

2024年度 漏洞态势分析报告



目 录

目 录.....	- 1 -
一、 前言	- 3 -
二、 2024 年度漏洞数据统计与分析	- 4 -
1. 基于主流漏洞库已公开披露漏洞的总体趋势分析	- 4 -
2. 基于主流漏洞库已公开披露漏洞的数据统计	- 5 -
3. 基于主流漏洞库已公开披露漏洞的趋势预测	- 10 -
4. 基于主流漏洞库已公开披露漏洞的厂商分析	- 11 -
5. 基于国产厂商分布数据的解析	- 13 -
6. 本章小结	- 14 -
三、 2024 年度 CWE 排行榜解读	- 15 -
1. CWE Top25 简介	- 15 -
2. 2024 年度 CWE Top25 排行榜单	- 15 -
3. 基于 2024 年度 CWE 数据的趋势分析	- 18 -
4. 本章小结	- 19 -
四、 2024 年度漏洞预警回顾	- 20 -
1. 年度严重漏洞 (CVSS3.1 评分 ≥ 9.0)	- 21 -
2. 基于 2024 年度漏洞预警数据的分析	- 25 -
3. 本章小结	- 27 -
五、 2024 年攻防演练高危漏洞回顾	- 28 -
1. 攻防演练期间常见漏洞类型分布概况	- 28 -

2. 攻防演练中常见漏洞利用类型	- 30 -
3. 本章小结	- 32 -
六、 AI 安全隐患与未来趋势分析	- 33 -
1. OWASP Top10 简介	- 33 -
2. OWASP LLM Top10 安全威胁解读	- 33 -
3. LLM 内生安全漏洞列表	- 37 -
4. LLM 应用过程中面临的风险场景	- 38 -
5. LLM 安全治理框架建设	- 40 -
6. 本章小结	- 43 -
七、 关于我们	- 45 -

一、前言

随着网络空间应用的普及和依赖程度的提升,随之而来的是日益复杂且严峻的网络安全挑战。网络安全问题已经成为全球关注的焦点,尤其是网络漏洞的存在,不仅威胁着企业和个人的数字资产安全,也对国家安全、社会稳定以及经济发展构成了极大威胁。

在过去的几年里,尽管安全技术不断进步,漏洞的数量和影响力依然没有得到有效遏制。从传统的操作系统和应用程序漏洞到新兴技术领域中的漏洞,如云计算、物联网、车联网、人工智能等,漏洞的类型和形式呈现出多样化和复杂化的趋势。在这些领域中,一些看似微小的漏洞,往往能够成为攻击者的突破口,造成系统瘫痪、数据泄露、甚至影响社会和国家的核心安全。

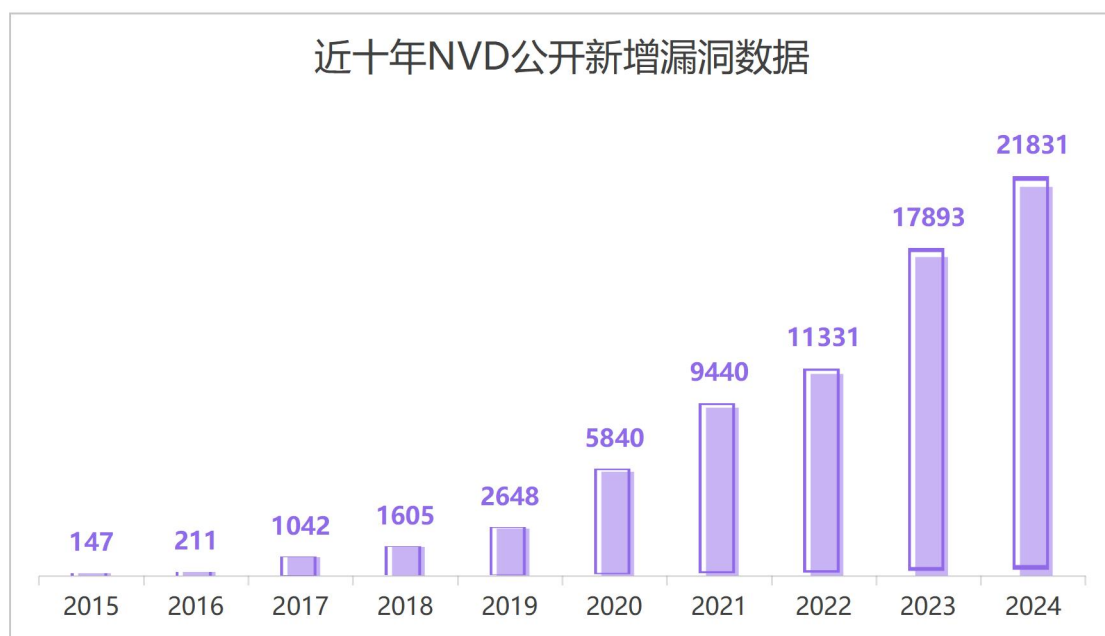
2024 年度漏洞态势分析报告旨在全面回顾和总结这一年内的漏洞数据,分析与揭示网络漏洞的发展趋势、关键特点以及潜在的风险点。通过深入分析来自国内外主流漏洞库的数据,报告全面梳理了 2024 年度漏洞的各项趋势指标,涵盖漏洞总数、漏洞等级分布、漏洞产生原因等方面。在数据分析过程中,我们通过对历史数据和 2024 年度漏洞数据的对比分析,揭示漏洞数量的变化趋势,以及漏洞等级分布情况、攻击类型以及漏洞背后可能带来的安全隐患。报告还梳理了 2024 年度高危漏洞的预警并整理出了 2024 年度需要注意的严重漏洞,旨在帮助各界及时发现和修补网络漏洞,减少潜在的安全风险。

当今网络安全问题已成为全球范围内急需解决的重要课题,通过加强漏洞监测、加强防护体系建设、提高安全防范意识,政府、企业及个人可以共同构筑一道坚实的网络安全防线,维护网络空间的安全与稳定,推动数字时代的健康发展。

二、2024 年度漏洞数据统计与分析

1. 基于主流漏洞库已公开披露漏洞的总体趋势分析

自 2024 年 1 月 1 日至 2024 年 12 月 17 日, 2024 年度安恒信息 CERT 监测到 NVD 已公开披露漏洞共计 21831 个, 较 2023 年同比增长 22.08%, 通过对历年 NVD 已公开披露漏洞数据的整理, 近十年 NVD 漏洞库已公开披露漏洞数量分布如下图所示:



图表 1 近十年 NVD 公开新增漏洞数据

从上述图表中可以看出, 近十年 NVD 公开的漏洞数量呈现出显著的逐年增长趋势, 尤其是在 2020 年之后, 漏洞数量的增长速度明显加快。2015 年 NVD 公开新增漏洞数量仅为 147 个, 而到 2024 年这一数字已达到 21831 个, 增长了约 150 倍。2017 年公开漏洞新增数量突破 1000 个, 达到 1042 个, 2020 年以后, 公开漏洞数量开始显著攀升, 2020 年到 2024 年的年均增长率高达 30% 以上。漏洞数量的快速增长表明系统和软件的复杂性在提升, 也反映了近年来信息安全领域的压力正在持续增大。

2. 基于主流漏洞库已公开披露漏洞的数据统计

漏洞各等级分布概况

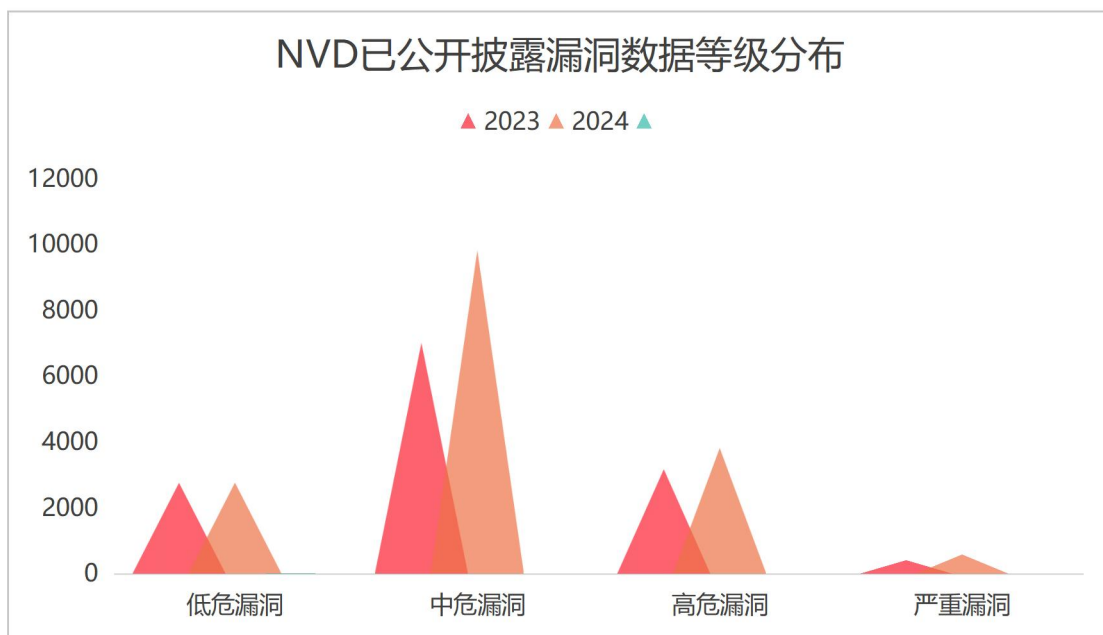
根据 NVD 2023 年度与 NVD2024 年度公开漏洞数据库中具 CVSS3.1 评分的漏洞数据分析，整体安全态势呈现出一定的上升趋势。数据详情如下：

低危漏洞：NVD 2024 年度低危漏洞数量为 2768 个，2023 年度低危漏洞数量 2763 个，2024 年度较 2023 年度同比增长 0.18%。

中危漏洞：NVD 2024 年度中危漏洞数量为 9821 个，2023 年度中危漏洞数量 7014 个，2024 年度较 2023 年度同比增长 40.01%。

高危漏洞：NVD 2024 年度高危漏洞数量为 3819 个，2023 年度中危漏洞数量 3175 个，2024 年度较 2023 年度同比增长 20.28%。

严重漏洞：NVD 2024 年度严重漏洞的数量为 586 个，2023 年度严重漏洞数量 413 个，2024 年度较 2023 年度同比增幅为 41.89%。



图表 2 NVD 已公开披露漏洞数据等级分布

漏洞产生原因分布

设计错误（3568 条，占比最高，约 60%），设计阶段的漏洞占绝对主导地位。这类漏洞往往是由于系统架构或逻辑设计不完善、需求分析不到位导致，后期修复成本较高。设计错误可能导致系统核心功能失效或被绕过，影响深远，覆盖面广，难以完全根除。

CWE-657: 不安全设计
CWE-628: 使用不正确指定参数的函数调用
CWE-306: 关键功能缺少身份验证
CWE-862: 缺少授权
CWE-326: 加密强度不足
CWE-327: 使用损坏或有风险的加密算法
CWE-20: 输入验证不当
CWE-391: 未检查错误条件
CWE-248: 未捕获的异常
CWE-833: 死锁
CWE-670: 始终不正确的控制流实施
CWE-16: 配置
CWE-276: 默认权限不正确

表 1 设计错误致漏洞产生对应的 CWE

输入验证错误（1194 条，占比约 20%），输入验证错误是由于开发过程中对用户输入的过滤和验证不足导致的漏洞，例如 SQL 注入、跨站脚本（XSS）等。这类漏洞通常成为攻击者的主要目标，因为其利用门槛较低、成功率较高，容易

导致数据泄露或系统控制。

CWE-20: 输入验证不当
CWE-89: SQL 注入
CWE-78: 操作系统命令注入
CWE-94: 代码注入
CWE-116: 输出编码不当或逃逸
CWE-118: 越界访问
CWE-601: URL 重定向
CWE-209: 通过错误消息暴露信息
CWE-190: 整数溢出
CWE-327: 使用损坏或有风险的加密算法
CWE-77: 命令注入

表 2 输入验证错误致漏洞产生对应的 CWE

边界条件错误 (605 条, 占比约 10%)，此类漏洞主要与程序对边界情况（如数组越界、内存溢出）处理不当有关。边界条件错误容易被攻击者利用进行缓冲区溢出攻击，导致系统崩溃或执行恶意代码。

CWE-120: 缓冲区溢出
CWE-125: 越界读取
CWE-787: 越界写入
CWE-129: 数组索引验证不当
CWE-190: 整数溢出
CWE-191: 整数下溢

CWE-680: 整数溢出导致缓冲区溢出
CWE-369: 除零错误
CWE-617: 可达断言
CWE-131: 缓冲区大小计算错误
CWE-805: 使用错误长度值访问缓冲区
CWE-22: 路径遍历
CWE-754: 未正确检查异常或特殊情况
CWE-124: 缓冲区下溢
CWE-134: 使用外部控制的格式化字符串

表 3 边界条件错误致漏洞产生对应的 CWE

访问验证错误 (114 条, 占比较低, 约 2%)，访问验证错误是由于对用户或系统访问权限的验证不足，例如缺乏身份认证或错误的授权配置。容易导致未授权访问，直接威胁系统安全性和敏感信息的保护。

CWE-285: 授权机制不正确
CWE-287: 身份验证不足
CWE-288: 身份验证绕过
CWE-302: 缺少授权
CWE-306: 缺少身份验证的关键功能
CWE-862: 缺少授权的功能
CWE-863: 权限验证不足
CWE-264: 权限分配错误
CWE-266: 权限和特权管理错误

CWE-269: 错误的特权管理

CWE-276: 不当的默认权限

CWE-732: 不当的文件权限分配

表 4 访问验证错误致漏洞产生对应的 CWE

其他错误 (11 条, 配置错误 1 条, 竞争条件 9 条), 竞争条件 (race conditions)

漏洞常见于多线程程序, 由于系统没有正确同步对共享资源的访问, 多个执行流在几乎相同的时间内争用资源, 导致错误的结果或未授权的访问。。

CWE-16: 配置错误

CWE-362: 并发访问共享资源时缺少适当的锁定

CWE-367: 时间和状态中的竞争条件

CWE-364: 信号处理程序中使用可中断的操作

CWE-665: 不当的同步

CWE-667: 共享资源的过度使用

CWE-821: 信号处理机制中的竞争条件

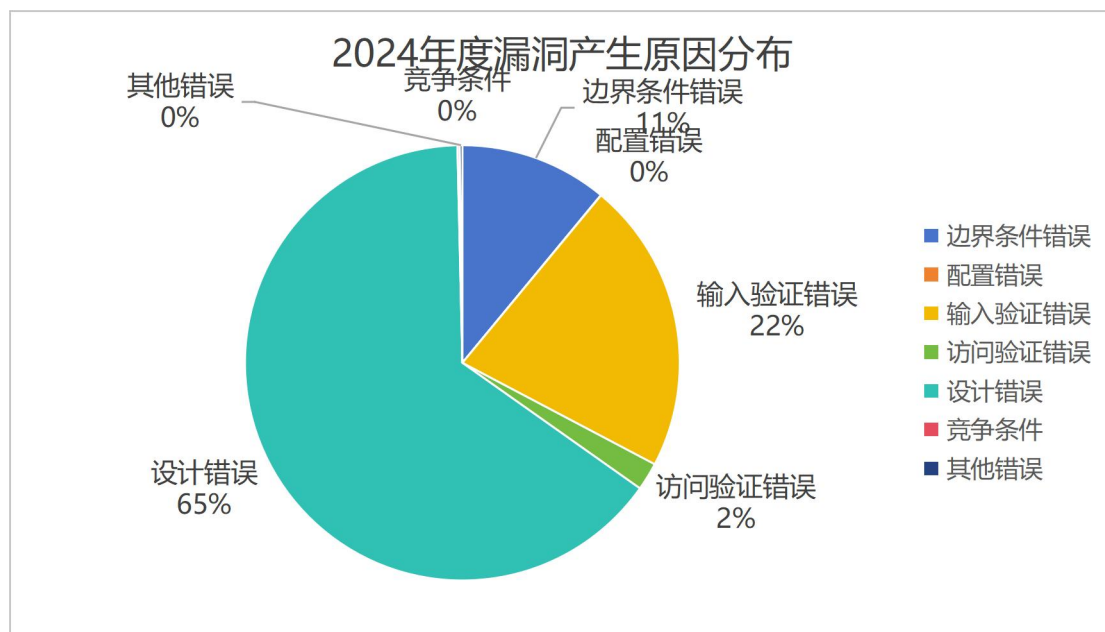
CWE-670: 总是被占用的资源

CWE-328: 使用竞争条件的非安全算法

CWE-667: 资源释放的竞争条件

CWE-755: 使用前的验证错误

表 5 其他错误致漏洞产生对应的 CWE



图表 3 2024 年度漏洞产生原因分布（注：数据来源 CNVD）

漏洞产生原因解析

首先从图中可以得出设计阶段问题占主导，设计错误占比高反映出系统在早期设计阶段缺乏全面考量，导致系统开发阶段未对安全问题进行重视出现较多漏洞。其次输入问题较多，开发人员在编码过程中对用户输入缺乏严格处理，而这类输入问题漏洞通常是攻击者的突破口。最后存在部分开发错误，此类开发错误引起的漏洞容易对多线程、高并发系统的稳定性和安全性造成严重影响。

3. 基于主流漏洞库已公开披露漏洞的趋势预测

随着软件系统复杂性和规模的增加，安全问题会更多地出现在设计阶段，设计错误的比例可能进一步上升。近年来，互联网的普及和 API、Web 服务的广泛应用，用户输入相关的攻击面不断扩大。输入验证类漏洞将依然是攻击者的主要目标，特别是在缺乏统一输入验证框架的系统中，输入验证类漏洞将会持续高发。云计算和分布式系统的高速发展，随之带来的问题也将愈发严重，例如高并发场

景中的竞争条件问题, 以及配置错误在自动化部署和微服务架构中也可能频繁出现。

4. 基于主流漏洞库已公开披露漏洞的厂商分析

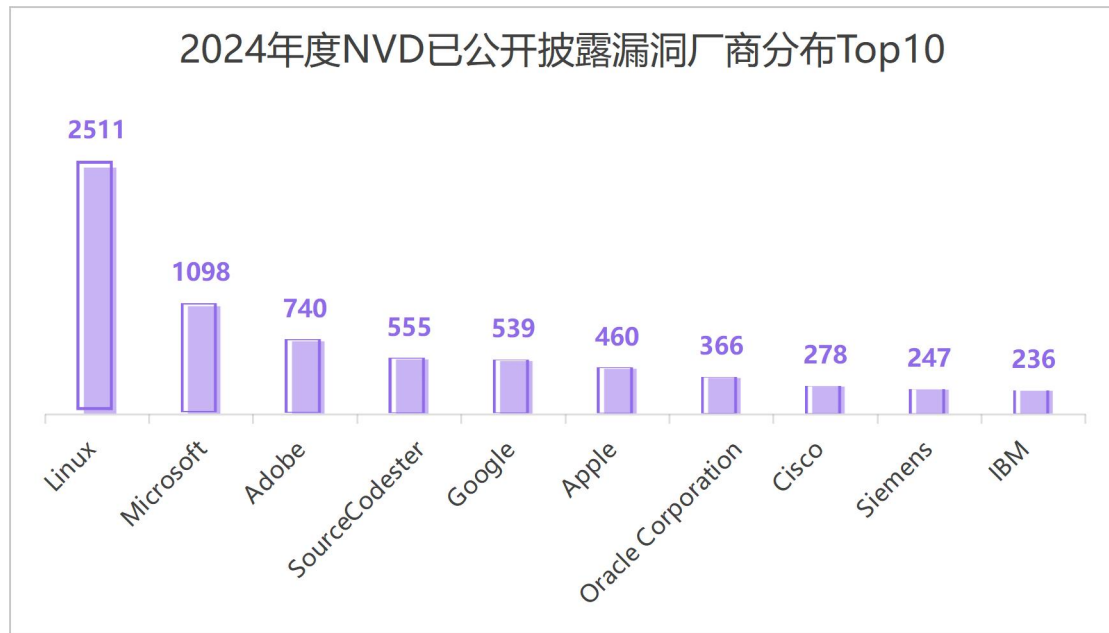
2024 年度漏洞数据的厂商分布概况

Linux 占据首位, Linux 生态系统在漏洞披露中处于显著位置, 可能与其广泛的开源应用和使用场景有关。Microsoft 排名第二, 漏洞数量为 1098 个, 这与其作为主流操作系统和企业软件供应商的市场地位密切相关。排名前两位的厂商漏洞数量远超其他厂商, 表明漏洞分布呈现明显的集中趋势。Adobe 位居第三, 漏洞数量为 740 个, 主要与其核心产品 (如 Acrobat、Photoshop) 相关。

SourceCodester 和 Google 分别以 555 个和 539 个漏洞排在第四和第五, 表明了开源项目和互联网巨头存在较大漏洞风险。Apple 和 Oracle Corporation 的漏洞数量分别为 460 个和 366 个, 位居中间, 表明其软件生态也存在安全隐患。Cisco 以 278 个漏洞位列第八, 体现其网络设备和安全解决方案的潜在风险。Siemens 和 IBM 分别以 247 个和 236 个漏洞排在第九和第十, 主要涉及工业控制系统和企业级解决方案的安全问题。

厂商涉及操作系统、开源项目、互联网服务、企业级解决方案、网络设备和工业控制系统, 说明漏洞分布覆盖广泛。随着软件和硬件生态的复杂化, 厂商需要加大对漏洞管理和修复的投入。加强对网络设备 (如 Cisco) 和工业控制系统 (如 Siemens) 的安全监测, 以防止潜在漏洞的利用。

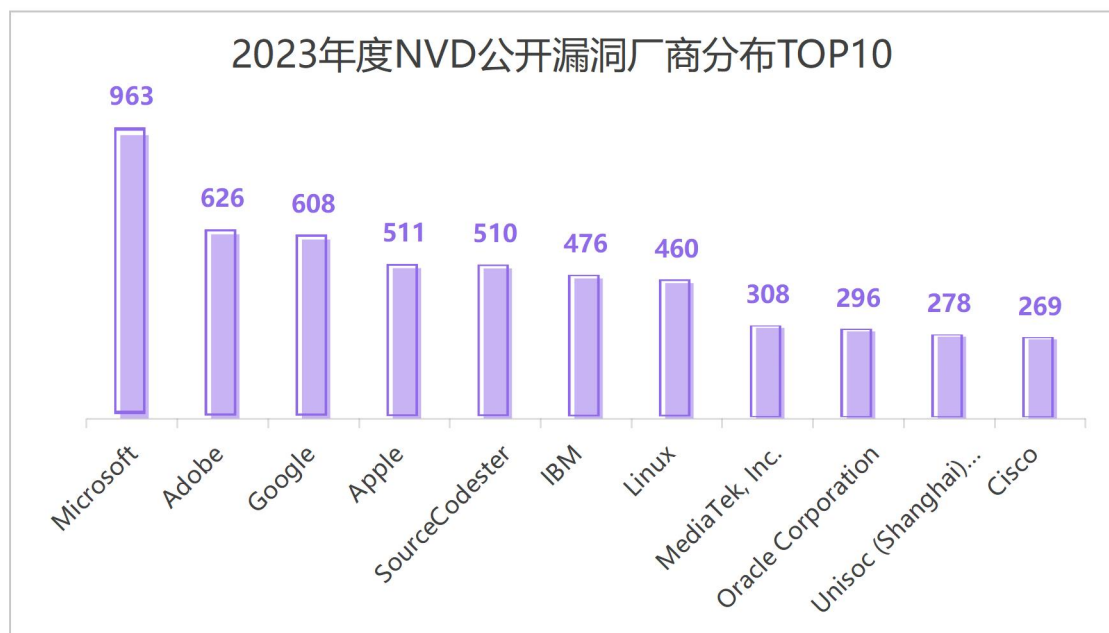
根据 NVD 已公开披露漏洞数据统计, 漏洞来源厂商排名 Top10 如下:



图表 4 2024 年度 NVD 已公开披露漏洞厂商分布 Top10

2024 年度厂商分布较 2023 年度数据解析

2024 年整体披露的漏洞数量较 2023 年显著增长，特别是 Linux 相关漏洞激增，可能与其广泛使用的生态有关。2024 年榜单中厂商覆盖从操作系统、互联网巨头到工业控制领域（如 Siemens），表明漏洞分布的行业覆盖面进一步扩大。



图表 5 2023 年度 NVD 公开漏洞厂商分布 Top10

Microsoft、Adobe、Google 等核心厂商在两年间持续占据榜单前列，反映出其复杂的产品生态长期面临安全挑战。工业控制系统厂商（Siemens）和开源项目（SourceCodester、Linux）在 2024 年的排名和数量提升。

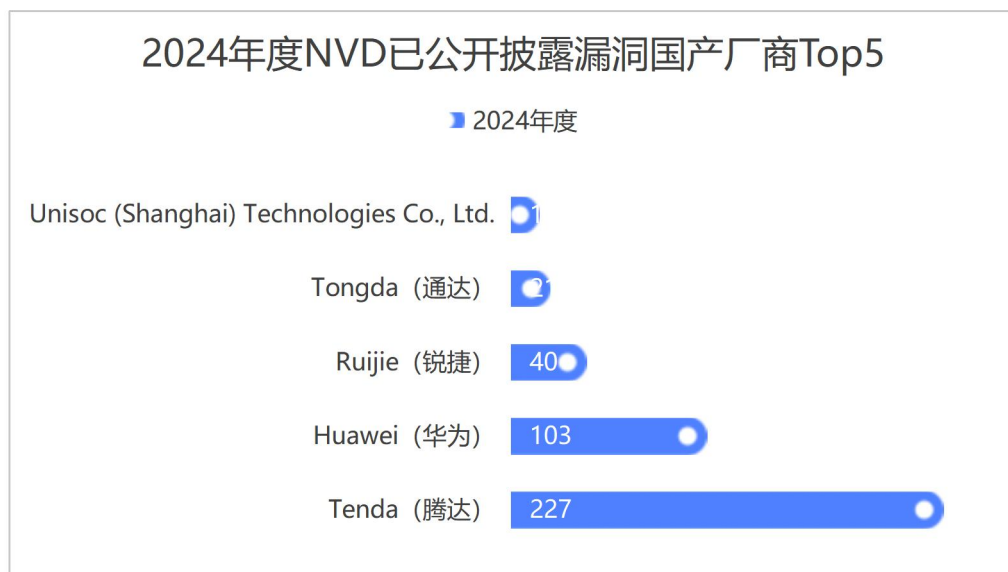
2024 年度厂商分布数据的分析总结

开源项目漏洞数量增长明显，应进一步加强对 Linux 等开源生态的安全研究和防护。工业控制系统厂商的进入反映出该领域潜在的安全威胁需引起重视。针对 Microsoft、Adobe 等厂商，应不断优化补丁管理和漏洞修复流程，以减少风险积累。

5. 基于国产厂商分布数据的解析

基于 NVD 已公开披露漏洞的数据，及国产厂商分布的数据解析中可以看出，通达的办公软件漏洞和紫光展锐的芯片漏洞尽管数量较少，但其影响力不可忽视，尤其是在企业办公场景和物联网应用中的潜在威胁。

2024 年度 NVD 已公开披露漏洞国产厂商 Top5 分布如下图所示：



图表 6 2024 年度 NVD 已公开披露漏洞国产厂商 Top5

基于上述 2024 年国产厂商的漏洞分布显示，消费级和企业级网络设备是安全问题的重灾区，而芯片和办公软件领域的漏洞也需得到更多重视。

(1) 消费级产品安全问题显著：腾达的高漏洞数量反映出小型路由器、交换机在 2024 年成为主要攻击目标。

(2) 企业级网络设备漏洞较多：锐捷和华为的高排名表明，企业级网络设备和通信产品在安全研究中占据重要地位，未来需加强企业级网络设备的防护措施，以防出现企业重大安全问题。

6. 本章小结

2024 年度的漏洞数据分析显示，全球信息安全威胁形势持续加剧，主流漏洞库记录了显著增长的漏洞数量，从漏洞的等级分布来看，中危和高危漏洞的增幅尤为显著，2024 年中危漏洞较 2023 年增长了 40.01%，高危漏洞增长了 20.28%。这些漏洞对网络安全带来的威胁不断加大，因此及时识别与修复中高危漏洞成为安全防护的重中之重。

在漏洞产生的原因方面，设计错误占据了绝大多数，达到约 60%。输入验证不足、边界条件错误、访问验证问题等也是常见的漏洞源。设计阶段的漏洞通常难以修复，影响深远，而输入验证类漏洞由于其较低的利用门槛，成为攻击者主要目标。

针对厂商分布情况，2024 年 Linux、Microsoft 和 Adobe 等厂商的漏洞数量居前，这表明主流操作系统和大型软件仍是漏洞的高发区域。此外，工业控制系统厂商的漏洞也需要引起重视。国产厂商方面，消费级产品和企业级网络设备的漏洞数量较多，需要着重关注这类厂商的安全公告发布。

三、2024 年度 CWE 排行榜解读

1. CWE Top25 简介

CWE Top 25（常见弱点枚举前 25 名）是由 CWE（Common Weakness Enumeration，常见弱点枚举）发布的一个年度报告，列出了在过去一年中最常见的、影响软件和系统的安全漏洞。CWE 是一个由 MITRE 组织管理的公共框架，用于识别和分类各种软件漏洞和弱点。CWE Top 25 排名是基于漏洞的严重性、曝光度和对安全影响的潜在风险等因素进行排序的。该排名通常包括软件开发中最常见的漏洞类型，这些漏洞被攻击者利用可能导致数据泄露、远程代码执行、服务拒绝等问题。了解 CWE Top 25 的排名可以帮助开发人员、运维人员和安全专家优先识别和修复软件中的高风险安全弱点。

CWE Top 25 的漏洞排名是基于漏洞在实际应用中的普遍存在程度、漏洞被利用后对系统的潜在影响，比如数据泄露、系统崩溃等、攻击者利用该漏洞的难易程度、修复该漏洞所需的技术难度、过往安全事件和漏洞报告中的数据，尤其是漏洞的曝光频率和漏洞利用案例，多维度出发，对漏洞进行评级排行。

2. 2024 年度 CWE Top25 排行榜单

CWE Top 25 的漏洞类型涉及多种攻击方式和漏洞利用手段，包括注入攻击（SQL 注入、命令注入）、认证与授权缺陷（身份验证不当、授权错误）、资源管理漏洞（缓冲区溢出、内存管理漏洞）等。随着技术的不断发展，新型漏洞（如代码注入、反序列化漏洞、敏感信息泄露）变得更加突出，而传统的漏洞（如缓冲区溢出）减少。

序号	ID	名称	较 2023 年相比
1	CWE-79	跨站点脚本	上升 1 位
2	CWE-787	越界写入	下降 1 位
3	CWE-89	SQL 注入	-
4	CWE-352	跨站请求伪造 (CSRF)	上升 5 位
5	CWE-22	路径遍历	上升 3 位
6	CWE-125	越界读取	上升 1 位
7	CWE-78	操作系统命令注入	下降 2 位
8	CWE-416	Use-After-Free	下降 4 位
9	CWE-862	缺少授权	上升 2 位
10	CWE-434	危险文件上传	-
11	CWE-94	代码注入	上升 12 位
12	CWE-20	输入验证不当	下降 6 位
13	CWE-77	命令注入	上升 3 位
14	CWE-287	身份验证不当	下降 1 位
15	CWE-269	权限管理不当	上升 7 位
16	CWE-502	反序列化	下降 1 位
17	CWE-200	敏感信息泄露	上升 13 位
18	CWE-863	授权错误	上升 6 位
19	CWE-918	服务器端请求伪造 (SSRF)	-
20	CWE-119	缓冲区溢出	下降 3 位
21	CWE-476	空指针解引用	下降 9 位

22	CWE-798	使用硬编码凭证	下降 4 位
23	CWE-190	整数溢出	下降 9 位
24	CWE-400	DDOS	上升 13 位
25	CWE-306	缺少关键功能的身份验证	下降 5 位

表 6 2024 年度 CWE Top25 排行榜单

总览分析

从 2023 年到 2024 年的排名变化来看，CWE 的前 25 大漏洞类型出现了不同的升降情况。总体来看，攻击面较大的漏洞类型有所上升，而那些技术实现较为过时或防范较为成熟的漏洞类型有所下降。根据提供的排名数据，各漏洞类型的主要变化分析如下：

上升排名的漏洞类型：

CWE-94 代码注入：上升 12 位，代码注入漏洞（命令注入、脚本注入等）在 2024 年成为更严重的问题。越来越多的应用程序存在不安全的代码执行或动态代码生成功能，攻击者能够利用这些漏洞进行远程代码执行攻击。

CWE-200 敏感信息泄露：上升 13 位，数据泄露方面的漏洞问题日益严重。企业对敏感数据的保护措施仍然存在不足，在涉及存储和传输敏感信息时，可能没有使用足够的加密和访问控制手段进行防护。

CWE-17 不受控制的资源消耗：上升 13 位，资源消耗的控制问题变得更加重要，攻击者可能利用不当的资源管理进行拒绝服务攻击（DoS）。

CWE-352 跨站请求伪造（CSRF）：上升 5 位，这一漏洞的上升可能与 Web 应用程序日益增多的跨域请求和不安全的认证机制有关。

CWE-869 授权错误：上升了 6 位，授权机制中的漏洞在 2024 年成为更为突出的安全问题，攻击者可以通过绕过授权机制获得未授权的访问权限。

下降排名的漏洞类型：

CWE-416 Use-After-Free：下降 4 位，随着开发人员越来越重视内存管理与资源清理，该漏洞的出现频率有所下降。

CWE-120 缓冲区溢出：下降 3 位，现代编译器和安全防护措施的进步（如栈保护、ASLR 等）相关，缓冲区溢出在现代应用中的威胁性有所降低。

CWE-787 越界写入：下降 1 位，随着编程语言和编译器的提升，越界写入漏洞的出现频率有所减少。

CWE-22 路径遍历：下降 3 位，开发人员在文件路径处理方面的安全性有了更好的提升。

CWE-400 不受控制的资源消耗：上升 13 位，拒绝服务（DoS）攻击的资源消耗问题变得更加突出。

未变动的漏洞类型：

CWE-89 SQL 注入和 CWE-434 危险文件上传等漏洞类型保持稳定，虽然这类漏洞依然存在，但随着开发人员对 SQL 注入防护技术（如 ORM 框架、参数化查询等）和上传文件的安全限制越来越重视，它们的排名变化不大。

3. 基于 2024 年度 CWE 数据的趋势分析

Web 应用程序中存在的不安全代码生成或缺乏输入验证，导致代码注入与执行漏洞有所增加。可以从 CWE-94（代码注入）的排名大幅提升得出，漏洞研究人员对于代码注入漏洞的挖掘力度有所加大。

从 CWE-200（敏感信息泄露）的排名大幅上升反映了当下的数据泄露问题。在当下大规模的数据存储和云计算环境中，数据泄露事件频发，数据保护措施不

足。

许多系统没有足够的授权验证措施，容易导致非授权访问的问题。正因这些身份认证和授权机制的薄弱，CWE-863（授权错误）排行提升。

CWE-400（不受控制的资源消耗）的上升可以表明拒绝服务攻击（DoS）问题仍然是网络安全中的一个重要问题，资源消耗类型的漏洞可能在互联网应用中日益增加。

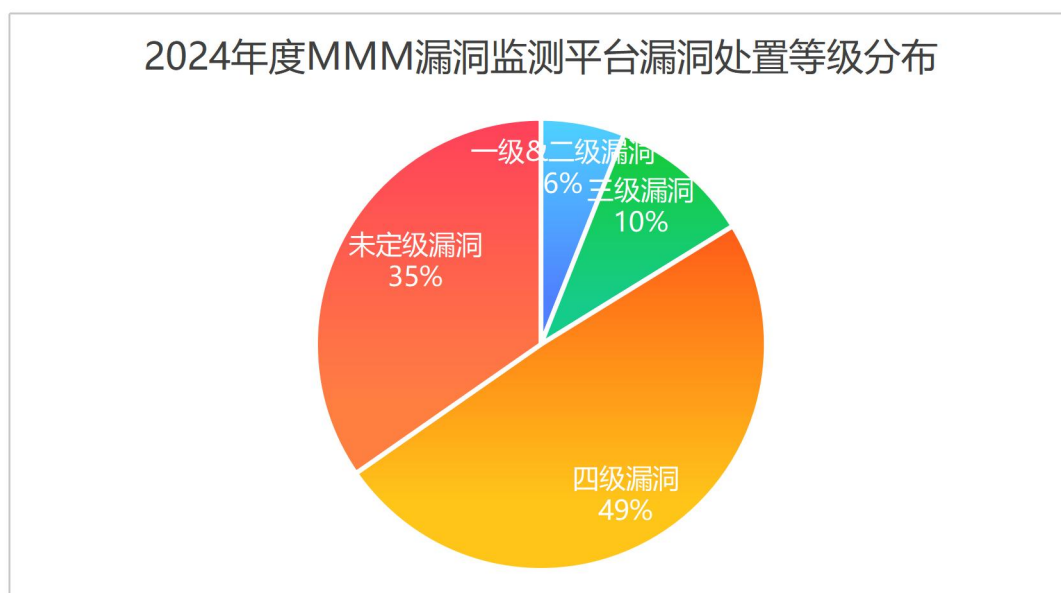
从 CWE-120（缓冲区溢出）和 CWE-416（Use-After-Free）等漏洞类型的排名下降不难看出，随着开发技术和编译器的改进，在针对传统漏洞的防护方面起到了积极作用。以及现代语言和框架提供了更强的防护措施，减少了这些漏洞被利用的可能性。

4. 本章小结

总的来说，网络安全的威胁正在从一些较为传统的漏洞转向更加复杂和难以发现的攻击方式。随着攻击技术的发展，跨站脚本、代码注入、敏感数据泄露等漏洞逐渐成为重点攻击目标。与此同时，现代开发技术的提升在一定程度上抑制了经典漏洞的上升趋势，但新型漏洞不断涌现，表明安全防护技术仍然需要不断创新与完善。不仅要加强对传统漏洞的防范，还要关注新兴漏洞，特别是资源消耗、身份认证等领域的安全问题，以提升整体的网络安全防护水平。

四、2024 年度漏洞预警回顾

MMM 漏洞情报监测平台是由安恒研究院自主研发并持续运营维护的专业漏洞平台。该平台涵盖了海量的漏洞数据，不仅包括各大权威漏洞数据库中的漏洞信息，还涵盖了无编号的在野漏洞数据。凭借其成熟的监测功能和完善的研判机制，平台有效支撑安恒 CERT 团队，能够在海量漏洞数据中迅速识别出严重且影响范围广泛的漏洞，并及时发布预警。同时，平台提供详细的漏洞信息和相应的解决方案，涵盖漏洞的各项特性（如利用状态、可用性等），为安全研究员提供充分的分析依据，帮助完善漏洞细节，确保漏洞处置的高效性与精准性。



图表 7 2024 年度 MMM 漏洞监测平台漏洞处置等级分布

2024 年，安恒 CERT 共监测并发现各类漏洞信息 39,226 条。经过 MMM 漏洞情报监测平台的智能研判和定级，其中 2,484 条漏洞被判定为处置等级较高的漏洞。经安全研究员进一步分析后，共发布了 103 条高危漏洞风险提示通告，涵盖漏洞 187 个，其中包括 80 条高危漏洞通告和 23 条严重漏洞通告。

1. 年度严重漏洞（CVSS3.1 评分 ≥ 9.0 ）

(1) GitLab 存在任意密码重置漏洞（CVE-2023-7028|DM-202312-003214）

用户帐户密码重置电子邮件可以发送到未经验证的电子邮件地址，攻击者可将重置帐户密码的邮件发送到未经验证的邮箱，在无需用户交互的情况下通过密码重置进行帐户接管。

(2) Atlassian Confluence Data Center and Server 存在远程代码执行漏洞(CVE-2023-22527|DM-202301-000034)

该漏洞允许未经身份验证的攻击者构造恶意请求，对影响版本实现远程代码执行。

(3) Jenkins 存在任意文件读取漏洞（CVE-2024-23897|DM-202401-002896）

未经身份验证的攻击者能够利用该漏洞读取 Jenkins 控制器文件系统上的任意文件。

(4) JetBrains TeamCity 存在身份验证绕过漏洞（CVE-2024-27198|DM-202402-002846）

该漏洞允许攻击者绕过身份验证执行管理操作，可能导致软件构建和部署方式的更改，或在构建中注入恶意代码，获得服务器管理控制权。

(5) FortiOS & FortiProxy 存在越界写入漏洞（CVE-2023-42789|DM-202309-001361）

FortiOS & FortiProxy 存在越界写入漏洞允许攻击者通过特制的 HTTP 请求执行未经授权的代码或命令。

(6) libzma / xz 库存在后门（CVE-2024-3094|DM-202403-002139）

xz 的上游 tarball 中发现了恶意代码，其存在的恶意代码可能允许对受影响

的系统进行未经授权的访问。liblzma 构建过程从源代码中存在的伪装测试文件中提取预构建的目标文件，使用该文件修改 liblzma 代码中的特定函数，产生一个修改后的 liblzma 库，任何链接到该库的软件都可以使用该库，拦截并修改与该库的数据交互。

(7) Rust 存在命令注入漏洞 (CVE-2024-24576|DM-202401-003634)

能够控制传递给派生进程的参数的攻击者可以绕过转义，传递恶意参数，导致任意的 shell 命令执行。

(8) Palo Alto Networks PAN-OS 存在命令注入漏洞 (CVE-2024-3400|DM-202404-001652)

未经身份验证的攻击者能够在防火墙上以 root 权限执行任意代码。

(9) Git 存在远程代码执行漏洞 (CVE-2024-32002|DM-202405-002232)

该漏洞允许攻击者在“克隆”操作期间对影响版本实现远程代码执行。

(10) PHP-CGI 存在远程代码执行漏洞 (CVE-2024-4577|DM-202405-001058)

由于 PHP-CGI 实现中输入验证不当，未经身份验证的攻击者可以通过特定的字符序列绕过 CVE-2012-1823 的防护，远程攻击者可以向应用程序发送特制的 HTTP 请求并在系统上执行任意操作系统命令。

(11) VMware vCenter Server 堆溢出漏洞 (CVE-2024-37079|DM-202406-000059)

具有 vCenter Server 网络访问权限的攻击者能够通过发送特制的数据包来触发该漏洞，从而可能导致远程代码执行。

(12) VMware vCenter Server 堆溢出漏洞 (CVE-2024-37080|DM-202406-000060)

具有 vCenter Server 网络访问权限的恶意参与者可能会通过发送精心编制的网络数据包来触发此漏洞，这可能会导致远程代码执行。

(13) GeoServer 存在远程代码执行漏洞 (CVE-2024-36401|DM-202405-004663)

由于不安全地将属性名称评估为 XPath 表达式，多个 OGC 请求参数允许未经身份验证的用户通过针对默认 GeoServer 安装的特制输入执行任意代码。

(14) GitLab 存在身份验证绕过漏洞 (CVE-2024-6385|DM-202406-003791)

该漏洞允许攻击者在某些情况下以其他用户的身份触发 pipeline。

(15) Windows 远程桌面授权服务远程代码执行漏洞 (CVE-2024-38077|DM-202406-001541)

该漏洞源于 Windows RDL 中（通常需要手动开启）的一个堆溢出问题。由于在解码用户输入的许可证密钥包时，未正确检验解码后数据长度与缓冲区大小之间的关系所导致的。攻击者可以通过精心构造的输入触发这个漏洞，并且在不需用户交互的情况下完全控制受害者的服务器。

(16) Windows TCP/IP 存在远程代码执行漏洞 (CVE-2024-38063|DM-202406-001527)

Windows TCP/IP 在处理 IPV6 数据包时存在问题，未经身份验证的攻击者可以通过发送特制的 IPV6 数据包实现远程代码执行。

(17) VMware vCenter Server 存在堆溢出漏洞 (CVE-2024-38812|DM-202406-002643)

vCenter Server 在 DCERPC 协议实施过程中存在堆溢出漏洞，具有 vCenter Server 网络访问权限的攻击者可以通过发送特制的网络数据包来触发此漏洞，可

能导致远程代码执行。

(18) Ivanti Endpoint Manager 存在远程代码执行漏洞 (CVE-2024-29847|DM-202409-001485)

Ivanti Endpoint Manager 存在远程代码执行漏洞，目前技术细节及 PoC 已公开，该漏洞允许受影响版本中的不受信任数据的反序列化允许远程未经身份验证的攻击者实现远程代码执行。

(19) Oracle WebLogic Server 存在反序列化漏洞 (CVE-2024-21216|DM-202410-003014)

未经身份验证的攻击者通过 T3/IIOP 进行网络访问来入侵 WebLogic 服务器，成功利用此漏洞后，攻击者可能执行任意代码从而完全控制服务器。

(20) Fortinet FortiManager 存在身份验证绕过漏洞 (CVE-2024-47575|DM-202409-002444)

FortiGate 与 FortiManager 协议 (FGFM) 中的关键功能缺少认证，攻击者通过注册未经授权的 FortiManager 或 FortiGate 设备获取系统配置数据，包含设备的 IP、凭据等敏感信息。这可能导致远程攻击者通过发送特制的请求从而执行任意代码或命令。

(21) Apache Solr 存在身份验证绕过漏洞 (CVE-2024-45216|DM-202408-003847)

该漏洞主要影响使用 PKIAuthenticationPlugin（通常在启用 Solr 身份验证时默认激活）的 Solr 实例。攻击者可以在任意 Solr API 路径的末尾添加伪造结尾，绕过身份验证，同时保持与原始 API 路径的契合。这一漏洞允许未经授权的用户获取敏感数据或进行修改操作，对系统安全构成较大风险。

(22) CyberPanel 存在远程命令执行漏洞 (CVE-2024-51567|DM-202410-0053

83)

未经身份验证的远程攻击者可以通过对缺乏充分验证和过滤的 upgrademysqlstatus 接口参数进行利用，从而绕过身份验证并通过构造恶意请求进行命令注入，执行任意命令。

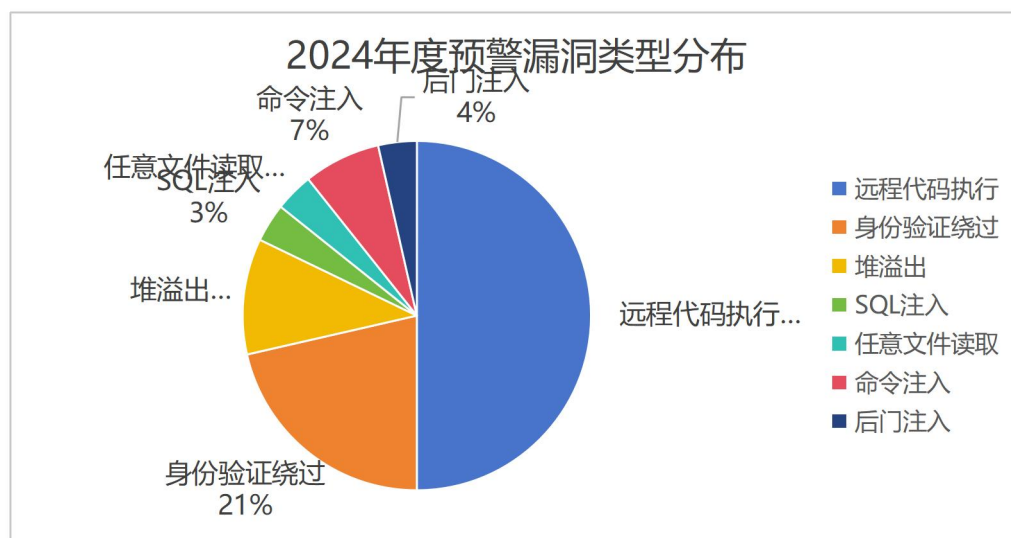
(23) Ivanti Endpoint Manager 存在 SQL 注入漏洞 (CVE-2024-50330|DM-202410-50330)

Ivanti Endpoint Manager 存在 SQL 注入漏洞，未经身份验证的攻击者可以利用该漏洞实现远程代码执行。

2. 基于 2024 年度漏洞预警数据的分析

基于 2024 年度漏洞预警数据的漏洞类型统计

根据对 2024 年度安恒 CERT 预警漏洞的统计分析，受影响的产品范围广泛，涵盖了常见的开发工具（如 Git、Jenkins 等）、企业关键业务系统（如 Fortinet、Ivanti、VMware vCenter Server 等）以及服务器平台（如 Apache Solr、PHP-CGI）



图表 8 2024 年度预警漏洞类型分布

这些漏洞大多数利用门槛较低，无需身份验证，且影响范围广泛，严重威胁系统安全。其中约 90% 的漏洞允许远程执行代码或绕过身份验证，攻击者能够通过绕过身份验证机制获得未授权的访问权限，进而执行恶意操作，且可能利用该类漏洞与其他漏洞进行组合利用，进一步扩大影响范围。其他漏洞类型，如敏感信息泄露、SQL 注入、命令注入、缓冲区溢出和任意文件读取等漏洞所占的比例较低，但这些漏洞不容忽视。敏感信息泄露漏洞可能导致用户隐私、金融信息以及企业机密数据的泄露；SQL 注入漏洞可以被利用获取数据库中的敏感数据，威胁数据安全；命令注入漏洞可能让攻击者在系统中执行恶意命令，甚至导致系统崩溃；缓冲区溢出漏洞能够导致内存破坏，进一步引发系统崩溃或恶意代码执行；任意文件读取漏洞则使攻击者可以访问服务器上的敏感配置文件、日志文件等，泄露系统信息。

基于 2024 年度漏洞预警数据的趋势预测

通过对 2024 年度安恒 CERT 预警漏洞数据的分析，可以得出涉及云端和虚拟化的漏洞逐渐增多，例如，VMware vCenter Server、Fortinet FortiManager 等，云端和虚拟化技术逐渐成为攻击热点。其次是开源组件和第三方库的风险渐增，例如，在本年度 xz 库、libzma 后门等相关漏洞的披露即可说明开源生态可能面临恶意代码植入和供应链攻击的挑战。

基于漏洞预警数据中的高发漏洞类型

远程代码执行 (RCE) 漏洞以及命令注入漏洞在 2024 年度漏洞预警数据中占比高达 57%，例如 Palo Alto Networks PAN-OS 存在命令注入漏洞 (CVE-2024-3400)

和 PHP-CGI 存在远程代码执行漏洞（CVE-2024-4577）等漏洞，这类漏洞攻击门槛低且可能导致完全失控。

身份验证绕过漏洞在 2024 年度漏洞预警数据中占比 21%，例如 GitLab 存在身份验证绕过漏洞（CVE-2024-6385），重点关注对关键业务系统（例如 FortiManager、Apache Solr）的保护，通常情况下攻击者对这类型漏洞进行利用的时候会结合其他漏洞进行组合利用，进一步扩大影响范围。

后门和供应链漏洞在 2024 年度漏洞预警数据中占比不高，但这类型的漏洞存在隐蔽性高且波及面广的特征，因此需要增加对此类漏洞的关注，并加强供应链安全检测，本年度漏洞预警中曾对此类漏洞进行预警，例如 libzma / xz 库存在后门（CVE-2024-3094）。

3. 本章小结

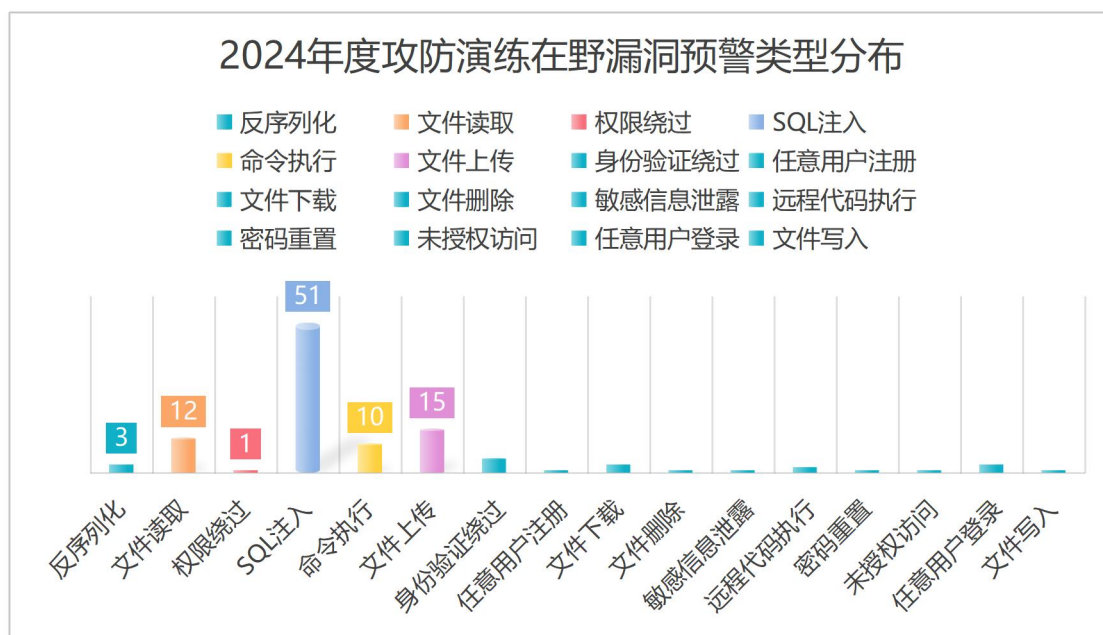
近年来漏洞趋势表现为影响范围扩大、攻击复杂度提升，尤其是远程代码执行和身份验证绕过漏洞需特别重视。建议加强系统更新与安全策略的制定，及时跟踪厂商安全公告以及应用补丁，从而有效应对潜在威胁。因面对这些潜在的安全威胁，企业和组织应持续强化安全防护措施，加强漏洞修补和安全审计，确保系统安全并保护敏感数据免受攻击。

五、2024 年攻防演练高危漏洞回顾

根据 2024 年度攻防演练期间安恒信息 CERT 监测到的漏洞数据,安恒研究院第一时间对相关漏洞进行研判,并针对高危和高频漏洞及时发布相关预警信息。并对演练期间捕获的在野 0day 漏洞及时将其档案添加至漏洞库并完善策略。现对这些漏洞进行分类和统计分析,以便更准确地评估实际攻击场景中最易被利用的漏洞类型,为安全防护提供数据支持。

1. 攻防演练期间常见漏洞类型分布概况

2024 年攻防演练期间,安恒信息 CERT 对在野漏洞进行了全面监测,新增在野漏洞档案共计 227 条,其中一级漏洞档案 114 条,二级漏洞档案 111 条。在本次演练期间捕获的在野 0day 漏洞中,SQL 注入漏洞成为攻击者最常使用的漏洞类型,占新增在野漏洞档案的 45.94%。经研判后,累计对 111 个漏洞发布漏洞预警,其漏洞类型分布如下:



图表 9 2024 年度攻防演练在野漏洞预警类型分布

SQL 注入漏洞，通过将恶意 SQL 语句注入到系统的数据库查询中，攻击者可以操控数据库，实现任意数据操作，包括读取、修改、删除数据，甚至完全控制数据库。根据数据统计，SQL 注入漏洞在监测到的漏洞中占据了较大比例，涉及产品包括：

XX 图书馆集群管理系统
XX 报表
XX ERP
XX OA
XX 管理系统
XX 企业管理系统
XX 人力资源信息管理系统等

远程命令执行漏洞，攻击者在受害系统上执行任意命令，通常用于控制远程主机，安装恶意软件或窃取数据。涉及该类漏洞的系统包括：

XX 财务系统
XX CMS
XX 业务协作平台
XX 堡垒机等
XX 网关

任意文件上传漏洞，通常允许攻击者上传恶意文件到系统中，这些文件可能包含 Web Shell 或恶意脚本。任意文件读取漏洞则允许攻击者访问本不应访问的文件，泄露敏感信息。涉及这类漏洞的系统包括：

XX 管理系统
XX CRM

XX 企业管理系统

XX 云平台

XX 视频平台等

身份验证绕过漏洞允许攻击者绕过正常的身份验证机制,直接获取用户权限。

这类漏洞常被用于提升攻击权限或获得未授权访问。涉及产品包括:

XX 协同办公管理平台

XX 企业管理系统

XX 企业级低代码平台等

反序列化漏洞发生在应用程序处理不受信任的数据时,攻击者可以通过特制的数据破坏程序的正常流程,执行任意代码。涉及产品包括:

XX 云财务系统等

命令注入漏洞允许攻击者将恶意命令注入到系统的命令行中,执行任意操作,通常用于执行恶意脚本或控制远程主机。涉及产品包括:

XX 多业务智能网关

XX 系统等

2. 攻防演练中常见漏洞利用类型

在攻防演练中,攻击者常常利用一些特定的漏洞类型来突破系统防线,并造成严重的安全威胁。经过上述的漏洞数据统计与分析中,得出了攻防演练中常见的几种漏洞利用类型,以及这些漏洞在实际攻击中的危害和影响。这些漏洞不仅是安全攻防的重点关注对象,也是漏洞修复和安全防御策略的重要参考。

(1) SQL 注入漏洞是典型的 Web 应用漏洞,在攻防演练中,攻击者经常通过 SQL 注入漏洞获取敏感数据、操控数据库,甚至执行恶意命令,它是最容易被利

用的漏洞之一。

(2) 远程命令执行漏洞危害极大，攻击者可以完全控制目标设备，执行任意操作，传播恶意软件或进行更广泛的攻击。这类漏洞在攻防演练中频繁出现，特别是在针对关键网络设备或基础设施进行渗透测试时。

(3) 身份验证绕过漏洞常用于获取管理员权限或访问特定受限资源。在攻防演练中，利用这类漏洞，攻击者可以直接获取系统权限，绕过安全防护。

(4) 命令注入漏洞允许攻击者绕过系统安全，执行系统命令，造成系统崩溃或远程控制。此类漏洞在一些嵌入式设备和管理平台中较为常见，通常被用于扩大攻击范围。

(5) 任意文件上传/读取漏洞常见于 Web 应用，攻击者通过上传恶意文件或读取敏感文件，可获取系统内部的信息或执行攻击。在攻防演练中，攻击者通常利用这类漏洞与其他漏洞组合利用，对系统进行入侵或进一步攻击。

(6) 反序列化漏洞，攻击者能够直接控制应用程序并执行恶意代码。这类漏洞。基于 Java 或 .NET 等技术栈的应用中较为常见，常被用于入侵和执行远程命令。

基于以上的漏洞分类，SQL 注入漏洞和远程命令执行漏洞在攻防演练期间最为常见且危害性最大。由于 Web 应用的普遍性，SQL 注入漏洞成为攻击者攻击的重点。攻击者通常通过注入恶意 SQL 代码来获取系统权限、窃取敏感信息或控制数据库。远程命令执行漏洞具有极大的破坏力，因为攻击者可以在受害主机上执行任意命令，完全控制目标设备。不仅如此，这类漏洞的利用可以让攻击者突破网络防护，进一步渗透到其他系统或网络中，扩大攻击规模。

3. 本章小结

在 2024 年度攻防演练期间，最常被利用的漏洞主要集中在 SQL 注入、远程命令执行和任意文件上传/读取等类型。Web 应用和网络设备，尤其是存在 SQL 注入和命令执行漏洞的系统，容易成为攻击者的攻击目标。因此，企业和组织应加强这些漏洞的防护措施，及时修补系统中的安全漏洞，并定期进行渗透测试和安全演练，以提升网络安全防护能力，降低潜在风险。

六、AI 安全隐患与未来趋势分析

1. OWASP Top10 简介

开放式 Web 应用安全项目 (OWASP) 是一个以 Web 应用安全为核心使命的国际非营利组织。OWASP 通过文档、工具、视频、会议和论坛提供一系列免费信息, 致力于帮助其他组织提高 Web 应用安全性。2023 年, OWASP 工作组启动了一个新项目, 创建一份重点关注大型语言模型 (LLM) 应用威胁的类似报告。OWASP 大型语言模型应用十大威胁报告确定了相关威胁, 提供了漏洞和实际攻击场景的示例, 报告重点介绍了安全专家指出的有关应用安全的 10 大关键风险。

2. OWASP LLM Top10 安全威胁解读

以下是 OWASP LLM Top10 的排名和标题:

LLM01: 提示词注入 (Prompt Injection)
LLM02: 不安全的输出处理 (Insecure Output Handling)
LLM03: 训练数据中毒 (Training Data Poisoning)
LLM04: 拒绝服务模型 (Model Denial of Service)
LLM05: 供应链漏洞 (Supply Chain Vulnerabilities)
LLM06: 敏感信息泄露 (Sensitive Information Disclosure)
LLM07: 不安全的插件设计 (Insecure Plugin Design)
LLM08: 过度代理 (Excessive Agency)
LLM09: 过度依赖 (Overreliance)
LLM10: 模型盗窃 (Model Theft)

表 7 OWASP Top 排行

1. LLM01: 提示词注入 (Prompt Injection)

黑客通过设计过的输入（提示词）操纵大型语言模型（LLM），提示词注入会覆盖系统提示词，而间接注入操纵外部数据源进行注入攻击，从而导致 LLM 执行意外操作。例如，攻击者向基于 LLM 的支持聊天机器人注入包含“忘记所有先前指令”和新指令，查询私有数据存储并利用后端函数的包漏洞和缺乏输出验证发送电子邮件，从而导致远程代码执行，获得未授权访问和权限升级。

2. LLM02: 不安全的输出处理 (Insecure Output Handling)

当 LLM 输出未经审查而被接受从而暴露后端系统，滥用可能会导致 XSS、SQL 注入、SSRF、权限提升或远程代码执行等严重后果。例如，某大型语言模型允许用户通过类似聊天的功能为后端数据库制定 SQL 查询。一个用户请求删除所有数据库表的查询。如果来自大型语言模型的查询语句没有受到输出处理，则所有数据库表将被删除。

3. LLM03: 训练数据中毒 (Training Data Poisoning)

训练数据中毒攻击是指攻击者故意修改或注入有害数据到 AI 模型的训练数据集中，从而破坏模型的性能或操控其行为。这种攻击方式不仅可能降低模型的准确性，还可能导致模型产生偏差，甚至被引导执行恶意行为。例如，训练数据没有经过正确过滤和处理，应用程序的恶意用户可能会尝试向模型注入有毒数据，以使其适应并输出有偏见和虚假的数据。

4. LLM04: 拒绝服务模型 (Model Denial of Service)

通过过度请求、滥用 API 或其他方式导致模型的资源耗尽或服务不可用。这类攻击可以使得 AI 模型在正常使用时无法响应用户请求，甚至直接使系统崩溃或瘫痪。例如，攻击者向 LLM 洪水般发送大量的超长输入，经过精心制作从而

以接近或达到上下文窗口的限制。通过用不同长度的输入淹没 LLM，攻击者旨在利用处理变长输入的任何低效之处。这些输入会对 LLM 的资源造成过多负担，可能导致性能下降，妨碍系统响应合法请求。

5. LLM05：供应链漏洞 (Supply Chain Vulnerabilities)

在生成式 AI 模型的开发、训练、部署和维护过程中，攻击者通过破坏或篡改软件、硬件、数据集、第三方库或服务，导致 AI 系统的安全性受到威胁。通常发生在 AI 模型的开发和部署流程中的某个环节，攻击者通过恶意修改或操控关键组件，从而影响整个模型的正常运行和安全性。例如，攻击者利用易受攻击的 Python 库来入侵系统，这在第一次 Open AI 数据泄露中发生过。

6. LLM06：敏感信息泄露 (Sensitive Information Disclosure)

由于生成式 AI 模型通常在大规模数据集上进行训练，某些模型可能会从训练数据中“记住”敏感信息，并在生成的输出中不小心暴露出来。LLM 可能会在其回复中泄露机密数据指生成式 AI 模型在生成内容或响应用户查询时，无意间泄露了本应保密的敏感信息等，从而导致未经授权的数据访问、隐私侵犯和安全漏洞。例如，毫不知情的合法用户 A 在与 LLM 应用程序进行非恶意交互时，通过 LLM 向其显示了某些其他用户的数据。

7. LLM07：不安全的插件设计 (Insecure Plugin Design)

LLM 插件可能具有不安全的输入和不足的访问控制，这些插件可能与主模型交互，扩展系统的功能，但如果设计不当，便可能会成为攻击者的入口，危及整个系统的安全。并可能导致远程代码执行等后果。例如，一个插件接受 SQL WHERE 子句作为高级过滤器，然后将其附加到过滤 SQL 中，攻击者将可以对其进行 SQL 注入。

8. LLM08: 过度代理 (Excessive Agency)

过度代理可能增加数据传输的复杂性,降低安全性,并增加潜在的漏洞暴露机会。在生成式 AI 系统中,代理层(如中介服务、API 层、模型调用层等)过多,导致系统过于复杂且安全性降低。系统可能在多层代理的过程中暴露更多的攻击面,攻击者能够利用这些额外的代理层进行操控或窃取数据。例如,一个基于 LLM 的个人助手应用程序通过插件被授予访问个人邮箱的权限,以便总结收件箱中的邮件内容。为实现此功能,邮件插件需要具备读取邮件的能力,但系统开发人员选择使用的插件还包含发送邮件的功能。LLM 可能受到间接提示注入攻击,构造恶意的入站邮件欺骗 LLM,使其命令邮件插件调用“发送邮件”功能,从用户的邮箱发送垃圾邮件。

9. LLM09: 过度依赖 (Overreliance)

生成式 AI 系统过度依赖某些组件、模型、外部服务或第三方资源,这种依赖可能导致系统在面对故障或攻击时变得脆弱。若过度依赖单一组件,可能在该组件发生故障、被攻击或被滥用时,整个系统的安全性和稳定性都会受到影响。尤其是在集成多个 AI 模型、API 或外部服务时,任何一个环节的漏洞都将可能影响整个系统的运行。例如,某软件开发公司使用 LLM 来协助开发人员,LLM 建议一个不存在的代码库或包,其开发人员信任 AI,无意将一个恶意包集成到公司的软件中,这凸显了交叉检查 LLM 的重要性,特别是涉及第三方代码或库时。

10. LLM10: 模型盗窃 (Model Theft)

未经授权的第三方通过逆向工程、数据抓取或其他恶意手段窃取训练好的 AI 模型及其知识产权的行为。由于生成式 AI 模型通常包含大量的训练数据、算法设计和专业知识,模型盗窃不仅是对技术的非法获取,还可能涉及侵犯知识产

权、商业机密泄露以及算法性能的滥用。例如，恶意攻击者绕过 LLM 的输入过滤技术和前文，执行侧信道攻击，将模型信息提取到受其控制的远程资源中。

3. LLM 内生安全漏洞列表

近年来，人工智能技术以飞快的步伐不断进步，并在各个行业中发挥着越来越重要的作用。大规模模型作为 AI 领域的关键技术之一，随着计算平台的算力提升、海量数据的积累以及深度学习算法的不断创新，展现出了更强的能力，并逐步在一些专业领域取得了显著的突破。与此同时，以大模型为基础的各种技术应用的兴起，也为计算机安全领域带来了诸多新的风险和挑战，模型安全性也成为了人工智能安全的关键构成部分。

在此将对本年度较为高危的 LLM 漏洞进行举例，详情如下：

CVE 漏洞编号	漏洞影响主体	漏洞描述
CVE-2024-42478	llama.cpp	llama.cpp 是一个开源软件库，可对 Llama 等各种大型语言模型进行推理。在 rpc_tensor 结构中，存在不安全的 data 指针成员，可能导致任意地址读取。
CVE-2024-42477	llama.cpp	llama.cpp 是一个开源 软件库，可对 Llama 等各种大型语言模型进行推理。在 rpc_tensor 结构中，不安全的 type 成员可能导致全局缓冲区溢出（global-buffer-overflow），可能导致内存数据泄露。
CVE-2024-32878	llama.cpp	Llama.cpp 是 C/C++ 中的 LLM 推理。gguf_init_from_file 中存在使用未初始化堆变量的漏洞，代码稍后将释放此未初始化的变量。在一个简单的 POC 中，它将直接导致崩溃。如

		果该文件是精心构建的，则可能会控制此未初始化的值并导致任意地址空闲问题。这可能会进一步导致被利用。导致 llama.cpp 崩溃 (DoS)，甚至可能导致任意代码执行 (RCE)。
CVE-2024-43598	LightGBM	篡改未知值可能会导致利用内存损坏的攻击。
CVE-2024-45758	H2O.ai	H2O 是 H2O.ai 开源的一个用于分布式、可扩展机器学习的内存平台。H2O3.46.0.4 及之前版本存在安全漏洞，该漏洞源于攻击者可以任意设置 JDBC URL，这可能导致反序列化攻击、文件读取和命令执行。
CVE-2024-48206	Chainer	Chainer 是一个基于 Python 的深度学习框架，ChainerMN 是 Chainer 深度学习框架的一个扩展，用于支持并行训练模型。 chainer.chainermn.communicators._communication_utility.py 文件下的 chunked_bcast_obj 方法，使用了 pickle.loads 反序列化数据，pickle 在反序列化时可以执行任意代码造成危害。

表 8 LLM 相关高危漏洞

上述多个框架中都存在着严重的安全问题，尤其是在内存操作和不安全反序列化方面，这些漏洞可能导致远程代码执行或数据泄露等危害，还可能间接导致社会、经济等方面的严重问题。随着模型与日常工作生活的结合，我们更应着重关注怎样去保护模型的完整性、保密性、可用性，同时防止模型遭到恶意利用或者攻击，从而在机器学习与人工智能模型于设计、训练、部署以及运行的整个过程中，有能力抵御各类安全威胁与攻击。

4. LLM 应用过程中面临的风险场景

在大语言模型（LLM）应用过程中，存在多种潜在的风险场景，这些风险可

能涉及安全性、隐私、道德规范以及模型性能等方面。以下是一些典型的风险场景：

(1) 大模型敏感信息泄露 (LLM Sensitive Information Disclosure)

大模型在许多实际应用场景中往往需要处理大量敏感数据和个人隐私信息，这使得隐私泄露和数据滥用的风险提高，例如用户的行为记录、社交互动、医疗数据和金融交易等数据被泄露或非法利用，将会严重侵害个人隐私权益，甚至被用于恶意行为，如身份盗用、金融诈骗以及社会工程攻击。以及在用户与 LLM 交互的过程中，用户可能会无意中输入敏感数据，这些数据随后可能会被 LLM 在其他地方的输出中返回。这可能导致未经授权访问敏感数据、知识产权、侵犯隐私和其他安全漏洞。

(2) 大模型训练数据中毒 (LLM Training Data Poisoning)

训练数据中毒是指通过操纵预训练数据或微调或嵌入过程中涉及的数据，以引入漏洞、后门或偏见，以损害模型的安全性、有效性或道德行为，从而导致大模型可能就此输出虚假信息，这将造成性能下降、下游软件被利用和声誉受损等风险。即使用户不信任有问题的人工智能输出，风险仍然存在，包括模型能力受损和品牌声誉受损。因此，篡改训练数据也被看做一种攻击，因为该行为将会影响模型正确预测的能力，从而输出虚假信息。在人工智能的加持下，大模型的强大功能也可能成为恶意攻击的工具，对抗性样本将变得更加隐蔽，攻击者可能利用自动化工具生成高级攻击模式，规避现有的检测和防御机制，通过生成对抗样本，攻击者对输入数据进行微调，诱使模型产生错误判断，从而被用于制造虚假信息、误导重要决策。

(3) 大模型提示词注入 (LLM Prompt Injection)

通过恶意构造提示词，攻击者通过覆盖泄露底层系统的提示，从而突破人工智能模型的安全限制或行为约束，以使其执行原本被禁止或限制的任务或操作，达到大模型“越狱”的效果。在这种情况下，LLM 实际上充当了攻击者的代理，在不触发常规防护措施或向最终用户发出入侵警报的情况下进一步实现他们的目标。

(4) 模型盗窃 (Model Theft)

大模型自身也可能成为攻击目标，其可能导致大模型的核心数据泄露。攻击者可能试图窃取模型参数和权重，这不仅会导致企业核心知识产权丧失，还可能使攻击者利用被盗模型执行不当行为。随着模型体量增大，模型窃取与滥用风险增加，被盗模型可能用于更多非法用途。

随着大语言模型的广泛应用，风险管理显得尤为重要。企业和开发者需要采取多重防护机制，确保 LLM 的使用符合标准，并在确保安全性和可靠性的基础上发挥其作用，这包括加强模型的安全性、增强数据隐私保护等。

5. LLM 安全治理框架建设

大模型和传统信息系统一样，在提供服务的过程中都面临着漏洞攻击、未授权访问、DDOS 等传统网络安全风险的威胁。在此基础上，由于大模型自身强大的内容生成能力，导致相关应用在内容安全上面临着更多的考验。在输入上需要注意提示词注入等大模型应用特有的攻击方式，在输出上要防范隐私泄露、道德伦理争议、违反法律法规等内容的出现。因此需要对模型做好内容安全评估，并加强对服务过程中输入和输出内容的检测，提升大模型应用的稳定性、安全性，更好的为用户提供高质量的服务。

(1) 供应链安全防护

大模型构建过程中涉及到大量开源数据及开源模型的使用，同样面临供应链攻击威胁，其中投毒攻击是大模型训练阶段面临的重要风险。对于此类攻击，应该优先从官方地址或已经被证明安全的来源下载模型。降低模型被篡改的风险，并减少恶意代码的植入可能性。其次，通过在模型传输和部署的过程中添加水印和签名，确保模型不会在传输过程中被恶意篡改，减少传输过程中被投毒的风险，验证模型的真实性和完整性。安全人员应不断更新对框架安全性的研究，确保对于框架特性的认知全面性。并对模型用到的相关代码进行审计，检查是否存在后门或恶意代码。对潜在的安全问题，进行相应的修复和改进。并使用专门的工具对开源模型相关组件进行漏洞扫描。通过检查组件中的漏洞，可以及早发现潜在的供应链安全隐患，并采取相应的修复和加固措施。在模型使用过程中，进行模型的运行行为检测，监控模型在进行预测时的活动。对于来源无法证明安全的模型，建议在沙盒环境中运行相关代码。沙盒环境提供了隔离和安全保护，可以防止恶意代码对系统的影响。

(2) 数据安全防护

大模型技术在模型训练、业务数据传输处理等过程中涉及大量的个人信息、公共数据，对数据采集、传输、存储、处理的全生命周期中的隐私保护与数据安全治理防护提出更高的要求与挑战。

从数据的采集阶段分析：大模型所需的训练和业务处理数据在采集阶段均需要保障采集环境的可信，可通过后续章节介绍的零信任身份鉴别体系完成人员、环境的可信接入。接入环境鉴别与鉴权完成后，针对采集的数据可以通过敏感数据识别工作，进行数据的分类分级，参考个人信息保护、公共数据防护的要求，

对数据中的敏感信息进行识别与分级，为后续存储、传输、处理等环节中的防护提供依据

从数据的传输阶段分析：数据在传输过程中，关注传输通道的可信。利用风险监测、通道加密、数字证书或其他访问控制策略，针对敏感数据参与大模型训练制定相关标准和安全策略。

从数据的存储阶段分析：大模型涉及的业务数据量大，敏感数据经过识别后需要进行针对性的存储加密，防止业务数据直接批量泄露。数据加密过程还需考虑符合商用密码能力，通过支持国密加密算法实现加密能力的可靠可控。

从数据的处理阶段分析：数据处理主要分为业务数据流转和运维运营人员的数据运维通道日常运维访问。业务数据的处理阶段需要构建数据流转的监测能力，通过综合应用数据库审计、API 接口监测能力的监测能力并融合关联分析能力，完整的刻画敏感数据的全链路流转，针对异常的业务访问调用进行告警和阻断。针对运维通道，除了使用传统的堡垒机、零信任机制对于运维人员接入应用和操作系统的访问通道进行身份识别和权限管控，针对数据库等数据资产还需要实现字段级的访问权限控制，与精细化的命令识别与拦截，对于单次请求的返回数据数量进行定义和控制等功能，防止敏感数据通过运维人员批量泄露。同时可以通过动态脱敏能力对于敏感数据的运维访问通路进行防护。

(3) 内容安全防护

大模型内容安全，可通过构建大模型输入检测引擎和输出检测引擎的方式实现，输入检测引擎主要实现对大模型提示词的安全检测，针对不安全提示词在输入层面进行拦截，不面向大模型进行提问。输出检测引擎主要实现对大模型输出内容的安全检测，实现违法信息过滤、不良信息过滤、恶意代码识别过滤、隐私

信息拦截等内容安全防护功能。

大模型提示词安全，需要对提示词进行评估，综合考虑模型、数据、提示词和应用场景结合度等多个方面。目前提示词注入主要集中在对抗性攻击后缀（Adversarial Attacks）和提示词越狱（Jailbreak）上。通常，这类问题可以通过 NLP 完整语义分析，防止生成结果和预期完全不符的输出，限制单个提示词和整体提示词长度，避免注入过多信息等方式来预防。

不安全内容输出风险需要结合应用的实际环境、上下游应用等多个环节进行综合考虑，建立有效的违规内容识别和过滤机制，对模型生成的内容进行监控和审核，及时发现和处理违规行为，确保模型生成的内容符合规定和法律要求。。

(4) 网络合规体系建设

随着大数据和人工智能等技术的发展，大模型业务在数据处理和数据交互等方面面临着越来越多的合规风险和安全隐患。为了确保大模型业务的安全性和合规性，建立网络合规体系是非常必要的。首先，明确等保合规体系建设的范围和目标，制定等保合规政策和制度，识别可能存在的合规风险和安全隐患。其次，建立健全的网络安全保障体系以及完善的等保合规管理制度和流程，确保数据安全和业务稳定与管理职责和权限的明确。最后，需要建立完善的内部控制制度和流程，及时发现和纠正违规行为并加强对大模型业务的人员培训和意识教育，提高员工的合规意识和安全意识。

6. 本章小结

大模型的技术进步为社会发展带来了巨大机遇，但也伴随前所未有的安全风险和隐患。面对日益复杂的安全威胁，需从大模型的供应链安全、数据安全、内

容安全、合规体系建设等多维度入手进行防护治理，才能在技术与风险的博弈中实现稳步发展。

七、关于我们

据本报告的安全大数据统计，近年来监测到的漏洞信息数量呈现爆发式增长，日均新增漏洞条目不断攀升。面对日益复杂的攻击手段与层出不穷的漏洞，传统的人工处理方式已经难以满足快速响应的需求，在这一背景下，漏洞处理由人工转向自动化，已成为企业提升安全效率与响应速度的必然趋势。为此，企业的安全能力和运维体系需要借助更先进的自动化平台，以便更为精准和高效地响应日益复杂的安全威胁。

MMM 漏洞情报检测平台是安恒信息推出的一款基于大数据和人工智能技术的漏洞情报检测与预警平台。该平台通过整合全球范围内的漏洞情报、威胁情报以及安全事件数据，为用户提供实时的漏洞预警、风险评估和修复建议。平台汇聚了海量的漏洞数据，涵盖了国家信息安全漏洞共享平台（CNVD）、美国国家漏洞数据库（NVD）、国家信息安全漏洞库（CNNVD）等权威漏洞数据库的数据以及当下的在野漏洞档案数据。截至本报告编写时间，MMM 漏洞情报监测平台已载有 28 万余条全量漏洞档案数据。MMM 漏洞情报监测平台结合先进的 AVPT 技术，AVPT（Attack Vector Prediction Technology，攻击向量预测技术）是安恒信息自主创新研发的安全分析及预测技术。AVPT 技术结合了大数据分析、机器学习和人工智能算法，能够对新爆发的漏洞在外界还没有披露过多信息时，对其进行攻击向量的预测。AVPT 技术能够整合来自全球的威胁情报，包括已知的漏洞、恶意软件、攻击者行为等，提高预测的准确性和全面性。基于预测结果，AVPT 技术能够持续学习和优化预测模型，随着新威胁的出现和防御策略的调整，不断提高预测的准确性和有效性。并将实时更新的动态漏洞信息反馈至 MMM 漏洞情报监测平台，相关档案同步更新，确保用户侧漏洞信息的时效性与准确性。

其次，每个漏洞都拥有完整的档案，详尽记录了其相关信息，不仅包括漏洞的详细信息，还涵盖了对漏洞的可利用性、可用性等多维度漏洞特性的评估，并利用这些特性对漏洞进行定级，旨在帮助用户准确评估漏洞的潜在威胁与风险。与此同时，平台将会通过这些漏洞数据来赋予相对应的漏洞状态，平台也支持通过对这些漏洞特性进行查询，方便用户在海量的数据中检索出需要的数据。



图 1 MMM 漏洞情报监测平台检索功能

平台不仅提供海量的漏洞数据，还为用户提供一整套从监测、修复、防护到缓解的操作指导。恒脑解决方案智能体为每个漏洞都提供了专业的缓解方案，结合深度学习与智能分析，确保用户能够第一时间获得具有针对性的漏洞安全防护建议。



图 2 MMM 漏洞情报监测平台解决方案展示

通过多维度的漏洞监测和智能化的信息补全以及漏洞缓解方案的及时供应，对于被研判为漏洞处置等级较高的漏洞，安全研究员会在第一时间对其细节进行深度分析，输出漏洞分析与复现的详细报告并赋能产品。安恒信息 CERT 团队正是依托强大的漏洞情报监测平台、先进的 AVPT 技术加持以及专业的安全研究团队从而具备卓越的漏洞运营能力。

为响应国家各级监管机构深化推进行业安全体系建设，安恒信息为客户提供全方位的安全管理解决方案，其涵盖但不限于针对软件供应链安全、软件开发安全以及安全数据等需求的智能化、定制化服务，助力企业快速响应与处理安全风险，为行业安全建设提供了强有力的技术支撑。这些平台能够将安全防护工作从被动响应转变为主动防御，为用户提供了更加可靠和高效的安全保障。

安全开发一体化平台是针对**日益复杂的应用软件漏洞管理问题**而设计的，旨在管控安全开发流程和分析风险处置优先级。通过**应用安全测试编排、漏洞研判、持续运营**等方式，一站式管理项目的开发生命周期的安全问题，优化研发团队和安全团队在静态源代码缺陷审计、开源组件风险分析、API 风险分析、主机扫描、容器镜像扫描等方面的安全工作，通过**安全漏洞模型和安全度量模型**持续提升应用软件的安全

安恒软件成分分析平台 (DAS-SCA) 是**安恒信息**面向软件开发安全需求研发的基于软件开发安全生命周期管理的软件组件检测系统。支持对源代码、二进制文件、特征文件及镜像中的组件进行静态解析，并基于机器学习技术，模拟开发过程中包管理的行为，通过算法、策略、模型对软件开发引用到的组件进行高效深层分析。

安恒研究院已推出全面的安全数据订阅服务，包含威胁情报数据、漏洞情报数据、安全策略规则、安全事件专题情报及网络安全周报/月报等数据，均实现服务化。若您欲动态掌握此类网络安全关键信息，欢迎订阅。您可通过以下方式获取支持：

联系人：贾腾飞、周飞军；电话：400-6059-110

期待与您合作，为您的网络安全保驾护航。



OLYMPIC COUNCIL OF ASIA

亚奥理事会官方合作伙伴
OFFICIAL PRESTIGE PARTNER OF THE OLYMPIC COUNCIL OF ASIA

安恒信息
DAS-Security

官网: www.dbappsecurity.com.cn

电邮: info@dbappsecurity.com.cn

客服专线: +86-400-6059-110

杭州总部地址: 杭州市滨江区西兴街道联慧街188号安恒大厦



扫码提交 > 需求直达

©安恒信息 V.20241224宣传品