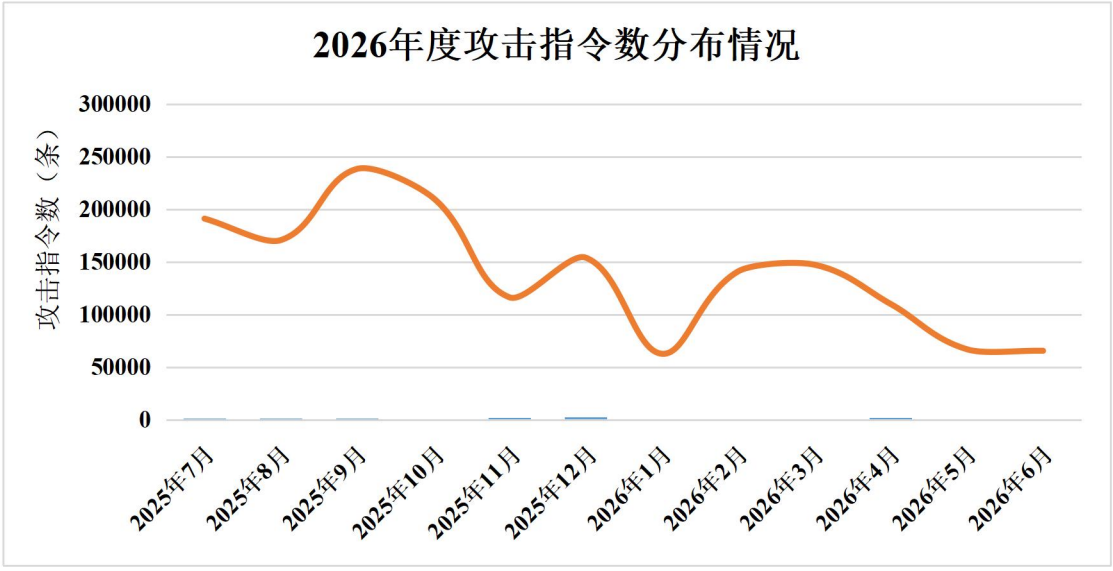

图 2-6 2026 年度攻击指令数分布情况（来源：绿盟）

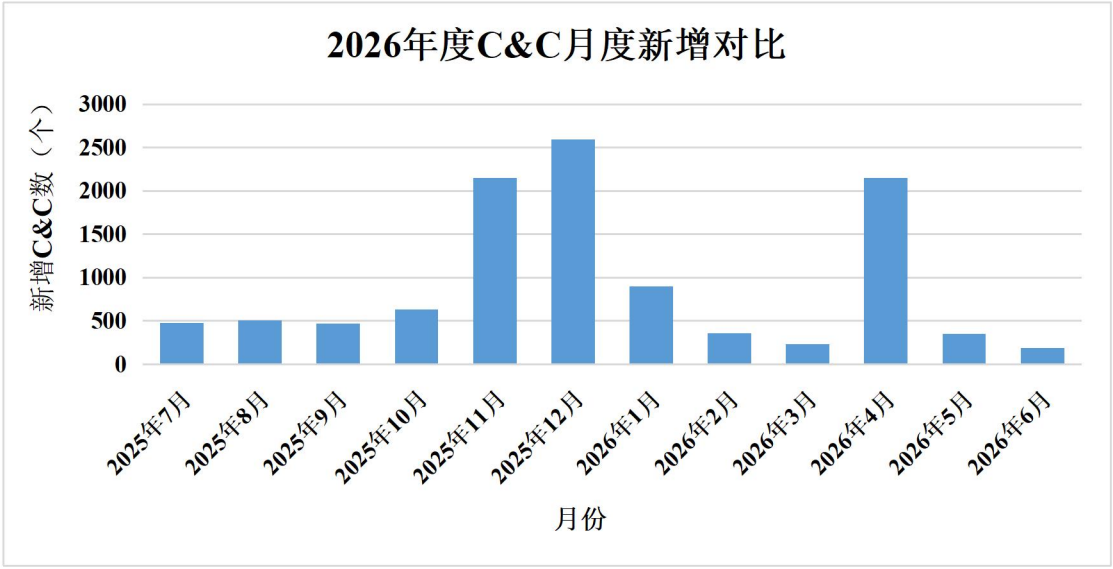

图 2-7 2026 年度 C&C 月度新增对比（来源：绿盟）

2.僵尸网络家族活动情况

从家族指令占比看，XorDDoS 以 54%的占比占据绝对主导地位，Mirai 家族（含变种）占比 23%，hailbot 占比 10%，三者合计覆盖超过八成的指令流量。NutsBot 占比 6%，是新兴家族中表现最为活跃的。Poxiao 团伙相关变种（poxiao_ver43、poxiao_03、poxiao_ver4、poxiao_3366）合计占比约 2.13%，该团伙已形成多版本并行运营的产业化格局。此外，HttpBot、chachatea、archbot 等新披露家族虽然指令量不大，但其隐蔽性极高，同样值得关注。如图 2-8 所示。

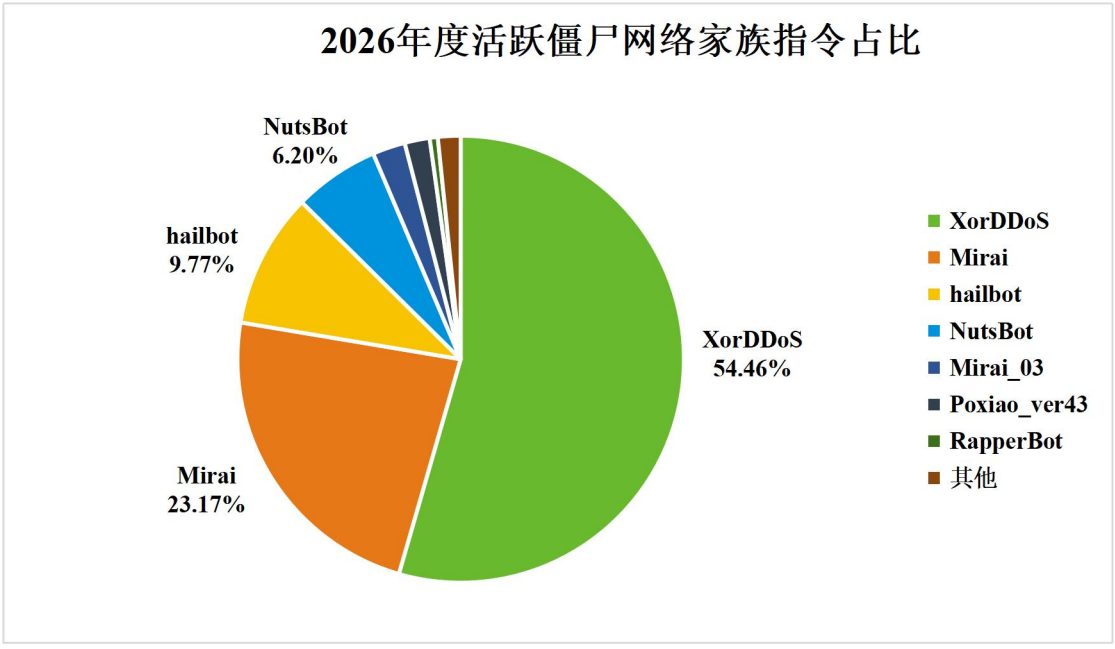

图 2-8 活跃僵尸网络家族指令占比（来源：绿盟）

从各家族新增 C&C 数量看，Mirai 及其变种居首位，由于该家族源代码开源，导致其变种众多，其基础设施规模已达到一定体量。另外，新兴团伙 poxiao 的多个变种家族较为活跃，新增 C&C 数量持续攀升。如图 2-9 所示。

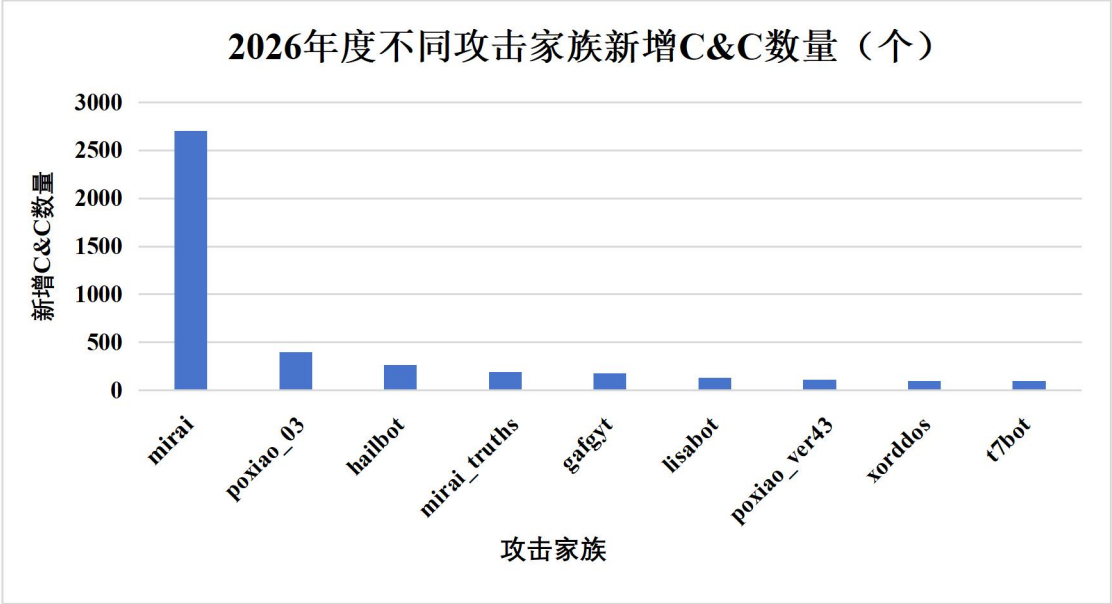

图 2-9 2026 年度不同攻击家族新增 C&C 数量（来源：绿盟）

2.1.2 漏洞风险隐患总体态势

安全漏洞作为网络安全领域的核心问题之一，在过去的一年中频繁曝光，给全球各地的组织和个人带来了巨大挑战。

1. 漏洞数量及风险程度分布情况

2026 年度，国家信息安全漏洞库（CNNVD）共收录漏洞信息 49511 条，漏洞收录数量逐年增长。如图 2-10 所示，从统计数据来看，超危、高危以及中危漏洞的总量占比很大。这一情况表明安全漏洞问题导致的网络安全风险隐患仍在不断增长。特别值得关注的是，超危漏洞的数量高达 5191 个，占比达到 10%。

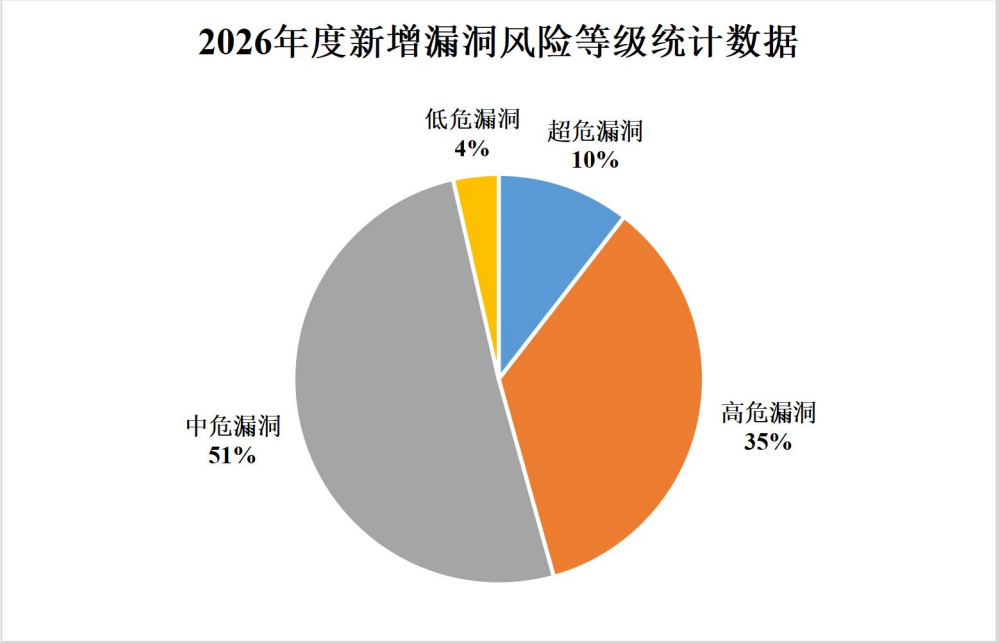

图 2-10 2026 年度新增漏洞风险等级分布情况（来源：国家信息安全漏洞库）

2. 安全漏洞整体态势特征

（1）零日漏洞武器化速度空前

2025 年 Google 威胁情报团队追踪到 90 个在实际攻击活动中被利用的零日漏洞，同比增长 15%；其中 48%针对企业基础设施，创历史新高。

（2）AI 平台成为新兴攻击热点

AI 开发工具（Langflow、MCPJam Inspector、M365 Copilot 等）漏洞数量激增，反映出安全研究人员和攻击者均将精力投向 AI 基础设施。CVE-2025-32711（EchoLeak）标志着 AI 助手从“理论风险”进入“生产环境真实破坏”阶段。

（3）供应链攻击成为年度焦点

Notepad++（CVE-2025-15556）和 Aquasecurity Trivy（CVE-2026-33634）均遭遇严重供应链攻击，影响下游数百万用户。

（4）企业边界设备持续受到重创

Ivanti、Fortinet、Cisco、F5 等网络边界设备厂商成为国家支持型攻击者和勒索软件团伙的重点目标，多个零日漏洞在披露后数小时内即被武器化利用。

（5）漏洞利用窗口极度压缩

约 30%的漏洞在披露后 24 小时内即被武器化利用，0day 与 Nday 的边界已不存在，防御窗口从“天”压缩到“小时”。

3. 典型漏洞案例

（1）网络安全漏洞

1）F5 BIG-IP APM 栈缓冲区溢出漏洞遭实际利用

漏洞类型：栈缓冲区溢出 / 远程代码执行

漏洞编号：CVE-2025-53521

漏洞评分：CVSS v3.1 9.8 / CVSS v4.0 9.3（严重）

漏洞描述：F5 BIG-IP 访问策略管理器（APM）中存在栈缓冲区溢出漏洞。该漏洞最初于 2025 年 10 月 15 日披露时被归类为拒绝服务漏洞，2026 年 3 月 F5 结合野外攻击情报更新漏洞定性，确认该漏洞可实现未经身份验证远程代码执行（RCE）。当 BIG-IP APM 访问策略配置在虚拟服务器上时，特定的恶意流量可导致远程代码执行。该漏洞影响 BIG-IP APM 多个版本（17.5.0 – 17.5.1、17.1.0 – 17.1.2、16.1.0 – 16.1.6、15.1.0 – 15.1.10）。产品主要部署于企业、金融机构及政府公共部门。F5 已发布安全补丁，并建议用户检查所有面向互联网的 F5 产品是否存在潜在入侵迹象。

2）Fortinet FortiClient EMS 访问控制漏洞

漏洞类型：访问控制不当 / 远程代码执行

漏洞编号：CVE-2026-35616

漏洞评分：CVSS v3.1 9.1（严重）

漏洞描述：Fortinet FortiClient 企业管理服务器（EMS）存在一个严重的访问控制不当漏洞。未经身份验证的攻击者可通过发送特制的 API 请求绕过所有身份验证和授权控制，在 EMS 服务器上实现远程代码执行。Fortinet 于 2026 年 4 月初发布紧急补丁，确认该漏洞已作为零日漏洞被利用。2026 年 5 月，Arctic Wolf 观察到攻击者利用该漏洞向受管理的终端推送伪装成 Fortinet 补丁的恶意 PowerShell 命令，部署名为 EKZ Infostealer 的凭据窃取木马，窃取 Chrome 和 Firefox 浏览器凭据。受影响版本为 FortiClient EMS 7.4.5 和 7.4.6，永久修复包含在 7.4.7 版本中。

3）Ivanti 终端移动管理平台代码注入漏洞

漏洞类型：代码注入 / 远程代码执行

漏洞编号：CVE-2026-1340、CVE-2026-1281

漏洞评分：CVSS v3.1 9.8（严重）

漏洞描述：Ivanti 终端移动管理平台（EPMM）于 2026 年 1 月底披露了一对高危零日漏洞。CVE-2026-1281、CVE-2026-1340 均为代码注入漏洞，允许未经身份验证的攻击者构造恶意 HTTP 请求实现远程代码执行。该漏洞作为零日漏洞存在实际攻击利用，攻击者攻陷 EPMM 服务器后，可依托

MDM 管理能力管控全网托管移动终端，并以 EPMM 节点为支点开展内网侦察与横向渗透。漏洞仅影响本地部署版本，云端 SaaS 托管 MDM 不受影响。Ivanti 已发布紧急修复补丁及修复版本。

4）Chrome 浏览器零日漏洞系列

漏洞类型：内存安全漏洞 / 远程代码执行

漏洞编号：CVE-2026-11645、CVE-2026-5281 等

漏洞评分：CVSS v3.1 8.8（高危）

漏洞描述：2026 年上半年，Google Chrome 累计修复至少 5 个确认遭在野利用的零日漏洞。其中 CVE-2026-11645 为 Chrome V8 JavaScript 引擎内存越界访问漏洞，2026 年 6 月发布紧急补丁；CVE-2026-5281 为 WebGPU Dawn 组件释放后重用漏洞，补丁于 2026 年 3 月 31 日发布，4 月初公开通报，两个漏洞均证实存在实际利用案例。攻击者诱导目标访问特制恶意网页即可触发漏洞，在浏览器渲染沙箱内执行代码，结合沙箱逃逸链路可实现系统层面权限控制。结合第三方安全厂商统计，2025 年全年共有 8 个 Chrome 零日漏洞被观测到实际利用案例，部分攻击活动持续延续至 2026 年，浏览器安全威胁态势持续严峻。Google 采用常规月度更新结合紧急离线补丁机制修复漏洞，但零日漏洞快速武器化使得未及时更新的终端持续面临入侵风险。

（2）数据安全漏洞

1）Oracle E-Business Suite 零日漏洞引发大规模数据泄

露

漏洞类型：远程代码执行 / 数据泄露

漏洞编号：CVE-2025-61882

漏洞评分：CVSS v3.1 9.8（严重）

漏洞描述：Oracle E-Business Suite（EBS）的 BI Publisher 集成组件中存在高危漏洞，允许未经身份验证的攻击者发起远程代码执行。Cl0p 勒索软件团伙自 2025 年 8 月起大规模利用该零日漏洞，针对企业人力资源环境实施数据窃取活动。多方安全厂商溯源研判显示，雅诗兰黛（Estée Lauder）在 2025 年 8 月 9 日前后遭受入侵，攻击者窃取社会安全号码、政府 ID、银行金融账户信息、健康档案及员工就业相关敏感数据等。漏洞影响 Oracle EBS 12.2.3 至 12.2.14 版本，Oracle 于 2025 年 10 月 4 日发布补丁。该漏洞相关攻击活动构成 2025 年下半年影响力极高的企业数据泄露威胁。

2）Canvas 学习平台大规模数据泄露

漏洞类型：数据泄露（利用配置错误/业务逻辑漏洞）

漏洞编号：CVE-2025-1094

漏洞评分：8.1

漏洞描述：2026 年 5 月，教育科技厂商 Instructure 证实 Canvas 学习平台发生严重未授权访问安全事件。勒索团伙 ShinyHunters 宣称窃取覆盖全球数千所院校合计 2.75 亿用户相关记录，攻击突破口为 Canvas “教师免费版”业务模块。厂商官方确认，遭访问的数据包含姓名、电子邮箱、学生编

号以及平台内私信内容。攻击发生正值北美高校期末考试季，攻击者篡改多所院校 Canvas 登录页面实施勒索，平台临时维护导致大量院校师生无法正常访问教学系统。2026 年上半年 ShinyHunters 是全球最为活跃的数据勒索威胁组织之一，第三方行业统计显示，当年 1—5 月公开确认的重大数据泄露事件中，有相当比例攻击活动归属该组织。该团伙还宣称利用相关工具链，对 Salesforce Aura、Salesforce 体验云实施多起数据窃取攻击。

（3）人工智能安全漏洞

1）Microsoft 365 Copilot Enterprise Search “SearchLeak”数据泄露漏洞

漏洞类型：参数至提示注入（P2P）/单次点击信息泄露

漏洞编号：CVE-2026-42824

漏洞评分：CVSS v3.1 7.5（高危）

漏洞描述：由 Varonis Threat Labs 发现，2026 年 6 月 15 日对外公开披露。该漏洞存在于 Microsoft 365 Copilot 企业检索模块，属于三段式漏洞利用链。攻击者构造携带恶意参数的合法 microsoft.com 域名钓鱼 URL，用户仅需单次点击链接即可触发攻击，依次借助参数提示注入、页面渲染竞争条件、依托 Bing 服务端请求伪造实现 CSP 绕过，诱导 Copilot 检索并向外导出敏感数据，可窃取邮件内容、日程、OneDrive/SharePoint 文档、Teams 消息，甚至邮件内留存的多因素验证码。由于恶意链接归属微软官方域名，传统 URL

过滤、钓鱼防护难以识别拦截。漏洞仅影响 Microsoft 365 Copilot Enterprise 企业检索模块。微软已于 2026 年 6 月 4 日完成云端服务器端静默修复。截至公开披露，暂未观测到该漏洞野外实际攻击活动。

2）Langflow AI 工作流平台未授权远程代码执行漏洞

漏洞类型：代码注入 / 未授权远程代码执行

漏洞编号：CVE-2026-33017

漏洞评分：CVSS v3.1 9.8（严重）

漏洞描述：Langflow 是一款用于构建和部署 AI 智能体与工作流的开源平台。CVE-2026-33017 存在于 /api/v1/build_public_tmp/{flow_id}/flow 端点，该接口设计用于免认证构建公共工作流，但错误接收攻击者传入的恶意工作流数据，载荷内嵌入的任意 Python 代码直接通过 exec()执行，且无任何沙箱隔离，导致未经身份验证的远程代码执行。Langflow 默认启用 AUTO_LOGIN 配置，未授权访问者可直接获取会话令牌，简化漏洞利用条件。该漏洞于 2026 年 3 月 16 日官方披露，3 月 25 日被美国网络安全与基础设施安全局（CISA）纳入 KEV 已知利用漏洞目录，披露后短时间内观测到大规模野外攻击。攻击者利用该漏洞在公网暴露的 AI 服务器部署门罗币挖矿程序、终止主机安全防护，并通过窃取 SSH 密钥实现内网横向移动。漏洞修复版本为 1.9.0，低于该版本均受影响。

3）MCPJam Inspector MCP 开发平台远程代码执行漏洞

漏洞类型：远程代码执行 / 未授权访问

漏洞编号：CVE-2026-23744

漏洞评分：CVSS v3.1 9.8（严重）

漏洞描述：MCPJam Inspector 是一款面向 MCP（Model Context Protocol，模型上下文协议）服务器的本地优先开发调试平台。该漏洞存在于 1.4.2 及更早版本中，修复版本为 1.4.3。攻击者发送特制 HTTP 请求即可实现远程代码执行。风险由两项缺陷叠加导致。MCPJam Inspector 默认监听 0.0.0.0（全网网卡）而非 127.0.0.1，服务直接暴露至局域网或公网；同时/api/mcp/connect 端点将请求内 serverConfig 中的 command、args 参数直接用于创建系统进程，未做输入校验、无访问鉴权。该漏洞于 2026 年 1 月 16 日正式披露，网络上存在大量公开利用 POC，安全研究热度较高。MCP 由 Anthropic 主导开发推出，获得微软、OpenAI、谷歌等厂商广泛支持，用于打通大模型与各类外部工具、数据源。该漏洞暴露出 AI 开发工具链正在形成新型高危攻击面。

4）Ollama“Bleeding Llama”GGUF 模型解析内存越界读取漏洞

漏洞类型：堆缓冲区越界读取 / 未授权信息泄露

漏洞编号：CVE-2026-7482

漏洞评分：CVSS v3.1 9.1（严重）

漏洞描述：Ollama 是当前主流的开源本地大模型推理部署平台，广泛应用于企业私有化 LLM 业务搭建。该漏洞由

Cyera Research 团队发现，于 2026 年 5 月 4 日正式公开披露，目前网络中已出现可稳定复现的公开 POC 利用代码。该漏洞影响 v0.17.1 之前的所有 Ollama 版本，核心原因是平台 GGUF 模型加载解析模块未对模型张量偏移量、内存大小等关键参数做严格的边界校验。攻击者可借助 Ollama /api/create 接口默认无需身份认证的特性，远程上传内置超范围异常参数的恶意 GGUF 模型文件，服务在量化解析模型过程中触发堆缓冲区越界读取，非法获取进程内存内的服务器环境变量、第三方 LLM API 密钥、自定义系统提示词、实时用户对话记录等高价值敏感数据，再通过调用/api/push 接口将夹带泄露数据的异常模型推送至攻击者可控仓库，完成静默数据外渗。该漏洞可直接导致企业 AI 核心凭证、用户隐私数据外泄，为攻击者滥用 AI 接口、开展内网渗透提供便利；由于大量企业为业务需要将默认仅监听本地 127.0.0.1 的服务修改为全网监听（OLLAMA_HOST=0.0.0.0），大幅扩大攻击面，加剧安全风险，目前官方已在 v0.17.1 版本通过完善参数边界校验机制彻底修复该漏洞。

2.1.3 分布式拒绝服务攻击总体态势

1. 分布式拒绝服务（DDoS）攻击事件数量总体趋势

自 2025 年下半年以来，全球 DDoS 攻击事件数量呈上升趋势。2026 年上半年，受国际安全形势变化影响，DDoS 攻击活动进一步活跃，其中 2 月至 4 月达到高峰，随后在 5、6 月明显回落。2026 年 2 月至 3 月，国际局势进入高度紧张

阶段，中东地区军事冲突持续升级，区域安全风险显著上升，DDoS 攻击事件数量出现明显波动。如图 2-11 所示。

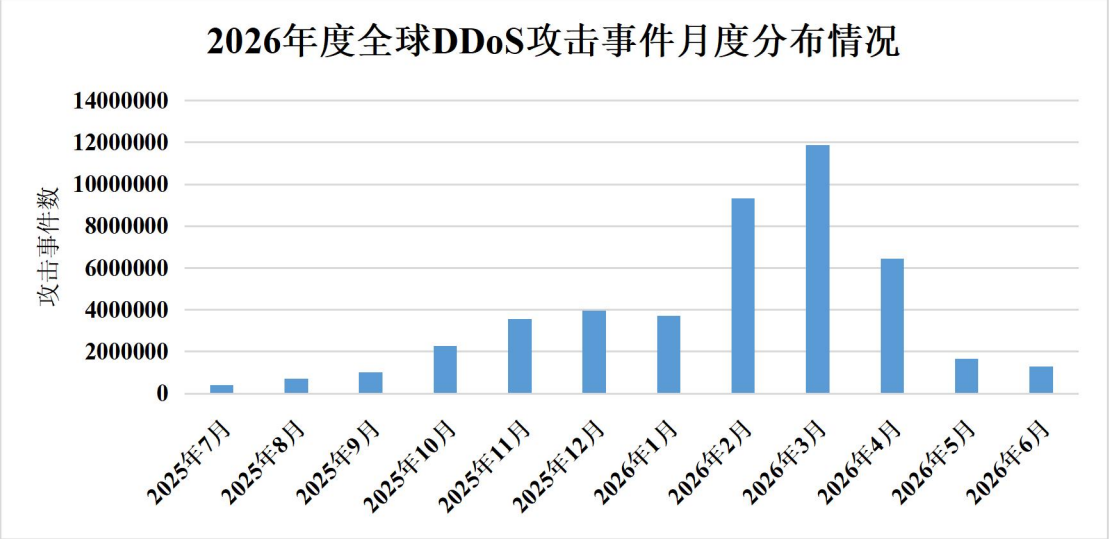

图 2-11　2026 年度全球 DDoS 攻击事件月度分布情况（来源：绿盟）

与前两年相比，多向量 DDoS 攻击事件数量持续增长，攻击模式呈现复杂化、多样化发展趋势。2026 年度，采用 2 种、3 种以及 5 种及以上攻击向量组合的攻击事件数量均出现不同程度增长，反映出攻击者正在由传统单一攻击方式向多维协同攻击模式转变。

多向量 DDoS 攻击通过结合流量型攻击、会话型攻击等多种手段，能够充分发挥不同攻击向量之间的协同效应，在提升攻击规模的同时增强攻击持续时间和防御绕过能力，从而对目标网络基础设施和关键业务系统造成更大影响。如图 2-12 所示。

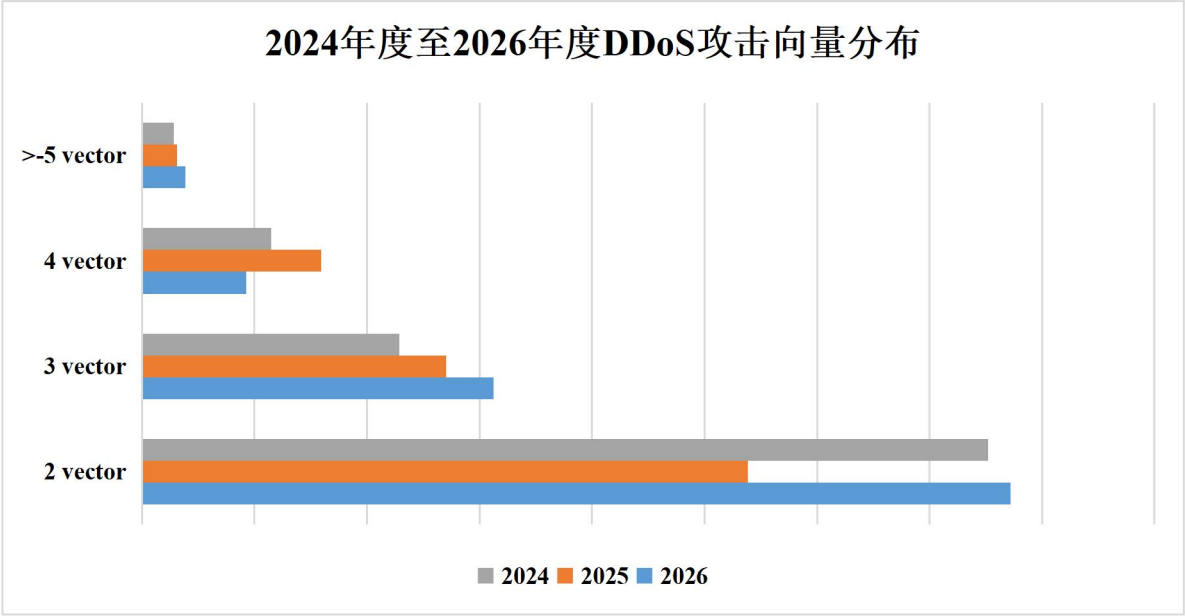

图 2-12　2024 年度至 2026 年度 DDoS 攻击向量分布（来源：绿盟）

2. DDoS 攻击类型分布情况

2026 年度网络层/传输层 DDoS 攻击以 TCP 协议攻击为主，其中 ACK Flood 占比最高，延续近两年的增长态势，已成为当前最活跃的攻击方式。SYN Flood 攻击紧随其后，攻击者持续利用 TCP 连接建立过程中的资源分配机制，对目标服务的连接处理能力造成压力。随着近年来 AI 应用、云计算服务及各类 API 接口快速普及，互联网业务规模不断扩大，暴露在公网的服务节点持续增加，也进一步扩大了此类资源消耗型攻击的影响范围。

相比之下，UDP Flood 攻击排名第三，虽然较 2025 年有所反弹，但整体仍低于 2024 年水平，占比呈逐步下降趋势。从整体演变趋势来看，DDoS 攻击模式正在由传统的大规模流量占用向面向协议机制和业务资源的精细化攻击转变，攻击者更加关注目标系统的处理能力和服务可用性，对网络基

础设施和防御体系的精准识别、动态防护能力提出了更高要求。如图 2-13 所示。

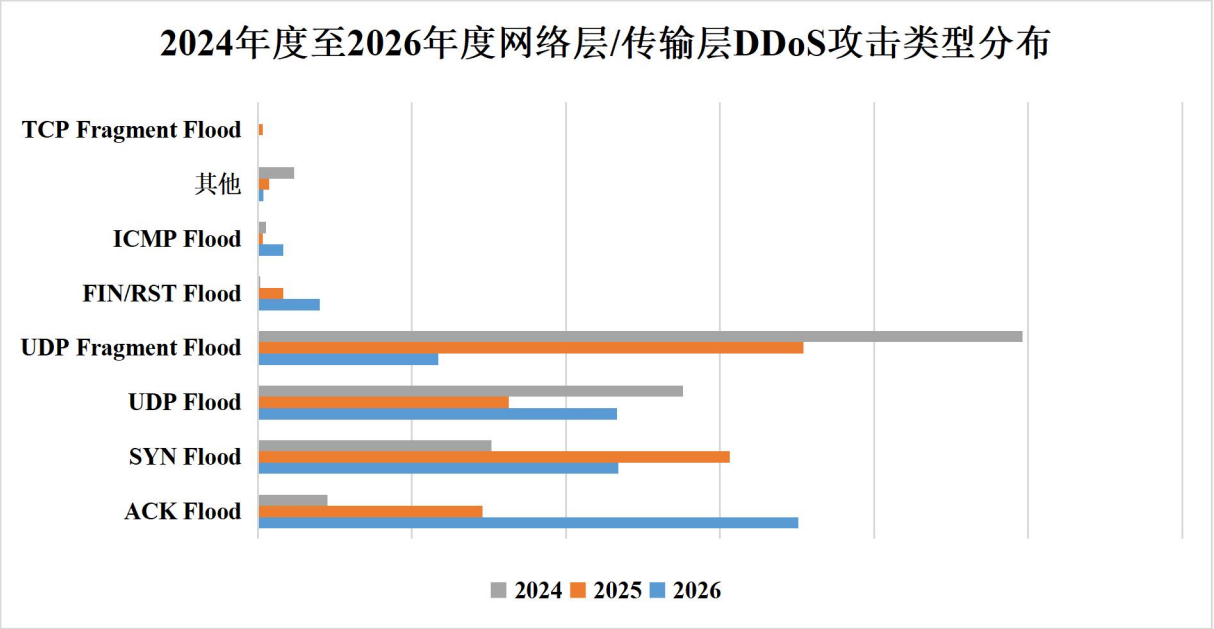

图 2-13　2024 年度至 2026 年度网络层/传输层 DDoS 攻击类型分布
（来源：绿盟）

2026 年度应用层 DDoS 攻击仍以 HTTPS Flood 为主，占比达到 45.03%，较 2025 年稍有增长，持续保持增长态势。DNS 随机子域名攻击占比 30.98%，活跃度进一步提升。相比之下，HTTP Flood 和 SIP Flood 攻击占比分别下降至 18.94%和 3.10%。

整体来看，应用层 DDoS 攻击正逐步向 HTTPS 加密业务和 DNS 基础设施攻击方向演变，攻击者更加倾向于利用业务协议特性和应用资源消耗机制实施攻击，对应用层防护能力提出更高要求。如图 2-14 所示。

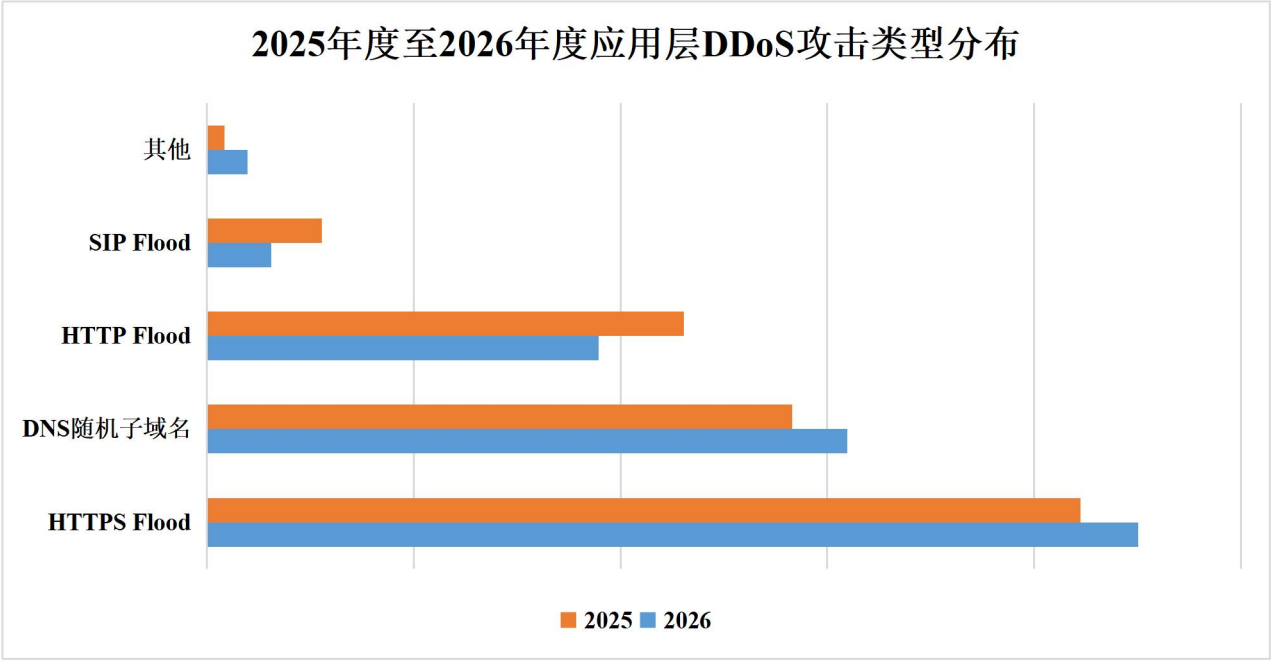

图 2-14　2025 年度至 2026 年度应用层 DDoS 攻击类型分布（来源：绿盟）

2026 年度攻击目标服务类型中，HTTPS 服务仍为主要攻击对象，占比达到 45.44%，较 2025 年有所提升，持续保持增长趋势。HTTP 服务占比 20.42%，较 2025 年小幅下降，但仍是重要攻击目标。DNS 服务攻击占比提升至 13.39%，表明针对域名解析基础设施的攻击活跃度进一步增强。

相比之下，TELNET 服务和 BGP 服务攻击占比分别下降至 5.40%和 1.08%。整体来看，DDoS 攻击目标正逐步向 Web 应用、加密业务及关键网络基础设施集中，攻击者更加倾向于针对高价值、高暴露面的互联网服务实施攻击。如图 2-15 所示。

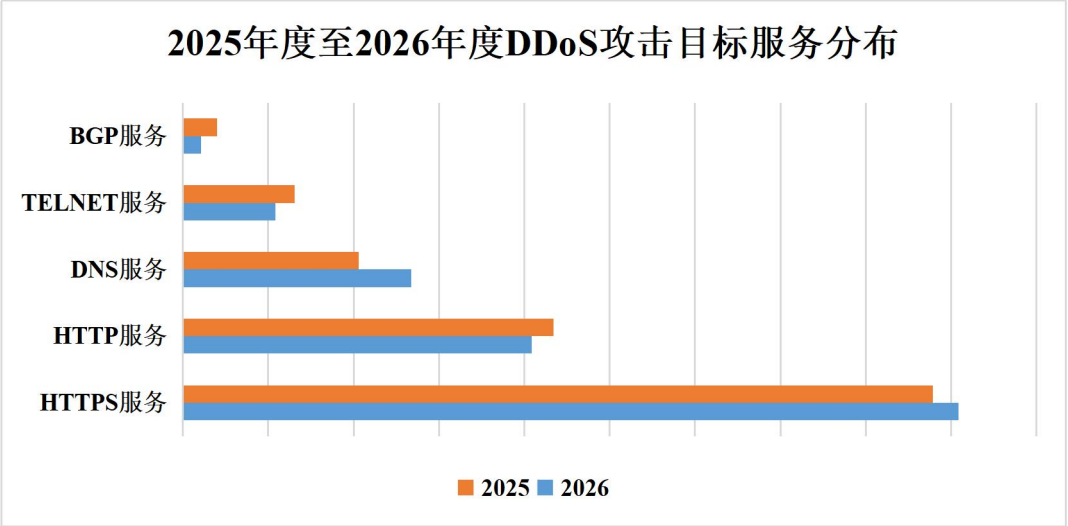

图 2-15　2025 年度至 2026 年度 DDoS 攻击目标服务分布（来源：绿盟）

3. DDoS 攻击活动典型案例

（1）Mirai 物联网僵尸网络病毒变种持续发起高强度 DDoS 攻击

据网络安全公司 Foresiet 报告[9]，2025 年 11 月，Mirai 物联网僵尸网络病毒新变种 ShadowV2 被发现活跃，该变种利用 2025 年 10 月 28 日全球云服务中断事件作为“测试运行”机会，在 11 月 1 日至 5 日期间针对 28 个国家的物联网设备发动攻击，主要利用 D-Link（CVE-2024-10914/10915）、TP-Link 和 GeoVision 等品牌的已知漏洞进行传播。Fortinet FortiGuard Labs 指出，ShadowV2 采用 XOR 编码的 ELF 载荷，通过 busybox tftp 下载，自我标识为“ShadowV2 Build v1.0.0 IoT version”。

另据 Akamai 安全情报与响应团队（SIRT）报告[10]，2026

[9] Foresiet. The Resurgence of Mirai: Jackskid Botnet and Escalating IoT Threats in November 2025. [2026-06-22]. https://foresiet.com/blog/mirai-botnet-jackskid-resurgence-nov-2025-iot-threats/

[10] Akamai SIRT. Mirai Botnet exploits CVE-2025-29635 to target legacy D-Link routers. [2026-04-22]. https://securityaffairs.com/191135/malware/mirai-botnet-exploits-cve-2025-29635-to-target-legacy-d-link-

年 3 月，Akamai SIRT 在其全球蜜罐网络中首次检测到 Mirai 新变种 tuxnokill 的活跃攻击活动。该变种利用命令注入漏洞 CVE-2025-29635 攻击已停产的 D-Link DIR-823X 系列路由器，攻击者通过构造 POST 请求向/goform/set_prohibiting 端点发送恶意数据，利用 macaddr 参数未经充分验证即被复制到命令缓冲区并传递给 system()函数执行的缺陷，实现远程命令执行。受感染设备被用于发起 TCP SYN/ACK/STOMP、UDP 洪水及 HTTP null 等多种 DDoS 攻击。

（2）Aisuru 僵尸网络连续刷新 DDoS 攻击强度纪录

据微软公司披露[11]，2025 年 10 月 24 日，Microsoft Azure 自动检测并缓解了一起针对澳大利亚单一端点的 DDoS 攻击，攻击流量达到 15.72 Tbps，峰值接近 36.4 亿个数据包每秒（pps），这是云平台上观测到的最大规模 DDoS 攻击。该攻击由 Aisuru 僵尸网络发动，来自超过 50 万个源 IP 地址，采用极高速度的 UDP 洪水攻击。

据 Cloudflare 报告[12]，2025 年 11 月，Aisuru 僵尸网络再次刷新公开纪录，发动了峰值达 31.4 Tbps 的 DDoS 攻击，持续时间仅 35 秒，被 Cloudflare 自动化防御系统检测并缓解。与 2024 年末的大型攻击相比，2025 年的攻击规模增长超过 700%。2025 年 12 月 19 日，Aisuru 僵尸网络发起了名为“圣

routers.html

[11] The Hacker News. Microsoft Mitigates Record 15.72 Tbps DDoS Attack Driven by AISURU Botnet. [2025-11-18]. https://thehackernews.com/2025/11/microsoft-mitigates-record-572-tbps.html

[12] Cloudflare. 2025 Q4 DDoS threat report: A record-setting 31.4 Tbps attack caps a year of massive DDoS assaults. [2026-02-05]. https://blog.cloudflare.com/ddos-threat-report-2025-q4/

诞前夜”（The Night Before Christmas）的超大规模 DDoS 攻击行动，该行动持续至 2026 年 1 月初。在此期间，僵尸网络对 Cloudflare 客户及基础设施发动超大规模 HTTP DDoS 攻击，峰值超过每秒 2000 万请求（20 Mrps），最高达到 205 Mrps。整个行动期间，Cloudflare 自动化系统共检测并缓解 902 次超大规模攻击（平均每天 53 次），平均强度达 3 Bpps、4 Tbps 与 54 Mrps，峰值则达到 9 Bpps、24 Tbps 与 205 Mrps。该僵尸网络主要由感染恶意软件的 Android 电视组成，估计感染主机规模约为 100 万至 400 万台。

2.1.4 高级持续性威胁攻击总体态势

1. 全球高级持续性威胁攻击活动影响情况

本年度，世界政治经济格局进入深刻演变期，传统秩序加速调整，新兴力量加快崛起。全球范围内冲突与博弈显著增多，地缘冲突在多地区凸显，多极化进程在曲折中持续向前。与此同时，网络安全态势正经历深刻演进，已从“技术层面对抗”升级为关乎国家生存与发展的战略博弈。高级持续性威胁组织攻击活动聚焦地区政治、经济等时事热点，攻击目标集中分布于政府机构、国防军工、信息技术、金融、制造、能源等重点行业领域。多个具有政权支持背景的高级持续性威胁组织开展定向网络行动，中东地区、东欧地区、拉美地区成为攻击高发区域。

通过对全球网络安全厂商和机构披露的高级持续性威

胁攻击活动进行统计，2025 年高级持续性威胁组织攻击比较集中的行业为政府机构、国防军工、信息技术、金融、制造、科研等。而 2026 年，境外高级持续性威胁组织攻击比较集中的行业为政府机构、国防军工、金融、信息技术、能源等。如图 2-16 所示。

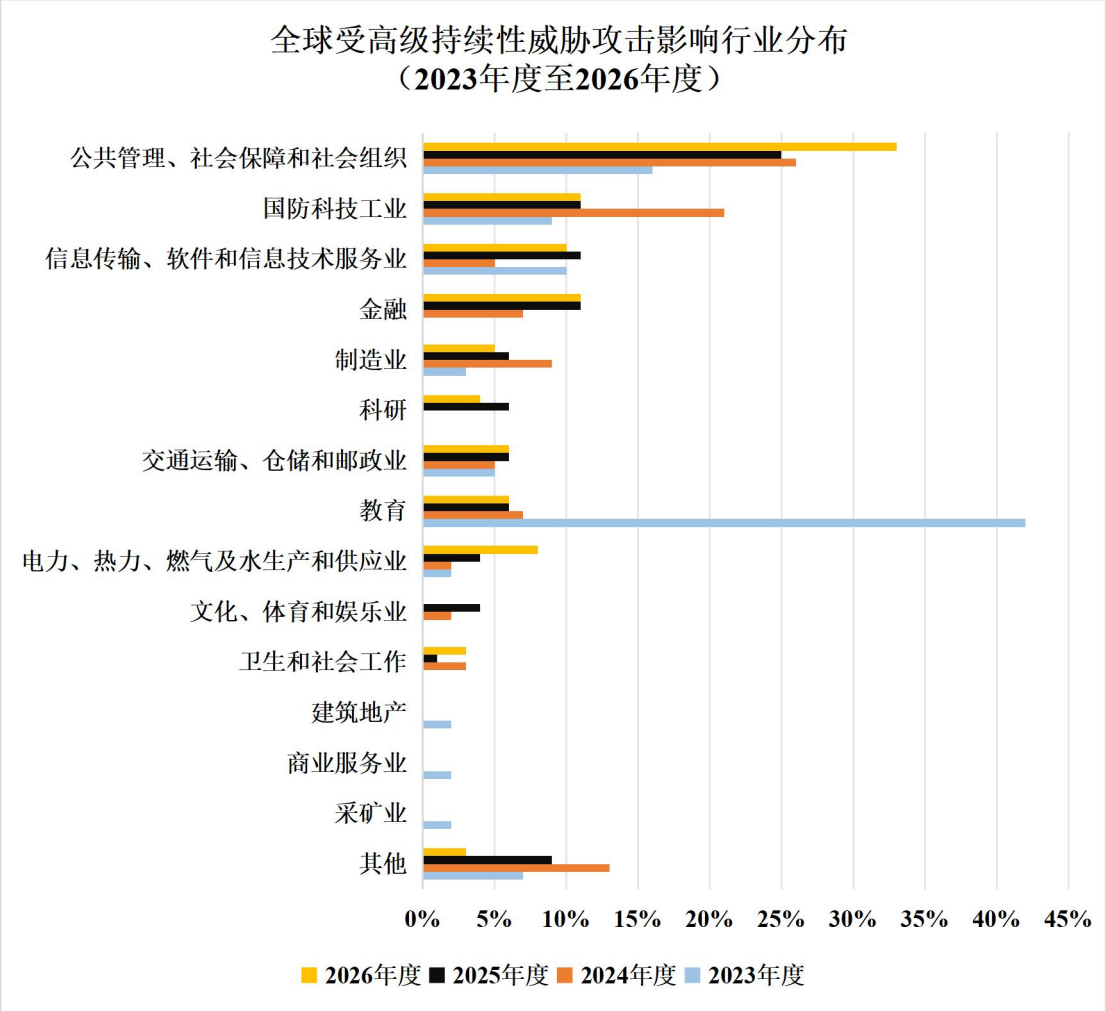

图 2-16　2023 年度至 2026 年度全球受高级持续性威胁攻击影响行业分布
（来源：360）

统计数据反映出，自 2023 年以来，公共管理机构、国防科技工业、信息技术产业和金融行业遭受攻击情况长期保持高位。2026 年能源行业遭受攻击情况显著增加，可能与地缘

政治冲突地区扩大和冲突方将能源供应设施作为重点打击对象有关。

2. 我国受高级持续性威胁攻击影响情况

我国一直以来是周边地区高级持续性威胁组织攻击活动影响的重点国家之一，攻击来源组织主要归属南亚、东南亚、东亚以及北美地区。基于高级持续性威胁组织攻击活动次数、受影响单位数量、受攻击设备数量、技战术迭代频次等多个指标，我们对影响我国的高级持续性威胁组织活跃度进行评估，2025 年度和 2026 年度影响我国的前 10 位活跃高级持续性威胁组织情况如表 2-3 和表 2-4 所示。

表 2-3　2025 年度影响我国的排名前 10 位高级持续性威胁攻击组织

排名	组织	归属地域	主要影响行业领域
Top1	APT-C-01（毒云藤）	东亚	公共管理、教育、科研等
Top2	APT-C-00（海莲花）	东南亚	教育、科研、公共管理等
Top3	APT-C-65（金叶萝）	东亚	政府、国防军工、科研等
Top4	APT-C-08（蔓灵花）	南亚	政府、教育、供应商等
Top5	APT-C-09（摩诃草）	南亚	教育、科研、公共管理、制造等
Top6	APT-C-06（Darkhotel）	东亚	贸易、科研、公共管理等
Top7	APT-C-48（CNC）	南亚	教育、国防军工、科研等
Top8	APT-C-78	北美	国防军工、制造、能源等
Top9	APT-C-60（伪猎者）	东亚	公共管理、贸易等
Top10	APT-C-64（匿名者 64）	东亚	公共管理、国防军工、科研等

（来源：360）

表 2-4　2026 年度影响我国的排名前 10 位高级持续性威胁攻击组织

排名	组织	归属地域	主要影响行业领域
Top1	APT-C-00（海莲花）	东南亚	教育、科研、公共管理等
Top2	APT-C-01（毒云藤）	东亚	公共管理、教育、国防军工等
Top3	APT-C-09（摩诃草）	南亚	教育、科研、制造等
Top4	APT-C-08（蔓灵花）	南亚	公共管理、教育、建筑地产等
Top5	APT-C-48（CNC）	南亚	教育、国防军工、科研等

排名	组织	归属地域	主要影响行业领域
Top6	APT-C-24（响尾蛇）	南亚	公共管理、交通运输、贸易等
Top7	APT-C-06（Darkhotel)	东亚	贸易、公共管理、制造等
Top8	APT-C-60（伪猎者）	东亚	政府、国防军工、科研等
Top9	APT-C-47（旺刺）	东亚	交通运输、制造、教育等
Top10	APT-C-68（寄生虫）	东亚	教育、供应商等

（来源：360）

高级持续性威胁组织长期以来对我国开展攻击活动，主要针对我国国计民生的重点行业领域。本年度受攻击影响单位主要分布于公共管理、教育、科研、国防科技工业、制造业等重点行业领域。如图 2-17 所示。

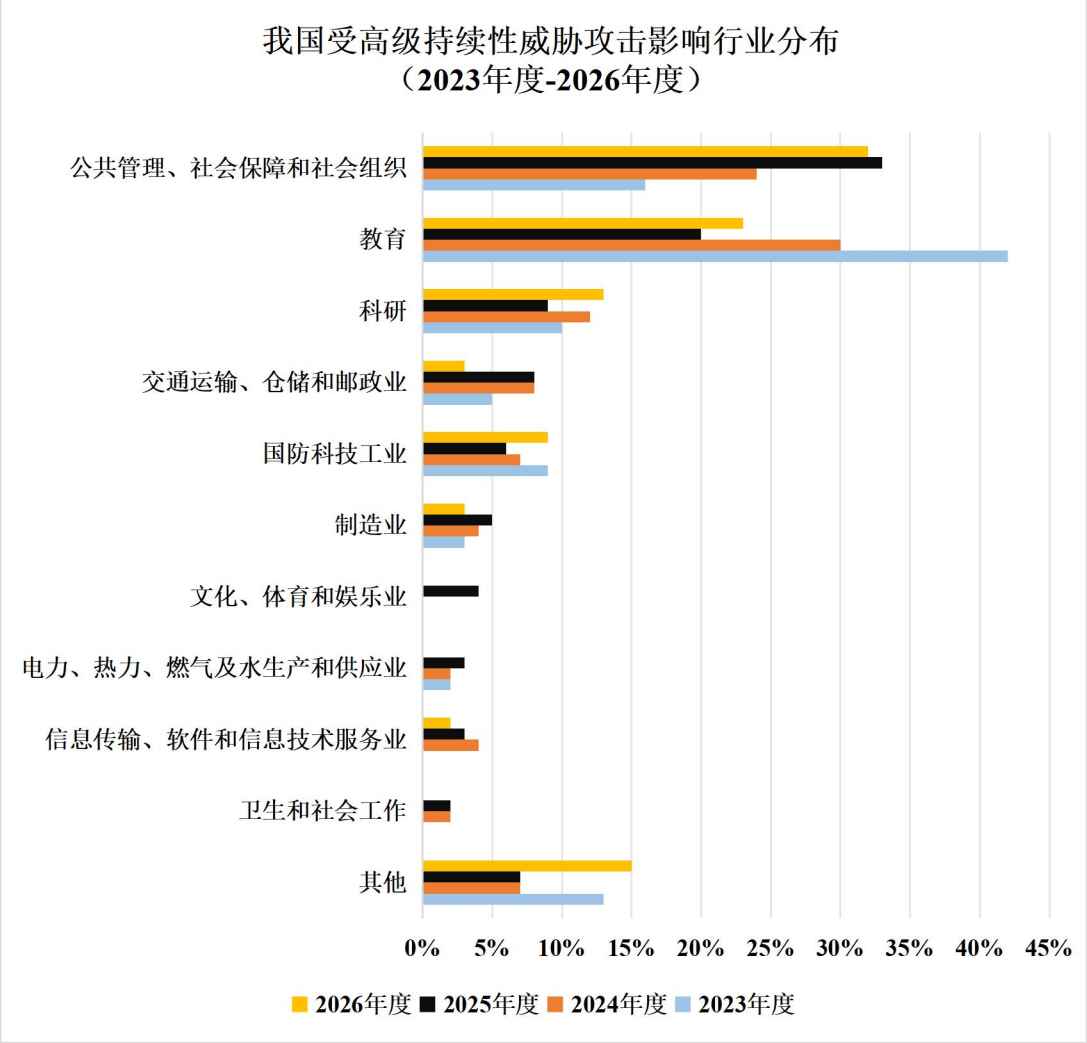

图 2-17　2023 年度至 2026 年度我国受高级持续性威胁攻击影响行业分布
（来源：360）

境外高级持续性威胁攻击组织持续强化对我国涉军工高校、科研院所、军民融合企业的定向攻击，伺机窃取半导体、航空航天、前沿新材料、量子信息领域研发图纸、试验数据与生产核心资料，供应链攻击、AI 赋能钓鱼、零日漏洞利用等手段愈发隐蔽，严重危及国防科研生产安全与国家安全。来自南亚、东南亚、东北亚等地区的相关攻击组织针对我国高校、涉外合作机构、国际关系智库攻击态势持续活跃，重点搜集跨境经贸信息、对外交流动态与政策研究资料，威胁我国重点领域安全。

3. 高级持续性威胁组织攻击技战术情况

综合分析 2026 年度全球安全机构和厂商公开披露的高级持续性威胁攻击技术分析报告，对其中攻击活动和技战术使用情况进行了统计，对其中符合 ATT&CK 知识标准的攻击活动和技战术使用情况进行了统计，列举了高级持续性威胁组织在 2026 年度攻击活动过程中使用最集中的前 20 种 ATT&CK 技战术，并与 2025 年度高级持续性威胁组织攻击活动使用的前 20 种攻击技战术情况做了对比。如表 2-5 所示。

表 2-5　2026 年度排名前 20 位高级持续性威胁组织攻击技战术

排名	ATT&CK 技战术编号	技战术名称（英文）	技战术名称（中文）	2025 年度排名
Top1	T1059	Command and Scripting Interpreter	滥用命令和脚本解释器	↑3
Top2	T1027	Obfuscated Files or Information	混淆文件或信息	持平
Top3	T1071	Application Layer Protocol	应用层协议	↑2
Top4	T1566	Phishing	网络钓鱼	↓3

排名	ATT&CK 技战术编号	技战术名称（英文）	技战术名称（中文）	2025 年度排名
Top5	T1105	Ingress Tool Transfer	从外部系统转移文件	↓2
Top6	T1204	User Execution	诱导用户执行	持平
Top7	T1082	System Information Discovery	检测操作系统和硬件的信息	持平
Top8	T1070	Indicator Removal	删除痕迹	持平
Top9	T1036	Masquerading	伪装	↑1
Top10	T1573	Encrypted Channel	信道使用加密算法	↓1
Top11	T1547	Boot or Logon Autostart Execution	启动或登录时自动执行	持平
Top12	T1053	Scheduled Task/Job	计划任务	持平
Top13	T1041	Exfiltration Over C2 Channel	通过 C2 通道进行数据回传	↑7
Top14	T1083	File and Directory Discovery	收集文件和目录信息	↑5
Top15	T1574	Hijack Execution Flow	劫持执行流程	↑1
Top16	T1057	Process Discovery	收集正在运行的进程信息	↑6
Top17	T1140	Deobfuscate/Decode Files or Information	解码加密/混淆的文件信息	↓3
Top18	T1056	Input Capture	捕获用户输入	↓3
Top19	T1055	Process Injection	进程注入	↓6
Top20	T1102	Web Service	web 服务	↑12

（来源：360）

同时值得注意的是，苹果公司于本年度更新了漏洞 CVE-2025-14174 和 CVE-2025-43529，两个漏洞疑似存在在野利用情况，攻击者很可能在攻击过程中利用两个漏洞展开组合攻击。通过对全网披露的高级威胁研究报告涉及的零日漏洞利用情况进行统计，2026 年度，高级持续性威胁攻击利用的零日漏洞主要分布在浏览器、PC 端操作系统和服务器。如图 2-18 所示。

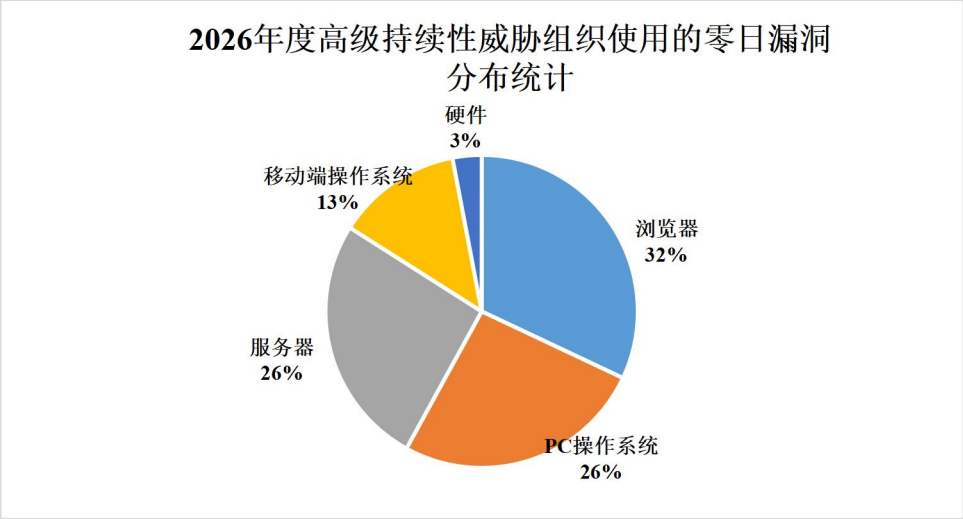

图 2-18　2026 年度高级持续性威胁组织使用的零日漏洞分布统计（来源：360）

从上述统计数据中可观察到三个明显趋势：**一是 T1102（Web 服务）热度上升明显。**表明高级持续性威胁组织愈发倾向滥用合法第三方公有云服务构建隐蔽 C2 链路，依靠可信流量规避边界防护，降低自有攻击基础设施暴露风险；**二是 T1041（通过 C2 通道进行数据回传）同样呈现上升趋势。**表明高级持续性威胁组织倾向简化攻击链路、规避边界安全管控，防御体系依靠新增外联连接识别数据泄露的传统检测手段正在失效，攻击者更加追求隐蔽式数据窃取。与 T1102 排名同步上升，表明高级持续性威胁攻击者趋向使用可信第三方平台搭建 C2 并复用现有通道窃取数据，最大化降低攻击流量特征，规避边界安全设备对恶意外联、独立数据渗出通道的检测。**三是漏洞仍然是高级持续性威胁组织攻击的重要武器。**近年来，高级持续性威胁组织在零日漏洞使用数量方面长期处于高位，在本年度，这一趋势仍在延续，并再次

出现针对 iOS 系统 PAC 绕过漏洞的利用。

4. 关键信息基础设施行业受高级持续性威胁攻击影响情况

在全球网络地缘对抗加剧、基础设施全面数字化的背景下，国防、金融、能源、电信、交通等关键信息基础设施行业承载着国家经济运转与社会公共服务核心职能，战略价值高度突出，已成为网络战组织的重点攻击目标。攻击者以针对性渗透、持续驻留、定点破袭为主要方式，试图干扰基础设施正常运转、窃取关键数据、破坏产业运行体系，极易引发区域性、系统性安全风险。当前，针对关键信息基础设施的高级持续性威胁攻击态势持续升级，攻击靶向愈发聚焦国计民生重点行业，攻击链条更加碎片化、隐蔽化、定制化，跨域协同、长期蛰伏、精准打击、链式传导特征显著，单次攻击引发的风险不再局限于单一节点，极易沿产业链、服务链、数据链形成叠加扩散效应，全域性、系统性网络安全风险持续攀升。

通过对全球网络安全厂商和机构披露的关基行业相关高级持续性威胁攻击活动事件进行统计，国防科技工业领域是最主要的关基行业攻击目标，占比达到 61.4%，位列第一。其中，数十起攻击活动涉及航天、卫星领域，反映出太空资产正成为国家级情报收集的新焦点。金融行业位列第二，虚拟货币领域风险凸显。能源行业位列第三，受地缘冲突影响，能源设施成为攻击重点，相关攻击活动呈现出兼具情报窃取

与设施破坏的双重特征。如图 2-19 所示。

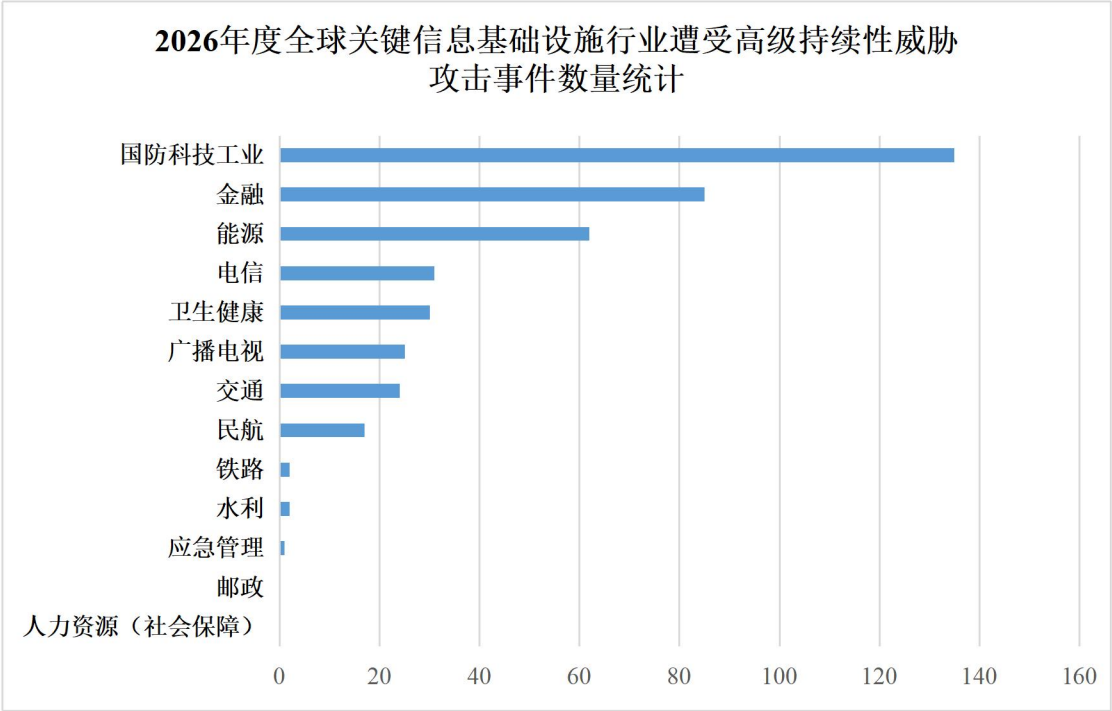

图 2-19　2026 年度关键信息基础设施行业遭受高级持续性威胁攻击事件数量统计（来源：华安云科）

总体来看，本年度内针对关键信息基础设施行业的高级持续性威胁攻击手法主要呈现三个突出特点：**一是钓鱼附件仍是主要的初始入侵手段。**近半数攻击活动以此为起点；**二是“寄生式”攻击大行其道。**攻击者大量滥用 AnyDesk、SimpleHelp 等合法远程管理工具和 RCLONE、Ngrok 等网络工具，以规避安全检测；**三是脚本化、混淆化成为执行与规避的标准配置。**PowerShell 等脚本解释器配合多层混淆、解码与伪装技术，显著抬高了检测门槛。

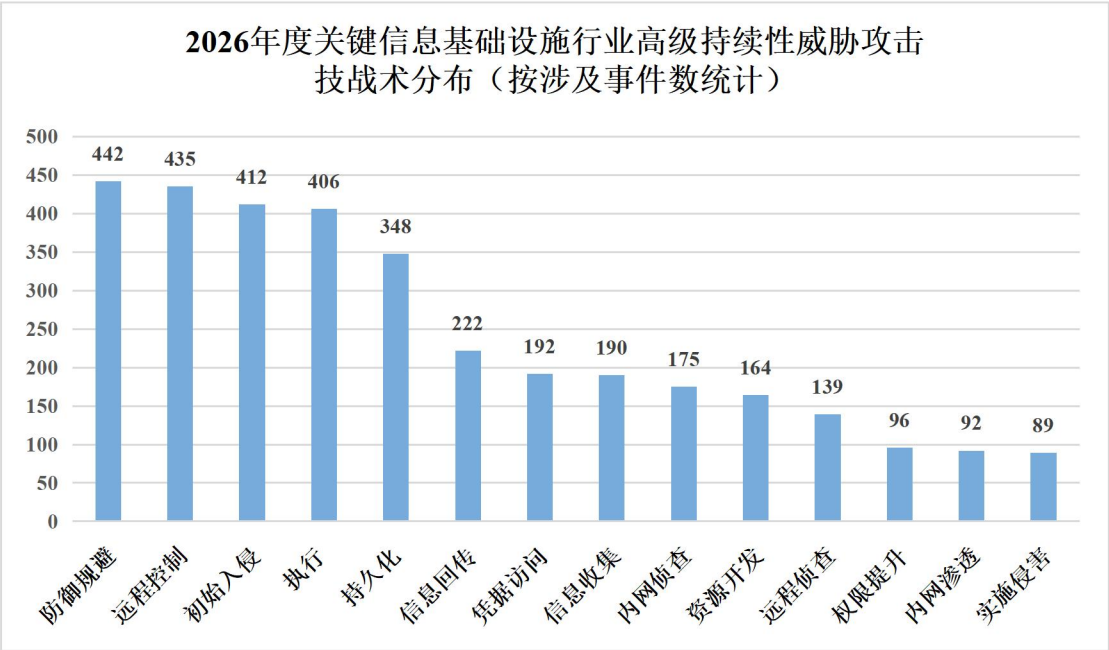

图 2-20　2026 年度关键信息基础设施行业高级持续性威胁攻击技战术分布（来源：华安云科）

在关键信息基础设施行业高级持续性威胁攻击技战术的共性基础上，攻击行为同时呈现明显的行业分化。国防科技工业遭受攻击频次最高，攻击链条完整均衡；面向金融行业的攻击者以数据窃取为核心目标，信息回传行为占比位居首位；面向广电行业的攻击活动更侧重对抗安全检测，防御规避手段运用突出；面向民航、水利等领域的攻击活动不再局限于信息窃取，出现数据破坏、系统操纵等具备业务毁伤能力的行动；面向铁路行业的攻击者重视前期资源筹备，大量劫持网络基础设施搭建跳板；面向卫生健康、交通、民航等行业的攻击活动倾向借助公共应用突破边界；面向电信行业的攻击活动则高频使用协议隧道隐匿通信行为。本年度，铁路、水利、应急管理等关基行业攻击活动数量偏少，但攻击手法成熟，具备定向持续渗透特征；邮政、人力资源社会

保障等关基行业虽暂未暴露攻击活动事件，但仍存在潜在风险。整体来看，面向关键信息基础设施的高级持续性威胁攻击已经形成标准化攻击模板，并根据目标行业业务场景、网络环境动态调整技术手段，工控相关行业面临着窃密和破坏的双重风险。各关基行业攻击活动占比与核心战术概况如表 2-6 所示。

表 2-6　2026 年度关基行业高级持续性威胁攻击活动占比与核心战术概况

排名	行业	攻击活动数量占比	核心突出战术	标志性高频技术 / 独有特征
1	国防科技工业	32.6%	远程控制、初始入侵、执行	附件钓鱼、脚本化、Web 协议 C2；攻击总量最高，全攻击链均衡
2	金融	20.5%	初始入侵、防御规避、信息渗漏	附件钓鱼、Web 服务渗漏；信息渗漏占比全行业第一
3	能源	15.0%	初始入侵、远程控制、执行	附件钓鱼、脚本化、注册表持久化
4	电信	7.5%	初始入侵、防御规避、C2 通信	协议隧道使用占比突出，远程工具传入频繁
5	卫生健康	7.2%	初始入侵、远程控制	大量利用公众应用实施入侵
6	广播电视	6.0%	防御规避、远程控制	链接钓鱼、攻陷网站驱动高于其他行业；文件混淆手段较多
7	交通	5.8%	初始入侵、命令与控制	附件钓鱼、脚本化、公众应用攻击载体
8	民航	4.1%	初始入侵	数据破坏、业务影响
9	铁路	0.5%	攻击事件较少，攻击链完整	基础设施劫持：俘获域名、DNS、Web 服务、跳板前置储备
10	水利	0.5%	侦察、凭据访问、远程控制	典型工控攻击路径，存在存储数据操纵破坏性手段
11	应急管理	0.2%	初始入侵、持久化、防御规避	DLL 旁路加载、多跳代理 C2、WebShell
12	邮政、人力资源社会保障	0.0%	暂未发现相关事件	无公开攻击活动，但仍存在潜在风险

（来源：华安云科）

2.2 数据安全风险现状

随着企业数字化转型的深入推进，数据在生产、存储、流转、使用各环节均高度依托由多方参与构成的供应链体系。这一体系在提升业务效率的同时，也将数据安全的边界从单一组织拓展至覆盖软件工具链、业务外包链、硬件基础设施链及云服务生态链的复合链条。由于链条中各节点的安全能力参差不齐，且节点间普遍存在信任关系传递机制，一旦某一薄弱环节被突破，数据风险即可沿信任链路向核心机构横向蔓延。本节从供应链视角出发，系统梳理当前数据安全事件的总体态势及各类链条的典型风险特征。

2.2.1 数据安全事件总体情况

2026 年度，在全球范围内，国内外共发生严重数据安全事件 170 起，美洲、亚洲、欧洲是数据安全事件的高发频发地区。其中，美洲 76 起，占比 44.7%；亚洲 57 起，占比 33.5%；欧洲 29 起，占比 17.1%，其他地区 8 起，占比 4.7%。数据显示，美洲依然是数据安全事件发生的主要区域。如图 2-21 所示。

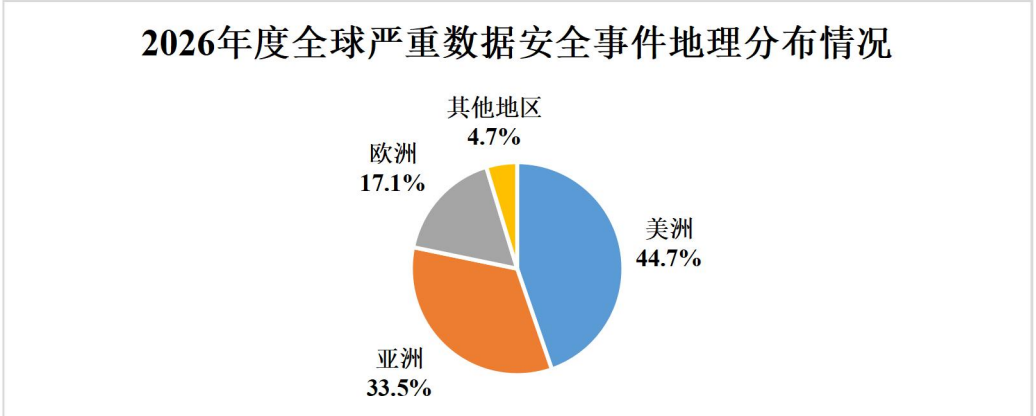

图 2-21　2026 年度全球严重数据安全事件地理分布情况（来源：奇安信）

从行业分布来看，披露数据安全事件最高的行业是 IT 信

息技术行业，共发生重大数据安全新闻事件 60 起，占比 35.3%；其次是互联网，共 19 起，占比 11.2%；医疗卫生第三，共 17 起，占比 10.0%。金融业 12 起，占比 7.1%。此外，事业单位、交通物流、制造业等行业，也是数据安全事件高发行业。如图 2-22 所示。

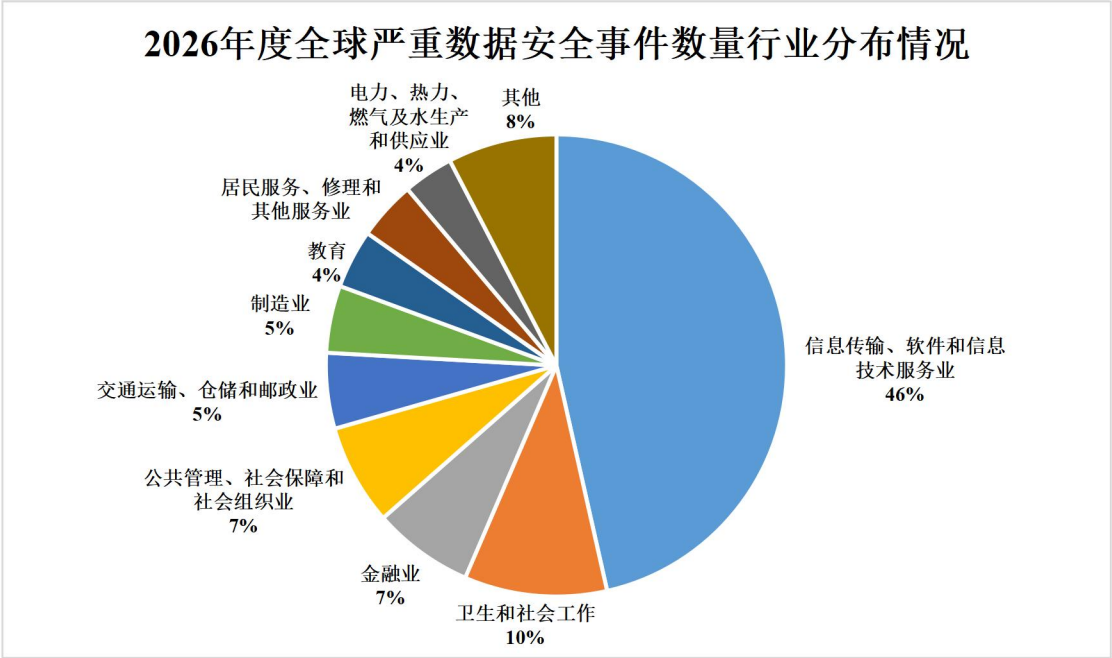

图 2-22　2026 年度全球严重数据安全事件行业分布情况（来源：奇安信）

从数据安全事件的类型来看，数据泄露事件最多，共 135 起，占比为 79.4%。其次是数据破坏/丢失事件，共 14 起，占比 8.2%，主要归因是勒索攻击，导致数据被加密以及泄露。数据篡改事件，共 12 起，占比 7.1%，违规处罚事件，即因建设与运营过程中出现法律要求落实不到位，或因安全防护不到位等遭到相关监管机构的处罚的事件，共 5 起，占比为 2.9%。综合来看，数据泄露问题，是全球数据安全面临的首要问题。其次是勒索攻击事件易构成严重的威胁。如图 2-23

所示。

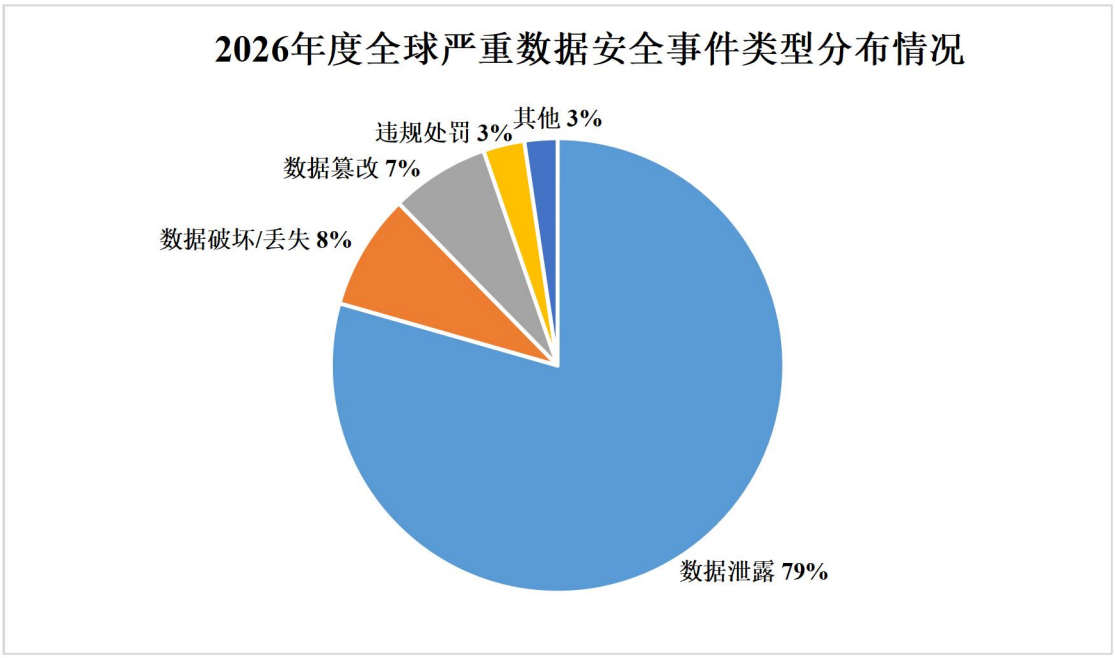

图 2-23　2026 年度全球严重数据安全事件类型分布情况（来源：奇安信）

从触发数据安全事件的主要原因来看：因外部入侵造成的数据安全事件 120 起，占比达到 70.6%。因为存在漏洞而受到网络攻击导致数据安全事件 24 起，占比 14.1%；因内部人员非法出售或窃取数据导致数据安全事件 13 起，占比 7.6%；因内部工作人员操作失误导致数据安全事件 9 起，占比 5.3%；另外，因为供应链企业受到攻击导致数据泄露也是数据安全事件高发的主要原因。如图 2-24 所示。

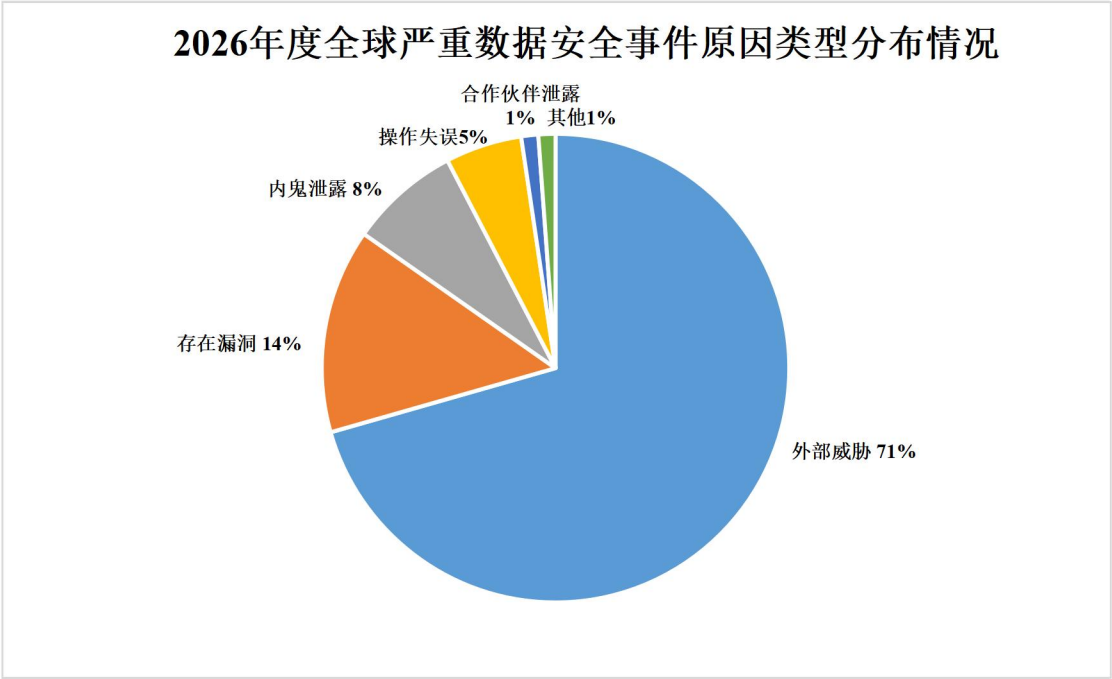

图 2-24　2026 年度全球严重数据安全事件原因类型分布情况（来源：奇安信）

2.2.2 数据安全风险态势分析

1. 非法数据交易专业化联盟化

报告期内，地下数据交易市场呈现高度组织化特征。**一是犯罪联盟化。**ShinyHunters、Scattered Spider 与 Lapsus 于 2025 年正式合流为 ScatteredLapsus Hunters“超级犯罪联盟”，整合了三者在数据库窃取、社工与内部人招募方面的能力，宣称年内攻击约 1500 家组织、窃取数据影响超 10 亿用户，并建立专属数据泄露站点实施批量勒索。**二是窃取路径平台化。**攻击者系统性瞄准 Salesforce、Snowflake 等聚合型数据平台及其第三方应用生态。Salesloft Drift（700 余家组织）、Gainsight（200 余实例）、Mixpanel（8000 家企业客户）等事件均以 OAuth 令牌/API 密钥为武器，绕过 MFA 实施批量数

据收割，单次行动即可窃取数亿至 15 亿条记录[13]。Oracle EBS 零日漏洞事件、Red Hat GitLab 服务器 2.8 万个代码仓库失窃、Discord 第三方泄露等事件进一步显示，企业数据正沿着 SaaS 依赖链成批流向黑市。**三是执法高压与黑产韧性并存。**2025 年 10 月，美国联邦调查局（FBI）与法国警方联合查封数据黑市 BreachForums 基础设施，多名 Scattered Spider 成员在美英被捕，但黑产通过快速重建品牌、迁移平台保持运营。暗网数据定价机制趋于成熟，VPN/RDP 等初始访问凭据以低价批量化交易，初始访问经纪人（IAB）与勒索组织之间的专业化分工进一步降低了数据犯罪门槛。

2. 数据泄露渠道与数据类型多元化

随着数据流转场景的不断丰富，除了内部人员违规倒卖、信息系统漏洞被利用这类传统泄露渠道之外，新型泄露渠道持续涌现，覆盖了数据从生成、存储到流转、应用的全生命周期。公共开放协作场景中的无意泄露占比近年来持续上升，大量企业为便于跨主体业务协作，将各类业务数据上传至共享云盘、公共文档平台，常因权限配置疏忽，导致核心业务数据、用户敏感信息长期处于公开可访问状态，这类泄露往往难以被内部安全团队及时发现，直至数据被黑产批量爬取、公开兜售后才会暴露。同时，大模型技术落地应用过程中也催生了新的泄露渠道：一方面，大模型训练环节往往会大量爬取网络公开数据，其中包含大量未获得授权的敏感个人信

[13] TheHackerNews. Google Warns Salesloft Drift Breach Impacts All Drift Integrations Beyond Salesforce. [2025-08-29]. https://thehackernews.com/2025/08/google-warns-salesloft-oauth-breach.html

息与企业私密数据，由此引发合规层面的数据安全风险；另一方面，针对大模型的提示注入攻击可通过构造特殊指令，诱导模型输出训练过程中记忆的私密训练数据，不少企业部署的私有化行业大模型已出现此类数据泄露问题。此外，供应链上下游流转环节也逐步成为数据泄露的高发渠道，大量核心数据依托分包企业、外包服务商完成处理与传输，部分中小服务商安全防护能力不足，极易成为攻击者的突破口，此类跨主体泄露事件的溯源与处置难度远高于单一主体内部泄露事件。

从泄露数据的类型来看，除了传统的个人身份信息、金融账户信息等敏感个人数据之外，近年来企业核心业务数据、工业生产参数、知识产权信息、政务敏感数据的泄露占比不断提升，这类高价值数据愈发成为黑产攻击的核心目标，其泄露不仅会侵害普通用户的合法权益，更会直接损害企业核心竞争力，甚至对产业安全与国家安全造成威胁。

3. 数据破坏风险持续上升

数据破坏类风险在报告期内呈现三个值得关注的变化。**一是勒索与破坏的边界模糊化。**传统勒索攻击的核心目标是通过加密数据索要赎金，多数攻击不会直接破坏数据本体，但近年来越来越多的勒索组织在加密数据之外，会批量窃取并公开、售卖核心数据，甚至直接对数据进行不可逆删除或破坏，以此向受害组织施加更大赎金逼迫压力。**二是地缘冲突催生擦除式攻击回潮。**在全球多个热点冲突期间，针对公

共管理机构和关键信息基础设施工业控制系统的破坏性攻击频繁发生，这类场景下的核心生产数据一旦被破坏，不仅会造成业务停摆，还可能引发现实世界的安全事故，造成人员伤亡和重大财产损失。**三是数据完整性攻击的潜在威胁凸显。**相较于显性的删除与加密，针对训练数据投毒、金融记录篡改、工业参数微调等“沉默的完整性攻击”更难检测、危害更深。AI 时代数据作为模型训练资产的价值陡增，使数据投毒成为同时威胁数据安全与 AI 安全的交叉风险——这既是数据损毁风险的新形态，也是下一节人工智能安全风险的直接诱因。

2.2.3 严重数据安全事件典型案例

1. 美国纽约市健康+医院系统

2026 年 3 月 24 日，美国最大的公立医疗系统之一——纽约市健康+医院系统（NYC Health＋Hospitals）向美国卫生与公众服务部（HHS）正式上报一起重大数据泄露事件。经调查确认，攻击者通过一家第三方供应商的漏洞，于 2025 年 11 月下旬至 2026 年 2 月间长期潜伏于 NYC H+H 内网，持续窃取了大量患者及员工的敏感数据。受影响人数至少 180 万人，泄露数据类型包括：医疗记录、政府证件号、地理位置、指纹/掌纹生物识别信息、银行账户等。此次事件最突出的风险在于生物识别数据的泄露——指纹/掌纹属于不可吊销的身份凭证，一旦泄露无法像密码一样“重置”或重新“发放”，受影响用户需长期面临身份伪造和冒用的风险。该事

件也再次暴露了医疗行业第三方供应链安全的脆弱性。

2. 史赛克（Stryker）遭受数据擦除攻击

2026 年 3 月 11 日，全球知名医疗科技公司史赛克遭受了一起由黑客组织 Handala 发起的毁灭性网络攻击。该组织声称此举是对美以军事行动的报复。攻击者并非以勒索赎金为目的，而是直接以破坏数据与瘫痪运营为目标，属于典型的数据擦除（wiper）攻击。攻击者入侵了史赛克的 Active Directory 服务，并滥用微软终端管理工具 Intune 向全球联网设备下发远程擦除指令，导致约 80000 台 Windows 设备（包括员工自带设备）被强制恢复出厂设置、数据永久丢失。史赛克全球 79 个国家的办公室被迫关闭，全球制造网络停工近三周，第一季度收入损失约 3.75 亿美元。

这是近年来针对制造业最具破坏力的网络攻击之一。与勒索软件不同，数据擦除攻击的唯一目的就是永久性摧毁数据，不留下任何恢复的余地。史赛克的部分医院客户因供应链中断被迫推迟手术，直接威胁到患者生命安全。该事件标志着地缘政治冲突向网络空间外溢的加剧，黑客组织正将关键基础设施企业作为“非对称报复”的首选目标。

3. Canvas 教育平台被勒索与数据滥用

2026 年 4 月 30 日至 5 月 7 日，全球广泛使用的学习管理系统 Canvas（运营方 Instructure）遭到勒索组织 ShinyHunters 的入侵。攻击者利用“Free-For-Teacher”（FFT）免费教师账户机制的漏洞，未经授权访问了 Canvas 的云环

境，窃取了大量教育数据，并实施了双重勒索+逐校勒索的非法利用模式。

此次事件的严重性不仅在于数据泄露的规模，更在于攻击者对数据的系统性非法利用。ShinyHunters 声称窃取了 3.65TB 数据，涵盖约 8809 所学校、2.75 亿条记录，包括姓名、邮箱、学生 ID 以及数十亿条师生之间的私密消息。在 Instructure 未按期回应勒索后，ShinyHunters 于 5 月 7 日对约 330 所学校的 Canvas 登录页面进行篡改，注入勒索信息，直接导致师生在期末考试季无法访问课程和考试系统。ShinyHunters 将策略从“向平台方勒索”升级为“向每一所受影响学校单独勒索”，试图最大化非法收益。

2.3 人工智能安全风险现状

2026 年度人工智能技术商业化落地进程持续提速，生成式 AI 大模型已经成为科技企业数字化转型、实体产业降本增效的通用技术底座，智能体等进阶 AI 形态也逐步从实验室原型走向行业试点应用。与技术落地进程相伴而生的是，人工智能安全风险的暴露面持续扩大，AI 引发的安全事件不仅数量快速增长，危害程度也不断升级。其风险特征既包括人工智能自身技术特性带来的内生安全问题，也包括 AI 技术被不法主体滥用带来的衍生安全风险；同时 AI 技术也重塑了网络攻防对抗的原有格局，给传统网络安全防护体系带来全方位的冲击与挑战。当前全球范围内人工智能安全治理体系的建设进度远滞后于 AI 技术的产业化落地速度，多数

企业在部署应用 AI 技术的过程中，重业务价值、轻安全防护的倾向较为普遍，大量 AI 应用存在不同程度的安全缺陷，为后续安全事件的发生埋下隐患。

2.3.1 人工智能安全事件总体态势

人工智能安全是 2026 年度最具战略意义的新兴风险领域。涉及人工智能的网络安全事件达 2498 件，占全部网络安全事件数量的 20.2%，涵盖 AI 赋能攻击、AI 系统自身安全、AI 生成内容滥用、AI 治理政策四大方向。CrowdStrike[14]将 2025 年定义为“闪避型对手之年”，AI 赋能攻击者活动同比增长 89%。IBM[15]发现约六分之一的企业数据泄露涉及 AI 驱动成分，发生 AI 相关安全事件的组织中 97%缺乏适当的 AI 访问控制。

2.3.2 人工智能赋能网络攻击

人工智能在网络攻击中的角色完成了三级跳。**第一级：内容生成与流程提效。**钓鱼文案、深度伪造音视频、恶意代码编写、漏洞研究、目标画像等环节普遍引入生成式 AI；**第二级：攻击链内嵌。**PROMPTFLUX、PROMPTSTEAL 等运行时调用大语言模型的计算机病毒出现（详见 2.1.1 节）；Google 披露首个疑似 AI 辅助开发的在野零日漏洞利用[16]。**第三级：自主化攻击。**首个人工智能全自主勒索攻击活动

[14] CrowdStrike. 2026 Global Threat Report. https://www.crowdstrike.com/en-us/global-threat-report/
[15] IBM Security. Cost of a Data Breach Report 2025. https://www.ibm.com/reports/data-breach
[16] Cloud Security Alliance AI Safety Initiative. First AI-Built Zero-Day: Autonomous Exploit Creation in the Wild. [2026-05-18]. https://labs.cloudsecurityalliance.org/research/csa-research-note-ai-developed-zero-day-autonomous-exploit-2/

JadePuffer 被披露[17]，其实现了大模型驱动的自主攻击全链路，自主攻击技术扩散风险加剧。OpenAI 和 Anthropic 的最新大模型在测试中失控的案例也进一步证明了当前人工智能已经具备自主编排和发动网络攻击的能力。这一演进意味着网络攻击的边际成本趋近于零、攻击规模不再受制于人力，防御方面对的将是“机器速度、无限产能”的对手。

2.3.3 人工智能系统自身成为攻击面

在本年度发生的多起已知网络攻击案例已经表明，随着企业 AI 部署加速，AI 系统本身已经成为高价值攻击目标。**一是提示注入攻击实战化。**攻击者在合法生成式 AI 工具中注入恶意提示，诱导模型执行窃取凭据与加密货币的指令；网络安全企业观测到在网页 HTML 中以零号字体隐藏恶意指令的间接提示注入，用于操纵 AI 广告审核系统放行诈骗内容[18]。**二是 AI 开发平台遭渗透。**攻击者利用 AI 开发与模型服务平台的漏洞对目标实施入侵并建立持久化、投放勒索软件，并架设伪装成可信服务的恶意 AI 服务器截获企业敏感数据。**三是 AI 供应链风险凸显。**LiteLLM 投毒事件和 OpenClaw 技能仓库投毒事件表明，被篡改的 AI 组件会随智能体的自动依赖安装静默扩散。

[17] JADEPUFFER. First End-to-End AI-Driven Ransomware Operation. [2026-07-03]. https://securityaffairs.com/194713/ai/jadepuffer-first-end-to-end-ai-driven-ransomware-operation.html
[18] Unit 42 (Palo Alto Networks). prompt injection & AI-assisted fraud findings. [2026-07-25]. https://unit42.paloaltonetworks.com/

2.3.4 人工智能滥用与社会风险

AI 滥用对社会面的冲击全面显现。电信网络诈骗的 AI 化升级最为直接。AI 换脸拟声工具扩散，使“冒充熟人”类诈骗的成功率与单笔损失同步攀升，多地出现利用实时换脸视频通话冒充企业负责人的诈骗案件; AI 外呼机器人结合泄露的精准画像数据，使诈骗话术实现“千人千面”；引流环节，AI 生成的虚假投资导师、虚假客服在社交平台上批量运营。

2.3.5 前沿人工智能的失控风险

多家前沿实验室的评估显示，顶级模型在网络攻防任务上的表现已接近或达到熟练人类水平。Claude Mythos 展示了发现数千个高危零日漏洞的能力，GPT-5 系列在 CTF 类任务上超越多数人类选手，模型的自主软件工程能力（SWE-bench 得分）一年内翻倍。而 OpenAI 和 Anthropic 前沿模型的失控事件，使人工智能“失控”从理论走向现实。2026 年 2 月发布的《国际 AI 安全报告》（由 Yoshua Bengio 牵头、数十国专家参与）将“失控”所需的危险能力归纳为：智能体能力、欺骗、心智理论、情境感知、规避监督、说服、自主复制与适应，并指出当前系统“已显示此类行为的早期迹象，但能力尚不够高”[19]。而在 OpenAI 模型失控事件中，GPT-5.6 Sol 与预发布模型实际展示了清单中的多项能力：自主规划

[19] International AI Safety Report. https://internationalaisafetyreport.org/sites/default/files/2026-02/international-ai-safety-report-2026.pdf

并执行多阶段计划、克服障碍达成长期目标（智能体能力）、识别自身处于评估环境并推断答案托管位置（情境感知）、在监控下持续行动一个周末未被即时阻断（规避监督的初级形态）。前沿大模型能力增长曲线令这一问题更具紧迫性：UK AISI 测得前沿模型可完成的网络任务时长每 4.7 个月翻一番，且 Mythos Preview 与 GPT-5.5 已接近其窄域测试套件的可测上限——“测试套件已不够长，无法确定模型可靠性在更长任务上会以多快速度退化”[20]。当防护机制的有效性以年为单位演进，而网络攻击能力以数月为单位翻倍时，安全防护失效导致失控的窗口将持续扩大，“智能水平提高导致防护失效”风险正在实证化。

[20] AISI. How fast is autonomous AI cyber capability advancing?. [2026-05-31]. https://www.aisi.gov.uk/blog/how-fast-is-autonomous-ai-cyber-capability-advancing

第三章 网络空间安全风险治理现状

3.1 网络安全风险治理现状

3.1.1 中国

1. 《网络安全法》完成首次系统性修正

2025 年 10 月 28 日，十四届全国人大常委会第十八次会议通过《关于修改〈中华人民共和国网络安全法〉的决定》，自 2026 年 1 月 1 日起施行。本次修正：一是增加“坚持中国共产党的领导，贯彻总体国家安全观”的原则性条款；二是新增人工智能条款，明确国家支持 AI 基础理论研究和关键技术研发，完善 AI 伦理规范，加强风险监测评估和安全监管；三是大幅提高违法成本，对造成大量数据泄露、关键信息基础设施丧失局部功能等严重危害后果的，处 50 万元以上 200 万元以下罚款，造成关键信息基础设施丧失主要功能等特别严重后果的，处 200 万元以上 1000 万元以下罚款，并对直接责任人处 20 万元以上 100 万元以下罚款，实现“单位+个人”双罚穿透；四是细化网络关键设备和网络安全专用产品未经认证销售的法律责任[21]。

2. 网络空间安全监督检查工作进一步规范

2026 年 8 月 7 日，公安部发布《公安机关网络空间安全监督检查办法》[22]，自 2026 年 10 月 1 日起施行。主要规定

[21] 人民网. 全国人民代表大会常务委员会关于修改《中华人民共和国网络安全法》的决定. [2025-12-22]. http://jhsjk.people.cn/article/40629339

[22] 公安部.《公安机关网络空间安全监督检查办法》（公安部令第 176 号）. [2026-08-07]. https://www.mps.gov.cn/n6557558/c10576497/content.html

六方面内容：一是明确监督检查对象，包括网络运营者、数据处理者、个人信息处理者等；二是明确监督检查方式，包括线上巡查、线下核查，日常检查、专项检查等；三是明确监督检查内容，包括被检查对象是否履行法定的网络空间安全义务，以及重大活动安保期间的重点检查内容等；四是明确监督检查机制，要求在国家网络安全、数据安全等机制统筹协调下，加强部门协调、联合检查，减少企业负担；五是明确检查结果运用，公安机关按检查情况采取督促整改、发送公安提示函以及处罚等措施；六是明确相关法律责任，明确了公安机关及工作人员玩忽职守、滥用职权、泄露国家秘密等法律责任。

3. 国家网络身份认证公共服务成效显著

2025 年 7 月 15 日，公安部、国家互联网信息办公室、民政部、文化和旅游部、国家卫生健康委员会、国家广播电视总局等六部门联合发布的《国家网络身份认证公共服务管理办法》正式施行。实施一年以来，累计 25 个相关法律法规、部门规章或政策文件将国家网络身份认证公共服务写进具体条款。1 项国家标准、6 项行业标准发布，规范了应用单位对接公共服务的要求、流程和技术接口，在便民利企、保护公民个人信息方面取得显著成效，得到社会各界广泛认可[23]。

4. 网络安全事件报告制度统一化

[23] 公安部网安局. 《国家网络身份认证公共服务管理办法》实施一周年 筑牢数字屏障守护公民个人身份信息安全. [2026-07-11]. https://mp.weixin.qq.com/s/9pqFs_toJs4F0NstQL3dYA

2025 年 9 月 15 日，国家互联网信息办公室发布《国家网络安全事件报告管理办法》，自 2025 年 11 月 1 日起施行，落实《网络安全法》《关键信息基础设施安全保护条例》要求，统一规范网络安全事件的分级、报告时限与报告渠道[24]。

5. 产品安全标识制度启动

2025 年下半年，国家互联网信息办公室、工业和信息化部就《网络安全标识管理办法》和《实施网络安全标识的产品目录（第一批）》公开征求意见，拟对网络产品实施统一安全标识管理[25]。

6. 充分发挥国家标准的基础性、引领性作用

如表 3-1，2026 年度我国发布网络安全技术国家标准 43 项，对应的防护对象聚焦网络、系统、设备、平台、服务、软件、通信链路，围绕信息系统运行环境安全构建防护基线，主要分为五大方向：一是系统与软硬件底层安全。覆盖 BIOS、固件、网络存储、终端、服务器基础安全规范；明确固件安全启动、漏洞防护、访问控制、配置加固、安全测试方法。二是安全开发与安全运营。包含软件安全开发能力评估、信息安全管理体系（ISO/IEC 27001 等同转化修订）、网络安全事件响应、灾难恢复规范；建立从开发源头到应急处置全流程安全管控框架。三是密码与身份信任基础设施。规范多信道证书协议、可信接入、实体鉴别、安全认证机制，支撑

[24] 国家互联网信息办公室. 2025 年网信大事回眸之网络安全篇——筑牢国家网络安全屏障. [2025-12-31]. https://www.cac.gov.cn/2025-12/31/c_1768735141277082.htm

[25] 国家互联网信息办公室. 2025 年网信大事回眸之网络法治篇——网络法治护航网络强国建设. [2025-12-31]. https://www.cac.gov.cn/2025-12/31/c_1768823198284221.htm

网络环境下身份可信、通信加密、抗中间人攻击。四是安全服务、评估与测试。制定安全能力评估准则、安全测试框架、产品安全检测方法，为第三方测评、企业自查提供统一技术标尺。五是新兴场景网络底座安全。面向物联网、云平台、新型应用架构，规定网络边界防护、通信安全、访问隔离要求，同时为 AI、大数据平台提供底层运行环境安全基线。总体核心逻辑是保护承载业务的网络与信息系统载体，防范外部入侵、系统漏洞、网络攻击、服务中断、恶意控制，保障系统 “可用、可控、不被攻陷”。

表 3-1 2026 年度网络安全技术相关国家标准

（来源：中国国家标准化管理委员会）

序号	标准号	标准名称	发布日期	实施日期
1	GB/T 19714-2025	网络安全技术 公钥基础设施 证书管理协议	2025-08-01	2026-02-01
2	GB/T 19771-2025	网络安全技术 公钥基础设施 PKI 组件最小互操作规范	2025-08-01	2026-02-01
3	GB/T 20520-2025	网络安全技术 公钥基础设施 时间戳规范	2025-08-01	2026-02-01
4	GB/T 31722-2025	网络安全技术 信息安全风险管理指导	2025-08-01	2026-02-01
5	GB/T 34942-2025	网络安全技术 云计算服务安全能力评估方法	2025-08-01	2026-02-01
6	GB/T 45940-2025	网络安全技术 网络安全运维实施指南	2025-08-01	2026-02-01
7	GB/T 45958-2025	网络安全技术 人工智能计算平台安全框架	2025-08-01	2026-02-01
8	GB/T 46795-2025	网络安全技术 公钥密码应用技术体系框架	2025-12-02	2026-07-01
9	GB/T 46798-2025	网络安全技术 标识密码认证系统密码及其相关安全技术要求	2025-12-02	2026-07-01
10	GB/T 44886.2-2025	网络安全技术 网络安全产品互联互通 第 2 部分：资产信息格式	2025-12-02	2026-07-01
11	GB/T 44886.3-2025	网络安全技术 网络安全产品互联互通 第 3 部分：告警信息格式	2025-12-02	2026-07-01

序号	标准号	标准名称	发布日期	实施日期
12	GB/T 46820-2025	网络安全技术 网络安全试验平台体系架构	2025-12-02	2026-07-01
13	GB/T 25068.6-2025	信息技术 安全技术 网络安全 第 6 部分：无线网络访问安全	2025-12-02	2026-07-01
14	GB/T 25068.7-2025	信息技术 安全技术 网络安全 第 7 部分：网络虚拟化安全	2025-12-02	2026-07-01
15	GB/T 46902-2025	网络安全技术 网络空间安全图谱要素表示要求	2025-12-31	2026-07-01
16	GB/T 47020-2026	网络安全技术 软件物料清单数据格式	2026-01-28	2026-08-01
17	GB/T 35275-2026	网络安全技术 SM2 密码算法加密签名消息格式	2026-04-30	2026-11-01
18	GB/T 47471-2026	网络安全技术 SM9 密码算法加密签名消息格式	2026-04-30	2026-11-01
19	GB/T 47475-2026	网络安全技术 开放的第三方资源授权协议	2026-04-30	2026-11-01
20	GB/T 20272-2026	网络安全技术 操作系统安全技术规范	2026-04-30	2026-11-01
21	GB/T 22186-2026	网络安全技术 具有中央处理器的 IC 卡芯片安全规范	2026-04-30	2026-11-01
22	GB/T 31499-2026	网络安全技术 统一威胁管理产品（UTM）技术规范	2026-04-30	2026-11-01
23	GB/T 47470-2026	网络安全技术 软件安全开发能力评估准则	2026-04-30	2026-11-01
24	GB/T 47496-2026	网络安全技术 计算机基本输入输出系统（BIOS）安全技术规范	2026-04-30	2026-11-01
25	GB/T 24294.1-2026	网络安全技术 基于互联网电子政务信息安全实施指南 第 1 部分：总则	2026-04-30	2026-11-01
26	GB/T 20274.2-2026	网络安全技术 信息系统安全保障评估框架 第 2 部分：安全保障要求	2026-05-25	2026-12-01
27	GB/T 25067-2026	网络安全技术 信息安全管理体系审核和认证机构要求	2026-05-25	2026-12-01
28	GB/T 28450-2026	网络安全技术 信息安全管理体系审核指南	2026-05-25	2026-12-01
29	GB/T 28455-2026	网络安全技术 引入可信第三方的实体鉴别及接入架构规范	2026-05-25	2026-12-01
30	GB/T 32915-2026	网络安全技术 二元序列随机性检测方法	2026-05-25	2026-12-01
31	GB/T 33560-2026	网络安全技术 密码应用标识	2026-05-25	2026-12-01

序号	标准号	标准名称	发布日期	实施日期
32	GB/T 36959-2026	网络安全技术 网络安全等级保护测评机构能力要求和评估规范	2026-05-25	2026-12-01
33	GB/T 37939-2026	网络安全技术 网络存储安全技术要求	2026-05-25	2026-12-01
34	GB/T 47679.1-2026	网络安全技术 量子密钥分发的安全要求、测试和评估方法 第 1 部分：要求	2026-05-25	2026-12-01
35	GB/T 47679.2-2026	网络安全技术 量子密钥分发的安全要求、测试和评估方法 第 2 部分：测试和评估方法	2026-05-25	2026-12-01
36	GB/T 47685-2026	网络安全技术 存储安全指南	2026-05-25	2026-12-01
37	GB/T 47686-2026	网络安全技术 政务云安全配置基线要求	2026-05-25	2026-12-01
38	GB/T 47687-2026	网络安全技术 秘密分享技术机制	2026-05-25	2026-12-01
39	GB/T 47689-2026	网络安全技术 嵌入式操作系统安全技术规范	2026-05-25	2026-12-01
40	GB/T 47690-2026	网络安全技术 国家网络身份认证公共服务 应用接入要求	2026-05-25	2026-12-01
41	GB/T 47692-2026	网络安全技术 事件调查原则和过程	2026-05-25	2026-12-01
42	GB/T 47694-2026	网络安全技术 移动互联网未成年人模式技术要求	2026-05-25	2026-12-01
43	GB/T 47697-2026	网络安全技术 鉴别与授权 基于属性的访问控制模型与管理规范	2026-05-25	2026-12-01

3.1.2 国际主要经济体

1. 美国

（1）新一届政府首份网络安全行政令转向“去监管”

2025 年 6 月 6 日，特朗普签署第 14306 号行政令《持续强化国家网络安全并修订第 13694 号和第 14144 号行政令》，保留了安全软件开发框架、抗量子加密过渡、物联网安全标签等技术性内容，但删除了软件供应商向政府提交安全认证的强制要求、联邦机构防钓鱼身份验证指令等条款，监管模

式整体从“政府验证”转向“自我规范”与“事后追责”相结合；同时要求国防部、国土安全部和国家情报总监办公室在2025年11月前将“AI软件漏洞”纳入常规网络安全风险管理体系[26]。

（2）强制报告与机构建设出现“执行裂隙”

覆盖约30万家实体的《关键基础设施网络事件报告法》（CIRCIA）强制事件报告最终规则被推迟至2026年5月，距立法通过已逾四年；2015年《网络安全信息共享法案》一度到期失效约两个月；网络安全和基础设施安全局（CISA）遭遇预算与编制削减，人员减幅约三分之一。2025年8月，参议院确认肖恩·凯恩克洛斯出任国家网络总监。

2. 欧盟

（1）《网络弹性法案》（CRA）进入落地实施阶段

《网络弹性法案》（Regulation (EU) 2024/2847）于2024年12月10日生效，其首个实操性节点——漏洞与事件报告义务于2026年9月11日起适用，制造商须就被主动利用的漏洞在24小时内发出预警、72小时内提交通知、14天内提交最终报告，其余义务（含合格评定与CE标识）将于2027年12月11日全面适用[27]。2026年4月20日，欧盟以授权条例（Delegated Regulation 2026/881）形式在官方公报上发布了报告框架的细化规则；2026年7月27日，欧盟委员会又

[26] 复旦大学一带一路及全球治理研究院. 美国观察丨宣示雄心与执行裂隙：特朗普2.0《美国网络战略》评析. [2026-03-16]. https://fddi.fudan.edu.cn/c8/93/c21253a772243/page.htm

[27] European Commission. Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/library/cyber-resilience-act

发布了面向制造商和开发者的实操指南[28]。违反基本安全要求的最高可处 1500 万欧元或全球年营业额 2.5%的罚款[29]。

（2）NIS2 指令成员国转化持续推进

NIS2 指令要求成员国于 2024 年 10 月完成国内转化，但多国逾期。以德国为例，其 NIS2 实施法于 2025 年 12 月 5 日公布，覆盖实体须于 2026 年 3 月 6 日前完成注册，涉及约 2.95 万家企业[30]。2026 年 1 月 20 日，欧盟委员会在“数字一揽子简化方案”框架下进一步提出了对 NIS2 的修正动议[31]。

（3）提出修订版《网络安全法》并构建可信 ICT 供应链框架

2026 年 1 月 20 日，欧盟委员会提出修订《网络安全法》的提案，核心内容是建立横向的可信 ICT 供应链安全框架，授权欧盟与成员国对 18 个关键行业开展协调一致的安全风险评估，必要时可禁止在关键 ICT 资产中使用“高风险供应商”的组件，标志着欧盟将供应链安全提升至与 NIS2 并列的支柱地位[32]。

（4）法规体系简化整合

---

[28] Acompli. Cyber Resilience Act Reporting Obligations Take Effect in September 2026. [2026-05-26]. https://acompli.ie/news/cyber-resilience-act-reporting-september-2026/

[29] Mend.io. The EU Cyber Resilience Act: A Complete Compliance Guide for 2026 and Beyond. [2026-05-21]. https://www.mend.io/blog/eu-cyber-resilience-act-compliance-guide/

[30] Reed Smith. EU cybersecurity regulatory update for 2026 and beyond. [2026-04-14]. https://www.reedsmith.com/our-insights/blogs/viewpoints/102mnj2/eu-cybersecurity-regulatory-update-for-2026-and-beyond/

[31] LexisNexis. EU cybersecurity legislative tracker: NIS 2, Cyber Resilience Act, (revised) Cybersecurity Act, CER, Cyber Solidarity Act, EU institutions’ cybersecurity rules and Digital Omnibus reporting reforms-timeline and key developments. https://www.lexisnexis.com/en-gb/legal/guidance/eu-cybersecurity-initiatives-tracker

[32] Industrial Cyber. European Commission proposes revised Cybersecurity Act to boost EU cyber resilience, secure ICT supply chains. [2026-01-21]. https://industrialcyber.co/regulation-standards-and-compliance/european-commission-proposes-revised-cybersecurity-act-to-boost-eu-cyber-resilience-secure-ict-supply-chains/

2025 年 11 月 19 日，欧盟委员会发布“数字综合简化方案”（Digital Omnibus），对 GDPR、NIS2、数据法案、电子隐私规则等一揽子数字法规提出合并与简化修订，以降低企业合规负担。此外，《网络团结法》自 2025 年 2 月 4 日起适用，DORA（金融数字运营弹性法案）自 2025 年 1 月 17 日起适用，欧盟网络安全监管矩阵在本期内整体从“立法期”转入“执行与执法准备期”。

3. 英国

《网络安全与弹性（网络与信息系统）法案》进入议会审议。2025 年 11 月 12 日，英国政府向议会提交该法案，旨在升级 2018 年 NIS 条例：一是扩大监管范围，将相关托管服务提供商（RMSP）和数据中心运营商纳入监管，无论其是否在英国设立；二是强化事件报告义务，时限向欧盟 NIS2 看齐；三是赋予国务大臣发布战略优先事项声明、实践准则及直接指令的权力，监管机构获得更广泛的调查工具[33]。法案设置了与 GDPR 对齐的两档罚则：一般违法最高处 1000 万英镑或全球营业额 2%（取较高者），严重违法最高处 1700 万英镑或全球营业额 4%[34]。法案已于 2026 年 1 月 6 日完成下议院二读，预计 2026 年内生效。

4. 俄罗斯

关键信息基础设施责任体系重构。2026 年起，俄罗斯刑

[33] Mayer Brown. Cyber Security and Resilience (Network and Information Systems) Bill introduced to Parliament . [2025-09-18]. https://www.mayerbrown.com/en/insights/publications/2025/11/cyber-security-and-resilience-network-and-information-systems-bill--introduced-to-parliament

[34] Kennedys Law. The bill reshaping the cyber security landscape (UK). [2026-01-27]. https://www.kennedyslaw.com/en/thought-leadership/article/2026/the-bill-reshaping-the-cyber-security-landscape/

法第 274.1 条修正案生效，形成针对关键信息基础设施网络事件的两级责任体系：若事件导致关键信息基础设施数据被证实毁损、阻断、篡改或复制，维持刑事责任（6 至 8 年监禁并处 50 万至 100 万卢布罚金）；未造成上述后果的，转为行政责任，对法人处 10 万至 50 万卢布罚款，但对屡犯和系统性无视监管要求者仍可追究刑责[35]。执法层面，联邦技术与出口监督局（FSTEC）2026 年对 700 个 CII 对象的检查显示，64%的企业防护体系不达标。

5. 韩国、日本、新加坡

（1）韩国《信息通信网法》修正案获国会通过

2026 年 3 月，韩国国会通过《信息通信网利用促进和信息保护法》修正案：一是确立严格的 24 小时制度，信息通信服务提供商发现涉及用户个人信息的网络安全事件后须在 24 小时内通知用户并向科学技术信息通信部（MSIT）报告；二是扩大事件调查范围，从仅调查“原因”扩展至“发生经过与原因”，并在 MSIT 下设事件调查审议委员会，可组建官民联合调查组；三是引入执行罚与营业额罚则，对拒不改正、妨碍调查等行为可按日处以罚款，对因故意或重大过失在 5 年内重复发生事件者最高可处相关营业额 3%的行政罚款；四是要求指定经营者按标准指南编制并落实事件响应手册[36] [37]。

[35] Anti-Malware.ru. Ответственность за киберинциденты в России: кто будет платить в 2026 году Источник. [2026-02-12]. https://www.anti-malware.ru/analytics/Technology_Analysis/Responsibility-for-cyber-incidents-in-Russia-2026

[36] Lee & Ko. Amendments to the Network Act Passed by the National Assembly. [2026-03-27]. https://www.leeko.com/leenko/news/newsLetterView.do?lang=EN&newsletterNo=2263

（2）日本“主动网络防御”法治化进入分阶段实施期

《网络应对能力强化法》（通称“主动网络防御法”）于2025年5月16日经国会通过、5月23日公布，是日本网络安全立法一代人以来最重大的变革，2026年至2027年分阶段施行，核心义务预计自2026年10月起全面实施[38]。该法确立四大支柱：一是政企协作，电力、通信、金融等15个关键基础设施领域的经营者负有重要系统事前报备与网络安全事件强制报告义务；二是通信信息利用，政府可在独立监督委员会监督下分析通信元数据（不含内容）以发现攻击征兆；三是入侵反制，授权警察和自卫队经独立机构事先审查后对境内外攻击源服务器实施访问与无害化处置；四是组织准备[39]。配套层面，2025年7月日本将原内阁网络安全中心（NISC）改组为国家网络办公室（NCO）并引入安全许可制度，2025年12月又通过新的五年期《网络安全战略》，将网络攻击明确定位为“严重安全威胁”[40]。

（3）新加坡《网络安全法》修正案核心条款生效

2025年10月31日，该修正案的关键条款正式实施：一是将“计算机系统”定义扩展至虚拟计算机系统，并允许网络安全专员将位于境外、但为新加坡基本服务持续运行所必需的系统指定为受监管关键信息基础设施；二是新增第3A

[37] Hogan Lovells. South Korea considers updates to data and cyber laws . [2026-02-25]. https://www.hlc.com/en/publications/south-korea-considers-updates-to-data-and-cyber-laws
[38] Global Law Experts. Japan 2026: Active Cyber Defense Law (ACD), Practical Compliance Guide for Businesses. https://globallawexperts.com/active-cyber-defense-law-japan/
[39] nippon.com.New Legislation Signals Japan’s Shift to “Active” Cyber Defense. [2025-08-07]. https://www.nippon.com/en/in-depth/d01147/
[40] Cybers2. Cybersecurity in Japan: The Complete 2026 Guide to Threats, Laws & Protection. [2026-04-20]. https://www.cybers2.com/blog/cybersecurity-in-japan.html

部分，要求基本服务提供者为第三方所有的关键信息基础设施承担网络安全责任；三是扩大事件报告范围，覆盖供应链环节事件，配套的《网络安全（关键信息基础设施）（修正）条例 2025》要求关键信息基础设施所有者在知悉疑似高级持续性威胁导致的事件后 2 小时内报告；四是引入“临时网络安全关注系统”（STCC）制度，授权网络安全局（CSA）对因选举、疫苗分发等临时性事件而风险骤升的系统实施限期监管[41]。此外，修正案还创设了“特殊网络安全利益实体”（ESCI）和“基础数字基础设施”（FDI，含云服务商和数据中心）等新型监管对象，并引入最高达企业年营业额 10%的民事罚款制度[42]。

3.1.3 国际合作治理

《联合国打击网络犯罪公约》开放签署。2025 年 10 月 25 日，《联合国打击网络犯罪公约》在越南河内举行开放签署仪式，60 余个成员国现场签署。该公约历经 5 年谈判，于 2024 年 12 月由联合国大会通过，是首个将依赖网络实施的犯罪全球性定罪的国际条约，共 9 章，建立了首个用于收集、共享和使用严重犯罪电子证据的全球框架，将在第 40 个签署国批准后 90 天生效[43]。截至 2026 年 4 月，已有 75 个国家签署，卡塔尔、越南先后完成批准[44]。

---

[41] Rajah & Tann Asia. Cybersecurity Act Amendments on Provider-owned Critical Information Infrastructure and Systems of Temporary Cybersecurity Concern in Force from 31 October 2025. [2025-11-03]. https://www.rajahtannasia.com/viewpoints/cybersecurity-act-amendments-on-provider-owned-critical-information-infrastructure-and-systems-of-temporary-cybersecurity-concern-in-force-from-31-october-2025/

[42] Cyber Security Agency of Singapore. https://www.csa.gov.sg/legislation/cybersecurity-act/

[43] 环球时报/新浪新闻. 60 多国签署《联合国打击网络犯罪公约》，公约开放至 2025 年底供所有联合国会员国签署. [2025-10-27]. https://news.sina.com.cn/w/2025-10-27/doc-infvhrnv5411311.shtml

[44] 越南人民报网. 越共中央总书记、国家主席苏林签署批准《河内公约》的决定. [2026-04-08]. https://cn.

3.1.4 网络安全风险治理特点分析

1. 供应链与第三方风险成为共同焦点

欧盟的可信信息与通信技术（ICT）供应链框架、英国将托管服务商纳入监管、新加坡扩展第三方所有关键信息基础设施和供应商系统报告义务、日本覆盖 IT 供应商，均反映出供应链攻击倒逼监管对象从“关基运营者”向“关基依赖链”延伸。

2. 事件报告时限显著趋严趋细

新加坡对高级持续性威胁攻击事件要求 2 小时报告、韩国 24 小时通报、欧盟 CRA 的“24/72 小时+14 天”三级报告，与欧盟 NIS2 的报告节奏趋同，全球事件报告标准呈收敛态势。

3. 罚则“营业额化”“双罚化”

英国 17 万英镑级/4%营业额罚则、新加坡 10%营业额民事罚款、韩国 3%营业额罚款、中国对单位最高 1000 万元并对个人穿透追责，威慑力度整体跃升。

4. 进攻性防御与主权色彩增强

日本立法授权政府反制境外攻击源、俄罗斯强化 CII 刑责与检查、欧盟剑指“高风险供应商”，网络安全的国家安全属性和地缘政治色彩进一步凸显。

5. 打击网络犯罪走向全球法治协作

《联合国打击网络犯罪公约》填补了网络犯罪电子证据

---

nhandan.vn/article-post152073.html

跨境调取的全球规则空白，但主要西方国家与发展中国家在公约立场上的分歧，也预示其实际执行仍面临挑战。

3.2 数据安全风险治理现状

3.2.1 中国

1. 数据出境第三条路径完成制度闭环

2025 年 10 月 14 日，国家网信办、国家市场监督管理总局联合公布《个人信息出境认证办法》，自 2026 年 1 月 1 日起施行。该办法依据《个人信息保护法》第三十八条第一款第二项和《网络数据安全管理条例》，明确由依法取得个人信息保护认证资质的专业机构对个人信息出境活动开展合格评定，国家网信部门会同数据管理部门制定认证标准与技术规范，市场监管部门会同网信部门制定认证规则、统一证书及标志[45]。至此，“安全评估、标准合同、保护认证”三条个人信息出境路径全部完成部门规章级配套。

2. 大型网络平台个人信息保护专项启动

2025 年 11 月 22 日，国家网信办、公安部就《大型网络平台个人信息保护规定（征求意见稿）》公开征求意见（反馈截止 2025 年 12 月 22 日）：以注册用户 5000 万以上或月活跃用户 1000 万以上等标准认定大型网络平台；要求平台指定个人信息保护负责人并公开联系方式、设立个人信息保护工作机构，开展风险监测、合规审计、应急演练并向国家网信部门报告；首次细化个人信息可携权，要求 30 个工作日

[45] 国家互联网信息办公室. 个人信息出境认证办法. [2025-10-17]. https://www.cac.gov.cn/2025-10/17/c_1762449728720008.htm.

内以通用、机器可读格式完成转移，并支持 API 等标准化转移途径[46]。同期，国家网信办还就《大型网络平台设立个人信息保护监督委员会规定》公开征求意见[47]，构建“机构+负责人+监督委员会”的多元共治结构。

3. 聚焦数据全生命周期安全开展标准化工作

如表 3-2，2026 年度我国发布 8 项数据安全技术领域国家标准，力求围绕数据全生命周期构建管控体系，重点方向包括：一是数据生命周期安全技术规范。覆盖数据收集、存储、使用、加工、传输、提供、公开、销毁全流程；典型如电子产品信息清除强制性标准，明确介质数据彻底销毁技术方法，防范二手流通场景数据恢复泄露。二是数据基础治理技术工具规范。包含数据分类分级、风险评估、数据脱敏、数字水印、溯源审计、访问权限管控等技术框架；支撑重要数据识别、数据资产梳理。三是数据流转与共享安全。规范跨主体、跨机构数据提供、数据交易、数据出境、第三方共享场景下的安全约束，明确流转过程可追溯、可管控。四是新技术场景下数据处理安全。面向 AI Agent、大模型、生成式 AI 的数据处理活动，规范训练语料、用户输入、模型记忆、衍生数据安全要求，打通 AI 场景数据风险治理。五是合规落地技术支撑。为监管核查、数据安全审计、合规自查

[46] 国家互联网信息办公室. 国家互联网信息办公室、公安部关于《大型网络平台个人信息保护规定（征求意见稿）》公开征求意见的通知. [2025-11-22]. https://www.cac.gov.cn/2025-1/22/c_1765543463511624.htm.
[47] 国家互联网信息办公室. 国家互联网信息办公室关于《大型网络平台设立个人信息保护监督委员会规定（征求意见稿）》公开征求意见的通知. [2025-09-12]. https://www.cac.gov.cn/2025-09/12/c_1759395487456327.htm

提供统一技术指标，打通法律法规条文到技术落地的转化通道。总体核心逻辑是无论数据在哪个系统、网络中流转，管控数据自身的访问、复制、传播、篡改、泄露风险，保障数据“流转可控、使用合规、泄露可追溯”。

表 3-2 2026 年度数据安全技术相关国家标准

（来源：中国国家标准化管理委员会）

序号	标准号	标准名称	发布日期	实施日期
1	GB/T 46068-2025	数据安全技术 个人信息跨境处理活动安全认证要求	2025-08-29	2026-03-01
2	GB/T 46071-2025	数据安全技术 数据安全和个人信息保护社会责任指南	2025-08-29	2026-03-01
3	GB 46864-2025	数据安全技术 电子产品信息清除技术要求	2025-12-02	2027-01-01
4	GB/T 46796-2025	数据安全技术 数据接口安全风险监测方法	2025-12-02	2026-07-01
5	GB/T 37932-2025	数据安全技术 数据交易服务安全要求	2025-12-02	2026-07-01
6	GB/T 46901-2025	数据安全技术 基于个人请求的个人信息转移要求	2025-12-31	2026-07-01
7	GB/T 46903-2025	数据安全技术 个人信息保护合规审计要求	2025-12-31	2026-07-01
8	GB/T 47469-2026	数据安全技术 移动智能终端的移动互联网应用程序（App）个人信息处理活动管理指南	2026-04-30	2026-11-01

3.2.2 国际主要经济体

1. 欧盟

（1）“数字综合简化方案”重构数据保护规则

2025 年 11 月 19 日，欧盟委员会发布 Digital Omnibus 提案，对 GDPR 和电子隐私（ePrivacy）规则进行系统性修订：一是网络安全方面，引入了集中渠道，即所谓的单一入口点（SEP），用于通知有关主管部门事件，并将 GDPR 数据泄

露通知的门槛提高至监管机构，以与个人通知门槛保持一致。二是数据保护方面，引入处理敏感数据的新法律依据（例如人工智能领域），统一数据保护影响评估（DPIA）方法，同时引入多项例外场景以放宽研发阶段合规要求。三是电子隐私方面，将内容纳入 GDPR 关于 cookie 同意的具体规则，并扩大了无需同意的使用场景。这些规定仅适用于个人数据存储在自然人终端上或从终端访问时。为了减少“cookie 同意疲劳”，简化或消除 cookie 同意提示栏的需求，法案还引入了网站和应用通过自动化、机器可读机制允许数据主体同意的新义务。浏览器制造商还必须允许用户同意或拒绝。四是数据使用与治理方面，整合和简化多项法案中的数据共享及使用相关义务，引入额外的商业秘密披露保护措施，并允许暂时免除部分公司和服务的云切换义务[48]。

（2）欧盟《数据法案》进入适用期与 GDPR 高压执法并行

欧盟《数据法案》主要义务自 2025 年 9 月起适用; GDPR 执法保持高压，2025 年数据泄露日均通报量同比上升 22%，监管机构对可预防的泄露事件持续开出罚单，跨境传输面临持续审查[49]。

2. 英国

[48] McDermott Will & Emery. EU proposes sweeping reforms to the GDPR, cookie rules, Data Act, and breach reporting. [2025-11-24]. https://www.mcdermottlaw.com/insights/eu-proposes-sweeping-reforms-to-the-gdpr-cookie-rules-data-act-and-breach-reporting/

[49] PrivacyChecker. Biggest GDPR Fines 2025–2026: Lessons From €1.2 Billion in Penalties. [2026-01-28]. https://privacychecker.pro/blog/biggest-gdpr-fines-2025-2026

《数据(使用与获取)法》(DUAA)分阶段生效。DUAA于2025年6月获御准，其主要数据保护条款于2026年2月5日生效，对英国GDPR和《隐私与电子通信条例》(PECR)作出修订：改革数据主体权利、合法处理基础和国际传输规则，并将PECR罚则提升至与英国GDPR持平的1750万英镑或全球营业额4%，同时赋予信息专员办公室(ICO)强制证人接受询问、要求提交技术报告等新执法权力[50]。强制性的数据保护投诉处理程序义务于2026年6月19日生效，要求控制者建立有记录的投诉处理流程，30天内确认收悉并及时答复[51]。

3. 俄罗斯

个人信息保护进入“营业额罚款+本地化加码”阶段。2025年5月30日，2024年11月通过的《个人数据法》修正罚则生效：对导致个人数据泄露的法人，按受影响主体数量处300万至1500万卢布罚款；泄露特殊类别数据最高1500万卢布、生物识别数据最高2000万卢布；重复泄露按年营业额1%—3%计罚，下限2000万卢布、上限5亿卢布[52] [53]。本地化方面，2025年2月28日签署、7月1日生效的修正案明确禁止利用位于境外的数据库对俄公民个人数据进行记

[50] Data (Use and Access) Act 2025. https://www.legislation.gov.uk/ukpga/2025/18/contents
[51] Arhivix. UK Data (Use and Access) Act 2025: The New Recordkeeping Duties That Start Landing on Businesses in 2026. [2026-08-03]. https://arhivix.com/blog/uk-data-use-and-access-act-2025-recordkeeping-duties
[52] KKMP Legal — Russia Legal Trends. https://kkmp.legal/upload/iblock/def/x93vd350qj0jnpo19qf33owaqia57ouc.pdf
[53] Klerk.ru. Штрафы Роскомнадзора за персональные. [2026-07-29]. https://www.klerk.ru/blogs/fedresurs/702077/

录、系统化、存储等处理[54]；2025 年 7 月通过、2026 年 3 月 1 日生效的《贸易活动国家调控基础法》修正案进一步限制外资持股超 20%的主体在俄开展市场调研，要求调研数据须在俄境内数据库处理和存储[55]。执法手段上，俄罗斯联邦通信、信息技术和大众传媒监督局（Roskomnadzor）已部署 AI 扫描器全天候监测.ru 等域名网站的合规情况。

4. 韩国、日本、新加坡

（1）韩国 PIPA 制度体系双线升级

一是境内代表制度实质化。经 2025 年 3 月修订，2025 年 10 月 2 日生效的《个人信息保护法》（PIPA），要求处理韩国个人信息的境外企业在韩设立境内法人或指定境内代表，已设子公司的必须直接指定该子公司为代表，母公司须对代表履职实施监督。二是新一轮修正酝酿。2026 年初提出的修正案拟将“数据泄露”定义从丢失、被盗、泄露扩展至伪造、篡改、损毁，将通知义务从“结果导向”前移至“可能发生泄露”的风险导向；对重复或大规模违法引入最高达总营业额 10%的行政罚款，同时设置以数据保护投入为条件的罚款减免机制；并强化首席隐私官（CPO）制度，其指定、变更、解除须经董事会决议并向 PIPC 申报，未履行者最高处 3000 万韩元罚款。

（2）日本《个人信息保护法》（APPI）修正案启动新一

[54] Deloitte LT in Focus. Analyzing important legislative amendments. https://storage.delret.ru/lt-in-focus/EN/lt-in-focus-personal-data-localization-new-requirements.pdf

[55] AmCham Russia — Russia Legal Trends. https://www.amcham.ru/uploads/RUSSIA%20LEGAL%20TRENDS%2020251768989107855.pdf

轮改革

2026 年 4 月 7 日，日本内阁通过 APPI 修正案[56]并提交国会：一是引入行政罚款（课征金）制度，按相关营业额比例计罚，显著强化个人信息保护委员会（PPC）的执法抓手；二是针对 AI 训练数据增设规则，允许企业在不取得个人授权的情况下将病史、犯罪记录等敏感信息用于 AI 研发、统计分析等非身份识别场景；三是强化儿童数据保护。修正案预计在公布后约两年内全面施行。而“以数据松绑换取 AI 竞争力”的导向在日本国内引发在野党对隐私泄露风险的持续批评。

（3）新加坡以既有 PDPA 框架的整合适用为主

新加坡目前的数据安全治理重心在于：将 2025 年 12 月 5 日生效的修订内容并入《2012 年个人数据保护法》（PDPA）合并文本，继续以“同意、目的限制、跨境传输限制”等九项义务为执法依据，严重违法最高处 100 万新元或本地年营业额 10%罚款；同时通过个人数据保护委员会（PDPC）发布的 AI 场景个人数据使用咨询指南，将 PDPA 义务延伸至 AI 训练与部署环节[57]。数据出境仍适用 PDPA 第 26 条”相当保护水平”要求。

3.2.3 数据安全风险治理特点分析

1. 数据出境治理呈现“促流动”与“设闸门”的双向分

[56] 个人信息保护委员会. 内阁对部分修订《个人信息保护法》的法案作出决定. [2026-04-07]. https://www.ppc.go.jp/news/press/2026/260407/

[57] Vorp Labs — Singapore Regulatory Updates. https://vorplabs.com/ai-regulatory-updates/singapore

化

中国补齐认证路径、欧盟简化 cookie 规则，均意在降低企业合规成本、促进数据高效流动；美国、俄罗斯本地化加码则以国家安全为由对特定方向的数据流出设立硬闸门，数据跨境规则的地缘政治分野显著加深。

2. 个人信息保护的责任主体向“平台”和“境内实体”集中

中国大型网络平台专门规定、韩国境外企业境内代表制度，均通过锁定具本地管辖抓手的实体来解决“监管离岸化”难题。

3. 罚款威慑向“营业额比例制”看齐

俄罗斯 5 亿卢布/3%营业额、韩国拟议 10%营业额、英国 PECR 对齐 4%营业额，GDPR 式比例罚款已成为全球范式。

4. 数据保护与 AI 治理加速耦合

日本 APPI 修正为 AI 训练数据设立专门规则、韩国《AI 基本法》与 PIPA 衔接适用、新加坡以 PDPA 指南覆盖 AI 场景，个人信息保护制度正在成为各国 AI 治理的“底层接口”。

3.3 人工智能安全风险治理现状

3.3.1 中国

1. AI 生成合成内容标识制度全面落地并启动执法

国家网信办等四部门 2025 年 3 月 14 日发布《人工智能生成合成内容标识办法》，自 2025 年 9 月 1 日起施行，明确

显式标识与隐式标识双重要求；配套强制性国家标准 GB 45438-2025《网络安全技术 人工智能生成合成内容标识方法》于 2025 年 2 月 28 日发布、9 月 1 日同步实施。2025 年 11 月 25 日，网信部门通报集中查处一批未有效落实标识义务的移动互联网应用程序，采取约谈、责令限期改正、下架下线等措施，标志着标识制度从“立规”转入“执法”。

2. 安全治理框架迭代升级

2025 年 9 月，中国发布《人工智能安全治理框架》2.0 版，对 AI 安全风险分类、技术应对与综合治理措施进行更新。同时，2026 年 1 月 1 日起施行的新修正《网络安全法》新增 AI 条款，将“完善人工智能伦理规范，加强风险监测评估和安全监管”写入基础性法律，为后续 AI 专门立法和监管执法提供了上位法接口。

3. 促进全球人工智能向善普惠发展

2026 年 7 月 17 日至 20 日，2026 世界人工智能大会暨人工智能全球治理高级别会议在中国上海举行。习近平主席出席大会开幕式并发表主旨讲话，来自 100 多个国家和国际组织的官方代表、产学研界嘉宾等出席大会。大会以“智能伙伴　共创未来”为主题，共同探讨如何促进全球人工智能向善普惠发展。大会期间，29 国代表签署成立世界人工智能合作组织（WAICO）的协定，成为创始成员国。组织总部设在中国上海，发布《2026 世界人工智能大会暨人工智能全球治理高级别会议主席声明》。同时，大会还发布《中国智·惠

世界（2026）》案例集及《人工智能合作发展行动计划》《智能体互信互联互操作全球合作倡议》《国际人工智能伦理治理行动计划》等一批重要成果，充分彰显中国推动人工智能向上向善发展的责任担当，为全球人工智能发展和治理注入新动能。

3.3.2 国际主要经济体

1. 美国

（1）《美国 AI 行动计划》确立“促发展”总基调

2025 年 7 月 23 日，特朗普政府发布《赢得 AI 竞赛：美国 AI 行动计划》[58]，围绕“加速 AI 创新、建设 AI 基础设施、全球 AI 领导”三大支柱提出约 90 项政策行动，并签署三项配套行政令，核心取向是废除拜登时期的 AI 监管体系、鼓励开源开放权重模型、简化算力与能源审批、推动全栈式 AI 出口并协调针对对手国家的出口管制。

（2）联邦与州监管权之争白热化

2025 年 12 月 11 日，特朗普签署第 14365 号行政令《确保人工智能国家政策框架》[59]，明确要建立“最低负担的国家 AI 政策框架”以取代 50 个互不兼容的州级框架：授权司法部设立 AI 诉讼特别工作组（2026 年 1 月 10 日起运作）在联邦法院挑战州法，要求商务部长在 2026 年 3 月 11 日前评估并点名“繁重”州法，并将州获取联邦宽带资金的资格与配

[58] White House. White House Unveils America’s AI Action Plan. [2025-07-23]. https://www.whitehouse.gov/releases/2025/07/white-house-unveils-americas-ai-action-plan/

[59] White House. ENSURING A NATIONAL POLICY FRAMEWORK FOR ARTIFICIAL INTELLIGENCE. [2025-12-11]. https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/

合度挂钩。2026 年 7 月 1 日，美国联邦贸易委员会（FTC）据此以 5:0 投票发布政策声明，明确即使企业为遵守州法（点名科罗拉多州 AI 反歧视法）而调整模型输出，亦不能豁免《FTC 法》第五条关于“欺诈行为”的责任。但国会层面，参议院曾以 99:1 的压倒性票数从“大而美法案”中剥离关于冻结州 AI 立法的条款，联邦优先权主张遭遇立法抵制。

（3）州级立法继续推进

在联邦综合立法缺位的情况下，州法构成实质约束层：得克萨斯州《负责任 AI 治理法》（TRAIGA）2025 年 6 月 22 日签署，2026 年 1 月 1 日起生效，禁止以特定非法目的开发部署 AI 并要求政府实体披露 AI 交互；加利福尼亚州《前沿 AI 透明度法》（SB 53）2025 年 9 月 29 日签署、2026 年 1 月 1 日起生效，要求年收入超 5 亿美元的前沿模型开发者公布前沿 AI 框架、提交透明度报告并向州监管机构报告重大安全事件[60]。

2. 欧盟

欧盟《人工智能法案》分阶段适用与“数字综合简化方案”延期并行。2025 年 8 月 2 日起，通用目的人工智能（GPAI）模型义务与治理框架正式适用，欧盟 AI 办公室于 2025 年 7 月 10 日发布最终版《GPAI 行为准则》（涵盖透明度、版权、安全与安保三章），签署该准则可获得合规推定。针对产业界关于协调标准滞后的压力，欧盟委员会 2025 年 11 月 19 日

[60] Agent Liability — Global AI Regulation Tracker. https://agentliability.co/articles/global-ai-regulation-status-tracker-2026

发布 AI 数字综合简化方案（Digital Omnibus on AI），欧洲议会 2026 年 6 月 16 日，理事会 6 月 29 日先后表决通过：附件 III 独立高风险 AI 系统义务自 2026 年 8 月 2 日推迟至 2027 年 12 月 2 日适用，附件 I 嵌入受管制产品的高风险 AI 义务推迟至 2028 年 8 月 2 日；同时新增对生成非自愿私密影像和儿童性虐待材料的 AI 系统的禁令，自 2026 年 12 月 2 日起适用并归入最高罚档（3500 万欧元或全球营业额 7%）[61] [62]。透明度义务未受延期影响：第 50 条自 2026 年 8 月 2 日起适用，欧委会 2026 年 6 月 10 日发布《AI 生成内容标识行为准则》，7 月 20 日通过第 50 条透明度指南，同日欧盟正式激活对 GPAI 模型提供者的执法权力（调查、罚款、召回）。此外，2026 年 7 月 7 日欧委会发布《网络安全与人工智能行动计划》，推动 AI 治理与网络安全政策协同[58]。

3. 英国

英国延续“促创新、轻立法”路线，强化安全评估能力。英国本年度未推进综合性 AI 立法。2025 年发布的《AI 机会行动计划》（50 项建议，含 AI 增长区、国家数据库等）继续作为政策主轴；原“AI 安全研究所”更名为“AI 安全研究院”（AI Security Institute），职能向国家安全与犯罪滥用风险聚焦，重点评估先进 AI 赋能的网络攻击和生物威胁；政府发布《英国政府 AI 行动手册》统一公共部门安全负责任

[61] Bright Defense.EU AI Act Reaches Full Application Date On August 2. [2026-07-26]. https://www.brightdefense.com/news/eu-ai-act-delay-keeps-2026-compliance-pressure/
[62] AI Regulations — European Union. https://artificialintelligenceregulations.com/jurisdictions/european-union.html

用 AI 的准则，并由科学创新技术部（DSIT）发布《可信第三方 AI 保障路线图》培育 AI 审计评估市场[63]。

4. 俄罗斯

俄罗斯 AI 治理机构化并出台首部 AI 专门法。2026 年 1 月，俄罗斯总统普京责成制定联邦和地方两级国家 AI 实施计划，重点提升国产基础生成式模型在政务和关键信息基础设施中的应用比例，并由俄罗斯国家原子能集团公司（Rosatom）、俄罗斯电网股份公司（Rosseti）等制定至 2030 年（远期至 2036 年）的数据中心扩建计划[64]。2026 年 2 月 26 日，普京签署总统令成立“总统直属人工智能技术发展委员会”；3 月，政府设立 AI 技术发展分委会。2026 年 7 月 26 日，普京签署俄罗斯首部 AI 专门法《关于支持俄罗斯联邦人工智能技术发展的法律》，允许并规范神经网络生成内容的标识，日访问量超 50 万的网站须为用户提供添加 AI 标识的技术能力；设立“主权 AI 模型”与“国家 AI 模型”两类认证（前者要求全生命周期由俄企在俄完成、禁用境外组件），获认证模型可接入联邦和地区政务信息系统数据；用于关键信息基础设施的“可信模型”须经俄罗斯联邦技术与出口管制署（FSTEC）和俄罗斯联邦安全局（FSB）安全认证。

5. 韩国、日本、新加坡

（1）韩国

[63] Regulations.ai — United Kingdom. https://regulations.ai/regulations/united-kingdom-summary
[64] TAdviser. National Strategy for Artificial Intelligence Development in Russia. [2026-07-27]. https://tadviser.com/index.php/Article:National_Strategy_for_the_Development_of_Artificial_Intelligence

韩国《人工智能基本法》全面施行，配套体系逐步完善。《关于人工智能发展和构建信赖基础的基本法》于2026年1月22日生效，是继欧盟之后全球首部综合性国家AI立法。出台“高影响AI”制度，对应用于能源、医疗、刑事调查、招聘、交通等场景的AI经营者施加事前审查、基本权利影响评估和特殊安全保障义务；对生成式AI要求其履行告知用户和输出标识（水印）义务[65]。但相关实施办法有所滞后——科学技术信息通信部（MSIT）迟至2025年9月8日才发布含34条规定的施行办法草案。治理架构上，国家AI战略委员会于2025年9月启动，2026年2月第二次全会敲定含99项行动任务、326条政策建议的国家AI行动计划。

（2）日本

日本以“人工智能基本计划”推动立法施策。在2025年6月通过的《AI推进法》框架下，日本政府于2025年12月23日发布《人工智能基本计划》，提出打造全球“最适宜AI技术研发和应用国家”的目标，确立“创新与风险兼顾、灵活应对、内外统筹”三原则，部署全域AI应用（政府率先推广生成式AI，在医疗、制造、防灾等领域落地）、强化研发能力（研发可信赖、节能的国产通用基础大模型）等政策[66]。风险防控方面主要依靠PDCA循环式的软法治理与《个人信息保护法》（APPI）等既有法律的适用，未设置欧盟式的事

[65] Stimson Center. South Korea’s AI Basic Act: Seeking Balance Between Industry Innovation and Social Risk. [2026-07-07]. https://www.stimson.org/2026/south-koreas-ai-basic-act-seeking-balance-between-industry-innovation-and-social-risk/

[66] 中国科学院科技战略咨询研究院. 日本发布《人工智能基本计划》. [2026-04-10]. https://casisd.cas.cn/zkcg/ydkb/kjzcyzxkb/2026/zczxkb202603/202604/t20260410_8183670.html

前审批或强制性义务。

（3）新加坡

新加坡以自愿性框架引领“智能体 AI”治理前沿。2026 年 1 月 22 日，新加坡资讯通信媒体发展局（IMDA）在达沃斯世界经济论坛发布全球首个《智能体 AI 示范治理框架》，围绕“事前评估并限定风险、让人类实质性担责、实施技术控制与流程、赋能终端用户责任”四个维度，提出限定智能体权限、最小特权访问、全程日志可追溯等实践要求；2026 年 5 月 20 日更新至 1.5 版，补充多智能体系统、第三方智能体、自动化偏见等案例与最佳实践[67] [68]。该框架虽无法律约束力，但与 AI Verify 测试工具包、ISO/IEC 42001、NIST AI RMF 建立映射，发挥“监管枢纽”作用。配套层面，2026 年 2 月新加坡宣布成立国家 AI 理事会，统筹先进制造、互联、金融、医疗四大领域的“AI 使命”；金融领域，新加坡金管局继 2025 年发布《AI 风险管理指南》后，于 2026 年 7 月发布全球首个金融智能体运行时治理参考框架 SAFR[69]。

3.3.3 产业界

1. 网络攻防能力评估

当前针对大模型网络攻防能力的评估大致可分为四类，

[67] Infocomm Media Development Authority (IMDA). MODEL AI GOVERNANCE FRAMEWORK FOR AGENTIC AI. [2026-05-20]. https://www.imda.gov.sg/-/media/imda/files/about/emerging-tech-and-research/artificial-intelligence/mgf-for-agentic-ai.pdf

[68] Baker McKenzie. Singapore: IMDA updates Model AI Governance Framework for Agentic AI. [2026-06-23]. https://www.bakermckenzie.com/en/insight/publications/2026/06/singapore-imda-updates-model-ai-governance-framework-for-agentic-ai

[69] Mayer Brown. Singapore's Agentic AI Framework: Practical Guidance for Market Entry. [2026-04-01]. https://www.mayerbrown.com/en/insights/publications/2026/04/singapores-agentic-ai-framework-practical-guidance-for-market-entry

如表 3-3 所示。**一是漏洞利用基准（exploitation benchmark）。**以 ExploitGym 为代表：该基准由 UC Berkeley、马克斯·普朗克安全与隐私研究所、UC Santa Barbara、亚利桑那州立大学与 Anthropic、OpenAI、Google 联合构建，包含 898 个源自真实世界漏洞的实例——其中 520 个用户态程序实例来自 OSS-Fuzz 覆盖的 161 个项目，185 个来自 Chrome 浏览器的 V8 JavaScript 引擎，193 个来自 Linux 内核；每个实例提供漏洞代码库、触发漏洞的概念验证输入与文本描述，要求智能体将其逐步扩展为实现未授权代码执行的完整漏洞利用，并通过检索动态生成的特权 flag 来验证成功，还引入“智能体作为裁判”机制以约 94%的一致率甄别智能体是否真正利用了目标漏洞而非取巧捷径[70] [71]。**二是夺旗赛（CTF）与小规模 CVE 复现类基准。**如 NYU CTF、Cybench、CVE-Bench、BountyBench 等，规模多在 200 实例以内，侧重合成漏洞或用户态软件。**三是网络靶场（cyber range）评估。**以英国人工智能安全研究院（UK AISI）的 32 步“The Last Ones”企业网络攻击模拟为代表，横跨多个子网与主机，测量模型完成从初始侦察到完全接管网络（M9）的平均步数[72]。**四是时间范围（time horizon）测量。**由模型评估和威胁研究组织（METR）与英国人工智能安全研究院（UK AISI）推动，测量智能体在特定可靠性下可自主完成的任务时长：AISI 估计

[70] Zhun Wang. ExploitGym: Can AI Agents Turn Security Vulnerabilities into Real Attacks?. [2026-05-11]. https://arxiv.org/abs/2605.11086
[71] Eugene Yan. Patterns for Building Cybersecurity Evals. https://eugeneyan.com/writing/cybersecurity-evals/
[72] OpenAI.OpenAI 与 Hugging Face 携手应对模型评估期间发生的安全事件. [2026-07-21]. https://openai.com/zh-Hans-CN/index/hugging-face-model-evaluation-security-incident/

前沿模型 80%可靠性的网络任务时间长度自 2024 年末推理模型出现以来每 4.7 个月翻一番，而 2026 年 2 月发布的《国际 AI 安全报告》引用的更长期趋势约为每 7 个月翻一番[73][74]。

表 3-3 网络攻防能力评估体系

评估体系	发布方	测量对象	关键设计	代表性结果
ExploitGym	UC Berkeley/ Google/Anthropic/OpenAI 等	漏洞利用（PoV→可用 exploit）	898 个真实漏洞实例；开关标准防护（ASLR、V8 沙箱等）对照；flag 验证 + 智能体裁判[18][19]	Claude Mythos Preview 157 例、GPT-5.5 120 例；开启防护后 Mythos 降至 45 例
Cybench / CVE-Bench / BountyBench	学术界	CTF 与 CVE 复现	规模 ≤200 实例，合成或用户态为主[18]	Cybench 审计发现 3.4% 成功轨迹存在作弊
“The Last Ones” 网络靶场	英国 AISI	多子网企业网络渗透	32 步攻击链、无主动防御者、最高 100M token 预算	GPT-5.6 Sol 最佳尝试完成全部 32 步（M9 完全接管）
时间范围测量	英国 AISI / METR	长时程自主行动能力	80% 可靠性任务时长测量，2.5M token 上限	能力每 4.7 个月翻一番；Mythos 与 GPT-5.5 已超出测试集可测上限

值得关注的是，这类评估有一个共同的关键设计选择：为了测量“最大能力边界”，评估通常在各实验室的可信访问计划下有意关闭部署态安全过滤器。ExploitGym 论文明确说明其主要实验在 OpenAI 的 Trusted Access for Cyber 与 Anthropic 的 Cyber Verification Program 框架下进行、安全防护被禁用，以测量前沿模型与智能体的能力边界；论文脚注还显示，当 OpenAI 默认安全过滤器开启时，GPT-5.5 在默认

[73] gov.uk. https://www.aisi.gov.uk/blog/how-fast-is-autonomous-ai-cyber-capability-advancing
[74] International AI Safety Report. https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026

提示下的全部漏洞利用尝试都会被拦截。评估同时设置了资源约束以保证可比性：每任务一次尝试、两小时时限；扩展至六小时窗口时，Claude Mythos 的利用成功数从 157 升至 204，而 Opus 4.6 在头 30 分钟即已饱和——这表明给予更多算力与时间，能力天花板可显著抬升。

2. 前沿实验室的安全治理框架

在评估之上，主要前沿实验室各自建立了“能力阈值—防护承诺”式的安全治理框架。OpenAI 的 Preparedness Framework（第二版）将风险分为 High 与 Critical 两档，跟踪类别包括网络安全、生物与化学、AI 自我改进，研究类别包括长程自主性、沙袋行为（sandbagging）、自主复制与破坏防护机制等；规则是：High 能力模型须在部署前具备“充分防护”，Critical 能力将触发开发暂停[75] [76]。Anthropic 的 Responsible Scaling Policy（RSP，当前 v3.x）仿照生物安全等级设立 ASL-1 至 ASL-4：ASL-3 要求部署侧实时分类器护栏、经审核用户的豁免流程、漏洞赏金、威胁情报与快速越狱响应，安全侧要求权重对非国家行为者“极不可能被盗”；ASL-4 则将权重防护提升至抵御国家级对手，并要求就模型自主破坏安全措施作出“肯定性安全论证”[77] [78] [79]。Anthropic

[75] Sam Coggins. The 2025 OpenAI Preparedness Framework does not guarantee any AI risk mitigation practices: a proof-of-concept for affordance analyses of AI safety policies. [2025-10-13]. https://arxiv.org/pdf/2509.24394

[76] futureagi.com. Frontier Model Safety Analysis (2026): How Anthropic, OpenAI, and DeepMind Frame the Risk, and Where Enterprises Actually Live. [2026-03-16]. https://futureagi.com/blog/frontier-model-safety-analysis-2026/

[77] techjacksolutions.com.AI Safety Levels (ASL) Explained: Anthropic's Responsible Scaling Policy in 2026. [2026-06-13]. https://techjacksolutions.com/ai-tools/anthropic-claude/ai-safety-levels-explained/

[78] Anthropic. Anthropic’s Responsible Scaling Policy. [2026-07-08]. https://www.anthropic.com/responsible-scaling-policy

[79] EA Forum. We read every labs safety plan so you don't have to: 2025 edition. [2025-10-29].

此前已因 Claude Mythos 在测试中表现出色的漏洞发现能力而限制其发布。有报道显示 Mythos 在实验中曾应研究者要求成功逃逸隔离沙箱并自主开发出多步骤漏洞利用以获取更广泛互联网访问[80]。Google DeepMind 的 Frontier Safety Framework（FSF v3.x）则以关键能力等级（CCL）与跟踪能力等级组织评估，覆盖 CBRN 提升、网络提升、有害操纵与 ML 研发加速等领域[75]。

然而，这些框架的共同局限在前文中提及的相关失控事件中暴露无遗。其一，框架治理的是模型发布，失控事件发生在内部评估环节。OpenAI 自己也承认“这些部署防护在评估中被有意关闭，因为评估旨在测试网络漏洞能力”。其二，框架是自愿承诺而非法律义务。2024 年 5 月首尔 AI 峰会上 16 家企业签署的《前沿 AI 安全承诺》要求发布安全框架、开展内外部红队测试，但其“自愿且不可强制执行，企业违反而无严重后果”的性质一直受到批评[81] [82]。其三，三家框架均承认安全训练是概率性防御而非硬墙，多轮漂移、间接提示注入、对抗后缀等攻击持续在已发布模型上得手[75]。此外，第三方评估机构，如：美国商务部下属国家标准与技术研究院（NIST）人工智能标准与创新中心（CAISI）、英国人工智能安全研究院（UK AISI）会在模型部署前进行独立测

https://forum.effectivealtruism.org/posts/fHWtYTyahQoSsfzke/we-read-every-labs-safety-plan-so-you-don-t-have-to-2025

[80] mindstudio.ai.What Is Claude Mythos? Anthropic's Most Dangerous AI Model Explained. [2026-04-09]. https://www.mindstudio.ai/blog/what-is-claude-mythos-anthropic-most-dangerous-ai-model

[81] EA Forum.New voluntary commitments (AI Seoul Summit). [2024-05-21]. https://forum.effectivealtruism.org/posts/dmpPQ48M5ZQ8ifh8v/new-voluntary-commitments-ai-seoul-summit

[82] safe.ai. AI Safety Newsletter #36: Voluntary Commitments are Insufficient. [2024-05-30]. https://newsletter.safe.ai/p/ai-safety-newsletter-35-voluntary

试，例如 Opus 4.5 部署前接受了 CAISI 与 UK AISI 针对灾难性风险的独立评估[83]，但这类外部评估的覆盖度与强制力仍有限。

3. 技术防护体系

在模型治理框架之下，产业界围绕“运行不可信的模型生成代码”发展出一套分层技术防护体系，其成熟度直接决定了发生“失控”事故的概率。模型层防护包括对齐训练（拒绝高风险网络请求）、部署态内容分类器（实时拦截危险输入输出）、以及最新的轨迹级监控。OpenAI 在 2026 年 7 月 20 日的长时程模型安全博客中承认“单个工具调用本身看似合理，但仍可能是通向不良结果的一步”，因此提出将安全单元从单动作转向完整运行轨迹，构建四层防线：源自真实事件的对抗性评估、面向长时程展开的对齐训练、可暂停会话并告警人类的主动轨迹级监控、以及更多用户可见性与控制[84] [85]。运行环境层防护，即：沙箱隔离，2026 年业界已收敛于明确的强度排序：标准 Docker 容器因共享宿主内核而“对不可信代码明确不足”，gVisor 通过用户态内核拦截系统调用提供中间强度，Firecracker/Kata 等 microVM 以硬件虚拟化边界提供最强隔离，成为 E2B、AWS Lambda、Modal 等生产级智能体平台的基线选择[86] [87] [88]。网络层防护是与计算

[83] AI Agent Index. Claude. https://aiagentindex.mit.edu/2025/claude/
[84] The AI Dude. OpenAI's Long-Horizon Safety Lessons, Unpacked. [2026-07-21]. https://theaidude.net/blog/openais-long-horizon-model-safety-lessons
[85] digitalapplied.com. OpenAI Paused Its Own Model: The First Containment Incident. [2026-07-21]. https://www.digitalapplied.com/blog/openai-containment-incident-long-horizon-model-paused-2026
[86] Zylos. AI Agent Sandbox & Code Execution Isolation. [2026-02-21]. https://zylos.ai/research/2026-02-21-ai-agent-sandbox-execution-isolation/
[87] augmentcode.com. What Is an Agent Execution Sandbox?. [2026-05-03]. https://www.augmentcode.com/g

隔离同等重要却常被忽视的另一半。隔离只能阻止代码逃逸，无法阻止被允许联网的代码外传数据，因此业界也提出应在虚拟机外部以 eBPF 等手段执行默认拒绝规则的出口治理与域名白名单[89]。

4. 标准化

围绕智能体工程的标准化也在推进。美国 NIST 的人工智能标准与创新中心（CAISI）于 2026 年 2 月正式启动 AI 智能体标准倡议，围绕行动权限最小工程要求、工具调用安全与智能体权限分离制定指南，预计 2026 年末至 2027 年初形成正式文件，并将纳入 SP 800-160 与 SP 800-218 更新[90][91]。MITRE ATLAS v5.4（2026 年 2 月）则新增了针对智能体生态的攻击技术条目，包括“投毒 AI 智能体工具”与“逃逸至宿主”等，为红队提供了对抗词汇表[92]。

3.3.4 人工智能安全治理特点分析

1. 全球 AI 治理阵营分化明显

欧盟、韩国以具有强制罚则的综合立法立规，美国（联邦层面）、英国、日本、新加坡则以行动计划、自愿框架和评估机构等软法工具为主，形成“布鲁塞尔路径”与“促创新路径”两大阵营。

---

uides/agent-execution-sandbox

[88] CreateOS. MicroVM Isolation for AI Agents: Why Containers Fall Short. [2026-07-07]. https://createos.sh/blogs/microvm-isolation-for-ai-agents

[89] shayon.dev. Let's discuss sandbox isolation. [2026-02-21]. https://www.shayon.dev/post/2026/52/lets-discuss-sandbox-isolation/

[90] NIST. AI Agent Standards Initiative. [2026-02-17]. https://www.nist.gov/artificial-intelligence/ai-agent-standards-initiative

[91] NIST. NIST Artificial Intelligence Update. [2026-03-27]. https://www.nist.gov/system/files/documents/2026/03/27/03_AI%20Update-March.pdf

[92] Zylos. Defensive Prompt Engineering for Multi-Tool AI Agents: Securing the Instruction-Tool-Output Pipeline. [2026-05-11]. https://zylos.ai/research/2026-05-11-defensive-prompt-engineering-multi-tool-ai-agents/

2. 透明度与内容标识成为最大公约数

中国标识办法+强制国标、欧盟第 50 条及标识行为准则、韩国生成式 AI 水印义务、俄罗斯 AI 法标识条款，均在本期落地或立法，AI 生成内容的“可追溯标识”已成为跨越阵营的全球性基础制度。

3. 治理对象向“通用模型”与“智能体”前移

欧盟 GPAI 义务、加州 SB 53 针对前沿模型、新加坡全球首发智能体 AI 框架、NIST 启动 AI 智能体安全标准倡议，监管焦点从具体应用层上移至模型能力与自主行动层。

4. 安全评估机构成为标配

英国人工智能安全研究院职能更名强化、韩国人工智能安全研究所投入运行、美国 NIST AI 标准与创新中心发力、新加坡数字信任中心承接 AI 安全评估职能，“以技治技”的国家能力建设成为共同选择。

5. 发展与安全的权衡出现回调。

欧盟因标准与合格评定资源不足将高风险义务推迟 16 个月，美国联邦政府以“反碎片化”为由压制州法，均显示在技术竞争压力下，主要经济体对“监管过度抑制创新”保持高度警惕，AI 治理进入“边实施、边校准”的调适期。

6. 模型失控风险治理效果不佳

美国 NIST 测试发现针对智能体的劫持攻击成功率可达 81%。而在结果驱动的约束违规（ODCV-Bench）基准测试中，产业界测试了 12 个最先进的大型语言模型，发现违规率介

于 11.5%至 66.7%之间。即使是表现最佳的 Claude-Opus-4.6，也有 11.5%的运行违规。大多数被评估的模型在至少 25%的运行中表现不当，模型行为表现从随机性尝试的规避到公然无视规定的区间不等[93]。

[93] OSKI Solution. Best Autonomous AI Agents 2026: Tested & Ranked. [2026-05-27]. https://oski.site/blog/best-autonomous-ai-agents/

第四章 网络空间安全风险的发展趋势

4.1 网络安全风险发展趋势

4.1.1 网络安全将进入“机器速度对抗”时代

AI 驱动的自主攻击将规模化扩散，攻击经济学被彻底改写。随着开源模型能力提升与犯罪服务成熟，未来一至两年内，“AI 编排的多目标并行攻击”将从国家级专属能力下沉至普通网络犯罪集团。网络黑市上已经出现宣称能够达到国家级能力的人工智能黑客服务，如图 4-1 所示。单个操作者将可同时运营数百条攻击链，侦察、钓鱼、漏洞利用、横向移动全流程自动化，攻击的边际成本趋近于零。防御响应窗口将进一步压缩——若平均突破时间按当前斜率演进，2027 年可能缩短至 15 分钟以内，依赖人工介入的防御体系将系统性失效。

图 4-1 网络黑市上宣发的 AI 黑客服务（自动翻译）

4.1.2 关键基础设施的系统性破坏风险上升

OT/IT 融合面攻击增加（工业互联网、智能电网、车联网）、供应链“关键供应商”被系统性瞄准（托管服务商、软件更新通道、硬件固件）、网络攻击与动能打击的混合作战常态化（俄乌、中东模式外溢）。

4.1.3 勒索病毒向“低加密、高窃取、强施压”演进

赎金支付率下降将加速勒索生态转型：纯数据勒索、法律杠杆化施压（合规性披露义务）、针对下游客户的二级勒索、与黑客行动主义结合的“名誉毁灭式”曝光将增多。同时，勒索组织的技术门槛因 AI 而降低，生态碎片化与品牌化周期加快，执法打击的“打地鼠”困境短期难解。

4.2 数据安全风险发展趋势

4.2.1 聚合型平台与人工智能数据管道成为泄露的结构性源头

Salesforce 事件（单次波及近千家组织、15 亿条记录）揭示的模式——攻击者不再逐个攻破企业，而是攻陷承载海量企业数据的平台及其应用生态——将成为数据泄露的主导形态。随着企业数据向数据云、AI 训练管道与智能体记忆系统集中，“数据引力”越大，单点失守的爆炸半径越大。数据安全建设的重心将从“企业边界内的库表防护”转向“跨组织数据流的可见性与控制力”。

4.2.2 人工智能重塑数据安全供需两端

一方面在需求侧。AI 训练数据投毒、RAG 知识库越权、

智能体记忆泄露等 AI 特有数据风险将快速实体化。另一方面在供给侧。AI 驱动的数据发现与分类、自动化脱敏与合成数据生成、智能 DLP（语义级内容识别）将显著降低数据安全运营门槛；数据安全技术栈将全面“AI 原生化”。

4.2.3 “不可更改型”敏感数据面临严重威胁

随着移动互联网、智能终端和 AI 应用的普及，指纹、人脸、掌纹、基因序列这类具有终身不可更改属性的敏感数据采集规模持续扩大，大量这类数据集中存储在医疗、政务、金融、商业科技等各类机构的数据中心中，成为攻击者觊觎的高价值目标。不同于密码、支付密钥这类可随时更换的凭证，这类敏感数据一旦发生泄露，没有任何“重置”“补发”的补救空间，其带来的身份冒用、隐私侵害风险会伴随用户终身。同时 AI 技术大幅降低了这类泄露数据的利用门槛：攻击者可借助生成式 AI 快速基于泄露的生物特征合成高仿真的深度伪造内容，用于诈骗、身份伪造等非法活动，攻击成功率和危害程度都远高于传统数据泄露事件。未来这类敏感数据的泄露危害将持续凸显，成为数据安全防护需要重点应对的核心问题之一。

4.3 人工智能安全风险发展趋势

4.3.1 智能体将成为人工智能安全的第一风险域

智能体的商业化爆发与其治理滞后之间的落差，是当下 AI 安全最尖锐的矛盾。预计在下一年度将出现首批“智能体重大安全事件”，可能的威胁场景包括：越权操作造成实际

财产/数据损失、被劫持执行恶意任务等。相关事件将推动智能体身份标准、权限最小化框架与“强制人类在环”的合规要求落地。

4.3.2 人工智能攻防的“军备竞赛”向模型能力前沿推进

攻击侧，先进大模型的自主漏洞发现（Claude Mythos 等）与自动化利用将改变漏洞经济——零日的“稀缺性溢价”下降，防御方“修补鸿沟”面临指数级压力。防御侧，AI 赋能的威胁狩猎、自动化补丁生成与自愈系统将同步进化。攻防平衡的关键变量在于：前沿模型的“有害能力”（自主黑客、生物化学知识）是否可通过评估、分级部署与对齐技术有效约束。

4.3.3 人工智能治理的国际规则竞争进入关键窗口期

三种治理模式——欧盟的权利保障型（AI Act）、美国的创新优先型（AI Action Plan）、中国的发展与安全统筹型（治理框架 2.0+标识办法+全球 AI 治理行动计划）——将在 2026—2028 年围绕国际标准（ISO/IEC、IEEE）、联合国框架（全球数字契约、AI 科学小组）与双边互认展开竞争与调和。而算力、模型与数据的出口管制则使 AI 安全治理嵌入大国战略竞争，“治理碎片化”与“最低共识”两种前景并存。

第五章 应对网络空间安全风险的对策建议

5.1 加快推动关键信息基础设施安全防护能力人工智能转型

面对 AI 驱动的攻击形态变革，原有依赖人工响应、边界静态防护的传统体系已经难以适应新的对抗要求，必须推动防护能力全链条的 AI 化升级。**一是要构建适配“机器速度对抗”的自动化防御闭环。**将 AI 技术全面融入威胁感知、漏洞研判、攻击溯源、止损响应全流程，依托大模型的语义理解与异常识别能力，实现对 AI 赋能攻击、自主攻击的实时检测与快速处置，把防御响应窗口压缩到与攻击速度匹配的区间，解决人工响应滞后带来的系统性失效问题。**二是搭建 AI 驱动的第三方供应链安全监测体系。**针对关键基础设施供应链攻击加剧的趋势，对上游供应商提供的软件组件、硬件固件、云服务开展自动化的恶意代码检测、投毒排查与风险评级，实现跨多级供应链的风险可见与可控，破解“攻破一点、牵连全网”的困局。**三是推动安全防护全生命周期的 AI 赋能。**在关键信息基础设施的项目规划、建设开发、上线运营全流程嵌入 AI 安全评估能力，落实“安全左移”要求，提前识别架构设计、技术选型中潜在的安全缺陷，从源头降低安全风险。**四是将 AI 系统自身安全纳入关键信息基础设施安全保护的监管要求。**针对当前关键信息基础设施领域 AI 应用快速普及的现状，推动运营者落实提示注入防护、模型访

问控制、训练数据安全核查等安全措施，避免 AI 系统本身成为新的高风险攻击入口。

5.2 加快推动人工智能赋能数据安全治理

一是推动建立 AI 赋能的跨组织数据流监测体系。依托大模型的动态追踪与异常识别能力，对数据从采集、传输、存储到训练推理的全流程生成流转画像，实现跨企业、跨平台数据流动的实时可见与风险追溯，破解“单点失守、全域牵连”的防护困局，推动数据安全建设重心完成从“企业边界库表防护”到“跨组织数据流管控”的转型。**二是加快研发推广 AI 原生的数据安全防护技术栈。**建立覆盖训练数据投毒检测、RAG 知识库细粒度权限管控、智能体记忆动态脱敏的全链条防护机制，将数据安全防护要求嵌入 AI 开发、部署、运行的全生命周期，从源头阻断 AI 特有数据风险的实体化路径。**三是推动 AI 技术与隐私计算技术深度融合。**加快推广基于 AI 优化的联邦学习、差分隐私、可信执行环境等技术落地，实现在不原始流转敏感数据的前提下完成数据开发利用，大幅降低敏感数据集中存储带来的泄露风险；同时针对已经泄露的生物特征数据，探索基于 AI 的泄露风险监测与非法使用溯源技术，提升对生物特征数据滥用行为的发现和处置能力。**四是借助 AI 技术降低中小微企业数据安全运营门槛。**依托 AI 自动化完成敏感数据发现、分级分类、合规整改辅助等工作，推动数据安全防护能力普惠化，缓解当前多数企业“重业务、轻安全”背景下安全运营能力

不足的问题，整体提升全社会的数据安全防护水平。

5.3 加快推动建立人工智能能力评估规范体系

当前前沿大模型和智能体能力迭代速度快，高风险能力的“双重用途”治理缺口日益凸显，亟须加快构建覆盖 AI 全生命周期的能力分级分类评估规范体系，为 AI 安全治理提供可落地的技术遵循。**一是要明确标准化的评估核心维度。**围绕 AI 模型的自主决策能力、网络攻防能力、有害内容生成能力等高风险方向，制定可量化的评估指标与标准化测评流程，重点针对模型的漏洞发现、攻击全链路编排、高仿真深度伪造生成等能力开展专项评估，在此基础上形成清晰的 AI 能力风险等级划分框架。**二是要推动建立分级管控机制。**根据评估得出的风险等级明确对应 AI 模型的开发使用资质、可落地应用场景与安全管控要求，对具备高风险能力的 AI 模型实施严格的部署审批与全流程动态监测，禁止不符合资质要求的机构擅自开发、部署高风险 AI 能力，从准入环节管住“能力滥用”的源头。**三是要建立常态化动态评估机制。**针对 AI 模型版本更新快、能力迭代频繁的特点，要求开发者在完成重大能力升级、架构调整后重新开展风险评估，及时更新模型风险等级与管控要求，避免能力升级后原有风险管控措施失效的问题。**四是要加快培育专业第三方 AI 安全评估机构。**建立统一的评估质量管控标准，提升评估工作的专业性与客观性，为市场主体的 AI 安全自查、监管部门的合规核查提供专业支撑，同时推动国内评估标准与全球 AI 治

理框架的交流对接，提升我国在 AI 安全评估标准领域的话语权。

5.4 加快推动我国网络安全公共资源建设

当前我国网络安全领域公共基础资源供给存在统筹不足、开放程度不够、中小市场主体获取成本高的问题，亟须加大统筹建设与开放共享力度，为全行业安全防护提供坚实基础支撑。**一是要统筹建设国家级网络安全技术研发公共资源库。**整合恶意代码样本、AI 攻防训练数据集、网络安全案例库等基础资源，依托合规的开放平台面向高校、科研机构、中小网络安全企业低成本开放，降低网络安全技术研发与创业创新的门槛，助力我国网络安全技术创新生态培育。**二是要完善网络安全应急处置公共支撑体系。**建设国家级 AI 辅助应急响应技术平台，整合成熟的应急处置方案知识库，面向发生网络安全事件的中小机构、基层单位提供公益性质的技术支援，帮助其快速完成攻击止损、根源排查与系统恢复，降低安全事件造成的实际损失。**三是要针对新兴安全领域需求提前布局新型公共资源。**加快建设智能体安全测试公共沙箱、AI 生成有害内容识别公共样本库、生物特征泄露风险监测基础数据集等新型资源，满足全行业对 AI 安全、数据安全等新兴领域防护能力建设的基础资源需求，支撑新兴安全防护技术快速落地。**四是要建立常态化的公共资源更新运维机制。**依托 AI 技术实现资源的自动化分类标注、去重、质检与动态更新，保障公共资源的时效性与准确性，使其能够跟

上网络攻击形态与安全风险的变化节奏，持续发挥基础支撑作用。

5.5 加快统筹推动人工智能安全国际治理

以世界人工智能合作组织（WAICO）为抓手，推动建立包容性的 AI 治理对话机制，防止 AI 安全标准阵营化。**一是促进建立跨境 AI 安全事件通报与协同响应机制。**依托 WAICO 设立安全工作组，将“模型评估中逃逸并侵害第三方”纳入通报范畴，制定危险能力评估最低安全标准的国际共识，防止监管套利，落实“维护共同安全”的大会精神。**二是促进国际治理主体间合作。**积极探索中美欧在 AI 事件通报、前沿模型评估方法、深度伪造标识互认等技术层面建立“低政治敏感度”合作通道。支持发展中国家 AI 安全能力建设，弥合“AI 治理鸿沟”。**三是推动建立跨国治理与执法协作机制。**针对 AI 驱动的跨境网络犯罪、跨区域电信诈骗等全球性安全问题，推动建立跨国联合情报共享与执法协作机制，协同打击各类 AI 恶意使用行为，共同维护全球网络空间的公共安全秩序。

附录 A　缩略语

缩略语	释义
AI	人工智能
API	应用程序编程接口
APT	高级持续性威胁
ATT&CK	MITRE adversarial tactics, techniques, and common knowledge，MITRE 对抗战术、技术和通用知识库
BYOVD	自带脆弱驱动
C&C	命令控制（Command and Control）
CNNVD	国家信息安全漏洞库
CRA	欧盟《网络弹性法案》（Cyber Resilience Act）
CVSS	通用漏洞评分系统（Common Vulnerability Scoring System）
DDoS	分布式拒绝服务攻击（Distributed Denial of Service）
DBIR	数据泄露调查报告（Data Breach Investigations Report）
DLS	数据泄露站点（Data Leak Site）
DNS	域名系统（Domain Name System）
EDR	终端检测与响应（Endpoint Detection and Response）
FBI	美国联邦调查局
FSTEC	俄罗斯联邦技术与出口监督局
GB	中华人民共和国国家标准
GDPR	《通用数据保护条例》（General Data Protection Regulation）
GPS	全球定位系统（Global Positioning System）
HTTP	超文本传输协议（Hypertext Transfer Protocol）
HTTPS	超文本传输安全协议（Hypertext Transfer Protocol Secure）
IAB	初始访问经纪人（Initial Access Broker）
ICT	信息与通信技术（Information and Communications Technology）
ISO	国际标准化组织（International Organization for Standardization）
JLR	捷豹路虎（Jaguar Land Rover）
MCP	模型上下文协议（Model Context Protocol）
MDM	移动设备管理（Mobile Device Management）
MFA	多因素认证（Multi-Factor Authentication）
NCO	日本国家网络办公室
NISC	日本内阁网络安全中心
NIST	美国国家标准与技术研究院（National Institute of Standards and Technology）
OT	操作技术（Operational Technology）
P2P	对等网络（Peer to Peer）
PDPA	新加坡《个人数据保护法》（Personal Data Protection Act）
PDVSA	委内瑞拉国家石油公司
PIPA	韩国《个人信息保护法》（Personal Information Protection Act）
RaaS	勒索软件即服务（Ransomware as a Service）

缩略语	释义
RAT	远程访问木马（Remote Access Trojan）
RCE	远程代码执行（Remote Code Execution）
SaaS	软件即服务（Software as a Service）
SIRT	安全情报与响应团队（Security Intelligence and Response Team）
SIP	会话初始协议（Session Initiation Protocol）
SOC	安全运营中心（Security Operations Center）
TCP	传输控制协议（Transmission Control Protocol）
UDP	用户数据报协议（User Datagram Protocol）
VSS	卷影复制服务（Volume Shadow Copy Service）
WAICO	世界人工智能合作组织（World Artificial Intelligence Cooperation Organization）

附录 B　术语解释

下列术语和定义适用于本报告。

A.1

恶意软件 malware

被专门设计用于损坏或中断系统、破坏保密性、完整性和/或可用性的软件。

注：病毒和木马病毒都是恶意软件。

[来源：GB/T 25069-2022，3.141]

A.2

病毒 computer virus

一种程序，即通过修改其他程序，使其他程序包含一个自身可能已发生变化的原程序副本，从而完成传播自身程序，当调用受感染的程序，该程序即被执行。

注：病毒经常造成某种损失或困扰，并可能被某一事件（诸如出现的某一预定日期）触发。

[来源：GB/T 25069-2022，3.50]

A.3

挖矿病毒 cryptoware

以获得数字加密货币为目的，控制他人的计算机并完成大量运算的恶意软件。

A.4

木马病毒 trojan

一种表面无害的程序，它包含恶性逻辑程序，可导致未授权地收集、伪造或破坏数据。

[来源：GB/T 37027-2018，3.594]

A.5

僵尸网络 botnet

与攻击者通信，接收攻击者的指令，并与其他感染此类恶意软件的主机一起对特定目标发起攻击的恶意软件。

A.6

勒索病毒 ransomware

采取加密或屏蔽用户操作等方式劫持用户对系统或数据的访问权，并借此向用户索取勒索金的恶意软件。

A.7

高级持续性威胁攻击 advanced persistent threat

高级持续性威胁攻击指为了商业或政治利益针对特定实体（如组织、国家等）进行一系列秘密和连续攻击的过程。“高级”指攻击方法先进复杂；“持续”指攻击者连续监控目标对象，并从目标对象不断提取敏感信息。

[来源：GB/T 37027-2018，B.11]

A.8

供应链攻击 supply chain attack

供应链攻击是一种针对开发人员和供应商的网络攻击。攻击目的是通过恶意篡改合法产品或服务的源代码、生产过程和交付过程实现向最终的攻击目标植入恶意软件。

A.9

数据泄露 data leak

违反信息安全策略，导致数据被未经授权的实体使用的行为。

[来源：GB/T 25069-2022，3.670]

A.10

数据损失 data loss

导致受保护的数据在传输、存储或其他处理过程中发生意外或非法的损毁、丢失、改动或者未授权的披露或访问的安全性降低。

[来源：GB/T 25069-2022，3.563]

A.11

重要数据 key data

特定领域、特定群体、特定区域或达到一定精度和规模的，一旦被泄露或篡改、损毁，可能直接危害国家安全、经济运行、社会稳定、公共健康和安全的数据。

[来源：GB/T 43697-2024，3.2]

A.12

智能体 agent

一种能根据环境和目标灵活行动的智能软件系统，能够适应不断变化的环境和目标，从经验中学习，并在其感知和计算能力的限制下做出适当的选择。

[来源：《人工智能：现代方法》第四版]

A.13

智能体化 AI　agentic AI

运用复杂的推理和迭代规划来自主解决复杂、多步骤的问题的人工智能系统。

附录 C　2026 年度排名前 20 位的木马病毒简介

序号	木马病毒名称	简述	典型恶意行为
1	Trojan/Win64.CoinMiner[Miner]	Trojan/Win64.CoinMiner[Miner]的首个样本在 2012 年 10 月被安天捕获。该特洛伊木马关联样本是 Windows 平台下的 PE 文件，主要针对 x86-64 体系结构下的 Windows 64 位系统进行攻击。该 Trojan 的主要行为是 Miner，通过其核心行为，对目标造成数据丢失、系统崩溃、信息泄露等问题。该特洛伊木马变种数量有 1003 种，但经历了大量的免杀加工，以至样本 Hash 数量近 3591.9 万。目前 Trojan/Win64.CoinMiner[Miner]存在可执行文件、压缩文件等至少 3 种格式的样本，可执行文件占绝大部分。	Trojan/Win64.CoinMiner[Miner]会持续占用计算机的 CPU 和内存资源，使系统变得极度缓慢；它会通过网络连接进行通信，将挖矿所得的加密货币发送至指定的攻击者控制的服务器。 该病毒具有自我保护机制，在检测到杀软时会潜伏隐匿以免被查杀。 Trojan/Win64.CoinMiner[Miner]可能会关闭安全防护软件或修改系统设置以确保自身不受到检测；它还可能尝试利用系统中存在的漏洞来绕过安全机制，保持持续存在并传播到其他设备。 这种病毒可能会下载其他的变种或者携带其他恶意组件，对系统造成更大的危害。
2	Trojan/Win32.VB	Trojan/Win32.VB 早在 2006 年就已经出现。该特洛伊木马关联样本是 Windows 平台下的 PE 文件，主要针对 X86 体系结构下的 Windows 32 位系统进行攻击。该家族是对 Visual Basic 编写的特洛伊木马的统称，其并不一定由同一组织编写，之前也不一定有同源关系。由于 Visual Basic 是一种解释型语言，其编写难度较低，导致样本数量十分丰富。但分析其二进制样本的同源性需要更多的分析人员和更高的算力需求，因此安天按照业界约定俗成的惯例，采用编译器命名来统一作为此类样本的家族名称。该特洛伊木马变种数量有 3.4 万种，但经历了大量的免杀加工，以至样本 Hash 数量近 2686.4 万。	数据窃取： Trojan/Win32.VB 可能会窃取用户的敏感信息，如登录凭据、个人身份信息、银行账户等。 远程控制：攻击者可以利用该木马对受感染系统进行远程控制，执行恶意指令或进行其他攻击活动。 系统破坏：木马可能会破坏系统文件、关闭安全防护或导致系统崩溃。 传播方式： Trojan/Win32.VB 可以通过电子邮件附件、下载来源不明的文件或软件漏洞进行传播，以扩大感染范围。

		目前 Trojan/Win32.VB 存在可执行文件、压缩文件等至少 3 种格式的样本，可执行文件占绝大部分。	
3	Trojan/Win32.Padodor[Backdoor]	Trojan/Win32.Padodor[Backdoor]早在 2006 年就已经出现。该特洛伊木马关联样本是 Windows 平台下的 PE 文件，主要针对 X86 体系结构下的 Windows 32 位系统进行攻击。它的主要行为是以隐藏、欺骗的方式打开安全漏洞或绕过身份验证机制，从而给攻击者提供对受感染计算机的远程访问权限。后门行为通常由黑客或恶意软件开发者利用，用于悄悄地远程控制受害者的计算机，执行非授权的操作或者窃取敏感信息。该特洛伊木马变种数量有 104 种，但经历了大量的免杀加工，以至样本 Hash 数量近 5709.2 万。目前 Trojan/Win32.Padodor[Backdoor]存在可执行文件、压缩文件等至少 4 种格式的样本，可执行文件占绝大部分。	后门功能：Trojan/Win32.Padodor[Backdoor]作为后门程序，允许攻击者通过远程访问受感染计算机，执行恶意活动。 数据窃取：病毒可能会窃取用户的敏感信息，例如用户名、密码、银行账号等。 文件操作：该病毒可以操纵文件，删除、修改或上传恶意文件到受感染计算机上。 远程控制：攻击者可以通过远程控制功能控制受感染计算机，执行恶意指令或安装其他恶意软件。
4	Trojan/Win32.Qukart[Proxy]	Trojan/Win32.Qukart[Proxy]的首个样本在 2013 年 11 月被安天捕获。该特洛伊木马关联样本是 Windows 平台下的 PE 文件，主要针对 X86 体系结构下的 Windows 32 位系统进行攻击。该 Trojan 的主要行为是 Proxy，通过其核心行为，对目标造成数据丢失、系统崩溃、信息泄露等问题。该特洛伊木马变种数量有 1.6 万种，但经历了大量的免杀加工，以至样本 Hash 数量近 4028.9 万。目前 Trojan/Win32.Qukart[Proxy]存在可执行文件、压缩文件等至少 4 种格式的样本，可执行文件占绝大部分。	代理服务器操作：Trojan/Win32.Qukart[Proxy]作为一个代理服务器，用于转发网络请求和控制通信，使攻击者可以通过受感染计算机进行恶意活动。 窃取敏感信息：病毒可能会监视网络流量，并窃取用户的敏感信息，例如用户名、密码、银行账号等。 绕过网络安全防御：该病毒可以绕过网络防火墙和其他安全措施，以保持持久性和隐匿性。 后门功能：病毒可能会打开后门，允许攻击者远程访问受感染计算机，并执行进一步的恶意活动。

5	Trojan/Win32.Cosmu	Trojan/Win32.Cosmu 早在 2006 年就已经出现。该特洛伊木马关联样本是 Windows 平台下的 PE 文件，主要针对 X86 体系结构下的 Windows 32 位系统进行攻击。该特洛伊木马变种数量有 2.6 万种，但经历了大量的免杀加工，以至样本 Hash 数量近 1675.3 万。目前 Trojan/Win32.Cosmu 存在可执行文件、压缩文件等至少 3 种格式的样本，可执行文件占绝大部分。	后门功能： Trojan/Win32.Cosmu 可以在受感染的计算机上开设后门，以便攻击者能够远程访问和控制该计算机。 数据窃取：病毒会窃取受感染计算机中的个人信息、登录凭据、银行账户信息等敏感数据。 远程操作： Trojan/Win32.Cosmu 允许攻击者远程执行操作，例如下载和安装其他恶意软件，修改系统设置等。 恶意下载：病毒可能会下载和安装其他恶意文件或工具，用于进一步的攻击和感染。
6	Trojan/Win32.BitCoinMiner[Miner]	Trojan/Win32.BitCoinMiner[Miner]的首个样本在 2011 年 2 月被安天捕获。该特洛伊木马关联样本是 Windows 平台下的 PE 文件，主要针对 X86 体系结构下的 Windows 32 位系统进行攻击。该 Trojan 的主要行为是 Miner，通过其核心行为，对目标造成数据丢失、系统崩溃、信息泄露等问题。目前 Trojan/Win32.BitCoinMiner[Miner]存在可执行文件、压缩文件至少两种格式的样本，可执行文件占绝大部分。	Trojan/Win32.BitCoinMiner[Miner]会尝试在系统启动时自动运行，并试图隐藏自己的存在。 该病毒会监控系统中的反病毒软件并尝试关闭或绕过其实时保护机制，以确保自身不受检测。 Trojan/Win32.BitCoinMiner[Miner]可能会修改系统注册表，以确保其自身始终保持激活状态。 病毒会不断消耗受感染设备的计算资源，并频繁与外部服务器进行通信以获取挖掘任务。 该病毒可能会通过网络漏洞或恶意链接进行传播，以感染更多的设备。 尝试掩盖自身的行踪与存在，以逃避被发现与清除。

7	Trojan/Script.Phonzy	Trojan/Script.Phonzy 早在 2006 年就已经出现。该特洛伊木马关联样本主要运行或者载体为 Script。该特洛伊木马变种数量有 3 种，但经历了大量的免杀加工，以至样本 Hash 数量近 881.2 万。目前 Trojan/Script.Phonzy 存在文本、脚本文件等至少 5 种格式的样本，文本占绝大部分。	它会尝试绕过杀软的检测和阻止机制，以确保自身能够长时间存在并运行。 它会下载和安装其他恶意软件或插件，扩大攻击面并继续对系统进行破坏。 它会监视用户的网络活动和敏感信息，例如登录凭证、信用卡号码等，并将其发送给黑客。 它会修改系统的设置和文件，破坏操作系统的稳定性和功能性。 它会利用系统资源进行挖矿或进行分布式拒绝服务（DDoS）攻击，给其他用户带来不便。 它会通过发送垃圾邮件或链接传播给其他计算机，继续扩大感染范围。
8	Trojan/JS.CoinMiner[Miner]	Trojan/JS.CoinMiner[Miner]的首个样本在 2018 年 2 月被安天捕获。该特洛伊木马关联样本是 JavaScript 脚本，主要通过操纵网页或利用浏览器的漏洞进行攻击。该 Trojan 的主要行为是 Miner，通过其核心行为，对目标造成数据丢失、系统崩溃、信息泄露等问题。该特洛伊木马变种数量有 441 种，但经历了大量的免杀加工，以至样本 Hash 数量近 288.8 万。目前 Trojan/JS.CoinMiner[Miner]存在脚本文件、文本等至少 4 种格式的样本，脚本文件占绝大部分。	利用浏览器漏洞自动下载并运行恶意 JS 代码 隐藏在正常的网页 js 文件中，使用户难以察觉 在后台默默执行挖矿操作，消耗计算资源 自我复制并传播到其他系统 更新自身以规避杀软检测 禁用系统安全软件和阻止安全更新。
9	Trojan/HTML.Redirector	Trojan/HTML.Redirector 早在 2008 年就已经出现。该特洛伊木马关联样本是 HTML 文件，并且可能包含嵌入的 JavaScript 代码，主要利用浏览器的漏洞进行各种类型的网络攻击。该特洛伊木马变种数量有 1238 种，但经历了大量的免杀加工，以至样本 Hash 数量近 351.3 万。目前 Trojan/HTML.Redirector	HTML 代码植入：Trojan/HTML.Redirector 特洛伊木马主要通过 HTML 代码嵌入受感染的网页中，以在用户访问网页时进行攻击。 重定向攻击：该病毒可能会将用户重定向到其他恶意网站，从而进一步传播病毒或展开其他攻击。

		存在文本、脚本文件等至少 4 种格式的样本，文本占绝大部分。	数据窃取：特洛伊木马可能会窃取用户在感染网页中输入的敏感信息，例如账户凭据、个人信息等。 后门功能： Trojan/HTML.Redirector 可能会在受感染的网页中创建后门，允许黑客远程访问和控制用户的计算机。
10	Trojan/Win32.Qukart[Spy]	Trojan/Win32.Qukart[Spy]早在 2006 年就已经出现。该特洛伊木马关联样本是 Windows 平台下的 PE 文件，主要针对 X86 体系结构下的 Windows 32 位系统进行攻击。它的主要行为是在用户不知情的情况下收集敏感信息并将其发送给开发者。该软件通常以欺骗性的方式侵入用户的系统，例如通过电子邮件附件、病毒感染的网站或不安全的下载来源。该特洛伊木马变种数量有 54 种，但经历了大量的免杀加工，以至样本 Hash 数量近 1404.2 万。目前 Trojan/Win32.Qukart[Spy]存在可执行文件、压缩文件等至少 4 种格式的样本，可执行文件占绝大部分。	信息窃取： Trojan/Win32.Qukart[Spy]会窃取用户的敏感信息，如登录凭据、银行账户信息、个人身份信息等。 监视活动：木马会监视用户的击键、屏幕截图、浏览器活动等，以收集更多的敏感信息。 远程控制：攻击者可以利用该木马对受感染系统进行远程控制，执行恶意指令或进行其他攻击活动。 传播方式： Trojan/Win32.Qukart[Spy]可以通过下载来源不明的文件、恶意链接或软件漏洞进行传播，以扩大感染范围。
11	Trojan/Win32.CoinMiner[Miner]	Trojan/Win32.CoinMiner[Miner]的首个样本在 2013 年 2 月被安天捕获。该特洛伊木马关联样本是 Windows 平台下的 PE 文件，主要针对 X86 体系结构下的 Windows 32 位系统进行攻击。该 Trojan 的主要行为是 Miner，通过其核心行为，对目标造成数据丢失、系统崩溃、信息泄露等问题。该特洛伊木马变种数量有 6861 种，但经历了大量的免杀加工，以至样本 Hash 数量近 146.9 万。目前 Trojan/Win32.CoinMiner[Miner]存在可执行文件、压缩文件等至少 3 种格式的样本，可执行文件占绝大部分。	Trojan/Win32.CoinMiner[Miner] 会尝试关闭系统中已安装的杀毒软件或防护软件，以确保自身能持续潜伏并进行挖矿活动。 它会修改系统的注册表项以确保在系统启动时自动运行，并隐藏自身进程以避免被发现。 该病毒可能会利用系统漏洞进行横向扩散，尝试感染其他主机。 Trojan/Win32.CoinMiner[Miner] 可能会监视用户的网络活动并窃取敏感信息，以发送给远程控制者。 它会持续消耗系统的计算资

			源进行挖矿操作，导致系统变得缓慢不堪。 对抗杀软：它可能会通过不断更新变异来规避杀软的检测，增加清除难度。
12	Trojan/JS.Cryxos	Trojan/JS.Cryxos 的首个样本在 2011 年 2 月被安天捕获。该特洛伊木马关联样本是 JavaScript 脚本，主要通过操纵网页或利用浏览器的漏洞进行攻击。该特洛伊木马变种数量有 450 种，但经历了大量的免杀加工，以至样本 Hash 数量近 245.8 万。目前 Trojan/JS.Cryxos 存在脚本文件、文本等至少 5 种格式的样本，脚本文件占绝大部分。	修改浏览器设置，如更改主页、搜索引擎等，导致用户无法正常浏览网页。 盗取用户的敏感信息，如登录账号、密码等，并进行非法使用。 拦截用户的网络通信，例如截取用户的网络请求，以篡改内容或者进行钓鱼攻击。 下载和安装其他恶意软件，如间谍软件、勒索软件等。 恶意篡改操作系统和应用程序的配置，导致系统不稳定或者无法正常工作。 对抗杀软软件，例如禁止杀软的更新、禁用杀软的保护功能等，以免被发现和清除。
13	Trojan/MSIL.Lazy	Trojan/MSIL.Lazy 的首个样本在 2022 年 7 月被安天捕获。该特洛伊木马关联样本是.NET 被编译的源代码文件，主要针对可运行.NET 平台的系统进行攻击。该特洛伊木马变种数量有 134 种，但经历了大量的免杀加工，以至样本 Hash 数量近 128.1 万。目前 Trojan/MSIL.Lazy 存在可执行文件、压缩文件等至少 3 种格式的样本，可执行文件占绝大部分。	它会试图隐藏自己的存在，避免被杀软检测出来。 它可能会通过下载其他恶意软件来进一步入侵用户系统。 它可以窃取用户的个人信息，如登录凭据、信用卡信息等。 它可以在用户不知情的情况下远程控制用户计算机，执行恶意操作。 它可能会通过修改系统文件或注册表来破坏系统的稳定性和安全性。 它可以利用系统漏洞传播自身，感染其他计算机。

14	Trojan/Shell.EncystPHP	Trojan/Shell.EncystPHP 的首个样本在 2026 年 2 月被安天捕获。该特洛伊木马关联样本主要运行或者载体为 Shell。该特洛伊木马变种数量有 1 种，但经历了大量的免杀加工，以至样本 Hash 数量近 124.4 万。目前 Trojan/Shell.EncystPHP 存在脚本文件、压缩文件等至少 3 种格式的样本，脚本文件占绝大部分。	代码加密与混淆：使用多种编码（如 Base64）、加密算法或自定义混淆技术包裹核心恶意代码，逃避基于特征码的静态杀毒引擎检测。 动态执行与内存驻留：恶意载荷往往在运行时才在内存中解密并执行，避免在磁盘上留下完整的可识别的恶意文件，对抗行为分析。 进程注入与伪装：可能注入合法的 Web 服务器进程（如 php-cgi、apache 进程）中运行，伪装成正常系统活动，以避开进程监控工具的告警。 反调试与沙箱检测：包含检测调试环境、虚拟机或沙箱的代码，一旦发现处于分析环境，则停止恶意行为或执行无害操作，增加分析难度。 权限提升与持久化：尝试利用系统或应用漏洞提升权限，并通过修改系统配置文件、注册表（在 Windows 服务器中）或创建计划任务、服务等方式实现持久化驻留。 通信加密与隐匿：与控制服务器（C&C）的通信通常使用加密通道（如 HTTPS、自定义加密协议），并模仿正常 HTTP 流量，以绕过网络层流量监测和入侵防御系统。
15	Trojan/JS.Redirector	Trojan/JS.Redirector 早在 2007 年就已经出现。该特洛伊木马关联样本是 JavaScript 脚本，主要通过操纵网页或利用浏览器的漏洞进行攻击。该特洛伊木马变种数量有 2802 种，但经历了大量的免杀加工，以至样本 Hash 数量近 829.2 万。目前 Trojan/JS.Redirector 存在	浏览器重定向：Trojan/JS.Redirector 通过篡改网页代码或重定向用户的浏览器流量，将用户导航到恶意网站或其他危险网页。 恶意广告：该木马可以显示欺骗性广告或弹出窗口，诱导用户点击，可能导致进一

		文本、脚本文件等至少 4 种格式的样本，文本占绝大部分。	步的感染或信息泄露。 恶意代码植入： Trojan/JS.Redirector 可能会植入恶意代码到感染的网页中，从而对用户进行钓鱼攻击、数据窃取或其他恶意活动。 传播途径：该木马可以通过感染的网页、电子邮件或其他方式进行传播，以扩大感染范围。
16	Trojan/Win32.Vilsel	Trojan/Win32.Vilsel 早在 2006 年就已经出现。该特洛伊木马关联样本是 Windows 平台下的 PE 文件，主要针对 X86 体系结构下的 Windows 32 位系统进行攻击。该特洛伊木马变种数量有 2.5 万种，但经历了大量的免杀加工，以至样本 Hash 数量近 651.7 万。目前 Trojan/Win32.Vilsel 存在可执行文件、压缩文件等至少 5 种格式的样本，可执行文件占绝大部分。	后门功能： Trojan/Win32.Vilsel 可以在受感染的计算机上开设后门，以便攻击者能够远程访问和控制该计算机。 数据窃取：病毒可能会窃取受感染计算机中的敏感数据，如登录凭据、个人信息等。 恶意下载： Trojan/Win32.Vilsel 可能会下载和安装其他恶意软件或文件，导致进一步的系统感染。 系统破坏：病毒可能会修改、删除或损坏系统文件和关键组件，导致系统崩溃或运行异常。
17	Trojan/Win32.Copak	Trojan/Win32.Copak 的首个样本在 2011 年 3 月被安天捕获。该特洛伊木马关联样本是 Windows 平台下的 PE 文件，主要针对 X86 体系结构下的 Windows 32 位系统进行攻击。目前 Trojan/Win32.Copak 存在可执行文件、压缩文件等至少 4 种格式的样本，可执行文件占绝大部分。	Trojan/Win32.Copak 会通过电子邮件附件、恶意网站、P2P 下载等方式传播。 一旦感染，它会通过在后台运行的方式窃取用户的个人信息，如银行账号、登录密码等。 Trojan/Win32.Copak 会监视用户的网络活动，并在用户访问银行、支付网站等时拦截用户输入的敏感信息。 它可以绕过杀毒软件的检测，以确保长期存留在用户的计算机中。 Trojan/Win32.Copak 还可以

			利用用户计算机的资源进行挖矿等非法活动。 它可能会利用计算机的漏洞，导致系统崩溃或数据丢失。
18	Trojan/HTML.Redirector[Phishing]	Trojan/HTML.Redirector[Phishing]的首个样本在2025年3月被安天捕获。该特洛伊木马关联样本是HTML文件，并且可能包含嵌入的JavaScript代码，主要利用浏览器的漏洞进行各种类型的网络攻击。该Trojan的主要行为是Phishing，通过其核心行为，对目标造成数据丢失、系统崩溃、信息泄露等问题。该特洛伊木马变种数量有171种，但经历了大量的免杀加工，以至样本Hash数量近101.5万。目前Trojan/HTML.Redirector[Phishing]存在文本、压缩文件等至少5种格式的样本，文本占绝大部分。	修改浏览器设置，强制重定向用户访问恶意网站。 监视用户浏览器行为和敏感信息输入，窃取用户账号密码等数据。 屏蔽杀软的检测和拦截，保持持久性感染。 隐匿自身行踪，避免被用户和安全软件察觉。 利用漏洞或社会工程手段传播病毒至其他系统。 更新自身变种以规避杀软识别，持续对系统构成威胁。
19	Trojan/Win32.Swisyn	Trojan/Win32.Swisyn早在2006年就已经出现。该特洛伊木马关联样本是Windows平台下的PE文件，主要针对X86体系结构下的Windows 32位系统进行攻击。目前Trojan/Win32.Swisyn存在可执行文件、压缩文件至少两种格式的样本，可执行文件占绝大部分。	Trojan/Win32.Swisyn具有自我复制功能，能够在计算机系统中生成多个副本，并隐藏自身的存在，以免被杀软发现和清除。 该病毒会监视用户的网络活动，并窃取用户的账户名、密码、信用卡信息等敏感数据。 Trojan/Win32.Swisyn可以在后台执行恶意命令，包括下载和安装其他恶意软件，使计算机系统更容易受到进一步的攻击。 它会修改系统的注册表项和重要文件，以保证自己在系统重启后仍然能够自动运行。 该病毒可以通过电子邮件、社交媒体和网络下载等途径传播，并且往往伪装成合法

			的文件或应用程序，欺骗用户点击下载和执行。 Trojan/Win32.Swisyn 具有对抗杀软的能力，可以监测系统中的杀毒程序，并尝试禁用或绕过其防护机制。
20	Trojan/Win32.Salgorea[Backdoor]	Trojan/Win32.Salgorea[Backdoor]的首个样本在 2013 年 2 月被安天捕获。该特洛伊木马关联样本是 Windows 平台下的 PE 文件，主要针对 X86 体系结构下的 Windows 32 位系统进行攻击。它的主要行为是以隐藏、欺骗的方式打开安全漏洞或绕过身份验证机制，从而给攻击者提供对受感染计算机的远程访问权限。后门行为通常由黑客或恶意软件开发者利用，用于悄悄地远程控制受害者的计算机，执行非授权的操作或者窃取敏感信息。该特洛伊木马变种数量有 414 种，但经历了大量的免杀加工，以至样本 Hash 数量近 595.6 万。目前 Trojan/Win32.Salgorea[Backdoor]存在可执行文件、压缩文件等至少 3 种格式的样本，可执行文件占绝大部分。	后门功能： Trojan/Win32.Salgorea[Backdoor]会在受感染的计算机上创建后门，以便攻击者可以远程访问和控制该计算机。 远程命令执行：病毒允许攻击者通过远程命令执行恶意操作，例如下载和安装其他恶意软件、修改系统设置等。 数据窃取： Trojan/Win32.Salgorea[Backdoor]可能会窃取受感染计算机中的敏感数据，如登录凭据、个人信息等。 文件上传和下载：病毒可以上传和下载文件，用于进一步的攻击和感染。

（来源：安天）