以零信任应对AI部署中的影子访问风险

CSA GCR cloud security GREATER CHINA REGION alliance®

CSA cloud security alliance®

身份和访问管理工作组的永久和官方地址是

https://cloudsecurityalliance.org/research/working-groups/identity-and-access-management

联盟简介

云安全联盟（Cloud Security Alliance,CSA）是中立、权威的全球性非营利产业组织，于2009年正式成立，致力于定义和提高业界对云计算和下一代数字技术安全最佳实践的认识，推动数字安全产业全面发展。

云安全联盟大中华区（Cloud Security Alliance Greater China Region, CSA GCR）作为CSA全球四大区之一，2016年在香港独立注册，于2021年在中国登记注册，是网络安全领域首家在中国境内注册备案的国际NGO，旨在立足中国，连接全球，推动大中华区数字安全技术标准与产业的发展及国际合作。

我们的工作

联盟会刊下载地址
了解联盟更多信息

加入我们

CSA大中华区官网
(https://c-csa.cn)

点击会员

加入联盟

填写相关申请信息

成为CSA会员

JOIN US

致谢

《以零信任应对 AI 部署中的影子访问风险》由 CSA 工作组专家编写，CSA 大中华区零信任工作和 IAM 工作组共同组织翻译并审校。

中文版翻译组成员

翻译组组长：

陈本峰

审校组成员：

戴立伟　于继万　谢琴　陈珊

研究协调员：

郑元杰　易利杰

贡献单位：

深圳竹云科技股份有限公司　　北京天融信网络安全技术有限公司

华为技术有限公司

英文版本编写专家

主要作者：

Venkat Raghavan

Steven Schoenfeld

Heinrich Smit

贡献者：

Ivan Djordjevic

Michael Roza

审校者：

Aakash Alurkar

Suramya Bakshi

Alan Curran MSc

Chris Kirschke

Arvin Reddy Jakkamreddy

Rangel Rodrigues

Lars Ruddigkeit

Akshay Shetty

Dr. Chantal Spleiss

Srinivas Tatipamula

CSA 全球员工：

Ryan Gifford

Claire Lehnert

Stephen Lumpe

在此感谢以上专家。如译文有不妥当之处，敬请读者联系 CSA GCR 秘书处给予改正！

联系邮箱 research@c-csa.cn；云安全联盟 CSA 公众号。

目录

序言

随着云计算、人工智能（AI）、生成式人工智能（GenAI）以及大型语言模型（LLM）等技术的快速发展，企业在数字化转型过程中面临的安全挑战日益复杂。影子访问，这种对系统和数据的意外或未经授权的访问现象，因现代技术环境的复杂性和访问权限管理不足而加剧，已经成为一个亟需解决的安全问题。与此同时，零信任安全理念作为应对新兴安全挑战的重要策略，逐渐受到各行业的重视。

本报告《以零信任应对 AI 部署中的影子访问风险》深入探讨了影子访问这一新兴安全问题与零信任安全框架之间的关系，特别是在生成式人工智能和大型语言模型广泛应用的背景下。报告还探讨了生成式人工智能技术引入的额外影子访问风险，并提出了应对这些新兴挑战的多种策略，如细粒度的访问控制、数据分类、身份验证和负责任使用等。这些内容为企业在日益复杂的技术环境中如何有效地管理和控制影子访问提供了宝贵的建议和指导。

在 AI 时代下，持续关注和深入探讨影子访问问题的重要性。通过不断探索和制定最佳实践，企业可以在复杂的技术环境中建立更健全、更安全的防护体系，确保信息系统的安全性和稳定性。

李雨航 Yale Li

CSA 大中华区主席兼研究院院长

摘要

影子访问是云计算中日益关注的问题，它指的是对系统和数据的意外或未经授权的访问，往往因现代技术环境的复杂性和访问权限管理不足而加剧。

零信任安全以“永不信任，持续验证”的理念为中心，被定位为应对影子访问的对策，提倡健壮的身份验证和严格的访问控制。然而，在影子访问非常普遍的云原生架构中，零信任的实际落地面临着挑战。

人工智能，特别是生成式人工智能（GenAI）、大型语言模型（LLMs）和检索增强生成（RAG）的集成，引入了影子访问风险的额外维度，如未经授权的访问、敏感数据暴露，以及人工智能和其他应用程序中的治理问题。本文档强调，由于人工智能导致影子访问脆弱性增加，所以在一个不断发展的环境中需要加强网络安全，同时传统的零信任方法必须适应生成式人工智能技术带来的细微差别。

本文档着重于讨论影子访问与当今最重要的两种技术的交叉点：零信任和人工智能。

什么是影子访问？

影子访问是对资源（如应用程序、网络和数据）的意外和/或非预期的访问。随着云计算、DevOps、云原生架构和数据共享的发展，这个新问题更加严重。

影子访问越来越成为一个云问题，这是由于各类云服务的聚合使用导致访问控制和用户权利要求增加，加上自动化的基础设施和软件开发，导致错误或意外地允许的帐户使用和资源开放。从小到大的组织经常会艰难地发现，曾经一个安全的起点已经默默地演变成了一个不安全的起点。此问题伴随着克隆帐户和权限（通常在入职或帐户创建过程中），为用户提供了非必要的访问，进而增加了影子访问问题。

零信任概述

零信任是新一代的网络风险战略，是一种基于“永不信任、持续验证”理念的方法，它可以大大减轻影子访问的可能性和影响。

零信任需要实现强身份验证（多因素身份验证（MFA）、密码学等），以及访问控制

（最小特权、基于角色的访问控制（RBAC）、基于属性的访问控制（ABAC）、基于上下文的访问控制（CBAC））。它还需要使 用网络分割和微隔离、持续认证和监控（用户和实体行为分析（UEBA）），以及审查、评估和审计。所有这些因素都有助于减少影子访问。

为什么存在影子访问？

零信任的基本原则是，不应该隐式或默认地授予任何访问权限，必须有意和明确地授予权限。因此，根据它的定义，零信任不允许影子访问，即意外或未经授权的访问路径。

那么，为什么会存在影子访问呢？

原因是很少有环境实现了所有的零信任原则。我们生活在一个不完美的世界里，人们会犯错误，应用程序 不断变化，有不同和不完整的流程和技术，数据的共享程度前所未有。此外，组织往往鼓励业务在审计与合规审查之前完成实施。影子访问往往普遍存在于现代云原生架构中的基础组件：对各类资源的授权过程。合乎逻辑的期望是，这应该可以简化零信任原则的实施。但是，架构是复杂的、可扩展的，并且分布在实例和控制逻辑中。这些问题，加上 IAM 流程和应用程序开发实践中的差距，是影子访问产生的根源。鉴于这些起点，问题的出现是不可避免的，这是可以理解的。事实上，这正是零信任核心原则的意图所在。零信任的这些要素强调了需要持续审查访问、路径和风险的必要性。

影子访问是可能被无意中授权或通过其他一些操作被授予为更大权限。零信任就是规避这一切。因此，无论是作为技术还是原则，影子访问的识别和删除都是零信任的一个组成部分，如果不能首先杜绝任何影子访问，就无法实现零信任的最终状态。

本文档进一步讨论了零信任和影子访问如何交叉，以及可以做些什么来确保一个更健壮和更安全的环境。

零信任原则可以如何减少影子访问

零信任最初通过其分配最小权限访问的核心原则来减轻影子访问。最小特权的访问权限意味着只为用户、设备、应用程序和工作负载提供他们绝对需要的访问权限。

零信任假设环境已经或者将出现问题，并通过对环境的持续监控，进一步降低影子访问的可能性。这将揭示并解决影子访问问题，无论是由于错误（显式操作或模板）还是由于应用程序更改导致的意外后果。

零信任原则通过一种基于风险的解决方案，进一步减轻了影子访问。对环境的这种理解使得能够优先解决最重要的影子访问问题。

GenAI、LLMs 等技术概述

GenAI（生成式人工智能）是一类设计用于创建新内容的人工智能模型和工具。它使用机器学习技术，如生成对抗网络（GAN）和 Transformer 模型，从大型的数据集中学习，并生成独特的输出。如果您想了解更多关于 GenAI 和 AI 治理、合规和责任的其他方面，请参考以下文件：

- 《AI 组织责任-核心安全责任》
- 《AI 弹性:AI 安全革命性基准型》
- 《从原则到实践:动态监管环境中的负责任 AI》

AI、GenAI、LLMs 和影子访问

以下是人工智能和大模型语言模型（LLMs）的新时代下创新者、安全、隐私和数据治理的领导者最关心的问题：

全面的清单、变更管理和运营控制

GenAI 引入了一类全新的技术资产：LLM 工具链、数据管道、RAG 系统、向量数据库、数据集、数据平台和知识库。首要任务是维护所有已批准的 GenAI 资产及其相

关元数据的详尽清单。企业 AI 资产管理（例如，企业 CMDB）成为事实的唯一来源，并成为建立操作控制和快速关闭行为不当的 AI 应用程序的基础。。

数据访问的可见性、可控性和合规性

确保 GenAI，特别是 RAG 对企业数据的透明、受控和合规的访问，RAG 将预训练的 LLM 与外部知识库集成在一起。鉴于 GenAI 的广泛应用，数据和业务所有者必须监督数据共享实践，确保符合数据隐私、存储地、主权和法规标准。

授权和治理框架

企业数据是人工智能和 LLM 时代最宝贵的资产。企业必须为企业数据和 AI 驱动的应用程序开发一个健全的授权和治理框架。模型访问策略在学习算法与企业数据交互时，建立了安全措施和指导方针，这通常需要严格的使用限制。这些政策对于确保在组织的生态系统中进行负责任且安全地处理、访问和共享敏感信息至关重要。

多模态数据的细粒度访问控制

随着人工智能领域的扩展，涵盖各种数据类型，如文本、图像、视频和非结构化数据，访问控制必须适应这种多模态环境。细粒度的授权机制使组织能够有效地管理对每个数据模式的访问。通过实施量身定制的访问策略，组织可以确保每种数据类型中的敏感信息得到适当的保护。

向量嵌入的访问控制

向量嵌入等衍生数据，如从企业数据仓库派生出的向量嵌入，对传统的访问控制机制提出了挑战。一旦创建了嵌入，即原始数据的派生表示，它就在自己的隔离环境中运行，向大型语言模型（LLMs）和可重用应用程序生成器（RAG）系统提供数据。与受企业访问控制管理的原始数据不同，嵌入独立运作，引入了需要理解的新访问控制漏洞。

企业必须认识到，LLMs 和 RAG 系统完全绕过了企业访问控制机制。风险在于关

键资产和数据可能被暴露，因为学习算法本质上规避了传统的访问控制机制，从而创造了大规模未授权访问暴露的环境，这为影子访问的滋生提供了肥沃的土壤。这就需要重新架构数据权限、分类和策略，从原始系统支持 LLM 堆栈的需求。

非结构化内容的数据分类

为非结构化内容建立数据分类模式对于有效地管理 LLM 应用程序的多样化需求至关重要。该模式支持基于内容类型、相关性、敏感性和上下文等各种参数对非结构化数据进行分类。通过实施一个健全的分类框架，组织可以简化数据处理流程，并确保 LLM 应用程序能够适当地访问和处理非结构化内容。

身份验证和负责任使用

实施身份验证和负责任使用协议对于确保只有授权用户和非人类实体（如 API）在其指定的责任和问责范围内使用生成式人工智能至关重要。通过验证用户和 API 的身份，组织可以降低未经授权的访问和滥用人工智能能力的风险。这种方法促进了人工智能使用的信任和透明度，同时防止潜在的滥用或伦理违规。

内容适宜性和授权

实施内容适宜性和授权措施对于验证和保证生成的内容符合用户期望并被授权分发至关重要。通过引入验证检查，组织可以根据相关性、合法性和伦理考虑等因素来评估内容的适宜性。

防止个人信息和机密数据泄露

实施保护措施以防止个人身份信息（PII）和机密数据在 LLM 堆栈中的泄漏对于维护隐私和安全至关重要。通过结合强加密、访问控制和数据匿名化技术，组织可以降低未经授权的访问和敏感信息泄露的风险。此外，实施输入-输出保护工具（如 Llama Guard 或内容过滤）可以降低泄漏的风险。

结论

人工智能时代正在带来了一波由大型语言模型（LLM）访问企业数据所推动的的变革性 AI 应用程序、助手和代理。这一时代标志着企业、私人和敏感数据在人类用户和 API 之间的前所未有的共享。随着这种 AI 驱动的互动激增，影子访问也在同时激增：附加在支持动态 AI 生态系统的人类和非人类身上的身份、凭证和权限，以前所未有的方式开启了影子访问。

生成式人工智能的集成，特别是由检索增强生成（RAG）驱动的集成，通过引入影子访问风险增加了复杂性，其中包括未经授权的访问、敏感的数据暴露和人工智能应用程序中的治理挑战。像 RAG 这样的新架构打开了访问企业数据的新路径，绕过了传统的访问控制和基于角色的访问控制（RBAC）。

对影子访问进行持续监控变得势在必行。这种监控将揭示影子访问风险，无论是由于错误（显式操作或模板）、新组件，还是 AI 引入的新路径所导致的。企业必须迅速适应并强化其监控实践，以应对影子访问的现实存在。

展望未来，我们的探索将深入研究这些新出现的挑战，并提供应对这些挑战的最佳实践见解，这标志着影子访问的持续讨论的新篇章的开始。

其他资料

- https://cloudsecurityalliance.org/artifacts/defining-shadow-access-the-emerging-iam-security-challenge
- https://csrc.nist.gov/pubs/sp/800/207/final
- https://cloudsecurityalliance.org/artifacts/zero-trust-principles-and-guidance-for-iam/
- https://cloudsecurityalliance.org/cloud-security-glossary
- https://cloudsecurityalliance.org/artifacts/confronting-shadow-access-risks-considerations-for-zero-trust-and-artificial-intelligence-deployments
- https://cloudsecurityalliance.org/artifacts/ai-resilience-a-revolutionary-benchmarking-model-for-ai-safety
- https://cloudsecurityalliance.org/artifacts/principles-to-practice-responsible-ai-in-a-dynamic-regulatory-environment