AMODE:多维效能驱动的欺骗元素自适应编排

陈墨楠[1)] 张云涛[2)] 刘欣然[1)] 孙岩炜[1)] 张天乐[1)]

[1)](北京邮电大学可信分布式计算与服务教育部重点实验室 北京 100876)

[2)](中央财经大学信息学院 北京 100081)

摘 要 随着网络威胁的不断发展,欺骗防御作为一种有效的主动防御手段应运而生。然而,传统欺骗防御技术面临欺骗元素的部署位置与语义内容优化脱节的问题,导致欺骗环境在攻击者视角下可见性不足或可信度缺失,严重制约防御效能。为此,本文提出一种多维效能驱动的欺骗元素自适应编排方法AMODE,旨在最大化可见性、可信性与诱导性三维协同的欺骗效能。本文首先构建了包含共现暴露概率、环境语义相容度与攻击吸引力三个因子的效能评估模型。其中,共现暴露概率因子通过分析攻击侦察路径评估元素与真实资产的共现可能性;环境语义相容度因子利用大语言模型自动生成与部署环境高度匹配的服务指纹与漏洞特征;攻击吸引力因子则基于层次分析法量化元素对攻击者的诱导价值。进而,基于该模型构建整数规划方程,在资源约束下自动生成最优编排方案。通过与现有代表性方法在量化指标和实战渗透测试场景下的系统对比,结果表明,AMODE生成的编排方案在综合欺骗效能指标上比对照方法平均提升了约33.4%;在渗透测试人因评估中,显著降低了攻击者的识别准确率与怀疑度,使得欺骗伪装效果(人因指标)提升了29.6%。AMODE显著增强了环境自适应性,为构建具备智能调度能力的主动欺骗防御系统提供了新的理论支撑与技术路径。

关键词 欺骗防御;自适应;多维协同优化;大语言模型;主动防御编排

中图法分类号 TP393 **DOI号** 10.11897/SP.J.1016.2026.01247

AMODE: Adaptive Multi-dimensional Orchestration of Deception Elements

CHEN Mo-Nan[1)] ZHANG Yun-Tao[2)] LIU Xin-Ran[1)] SUN Yan-Wei[1)] ZHANG Tian-Le[1)]

[1)](Key Laboratory of Trustworthy Distributed Computing and Service (BUPT), Beijing 100876)

[2)](School of Information, Central University of Finance and Economics, Beijing 100081)

Abstract As cyber threats evolve into sophisticated, multi-stage campaigns involving lateral movement and persistent reconnaissance, traditional passive defenses are increasingly insufficient. Deception defense has emerged as a critical active defense paradigm, aiming to disrupt the attacker's "Observe-Orient-Decide-Act" (OODA) loop by injecting false information. However, existing techniques suffer from a disconnect between the optimization of decoy deployment locations and the generation of semantic content. Previous approaches focus on a single dimension: either utilizing game-theoretic models to optimize topological placement while ignoring semantic consistency, or employing generative models to create high-fidelity content without considering the deployment context. This fragmentation results in deception environments that lack sufficient Visibility (decoys are not easily discovered by attackers) or Credibility (decoys are easily identified as fake due to context mismatches), thereby significantly diminishing the overall

收稿日期:2025-10-14;在线发布日期:2026-02-12。本课题得到国家重点研发计划项目(2024YFB31NL00102)、海南省院士创新平台科研项目(YSPTZX202506)、海南省方滨兴院士工作站(YSGZZ2023003)资助。**陈墨楠**,硕士,主要研究领域为网络欺骗防御、AI赋能网络安全等。E-mail:chenmonan@bupt.edu.cn。**张云涛(通信作者)**,博士,助理教授,主要研究领域为网络攻防、区块链安全与人工智能安全等。E-mail:zhangyt@cufe.edu.cn。**刘欣然**,博士,研究员,博士生导师,主要研究领域为网络安全、信息内容安全等。**孙岩炜**(通信作者),博士,讲师,硕士生导师,主要研究领域为网络威胁发现、态势感知、攻击溯源、威胁情报运营等。E-mail:sunyw@bupt.edu.cn。**张天乐**,博士,副教授,硕士生导师,主要研究领域为计算机网络、网络安全、移动互联网、智能电网。

defensive efficacy. To address these challenges, this paper proposes AMODE (Adaptive Multi-dimensional Orchestration of Deception Elements), a novel efficacy-driven orchestration framework that autonomously generates optimal deployment plans. The core of AMODE is a multi-dimensional evaluation model that unifies three critical factors: Visibility, Credibility, and Attractiveness (V-C-A). First, to quantify Visibility, we introduce the Co-occurrence Exposure Probability (CEP) factor. Unlike simple random placement, CEP models the attacker's network scanning behavior (e. g. , Nmap logic) and calculates the probability of a decoy being discovered alongside real assets based on IP address proximity and network topology structure. Second, to enhance Credibility, we propose the Contextual Semantic Fusion (CSF) factor. By leveraging the semantic understanding capabilities of Large Language Models (LLMs), CSF dynamically analyzes the fingerprints of surrounding real assets (e. g. , service banners, protocol versions) and generates highly compatible decoy configurations. Crucially, a constraint mechanism is applied to the LLM to prevent hallucinations, ensuring the generated vulnerabilities and service descriptors are logically consistent with the operational environment. Third, to maximize Attractiveness, the Attack Attractiveness (AA) factor is quantified using the Analytic Hierarchy Process (AHP), which measures the potential inducement value of a decoy based on its service type and data sensitivity levels. Based on this comprehensive evaluation model, we formulate the deception orchestration problem as an Integer Programming optimization model. This model automatically solves for the optimal combination of decoy types and locations under resource constraints (e. g. , limited IP availability) and business rules (e. g. , port conflict avoidance). We implemented a prototype system using containerization technology to support the automated lifecycle management of decoys. Extensive experiments and double-blind penetration tests were conducted in three distinct network scenarios: an enterprise office network, a development environment, and an industrial control system. The results demonstrate that compared with existing representative methods, AMODE's generated plans improve the comprehensive efficacy index by approximately 33.4%. Furthermore, in human-involved penetration testing, AMODE effectively misled professional testers, improving the human factor evaluation index by 29.6% (specifically, significantly reducing the Deception Salience Score and Precision Marking Rate). These findings confirm that AMODE significantly enhances both the environmental adaptability and the practical deceptiveness of active defense systems, providing a solid theoretical and technical foundation for intelligent cyber defense.

Keywords deception defense; adaptive; multi-dimensional coordinated optimization; large language model; active defense orchestration

1 引 言

近年来，网络攻击手段不断演化，呈现出链式、多阶段的复杂特征，攻击者往往通过资产发现、横向渗透等行动逐步扩大战果并达成攻击目标[1]。在此背景下，传统依赖规则匹配与边界封锁的被动防御体系逐渐暴露出瓶颈[2-3]。面对新时期网络安全保障体系面临的挑战，我国安全保障体系正面临从以“自卫模式”为主向以主动防御和威胁感知为核心的“护卫模式”转型的迫切需求，以全面提升国家网络安全防护水平[4]在此背景下，以“主动性”和“欺骗性”为核心理念的防御思路日益受到关注，其通过在攻击路径中合理布设欺骗元素，引导攻击者触发高置信度告警并实现对攻击的主动引导[5]。具体而言，一套理想的欺骗防御方案应当能够达成“4D”战略目标，即：将攻击者引向虚假目标(Direct)、扭曲其对网络环境的感知(Distort)、消耗其攻击资源与成本(Deplete)，以及发现其深层动机与战术技术(Discover)。

欺骗防御的实际防御效能并非单纯依赖欺骗元素数量，而在于防御者能否建立攻防不对称优势。这一优势要求防御者能够站在攻击者视角，预测其

侦察行为、评估其目标选择并影响其后续决策链条。从攻防博弈角度看，攻击者通常经历“发现—识别—利用”三个阶段；若欺骗元素在任一环节失效，整体欺骗链条便会被中断。鉴于此，系统化提升欺骗效能必须同时协同兼顾三个关键维度：可见性(Visibility)[6]，即欺骗元素必须容易被攻击者发现；可信性(Credibility)[7]，即攻击者需要相信其真实有效；诱导性(Attractiveness)[8]，即该元素足够有价值和吸引力以至于值得攻击者投入宝贵的攻击资源。下文将这三个关键维度合称为V-C-A。

这一多维协同的理念在学术研究与行业实践中均已得到广泛支持。在学术界，研究者普遍认为，实现高效的欺骗防御必须在可见性、可信性与诱导性三个关键维度上形成协同效能。首先，可见性直接决定攻击者能否注意到欺骗元素。近年来的实证研究表明，蜜罐的部署位置与网络拓扑结构显著影响其被发现的概率和引入的攻击流量，从而决定欺骗的触发效率[9-10]。其次，可信性依赖于欺骗元素在语义和环境上的相容性。在认知科学与攻击者心理层面，研究显示，只有当欺骗元素在语义上与周围环境高度契合时，才更容易被攻击者视为真实目标，进而显著提升欺骗效果[11-12]。最后，诱导性则决定攻击者是否愿意投入资源继续利用。博弈论与优化建模的最新成果揭示，攻击者在收益—成本权衡下，会优先选择价值较高、可利用性更强的资源，这与欺骗元素所具备的吸引力直接相关[13]。相应地，行业实践也提出了类似的思路。例如，DeceptIQ的网络欺骗成熟度模型强调，高级欺骗策略的核心在于同时兼顾对威胁的曝光度、资产的环境真实性以及对攻击者行动的成本-收益影响等关键因素，旨在实现多维度的协同欺骗效能。

然而，现有研究多聚焦于单一维度的优化，存在局限性。一类方法侧重部署位置建模，常基于攻击图或博弈论在关键节点放置欺骗元素[14-15]，但通常假设所有元素等效，忽视了语义一致性与差异化吸引力的影响。另一类研究集中于内容仿真，利用生成模型增强欺骗资源的拟真度[16-18]，但缺乏与部署路径结合，导致高仿真元素可能被放置在攻击者难以触达的位置。总体而言，现有方案未能在V-C-A三个维度实现协同优化。

为解决这一关键问题，本文提出一种多维效能驱动的欺骗元素自适应编排方法(Adaptive Multi-dimensional Orchestration of Deception Elements，AMODE)。该方法在统一框架下引入V-C-A三维效能，并进一步形式化为三个量化因子：共现暴露概率因子，用于刻画部署位置的暴露可能性；环境语义相容度因子，用于衡量元素与环境在语义层面的契合度；攻击吸引力因子，用于刻画元素的吸引力。在此基础上，AMODE通过整数规划在资源约束下生成最优部署方案，从而在提升诱导性的同时显著增强可见性与可信性。本文的主要贡献包括：

(1)形式化构建了V-C-A三维协同的欺骗效能建模框架。该框架将可见性、可信性与诱导性在理论上统一，为后续的量化评估和优化编排奠定了理论基础。

(2)提出了一种多维效能驱动的自适应编排方法AMODE。该方法通过将共现暴露概率、环境语义相容度和攻击吸引力三个核心因子集成为优化目标，并融入整数规划模型中求解，实现了在资源约束下欺骗元素最佳部署方案的自动生成。

(3)构建了AMODE编排系统，并采用效能因子量化评估与实战测试驱动的人因评估相结合的综合验证体系。对比结果表明，AMODE生成的方案在确保诱导性的前提下，显著提高了欺骗元素在攻击者视角下的可见性与可信性，有效提升了整体欺骗防御效能。

综上，本文在欺骗防御部署策略中填补了位置选择、语义适配与诱导吸引的协同优化空白，为构建具备智能化与环境适应性的主动欺骗防御系统提供了新的理论与技术路径。

本文的组织结构如下：第2节回顾欺骗编排的相关工作和研究现状。第3节提出AMODE整体架构，并构建基于V-C-A三维协同的欺骗效能评估方法和欺骗元素编排模型。第4节详细阐述AMODE编排策略的生成机制，包括基于效能因子的量化计算与整数规划优化模型的构建。第5节进行系统与场景搭建，并通过对比实验验证所提方法在增强防御效能和环境适应性上的有效性与优越性。最后，第6节对全文进行总结与展望。

2 欺骗编排相关工作

随着网络攻击行为日趋复杂，欺骗防御作为一种以混淆攻击者感知、主动诱捕威胁为核心的主动防御手段，近年已成为研究热点。为提升防御系统的实效性与智能化水平，已有大量研究围绕欺骗元素的编排部署策略展开，主要聚焦于系统自动化、资源语义仿真与部署位置优化等方向。

为提升欺骗编排系统的灵活性与自动化水平，部分研究尝试通过引入人工智能与容器化技术增强部署效率与可控性。例如，Pagnotta G等人提出的DOLOS框架[19]，融合移动目标防御和网络欺骗技术，通过攻击面随机变换增强主动防御能力；Sajid M等人设计的symbSODA系统[20]，基于多路径符号执行生成防御策略；Kahlhofer M等人的Koney框架[21]利用容器实现云原生蜜罐调度；Ahmed S等人的SPADE系统[16]引入生成式AI拓展欺骗剧本内容。这类方法提升了系统在部署效率和内容多样性上的能力，但通常未建立部署位置与欺骗元素语义之间的协同机制，难以支撑针对性强、诱导力高的引导策略。

在欺骗元素内容语义优化方面，部分研究专注于提升资源的仿真度与语义质量，以增强其在攻击者视角的可信度。在服务仿真方向，Sezgin A等人的DecoyPot[22]借助大语言模型生成具备上下文语义的API响应，王瑞等[23]也提出利用大模型模仿周围资产的语义特征，生成高相容度的欺骗服务，在内容的拟真性和环境相容度方面取得了显著进展；Gabrys R等人的HoneyGAN[17]基于生成对抗网络（Generative Adversarial Network，GAN）动态生成网络设备的诱饵配置。在文件欺骗方向，H Li等人提出的EDGE模型通过生成在内容上极具诱惑力的虚假文件内容来迷惑攻击者[24]，Timmer R C等人设计的Honeyfile系统[18]使用语义向量优化文件命名；凭据仿真方面，Dionysiou A等人的HoneyGen[25]通过词向量近似构造弱密码，Reti D等人提出基于Large Language Model（LLM）的蜜饵生成方案[26]，可生成语法合理、形式多样的蜜令牌。这类研究在资源本体的仿真性方面取得显著进展，但大多脱离具体部署环境进行内容生成，缺乏针对部署位置与攻击路径的有效关联建模，使得高可信度欺骗元素的优化成果难以在复杂网络中有效触达，限制了其实际可见性。

在部署位置优化上，部分研究则以提升欺骗元素在网络结构中的可见性为目标，专注于部署位置的优化建模。例如，张琳等[27]针对物联网环境，提出了DPOA算法，其核心思想就是通过优化网络拓扑，最大化真实物联网节点到欺骗元素的路径数，以提升欺骗元素的可见性。De Gaspari F等人[28]、Anwar A H等人[14]与Sayed M A等人[15]基于攻击图和博弈论模型，在攻击路径关键节点部署蜜罐以最大化防御覆盖；Zambianco M等人[29]将部署问题形式化为整数规划，在资源受限下优化路径覆盖能力。为应对云计算环境的动态变化与复杂的攻防博弈，研究者引入了强化学习方法，如Li H等人[30]和Kong G等人[31]基于Q策略进行动态部署，何威振等[32]则进一步提出了基于深度强化学习的多阶段Flipit博弈拓扑欺骗防御方法，以适应云原生网络下的时空多维度拓扑欺骗攻防场景。这类研究显著提升了欺骗元素的路径暴露概率，但普遍假设所有资源具有等效欺骗性，未能区分资源类型在不同部署环境中的语义适用性，同时忽略了语义融合对欺骗元素可信度的影响。

也有部分研究初步探索了部署位置与语义内容的协同优化机制，通常集中在特定资源或场景中进行嵌入式部署。例如，Prabhaker N等人[33]在数据库中嵌入伪造记录，确保统计特征与真实数据保持一致，从而实现语义拟真与部署融合；Bartwal U等人[34]在异常IP区段中部署蜜罐，以规则驱动的方式形成动态响应策略。这些尝试验证了部署位置和内容语义联合优化的可行性，但由于依赖特定规则、资源类型或平台特性，缺乏可扩展性，难以推广至更广泛的元素与应用环境。总体来看，现有研究虽然在自动化部署、内容仿真和位置优化等方面取得了显著进展，但大多停留在单一维度的优化：要么侧重部署位置的可见性，要么强调资源语义的可信性，缺乏统一的整体建模。当前方法仍难以将可见性和可信性有效地结合，从而限制了欺骗元素在真实攻防环境中的诱导效能。

针对这一不足，本文提出的AMODE方法首次在统一框架下，在兼顾欺骗元素诱导性的前提下，系统性地解决了部署可见性与内容可信度脱节的问题。AMODE以可见性、可信性与诱导性三维协同的欺骗效能为优化目标，构建了一个包含共现暴露概率（V）、环境语义相容度（C）与攻击吸引力（A）三个因子的效能评估模型作为核心驱动。现有研究侧重于提升可见性或可信性，但其最终目的均是为了增强欺骗元素的诱导潜力。AMODE正是将这种隐性的诱导潜力显性化和量化，与可见性、可信性协同作为优化目标，使得欺骗元素不仅能被发现（高V）、可被相信（高C），更能有效地被利用（高A）。AMODE通过效能因子量化与整数规划建模，能够在资源约束下生成更契合业务语义、同时具备较高暴露概率与吸引力的部署方案。在实验中，AMODE不仅在综合效能指标上显著优于近年的代表性工作，也在多种区域场景环境中展现出稳定的

适应性。这表明AMODE不仅在理论上实现了多维度协同优化创新，也在实用性上具备较强的可落地性。

3　AMODE方法概述

3.1　模型架构与机理

本节从宏观思路出发，构建对AMODE方法的统一建模视角，相关主要符号如表1所示。AMODE的核心目标是在复杂环境中，通过对多维度的联合建模与优化，生成一组具备高可见性、高可信度与高诱导性的元素部署方案，以最大化对攻击者的欺骗效果。

方法整体由两个核心模块构成：欺骗效能评估器与欺骗方案编排器，前者用于评估任意部署方案的欺骗效果，后者在约束条件下搜索最优部署策略。两者形成“评估-反馈-优化”的逻辑闭环，支撑系统智能生成部署策略。

表1　主要符号定义

分类	符号	概述
输入集合	Res	候选资源集合，$Res=(H,L)$
	Ctx	环境上下文集合，$Ctx=\{r_1,r_2,\cdots\}$
输出结果	$Plan$	最终生成的欺骗元素编排方案
	$DEV(\cdot)$	编排方案的宏观整体欺骗效能值
核心效能因子	$CEP(h,l)$	共现暴露概率，衡量部署位置的可见性
	$CSF(h,r)$	环境语义相容度，衡量欺骗元素与环境的语义一致性
	$AA(h)$	攻击吸引力，衡量元素本身的诱导潜力

AMODE以真实资产集合作为环境上下文，并将候选欺骗元素集合与候选部署位置集合作为可调配资源输入，借助评估器和编排器两大核心模块，融合部署评估与策略优化机制，最终输出一组欺骗效能最优的欺骗元素编排方案。整体建模架构如图1所示。

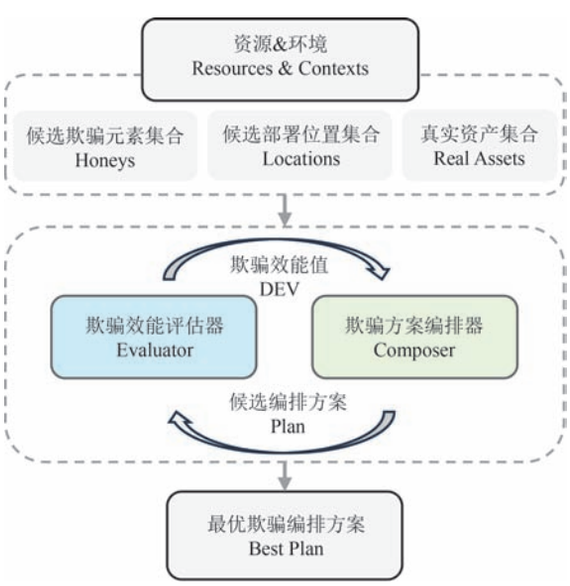

图1　AMODE架构图

为支持该方法的形式化建模，本文从V-C-A三个核心维度——部署位置可见性、环境语义可信性与欺骗元素诱导性——抽象出欺骗元素编排问题中的关键要素。这些要素构成部署空间、效能评估的基础以及策略搜索的范围，其形式化定义如下：

候选欺骗元素集合（Honeys，H）：表示系统可供部署的欺骗元素，形式可以是服务、文件等。

$$\mathcal{H}=\{h_1,h_2,\cdots\}$$

候选部署位置集合（Locations，L）：表示可以放置欺骗元素的位置，可以是IP、文件目录等。

$$\mathcal{L}=\{l_1,l_2,\cdots\}$$

资源空间（Resources，Res）：表示由候选欺骗元素集合与候选部署位置集合构成的可调配资源。

$$\mathcal{R}\mathrm{es}=(\mathcal{H},\mathcal{L})$$

环境信息集合（Contexts，Ctx）：表示当前环境中已知的真实资产信息，如业务服务、重要文件等。集合中的元素$r\in Ctx$特指网络中处于活动状态且运行具体业务的特定业务服务实例。

$$Ctx=\{r_1,r_2,\cdots\}$$

候选部署对集合（Candidate Set，C）：表示所有可能的“欺骗元素-部署位置”组合，作为效能评估与方案筛选的基础空间。

$$\mathcal{C}=\{(h,l)\,|\,h\in\mathcal{H},\ l\in\mathcal{L}\}$$

欺骗编排方案（Orchestration Plan）：表示一组选定的部署对子集，是输出的元素配置结果。

$$Plan\subseteq\mathcal{C}$$

在上述要素基础上，AMODE的部署优化过程由两个核心功能模块驱动：

欺骗效能评估器(Evaluator):基于可见性、可信性、诱导性三个维度,计算任意部署方案在当前上下文下的整体欺骗效能值。

欺骗方案编排器(Orchestrator):在资源约束条件下,基于效能评估结果搜索最优部署方案,使系统整体诱导能力达到最大化。

二者构成闭环机制,其协同作用可形式化为如下映射关系:

欺骗效能评估器。对任意部署方案$Plan\subseteq C$,在资源空间$\mathcal{R}es$与环境上下文Ctx下,输出该欺骗元素编排方案的整体欺骗效能值(Deception effectiveness value, DEV)。

$$\text{Evaluator}(Plan,\mathcal{R}es,Ctx)\rightarrow \text{DEV}。$$

欺骗方案编排器。基于效能评估结果与资源条件,返回满足约束且效能最优的部署方案$Plan^*$。

$$\text{Orchestrator}(\text{DEV},\mathcal{R}es)\rightarrow Plan^*。$$

综合上述过程,AMODE方法的核心优化目标可形式化为

$$\begin{aligned}&\underset{Plan\subseteq C}{\text{maximize}}\quad \text{DEV}(Plan,\mathcal{R}es,Ctx)\\&\text{subject to}\quad \text{Constraints}\end{aligned}\tag{1}$$

该优化模型体现了“评估-反馈-优化”三阶段联动机制,是AMODE方法实现部署策略智能编排的理论基础。后续章节将分别对效能评估函数的建模方法与策略搜索算法进行详细展开。

3.2 欺骗效能评估思路概述

衡量元素部署方案的效能是实现精准部署和有效诱导的关键前提。本文从微观层面出发,针对每个部署对$(h,l)\in C$,即欺骗元素与其部署位置的组合,设计量化的欺骗效能评估方法,并通过对所有部署对的微观效能进行累积,构建欺骗效能的评价方法,如图2所示。

具体而言,本文构建了基于可见性(V)、可信性(C)与诱导性(A)的V-C-A三维协同欺骗效能评估方法,其理论动因在于对攻击者完整交互链(即发现-识别-利用)的系统化解构和建模。为确保欺骗效能覆盖攻击全过程,分别为三个维度设计了三个量化效能因子:

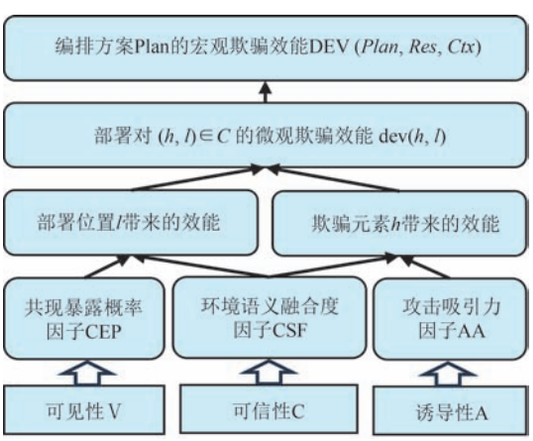

图2 基于V-C-A的多维欺骗效能评价方法

共现暴露概率(Co-discovery Exposure Probability, CEP),锚定攻击者的侦察与发现行为,衡量元素部署位置的可见性;

环境语义相容性(Contextual Semantic Fusion, CSF),针对目标的识别与验证行为,反映元素与周边环境的一致性,决定其可信性;

攻击吸引力(Attack Attractiveness, AA),聚焦于利用与攻击决策,描述元素本身对攻击者的诱惑力,即诱导性。

下面分别阐述各因子的设计思路及定义:

(1) 共现暴露概率因子(CEP)

部署位置的合理性决定元素能否“自然”暴露。攻击者的侦察和横向移动沿真实资产路径展开,部署位置l与真实资产位置l_r越邻近,二者被同时发现的可能性越大。形式化定义为

$$\text{CEP}_{l,l_r}=F_{\text{CEP}}(l,l_r):\mathcal{L}\times\mathcal{L}\rightarrow[0,1]\tag{2}$$

(2) 环境语义相容性因子(CSF)

欺骗元素h只有在语义层面融入由若干真实资产r构成的环境,才更有可能被攻击者视为可信的目标。语义层面包括功能逻辑、场景配套、脆弱性等多个维度。语义层面的相容性越高,攻击者越难识破其伪装属性,对欺骗元素的信任感越强。形式化定义为:

$$\text{CSF}_{h,r}=F_{\text{CSF}}(h,r):\mathcal{H}\times Ctx\rightarrow[0,1]\tag{3}$$

(3) 攻击吸引力因子(AA)

欺骗元素自身的诱导潜力对激发攻击者兴趣至关重要。不同元素因潜在价值不同,对攻击者吸引力差异显著。欺骗元素h本身吸引力越高,越具备对攻击者的诱导能力。形式化定义为

$$\text{AA}_h=F_{\text{AA}}(h):\mathcal{H}\rightarrow[0,1]\tag{4}$$

上述三个因子共同作用,决定了单个部署对(h,l)的欺骗效能,只有同时具备这三方面优势,才能达成最佳的欺骗效果。

将一个部署对(h,l)微观欺骗效能定义为

$$\text{dev}(h,l)=\text{AA}_h\times\sum_{r\in\text{Ctx}}[\text{CEP}_{l,l_r}\times\text{CSF}_{h,r}]\tag{5}$$

对于整个部署方案$Plan\subseteq C$,方案整体欺骗效能由所有部署对的微观效能值累积而成:

$$\mathrm{DEV}(Plan, Res, Ctx) = \sum_{(h,l)\in Plan} \mathrm{dev}(h,l) \quad (6)$$

该宏观欺骗效能值函数作为部署优化模块的目标函数，衡量编排方案整体的宏观欺骗效能，该评估方法是AMODE中方案编排的核心评价依据。

该评估方法实现了对不同类型欺骗元素的广泛通用性与可扩展性。这种源自攻击链分解的统一理论基础，使得该评估模型具有较好的跨场景扩展能力，不局限于某一类特定欺骗资产。如表2所示，无论是欺骗服务（HoneyService）、欺骗文件（HoneyFile）还是欺骗凭据（HoneyToken），AMODE方法都能基于这套统一的V-C-A理论框架，针对每种元素的独特特性，在可见性、可信性和诱导性三个维度上进行差异化且可量化的评估与优化。这种灵活的机制充分论证了模型在多场景下的强大通用性和适应能力，为各类欺骗元素部署提供了统一、完备且具备理论支撑的优化指导。

表2　各类欺骗元素的效能因子优化维度

欺骗元素类型	可见性(V)-CEP因子	可信性(C)-CSF因子	诱导性(A)-AA因子
HoneyService	与真实服务同子网、同主域名	相同内容逻辑、场景互补、共通脆弱性	SSH/Redis等高价值服务，诱导口令
HoneyFile	部署在高频访问路径	文件格式与命名伪装、内容结构模仿	私钥、密码、敏感关键词诱导
HoneyToken	嵌入真实页面、真实数据库	字段格式一致、访问路径自然	看似有效的访问令牌、虚假账号凭据

3.3　编排方案生成机制概述

在完成对部署对(h, l)的微观欺骗效能$\mathrm{dev}(h, l)$的量化评估后，本节进一步构建欺骗编排方法生成机制，旨在部署资源与条件约束下，从候选部署空间中抽取一组最优部署对集合，形成整体效能最优的欺骗部署方案。

首先，所有部署对$(h, l)\in C$的微观欺骗效能评估结果可组织为一个部署效能分布矩阵，其中$|H|$与$|L|$分别表示候选欺骗元素与可部署位置的数量，矩阵中的元素$\mathrm{D}_{h,l}$示将资源h部署于位置l所具有的欺骗效能值。

$$\mathbf{D}\in\mathbb{R}^{|\mathcal{H}|\times|\mathcal{L}|},\quad \mathbf{D}_{h,l}=\mathrm{dev}(h,l)$$

为统一不同资源部署倾向的量纲影响，本文对部署效能分布矩阵引入Softmax归一化操作，对每个资源在不同位置上的部署得分进行标准化处理，得到欺骗效能值的归一化分布密度：

$$\mathbf{P}_{h,l}=\frac{e^{\mathbf{D}_{h,l}}}{\sum_{l'\in\mathcal{L}} e^{\mathbf{D}_{h,l'}}},\quad \forall h\in\mathcal{H}, l\in\mathcal{L} \quad (7)$$

引入Softmax归一化的原因在于：原始效能矩阵$D_{h,l}$中的数值由多个异构因子合成，其分布可能较为平坦。Softmax函数具有非线性归一化特性，能够增强高适配部署对之间的分布区分度，同时抑制低效能组合的影响，从而为后续部署优化提供更稳定的评分依据。

部署过程中需满足多种现实约束，常见包括总体部署数量限制、每个部署位置不重复部署、欺骗元素类型间的数量比例控制等。

在归一化部署评分基础上，本文将部署方案的选取问题转化为受限空间下的组合优化任务。该任务可采用多种方式求解：对规模有限的场景，可借助整数规划进行精确求解；对大规模或需动态响应的部署任务，则可引入贪心算法、启发式搜索等高效策略进行近似最优解计算。

上述机制实现了从部署评分评估到部署策略输出的系统映射，使AMODE方法具备清晰的优化逻辑路径与高度的可执行性。

4　欺骗元素编排策略生成

在第3节系统构建了AMODE方法的整体建模框架与核心机制之后，本节将进一步聚焦其在欺骗元素编排环节的具体实现，围绕由欺骗效能评估器与部署策略编排器组成的核心模块，展开模型落地机制的具体方案设计。

值得强调的是，AMODE具备良好的通用性与可扩展性，可适用于服务类（HoneyService）、文件类（HoneyFile）与凭据类（HoneyToken）等多种类型的欺骗元素，并可针对不同网络场景灵活定制对应的评估与编排策略实现。在本文中，为便于分析展开与实验验证，研究对象聚焦于欺骗服务类欺骗元素，围绕其部署效能建模、评估方法与优化机制展开具体设计与实现。

在已完成资源空间$Res=(H, L)$和环境上下文Ctx的构建工作的前提下，进入策略生成阶段。其中，H与L分别候选欺骗服务集合与可部署IP位置集合，Ctx表示资产探测获取的真实服务，三者共同构成策略优化所依赖的输入空间。

在此基础上，本节将围绕V-C-A三个维度，设计部署效能评估函数，并推导最优欺骗服务编排方案，形成从建模理论到部署实践的闭环路径。

4.1 欺骗效能评估模块

为支持欺骗编排方案的优化生成，AMODE在前文中构建了以V-C-A三维效能评价为核心的三因子欺骗效能评估方法。本节将在此基础上，进一步细化实现在欺骗服务场景（HoneyService）下的欺骗效能评估模块的设计，系统刻画每个候选部署对(h,l)在特定环境中的欺骗效能。

相较于传统基于规则匹配、预定义模板或威胁知识库的欺骗策略生成方式，本模块融合了大语言模型的语义解析与知识抽取能力，具备更强的上下文建模能力，能够结合环境上下文实现更具适应性的部署效能评估。

4.1.1 共现暴露概率因子CEP

在虚假服务部署中，策略设计的首要考量在于提升欺骗元素在攻击者自然侦察过程中的被发现概率。尤其在攻击者处于资产探测、横向移动阶段时，往往会使用自动化工具（如Nmap）对目标网段进行大范围探测，此时虚假服务的“可见性”成为关键因素。共现暴露概率因子CEP在此场景下可被设计为度量欺骗服务h在部署于某一IP时，其在攻击者进行资产扫描发现时，同真实服务r一起被扫描到的概率，即共同出现在攻击者视角中的概率，相关符号如表3所示。

表3 共现暴露概率（CEP）因子符号定义

名称	符号	概述
前缀匹配度	$NPM(l)$	基于网络号的距离度量
后缀邻近度	$NSP(l)$	基于主机号的距离度量
权重系数	λ	调节NPM与NSP的相对重要性
响应控制参数	α,β	控制Sigmoid响应的中心与衰减速度

该因子的设计基础是对攻击者扫描行为模式的建模，主要包括以下两个关键特征：

批量扫描倾向：攻击者倾向于以连续IP块（例如/24子网）为单位执行扫描任务，因此位于同一CIDR子网的地址被同时扫描的概率更高。

地址邻近性效应：尽管有时存在扫描端口的随机化机制，但在实际实现中，IP地址的顺序通常仍具有一定的连续性，数值上更接近的地址更可能在时间或逻辑上被一起访问。

为对上述因素建模，定义如下两个评分函数：

（a）前缀匹配度NPM（l_h，l_r）：用于衡量部署地址l_h与真实服务地址l_r共享多少前缀比特（即是否可能处于同一小子网中），其值域为（0，1）。具体地，设部署地址l_h和真实地址l_r转换为整数后，其按位异或结果的前导0数量即为共享前缀长度SPL，其中$\sigma(x)$为Sigmoid函数，α为控制响应中心的偏移参数（如$\alpha=20$，对应/24子网）。

$$NPM(l_h,l_r)=\sigma(SPL-\alpha) \tag{8}$$

（b）后缀邻近度$NSP(h,r)$：用于衡量衡量部署地址l_h与真实服务地址l_r在排除共享前缀之外的低位地址上的数值差距。差距越小，越可能被短时间内一起扫描到，其中Δ_{suffix}表示IP地址低位的数值差异，β控制评分下降速度。

$$NSP(l_h,l_r)=\sigma\left(-\frac{|\Delta_{\text{suffix}}|}{\beta}\right) \tag{9}$$

最终，二者加权融合得到CEP值，其中$\lambda\in[0,1]$为权重因子，用于调节“是否属于同一扫描范围”和“是否在地址上相邻”两种影响的相对重要性。本场景最终CEP因子计算公式如下：

$$CEP(l_h,l_r)=\lambda\cdot NPM(l_h,l_r)+(1-\lambda)\cdot NSP(l_h,l_r) \tag{10}$$

该设计使得欺骗服务在部署位置选择时，能够优先贴近那些扫描高频的真实资产位置，从而在不显眼暴露的前提下，显著提高被动出现在攻击路径的可能性，为后续交互创造更多触发机会。

4.1.2 环境语义相容性因子CSF

在服务欺骗场景中，欺骗元素常以虚假网络服务的形式部署于空闲IP，若该服务与周边真实服务在语义层面存在功能、角色、或脆弱性上的明显不一致，极易被攻击者识破，导致部署效果失效。为此，设计环境语义相容性因子$CSF(h,r)$，用于量化欺骗服务h与邻近真实服务实例r在当前环境中的语义相容程度，从而评估其“是否伪装自然”，相关符号如表4所示。

$CSF(h,r)$由三个子因子组成，分别从三个角度借助大语言模型衡量欺骗服务h和真实服务r部署在一起的合适程度，认为当h任一子维度与r显著相容时，可视为二者具备较好的语义相容性。

为确保三个子因子中生成式内容的逻辑严谨性与技术可靠性，本文构建了一套统一的结构化指令框架（详见附录A）。该框架通过预设专家角色与多步推理约束，强制模型在生成服务描述、配套场景及脆弱性列表时，严格遵循网络安全领域的业务逻辑

表4　环境语义相容度(CSF)因子符号定义

名称	符号	概述
仿真相似性	$ES(h,r)$	评估欺骗元素 h 与真实资产 r 的指纹相似度
场景合理性	$SP(h,r)$	评估欺骗元素 h 的存在是否符合业务场景逻辑
脆弱性相似性	$VS(h,r)$	评估欺骗元素 h 与真实资产 r 的漏洞特征吻合度
响应控制参数	θ,k	控制语义相似度的激活阈值与曲线陡度
生成函数	Γ	基于LLM实现的功能函数
嵌入函数	Φ	基于向量化嵌入模型的函数

与依赖关系，而非单纯的概率续写，从而有效抑制模型幻觉，保证了后续量化计算的准确性。

(a)仿真相似性因子ES

该子因子衡量欺骗服务 h 是否在功能描述、行为表现等方面，能高度仿真真实服务 r 的核心特征。其核心思想是模拟攻击者在扫描识别阶段的“指纹验证”过程：若 h 与 r 的服务类型、响应方式等特征等具备较高相似性，则难以被识破。

实现流程如下，首先是服务功能描述生成，利用结构化提示约束的大语言模型生成机制，构建服务描述函数 Γ_{desc}，从探测结果(如Nmap工具输出)生成标准化服务功能描述：

$$d_h=\Gamma_{\text{desc}}(h),\quad d_r=\Gamma_{\text{desc}}(r)$$

然后将文本编码为语义向量，使用Sentence-BERT(下文称SBERT)模型 Φ_{SBERT} 对描述文本进行语义嵌入：

$$\mathbf{v}_h=\Phi_{\text{SBERT}}(d_h),\quad \mathbf{v}_r=\Phi_{\text{SBERT}}(d_r)$$

最后计算仿真相似度：

$$\text{ES}(h,r)=\frac{\mathbf{v}_h\cdot\mathbf{v}_r}{\|\mathbf{v}_h\|\cdot\|\mathbf{v}_r\|} \tag{11}$$

(b)场景合理性因子SP

该因子评估欺骗服务 h 在真实服务 r 所在环境中是否“合理存在”。攻击者通常具备对服务生态组合的认知，例如：存在WordPress，旁边有MySQL、phpMyAdmin是合理的；但若出现Modbus工控协议服务则可能引发怀疑。

实现流程如下：首先是配套服务生成，利用语言模型 $\Gamma_{companions}$ 推理 r 的常见配套服务集合：

$$\mathcal{A}_r=\Gamma_{companions}(r)=\{a_1,a_2,\cdots,a_k\}$$

之后是语义匹配函数，通过SBERT向量模型计算 h 的服务名称 n_h 与每个 a_i 的语义相似度。其中，σ 为Sigmoid归一化概率评分函数，θ 控制相似度阈值，k 控制陡度。

$$\Psi_{\text{soft}-\text{in}}(h,a)=\sigma\big(k\cdot(\cos(\Phi(n_h),\Phi(n_a))-\theta)\big)$$

最后取最大匹配得分作为场景合理性得分：

$$\text{SP}(h,r)=\max_{a\in\mathcal{A}_r}\Psi_{\text{soft}-\text{in}}(h,a) \tag{12}$$

(c)脆弱相似性因子VS(h,r)

该因子衡量 h 与 r 在攻击者视角下是否具备相似攻击价值，即是否存在共通或近似的可利用脆弱性。在横向移动、漏洞利用等阶段，若 h 与 r 在攻击面上高度重合，则更易被一同攻击。

实现流程如下，首先是漏洞集合生成，利用语言模型 Γ_{vuln} 生成各自服务常见脆弱性列表：

$$\mathcal{V}_h=\Gamma_{\text{vuln}}(h),\quad \mathcal{V}_r=\Gamma_{\text{vuln}}(r)$$

此处生成的脆弱性列表 V_h 与 V_r 包含的是宏观层面的“攻击向量(Attack Vectors)”描述(如“Remote Code Execution”，“SQL Injection”等)，而非具体的CVE编号。这种粗粒度的描述不仅容错性更强，而且通过SBERT向量化后的语义空间中，能够有效识别“弱口令(Weak Password)”与“凭证泄露(Credential Leak)”等近义概念的潜在关联，从而为后文的计算提供更稳定的语义表示。

之后是软交集匹配：计算两集合中语义相似度高于阈值 τ 的对数：

$$\mathcal{V}_h\otimes\mathcal{V}_r=\{(v_h,v_r)\,|\,\cos(\Phi(v_h),\Phi(v_r))\geqslant\tau\}$$

最后进行相似度计算：

$$\text{VS}(h,r)=\frac{2\cdot|\mathcal{V}_h\otimes\mathcal{V}_r|}{|\mathcal{V}_h|+|\mathcal{V}_r|} \tag{13}$$

融合策略采用最大值选择策略。考虑到三个维度分别对应攻击者的不同识别角度，只需任一维度具有较高的融合度，即可视为部署有效。因此，定义最终融合度如下：

$$\text{CSF}(h,r)=\max\big(\text{ES}(h,r),\text{SP}(h,r),\text{VS}(h,r)\big) \tag{14}$$

采用最大值融合策略(Max-Pooling)而非加权平均，该策略主要基于攻击者机会主义侦察行为的经验性假设。在侦察阶段，攻击者往往聚焦于寻找一个可利用的切入点。只要欺骗元素在功能拟真、场景合理或脆弱性暴露任一维度上与环境高度契合(即得分较高)，就足以建立初步信任并诱发攻击尝试。平均值策略可能会因为某一维度的平庸(如名称不突出)而掩盖了其在脆弱性上的高相容优

势，不符合攻防实战中的机会主义特征。

CSF因子综合考虑欺骗服务的功能匹配、部署合理性与攻击面伪装效果，依托大语言模型生成语义内容，结合语义编码模型实现相似度度量。通过离线结构化知识生成和在线向量快速匹配的设计，兼顾了精度与实时性，为欺骗部署的“自然伪装”提供了量化依据和自动化能力支撑。

4.1.3 攻击吸引力因子AA

攻击吸引力函数AA(h)用于描述不同类型欺骗服务对攻击者潜在诱导倾向的相对差异。其理论基础在于：不同类型网络服务因数据敏感性、漏洞严重性或与攻击TTP的契合程度，天然具备不同程度的诱导价值。基于MITRE ATT&CK知识库统计与攻击案例分析，本文将欺骗集合划分为六类具有明显攻击诱导差异的服务类型，如表5所示。

表5 归纳的欺骗服务类型

类别	典型服务实例	攻击者吸引力来源
身份认证服务	Kerberos	凭证窃取、权限提升
数据存储服务	MySQL	敏感数据泄露、未授权访问
运维管理服务	Jenkins	供应链攻击、容器逃逸
业务应用服务	ERP, CRM	商业数据窃取、逻辑漏洞利用
网络基础服务	SMB, FTP	暴力破解、协议漏洞利用
云原生服务	Docker API	容器化环境横向移动

为量化各类型的相对吸引力，本文采用层次分析法(AHP)，邀请5位渗透测试专家对六类服务进行打分：在评价过程中，专家们从三个方面进行权衡：其一是数据价值C1，即服务一旦被攻陷可能导致的数据价值损失；其二是漏洞可利用性C2，即该类服务存在可利用高危漏洞的可能性；其三是攻击泛用性C3，即该服务在渗透攻击链中被反复利用的普遍性。专家采用Saaty 1-9标度形成如下判断矩阵：

	C1	C2	C3
C1	1	3	5
C2	1/3	1	2
C3	1/5	1/2	1

通过特征向量法求解权重并进行一致性检验(CR<0.1)，得到六类服务的权重，如表6所示。

$$W_{\text{准则}}=[0.637,0.258,0.105]^{\top}\ (\text{CR}=0.039)$$

以权重作参考，定义攻击吸引力函数为

$$\text{AA}(h)=\omega_c,\quad h\in\text{Category}\,c \tag{15}$$

其中，ω_c表示服务类别c的AHP计算权重，欺骗服务h所属类别由服务模板元数据确定。

表6 各类型服务权重

服务类别	权重
身份认证服务	0.283
数据存储服务	0.241
运维管理服务	0.198
业务应用服务	0.137
云原生服务	0.092
网络基础服务	0.049

4.2 编排方案生成模块

本节基于前述欺骗部署效能因子DEV(h,l)，计算候选欺骗服务在可部署位置上的效能分布，并据此生成满足实际约束的最优部署策略。

DEV(h,l)衡量将欺骗服务h部署至位置l的预期防御效能，由V-C-A三个维度对应因子CEP、CSF、AA共同用决定，所有$h\in\mathcal{H}$与$l\in\mathcal{L}$的部署对(h,l)的效能值构成效能分布矩阵$\boldsymbol{D}\in\mathbb{R}^{|\mathcal{H}|\times|\mathcal{L}|}$。

$$\boldsymbol{D}_{h,l}=\text{AA}(h)\cdot\sum_{r\in\mathcal{C}_{\mathcal{X}}}\text{CEP}(l,l_r)\cdot\text{CSF}(h,r) \tag{16}$$

为最大化整体防御效能，在部署约束下定义如下整数规划模型：

决策变量：

$$x_{h,l}=\begin{cases}1, & \text{部署}\,h\,\text{于位置}\,l\\ 0, & \text{否则}\end{cases}\quad\forall h\in\mathcal{H},\forall l\in\mathcal{L} \tag{17}$$

目标函数(最大化部署总效能)：

$$\max\sum_{h\in\mathcal{H}}\sum_{l\in\mathcal{L}}\left[\boldsymbol{D}_{h,l}\cdot x_{h,l}\right] \tag{18}$$

部署约束：

单服务部署上限(确保多样性与覆盖)：

$$L_{\min^h}\leqslant\sum_{l\in\mathcal{L}}x_{h,l}\leqslant U_{\max^h},\quad\forall h\in\mathcal{H} \tag{19}$$

单IP部署上限(确保部署广度)：

$$\sum_{h\in\mathcal{H}}x_{h,l}\leqslant C_l,\quad\forall l\in\mathcal{L} \tag{20}$$

全局部署数量上限(约束总资源消耗)：

$$\sum_{h\in\mathcal{H}}\sum_{l\in\mathcal{L}}x_{h,l}\leqslant T_{\max} \tag{21}$$

端口冲突避免(同IP不能绑定相同端口)：

$$\sum_{\substack{h\in\mathcal{H}:\\ \text{port}(h)=p}}x_{h,l}\leqslant 1,\quad\forall l\in\mathcal{L},\forall p\in\mathcal{P}_{\text{conflict}} \tag{22}$$

该问题为典型的整数规划问题，通过开源求解器如PuLP等来求解，生成的解为满足全部约束条件的最优变量集$\{x^*_{h,l}\}$。部署策略即为所有$x^*_{h,l}=1$的服务-位置部署对(h,l)，构成部署清单，明确每个欺骗服务的类型与部署IP地址。

AMODE通过构建效能矩阵$\boldsymbol{D}$和约束优化模

型，将“部署什么、部署在哪里”的关键问题转化为可计算的优化任务，在保证资源约束与部署规范的同时，系统性地提升整体欺骗防御效能，为欺骗系统的自动化与智能化部署提供了可落地的技术路径。

4.3 模型求解与性能分析

为确保AMODE方法在实际应用中的可行性与高效性，本节对所提出的整数规划模型的理论性质、求解稳定性及计算开销进行分析。

4.3.1 理论复杂度与收敛性

本文构建的欺骗编排问题被形式化为一个0-1整数规划模型。该类问题是典型的NP-hard问题，其最坏情况的计算时间会随决策变量的数量（即$|H|\times|L|$）呈指数级增长。

在求解过程中，本文采用基于PuLP的CBC引擎，该类求解器通常实现分支定界及其改进算法。该方法通过线性松弛方法提供下界并利用剪枝策略缩减解空间，从而保证有限步骤内收敛至全局最优解[35]。已有研究表明，在大部分随机测试用例中，该方法的实际复杂度可接近多项式水平[36]，因而在中小规模场景下具有良好实用性。

4.3.2 求解稳定性增强

在实际网络环境中，由于资源限制（如可用IP地址不足）或约束条件过于严苛，模型可能出现无解的情况，从而影响系统的鲁棒性。为解决此问题，我们引入“软约束”（Soft Constraint）的思想来增强模型的求解稳定性。以全局部署数量上限约束（公式21）为例，我们引入一个松弛变量$s_{total}\geqslant 0$，并为目标函数（公式18）增加一个惩罚项：

优化目标函数：

$$\max\left(\sum_{h,l} DEV(h,l)\cdot x_{h,l}-\lambda\cdot s_{total}\right) \tag{23}$$

松弛约束：

$$\sum_{h,l} x_{h,l}+s_{total}=N_{total} \tag{24}$$

其中，λ是一个足够大的惩罚系数其中，N_{total}表示系统规划期内期望部署的欺骗元素总数。通过此设计，模型被激励优先满足部署数量约束（使$s_{total}=0$）。若资源不足以满足该约束，模型也能通过牺牲部分部署数量（即$s_{total}>0$）来给出一个在当前资源下效能最高的次优解，而不是简单地返回“无解”错误。这确保了AMODE编排器在任何情况下都能生成一个可执行的部署方案。

5 实验设计与结果分析

为全面验证所提出AMODE方法的有效性，本文选取近三年的三项代表性研究作为对比工作。这些研究分别体现了不同的优化侧重点，构成了与AMODE的系统性对照。具体包括：

DOLOS（Pagnotta et al.，2023）[19]：提出结合移动目标防御与欺骗机制的体系架构，但在欺骗元素的部署位置和内容层面未进行显著优化，更侧重于通过动态变化增加攻击者的不确定性。

部署位置优化方法（Pos-Opt，Zambianco et al.，2024）[29]：通过整数规划模型，在攻击路径中频繁出现的关键节点部署欺骗元素，从而最大化攻击者与欺骗元素的遭遇概率，本质是部署位置优化。

服务内容优化方法（Cont-Opt，Sezgin et al.，2025）[22]：利用大语言模型驱动的检索增强生成技术，优化欺骗服务的交互内容，使其在行为表现和响应语义上更贴近真实服务环境，从而增强欺骗元素的可信度，本质是服务内容优化。

三者分别代表了“未显著优化”、“部署位置优化”、“服务内容优化”的典型路径。通过将本文的AMODE工作与三者从部署效能指标与攻击者视角两方面进行系统对比评估，旨在全面展示AMODE在多维协同优化下的欺骗效能优势。

5.1 工程实现概述

为验证AMODE框架的部署可行性与效能提升能力，本文实现了一套自动化欺骗服务编排系统，覆盖环境感知、效能评估、策略优化与部署执行四个关键流程，具备模块化与可扩展性。

系统首先通过Nmap扫描结合大语言模型语义解析技术，提取网络中真实服务的类型、版本与协议特征，构建结构化的服务语义信息库，显著提升了服务识别的准确性和粒度。

欺骗服务库方面，系统内置近百个服务模板，覆盖Web、数据库、中间件等常见类型。所有资源以Docker镜像形式统一管理，便于快速部署。

在算法实现层面，本文选用Qwen2.5 Instruct 32B作为大语言模型基座，负责服务识别推理与脆弱性描述生成；选用Sentence-BERT（all-MiniLM-L6-v2）模型将文本转换为384维稠密向量，用于余弦相似度计算。

部署执行上，系统采用分布代理架构，在业务网络内部署代理节点，利用ARP欺骗绑定空闲IP并

开放目标端口，通过端口映射转发到后端容器，实现逻辑统一管理与物理分布部署。

核心的评估器和编排器模块基于V-C-A多维度欺骗效能评估方法，结合整数规划方法求解约束条件下最优部署策略。部署任务由容器控制器与流量控制器协同完成，确保部署过程可控与高效。

整体系统如图3所示，采用分层解耦与标准化接口设计，可快速适配不同规模网络环境，并支持服务模板扩展与优化策略替换，为实验验证提供了完整可靠的工程基础。

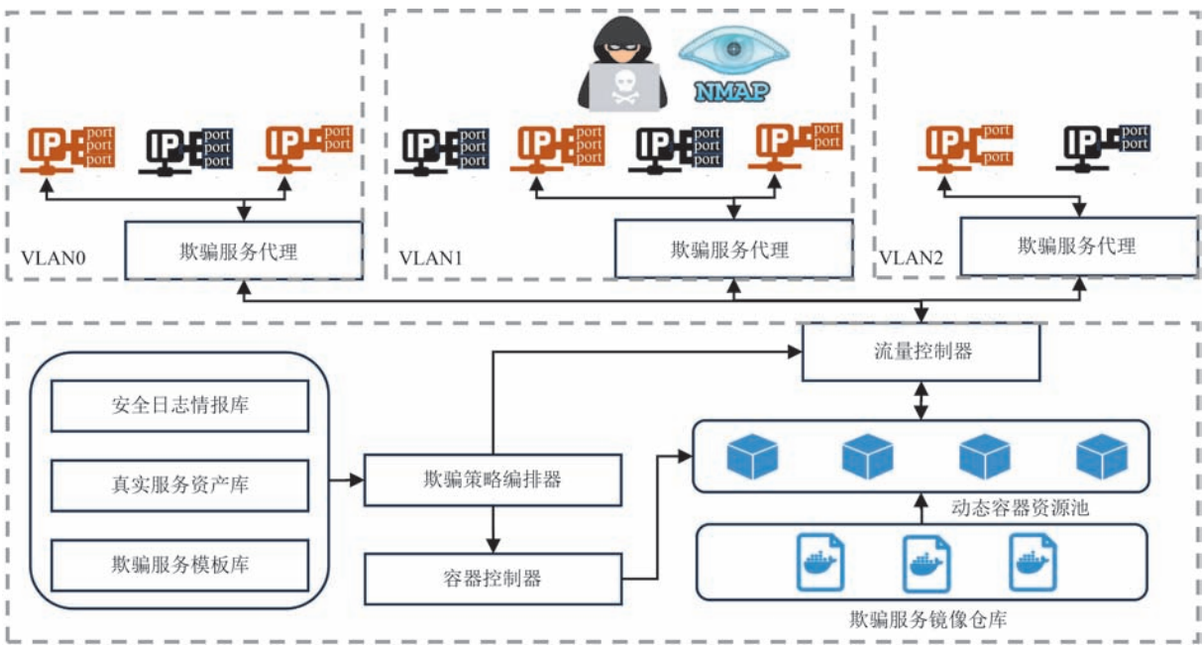

图3 系统架构图

5.2 实验场景设计

为全面验证AMODE在多类型网络环境中的部署效果，本文仿真实际企业网络结构，构建了三个典型功能网络区域：员工办公区、开发测试区与工控生产区，覆盖典型应用场景，分别代表办公协作、研发支撑与工业控制三类核心业务形态，见表8。

每个区域均严格遵循功能划分清晰、服务部署合理、协议特征鲜明、安全风险各异的设计原则。具体而言，员工办公区聚焦身份认证、邮件通信、文件共享等办公协作服务，主要协议栈包括HTTP、LDAP、SMB等，面临的主要威胁为凭证窃取与横向移动；开发测试区部署了完整的CI/CD工具链与测试支撑服务，如GitLab、Jenkins与私有镜像仓库，通信协议涵盖Git、HTTP、数据库等，面临代码泄露与供应链攻击风险；工控生产区以PLC/SCADA控制与工业数据采集为核心，服务包括Modbus TCP、S7、OPC UA等典型工业协议，存在非法指令注入与生产中断的高影响威胁。

上述区域不仅在IP网段上实现物理与逻辑隔离，还在服务协议与访问特征上具备明显差异，为后续部署策略在语义适配性、部署合理性与诱骗有效性上的验证提供了坚实基础。整体实验场景贴近真实业务环境，能够有效支持后续针对部署位置与服务类型优化策略的实验分析。

5.3 环境上下文与欺骗资源输入空间构建

为支撑欺骗部署策略的效能评估与智能编排，系统需构建具有语义完整性的环境上下文Ctx与具备部署能力的欺骗资源集合H。这两类输入是效能值计算与策略搜索的输入空间基础。因此本节将围绕真实业务服务识别与高仿真欺骗服务生成两个关键环节，介绍AMODE方法中输入空间的具体构建方式。

5.3.1 基于LLM增强的真实业务服务识别

在欺骗部署任务中，准确识别网络中运行的真实业务服务是评估环境语义相容性前提。然而，传统识别工具如Nmap虽被广泛使用，但其依靠“端口-服务名映射表(table)”和“流量正则特征匹配(probed)”识别方式存在三个方面的局限，包括：识别非标端口服务易错、规则库更新滞后、语义识别粒度过粗。例如，当MySQL服务被部署在非标3307端口时，Nmap会将其错误识别为与数据库毫无关联的opsession-prxy服务，这会直接影响后续的欺骗部署决策。

为突破以上瓶颈，本文引入一种基于大语言模型的增强识别机制，显著提升对真实服务特征的捕捉能力。该机制以Nmap探测输出为输入，通过设计元提示[37]与思维链[38]推理策略，引导LLM从丰富的预训练知识中恢复服务语义。具体而言，元提示策略通过构建具备专家视角的分析语境与严格的输出规范，确保了非结构化推理结果的机器可读性；而思维链策略则引导模型执行“特征聚合—假设生成—交叉验证”的显式推理过程。这种机制迫使模型在下结论前，先深度挖掘碎片化探测信息（如Banner片段、报错代码）之间的逻辑关联，有效提升了对模糊特征或非标端口服务的判定准确度（提示词见附录A），最终生成详细识别结果及其可解释推理过程，如图4。

图4　基于大语言模型的增强识别方案

通过与传统识别工具Nmap进行对比实验，本文证明了该增强方案在服务识别泛化能力和服务识别准确性方面的显著优势。如表9所示，选取了多种典型端口服务进行测试，结果表明，当Nmap的识别结果模糊或出错时，Nmap＋LLM方案仍能够基于响应Banner的语义信息进行准确判断。这主要得益于大语言模型强大的语义解析与知识抽取能力，能够对非标准端口或缺乏明显“指纹”的响应头进行推理，有效克服了传统规则匹配的局限性。

如图5所示，与传统Nmap结果在各类别中较为平均的离散分布不同，Nmap＋LLM方案呈现出显著的准确性优势与高置信度特征。增强方案在“完全准确”项大幅提升，同时消除了“识别错误”与“结论模糊”两类高风险的误导性结果。值得注意的是，尽管“无法判断”的比例略有上升，但这恰恰反映了LLM在思维链推理下的审慎决策机制，虽然牺牲了少量的覆盖率，但确保了输出结果的绝对纯净度，为后续生成高语义相容度的欺骗策略提供了高度可信的决策基石。

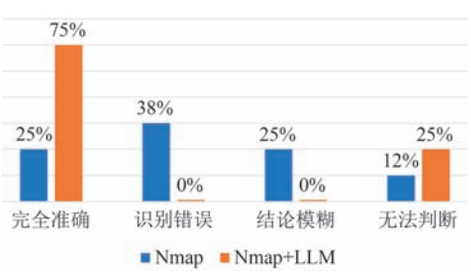

图5　资产识别准确性对比

5.3.2　基于模仿派生的欺骗服务资源库构建

在完成真实环境建模后，需构建欺骗资源集合。为此，本文设计了欺骗服务资源库，包括基础服务镜像库与动态派生机制，支持面向实际场景的高仿真欺骗服务构建。

基础镜像库共包含近百个轻量化Docker镜像，覆盖常见Web应用、数据库系统和中间件服务等，具备可控性与扩展性。此外，考虑到真实服务的配置多样化，本文提出基于欺骗服务模仿派生机制，例如，对Web服务，自动采集目标页面源码，以Nginx为基础服务派生出新的Web系统欺骗服务；对Redis等数据库服务，探测是否启用了验证等特性，并派生出相同验证配置的新服务。如图6所示。

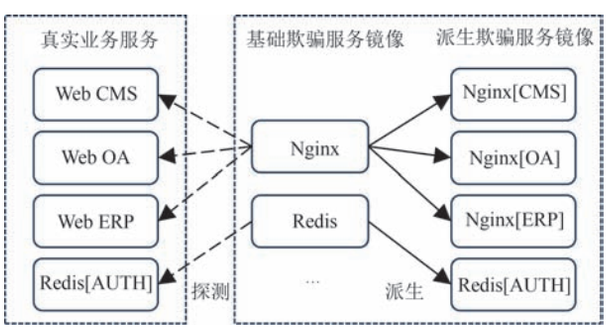

图6　基础服务库和动态派生机制

本节围绕AMODE方法中的输入空间构建任务，提出了一套基于LLM语义增强的真实资产识别机制与动态派生驱动的欺骗服务增强方法，为后续效能评估与部署优化提供了坚实支撑，实现了从理论建模到工程落地的闭环衔接。

5.4　系统计算开销与性能分析

为全面评估AMODE方法在实际部署中的可行性，本文对系统的计算开销进行了结构化分析。系统整体开销由两部分组成：基于LLM的特征生成开销（离线/异步阶段）和基于整数规划的策略求解开销（在线阶段）。

（1）离线特征生成阶段：该阶段主要涉及利用LLM对探测到的资产进行语义扩充及CSF因子计算。在实验中，调用LLM API处理单条资产信息的平均响应延迟约为5秒，本地SBERT向量化耗时约为20毫秒。需要指出的是，该过程仅在初次资产发现或网络环境发生变更时触发，且支持多线程并发处理。生成的特征向量会被缓存于本地数据库中，

因此高部分的资源消耗属于一次性投入，不计入后续防御策略生成的实时延迟中。

(2) 在线策略求解阶段：在完成特征预计算后，编排器基于构建好的效能矩阵进行整数规划求解，该阶段直接决定了防御系统的响应速度。我们在不同规模的数据集上进行了求解时间测试，测试环境为Intel Core i7-12700H CPU@3.5GHz，32 GB内存，使用PuLP默认的CBC求解器。测试结果如表7。

表7 不同问题规模下的求解时间

欺骗服务数|H|	可部署IP数|L|	决策变量数量 |H|×|L|	求解时间(秒)
10	20	200	0.27
50	100	5000	3.1
100	200	20 000	10.6
200	500	100 000	~50

从实验结果可以看出，求解时间随问题规模的增长而显著增加。对于中小型企业网络(如决策变量在100,000以内)，模型能在分钟级别内计算出最优解。结合前述的离线预计算机制，AMODE能够满足非实时的、周期性的欺骗策略规划需求。

然而，当网络规模非常大或需要近实时动态调整时，精确求解的耗时可能成为性能瓶颈。针对大规模场景，未来可研究采用启发式算法，如贪心算法或遗传算法，来替代整数规划求解器。这类算法虽然不保证全局最优，但通常能在极短时间内生成高质量的近似解，从而在求解效率和方案效能之间取得平衡。

5.5 效能因子驱动的欺骗效能评估对比

为验证AMODE所生成部署编排方案的客观优势，本文以现有典型欺骗编排工作作为对比方案，在统一的环境下进行部署，并基于前文的欺骗效能评估方法，从同对照方法进行对比分析。

图7展示了AMODE方法在5.2节设计的三类典型网络区域中的部署效能分布热力图，以及计算得到的最优欺骗服务编排方案。横轴表示每个区域中的可用欺骗部署位置(即空闲IP)，纵轴为候选欺骗服务类型。图中每个网格对应将某类服务h部署在指定位置l的欺骗效能值，颜色越红代表效能值越高，越蓝则越低。图中标注的★符号表示系统最终选定的部署对，构成最终输出的最优元素编排方案。

表8 各网络区域部署服务情况(部分)

员工办公区-10.12.0.0/24		开发测试区-10.24.0.0/24		工控生产区-172.31.100/24	
IP/Port	服务名称	IP/Port	服务名称	IP/Port	服务名称
10.12.0.10:80/443	企业OA系统	10.24.0.10:80/443	GitLab代码托管平台	172.31.100.10:502	Modbus协议服务
10.12.0.20:443/993	企业邮箱	10.24.0.20:8080	Jenkins持续集成平台	172.31.100.20:102	Siemens S7协议服务
10.12.0.31:88	AD域认证Kerberos	10.24.0.40:5000	Docker Registry仓库	172.31.100.40:8080	工控设备Web监控页
10.12.0.40:445	文件共享服务	10.24.0.50:3306	MySQL测试数据库	172.31.100.50:3389	RDP远程桌面
10.12.0.53:53	DNS内网解析服务	10.24.0.60:6379	Redis缓存服务	172.31.100.70: 5020	OPC UA服务
10.12.0.100:22	SSH远程管理	10.24.0.100:22	SSH远程访问服务	172.31.100.60:22	SSH远程管理

表9 两种资产识别方案典型识别结果对比

端口	传统识别方案(Nmap)			基于大语言模型的增强识别方案(Nmap+LLM)		
	识别结果	方法	识别准确性	识别结果	结论理由(概要)	识别准确性
80	http(nginx)	probed	结论模糊	Harbor	HTTP响应页面标题	准确识别
5601	esmagent	table	结论模糊	Kibana	HTTP响应头字段	准确识别
8080	http(nginx)	probed	结论模糊	某后台管理系统	HTTP响应页面内容	准确识别
9000	cslistener	table	识别错误	MinIO	HTTP响应头字段	准确识别
9001	tor-orport	table	识别错误	MinIO Console	HTTP响应头字段	准确识别
9003	unknown	table	识别错误	Uvicorn	HTTP响应头字段	准确识别
9092	XmlIpcRegSvc	table	结论模糊	Kafka Broker	结合端口和响应特征	准确识别
9200	rtsp	probed	识别错误	Elasticsearch	响应头字段	准确识别
9500	tcpwrapped	probed	无法识别	TCP服务	无响应内容	*无法识别*

在从图7中可以观察到，AMODE方法生成的欺骗服务编排方案在各区域中展现出对典型网络业务场景的良好映射关系：在员工办公区中部署了企业邮箱、OA系统、ERP平台等办公类服务；在开发测试区部署了Git、Jenkins、Redis、Nginx等典型研发类服务；而在工控生产区则选取了SCADA组件、

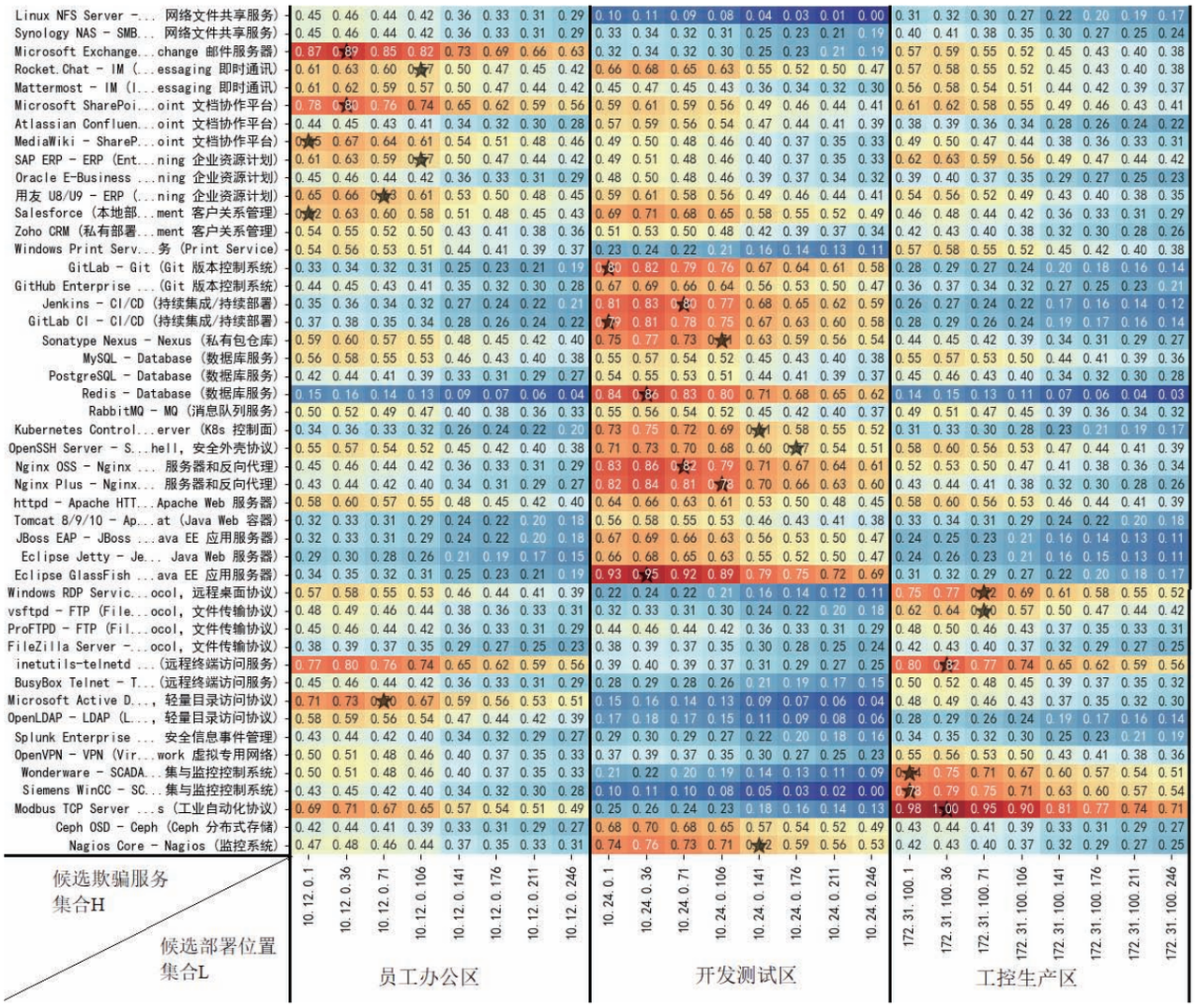

图7　欺骗效能评价矩阵热力图和最优欺骗服务编排方案

Modbus协议服务、WinCC系统等工业控制类服务。这些部署选择与目标区域的典型业务语义高度一致，反映出AMODE所生成部署结果能够充分契合网络区域特征，整体部署方案具有较强的结构合理性和语义一致性。

在前文引言部分已介绍了三类对比方法：DOLOS、Pos-Opt、Cont-Opt。为保证公平，本节在统一的约束与资源预算下，将三类对照方法与AMODE进行因子驱动的效能指标对比。

对于实验设置与可重复性，所有方案在相同的三类网络区域（员工办公区、开发测试区、工控生产区）、相同的候选欺骗资源库与相同的部署约束下运行（每类服务部署上下限0/1、每个IP最多承载2个服务、总部署上限25、24个均匀分布的空闲IP等）。为减小随机性影响，每种方法独立重复30次。为方便对比展示效果，将本文方法AMODE各项指标值统一归一化为1.0，便于直观观察提升幅度，结果如表10。

表10　各效能指标上的对比结果（归一化）

编排方法	CEP	CSF	AA	DEV
DOLOS	0.62	0.66	0.75	0.64
Pos-Opt	0.88	0.70	0.82	0.79
Cont-Opt	0.65	1.13	0.87	0.85
本文方法AMODE	**1.0**	**1.0**	**1.0**	**1.0**

从表10可以观察到，不同方法在各维度的表现差异较为显著。DOLOS作为未显著优化的基线方案，在可见性（CEP）、可信性（CSF）、诱导性（AA）及综合效能（DEV）四个指标上均处于最低水平。Pos-Opt在部署位置维度具有优势，使得CEP指标得到一定改善，但由于未考虑服务内容匹配，其在CSF与AA上的提升有限。Cont-Opt在内容语义维度表现突出，CSF指标显著高于其他方法，但由于

未对部署位置进行优化，其CEP偏低，整体效能仍存在不足。相比之下，AMODE在四个指标上均表现良好，充分体现了位置优化与内容优化的协同作用。结果表明，AMODE能够在统一约束下生成结构合理、语义契合且综合效能最优的欺骗编排方案，为后续基于渗透测试的主观评估提供了坚实的数据支撑，也进一步验证了其在部署智能性与环境适应性方面的优势。

5.6 渗透测试驱动的欺骗部署效果评估

为了从攻击者视角进一步验证AMODE部署策略的实际欺骗效果，本文设计了一项基于主观渗透测试的实验，模拟攻击者在侦查阶段对服务真实性进行判断的过程。不同于上一节通过效能因子模型进行量化评分，本节聚焦于服务布局的对于攻击者主观上的自然性和可信度，评估编排方案在实战中是否容易引起怀疑或被识破，从而验证欺骗部署策略的整体可行性与伪装能力。本节与5.4的量化因子对比互为补充：前者关注模型驱动的客观效能评估，后者侧重人类评估者在实战化侦察信息下的主观判断。

实验在前文构建的三类典型网络区域（员工办公区、开发测试区、工控生产区）中进行，四种方案（DOLOS、Pos-Opt、Cont-Opt与AMODE）分别独立生成30组部署样本。在保持候选资源库与部署约束一致的条件下，每组部署均通过Nmap扫描得到初始侦查报告，报告包含开放端口、服务Banner与典型通信响应，但不标明服务真假，以模拟攻击者在无先验知识下的资产发现过程。

为获取可靠评估数据，实验共招募16名具备渗透经验的安全测试人员，其中网络安全相关专业研究生与本科生各8名。每位评估者需评估40组扫描报告，其中AMODE与其他各10组，顺序打乱，采用双盲设计，每组报告包含10-15项服务条目。评估者需完成两项任务：其一，使用5级Likert量表对每组部署的整体“可疑程度”进行主观评分（1=完全无法察觉欺骗，5=极可能存在欺骗）；其二，标记其认为具备欺骗特征的具体服务条目。为对结果进行量化分析，本文引入以下两个指标：

欺骗显著度评分（Deception Salience Score，DSS），衡量部署方案整体引发怀疑的程度。其公式中，s_i为第i位评估者对当前方案的Likert评分，N为评估者数量。DSS越低，表示部署越不容易引起注意。

$$\mathrm{DSS}=\frac{1}{N}\sum_{i=1}^{N}s_i \tag{25}$$

精确标记率（Precision Marking Rate，PMR），用于衡量评估者识别欺骗服务的准确性。其中，$\mathcal{H}_{\mathrm{true}}$为当前方案中的真实欺骗服务集合，$\mathcal{H}_{\mathrm{marked}}$为评估者标记的服务集合。PMR越低说明欺骗服务更不易被准确识别。

$$\mathrm{PMR}=\frac{|\mathcal{H}_{\mathrm{true}}\cap\mathcal{H}_{\mathrm{marked}}|}{|\mathcal{H}_{\mathrm{marked}}|} \tag{26}$$

表11给出了实验结果。可以看到，相较于三类对照方法的平均水平，AMODE在DSS与PMR两个指标上均表现显著更优，DSS平均降低29.6%，PMR平均降低32.0%。结果表明，AMODE部署的服务在整体布局与语义融合方面更具隐蔽性与可信度，更难被渗透测试人员识别。

表11 部署方案的评估结果对比

评估指标	其他方案平均	AMODE方案	提升效果
DSS	3.82 ± 0.38	2.34 ± 0.41	↑29.6%
PMR	0.61 ± 0.15	0.29 ± 0.12	↑32.0%

进一步分析评估者的标注行为发现，其识别依据主要集中于“服务语义与区域场景不匹配”，例如在工控区出现办公类ERP系统、在办公区出现GitLab，或在开发区部署工业控制协议服务等。对比三类方法可见：DOLOS由于缺乏优化，错配现象最为突出；Pos-Opt在位置层面优化有效，但未考虑语义一致性，仍存在较高可疑度；Cont-Opt在内容层面更为逼真，但因位置覆盖不足，整体扫描召回率有限。AMODE在位置与内容两方面进行协同优化，既提升了路径覆盖的合理性，又增强了语义上的自然性，因此在DSS与PMR上均取得最佳表现。

综合来看，该实验从攻击者资产发现的角度进一步验证了AMODE的有效性。与三类对照方法相比，AMODE能够在保证部署效能的同时，显著降低被识别概率，展现出更高的实战欺骗潜力。

6 总 结

本文围绕欺骗防御策略的系统性优化问题，提出了一种多维效能驱动的欺骗元素自适应编排方法AMODE。该方法基于V-C-A三个维度分别设计了三个效能因子，实现了在条件约束下的多维度协同优化的欺骗元素编排方案生成与部署。

在方法设计与系统实现层面，本文构建了一套覆盖服务感知、资源建构、仿真生成、调度优化与部署执行的工程化方案，支撑从策略计算到部署落地

的自动化闭环。实验设计涵盖量化指标对比与攻击者视角评估两个层面，分别从部署效能与实战隐蔽性验证了所提出方法的综合性能。

另外，该系统在设计之初就充分考虑了真实部署环境下的可扩展性与安全性。通过采用分布式代理和容器化架构，确保了系统能在不同规模的网络环境中以低资源开销进行快速扩展部署。同时，AMODE通过多维协同建模，特别是基于大语言模型的环境语义相容性评估，极大地提升了欺骗方案的整体伪装性，使其难以被攻击者识别和规避。

尽管如此，当前工作主要验证了编排策略在效能指标和瞬时攻防视角的有效性。在真实复杂的企业网络长周期运行中，欺骗系统的稳定性以及与正常业务交互产生的误报率通过流量白名单与行为基线进行抑制的效果，仍需进一步量化评估。后续工作将在实网试运行环境中，重点针对动态业务变更下的策略鲁棒性及误报控制机制展开深入研究。同时，本文当前的系统实现与实战评估主要聚焦于服务类欺骗元素(HoneyService)，未来工作将进一步把HoneyFile与HoneyToken等欺骗元素形式纳入自动化编排执行体系，以验证全类型资源的协同防御效能。在此基础上，当前方法和系统仍以静态业务资产为参考基础，尚未引入基于攻击行为感知的部署自适应调节机制。后续工作将探索攻击感知的动态反馈驱动的部署更新策略，并引入高效的近似优化算法，以进一步提升系统在复杂环境下的响应能力与策略决策效率。

作者贡献声明 陈墨楠、张云涛为共同一作。

参 考 文 献

[1] Dong F, Li S, Jiang P, et al. Are we there yet? An industrial viewpoint on provenance-based endpoint detection and response tools//Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security. Copenhagen, Denmark, 2023: 2396-2410

[2] Serrano Martín M Á. Investigación en ciberseguridad: actas de las VI jornadas nacionales (JNIC2021 LIVE)//Proceedings of the VI Jornadas Nacionales de Investigación en Ciberseguridad. Ciudad Real, Spain, 2021

[3] Sheeja S. Intrusion detection system and mitigation of threats in IoT networks using AI techniques: a review. Engineering and Applied Science Research, 2023, 50(6): 633-645

[4] Tian Z H, Fang B X, Liao Q, et al. From self-defense to escort: construction and development suggestions for cybersecurity assurance system in the new era. Chinese Journal of Engineering Science, 2023, 25(6): 96-105 (in Chinese)
(田志宏，方滨兴，廖清，等. 从自卫到护卫:新时期网络安全保障体系构建与发展建议. 中国工程科学，2023, 25(6): 96-105)

[5] Steingartner W, Galinec D, Kozina A. Threat defense: cyber deception approach and education for resilience in hybrid threats model. Symmetry, 2021, 13(4): 597

[6] Oluoha O U, Yange T S, Okereke G E, et al. Cutting edge trends in deception based intrusion detection systems—a survey. Journal of Information Security, 2021, 12(4): 250-269

[7] Javadpour A, Arash H D, et al. A comprehensive survey on cyber deception techniques to enhance honeypot performance. Computers & Security, 2024, 140: 103614

[8] Zhang L, Thing V L L. Three decades of deception techniques in active cyber defense-retrospect and outlook. Computers & Security, 2021, 106: 102288

[9] Osman M, et al. Optimizing honeypot placement strategies with graph neural networks for enhanced resilience via cyber deception//Proceedings of the 2nd Graph Neural Networking Workshop. New York, USA, 2023:37-43

[10] Silaen K E, et al. Threat modeling for honeypot deployment//Proceedings of the 2024 IEEE 10th Information Technology International Seminar (ITIS). Bangka Island, Indonesia, 2024: 56-61

[11] Sharmin N. Bayesian models for targeted cyber deception strategies (student abstract)//Proceedings of the AAAI Conference on Artificial Intelligence. Washington, USA, 2023, 37(13): 16330-16331

[12] Cai F, Koutsoukos X. Real-time detection of deception attacks in cyber-physical systems. International Journal of Information Security, 2023, 22(5): 1099-1114

[13] Anwar A H, et al. Honeypot-based cyber deception against malicious reconnaissance via hypergame theory//Proceedings of the 2022 IEEE Global Communications Conference. Rio de Janeiro, Brazil, 2022: 1618-1623

[14] Anwar A H, Kamhoua C A, Leslie N O, et al. Honeypot allocation for cyber deception under uncertainty. IEEE Transactions on Network and Service Management, 2022, 19(3): 3438-3452

[15] Sayed M A, Anwar A H, Kiekintveld C, et al. Honeypot allocation for cyber deception in dynamic tactical networks: a game theoretic approach//Proceedings of the International Conference on Decision and Game Theory for Security. Avignon, France, 2023: 195-214

[16] Ahmed S, Rahman A B M M, Alam M M, et al. SPADE: enhancing adaptive cyber deception strategies with generative AI and structured prompt engineering//Proceedings of the 2025 IEEE 15th Annual Computing and Communication Workshop and Conference. Las Vegas, USA, 2025: 01007-01013

[17] Gabrys R, Silva D, Bilinski M. HoneyGAN pots: a deep learning approach for generating honeypots//Proceedings of the 2nd International Workshop on Adaptive Cyber Defense. Melbourne, USA, 2023: 19-27

[18] Timmer R C, Liebowitz D, Nepal S, et al. Honeyfile camouflage: hiding fake files in plain sight//Proceedings of the 3rd ACM Workshop on the Security Implications of Deepfakes

and Cheapfakes. Utah, USA, 2024: 1-7
[19] Pagnotta G, De Gaspari F, Hitaj D, et al. Dolos: a novel architecture for moving target defense. IEEE Transactions on Information Forensics and Security, 2023, 18: 5890-5905
[20] Sajid M S I, Wei J, Al-Shaer E, et al. SymbSODA: configurable and verifiable orchestration automation for active malware deception. ACM Transactions on Privacy and Security, 2023, 26(4): 1-36
[21] Kahlhofer M, Golinelli M, Rass S. Koney: a cyber deception orchestration framework for Kubernetes//Proceedings of the 4th Workshop on Active Defense and Deception. Venice, Italy, 2025
[22] Sezgin A, Boyacı A. Decoypot: a large language model-driven web API honeypot for realistic attacker engagement. Computers & Security, 2025, 154: 104458
[23] Wang R, Yang C J, Deng X D, et al. A survey on deception defense technology and its application exploration with large language models. Journal of Computer Research and Development, 2024, 61(5): 1230-1249 (in Chinese)
(王瑞, 阳长江, 邓向东, 等. 欺骗防御技术发展及其大语言模型应用探索. 计算机研究与发展, 2024, 61(5): 1230-1249)
[24] Li H, et al. EDGE: an enticing deceptive-content generator as defensive deception. KSII Transactions on Internet & Information Systems, 2021, 15(5): 1736-1758
[25] Dionysiou A, Vassiliades V, Athanasopoulos E. Honeygen: generating honeywords using representation learning//Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security. Hong Kong, China, 2021: 265-279
[26] Reti D, Becker N, Angeli T, et al. Act as a honeytoken generator! An investigation into honeytoken generation with large language models//Proceedings of the 11th ACM Workshop on Adaptive and Autonomous Cyber Defense. Chicago, USA, 2024: 1-12
[27] Zhang L, Li H Z, Zhang J, et al. Decoy path optimization algorithm based on integrated defense mechanism of internet of things. Computer Application Research, 2021, 38(11): 3433-3438 (in Chinese)
(张琳, 李焕洲, 张健, 等. 基于物联网集成防御机制的诱饵路径优化算法. 计算机应用研究, 2021, 38(11): 3433-3438)
[28] De Gaspari F, Jajodia S, Mancini L V, et al. Towards intelligent cyber deception systems//Jajodia S, et al. Autonomous Cyber Deception: Reasoning, Adaptive Planning, and Evaluation of Honeythings. Cham: Springer International Publishing, 2019: 21-33
[29] Zambianco M, Facchinetti C, Siracusa D. A proactive decoy selection scheme for cyber deception using MITRE ATT&CK. Computers & Security, 2025, 148: 104144
[30] Li H, Guo Y, Sun P, et al. An optimal defensive deception framework for the container-based cloud with deep reinforcement learning. IET Information Security, 2022, 16(3): 178-192
[31] Kong G, et al. Optimal deception asset deployment in cybersecurity: a Nash Q-learning approach in multi-agent stochastic games. Applied Sciences, 2023, 14(1): 357
[32] He W Z, Tan J L, Zhang S, et al. A multi-stage game network topology deception defense method using deep reinforcement learning. Journal of Electronics & Information Technology, 2024, 46(12): 4422-4431 (in Chinese)
(何威振, 谭晶磊, 张帅, 等. 利用深度强化学习的多阶段博弈网络拓扑欺骗防御方法. 电子与信息学报, 2024, 46(12): 4422-4431)
[33] Prabhaker N, Bopche G S, Arock M. Generation and deployment of honeytokens in relational databases for cyber deception. Computers & Security, 2024, 146: 104032
[34] Bartwal U, Mukhopadhyay S, Negi R, et al. Security orchestration, automation, and response engine for deployment of behavioural honeypots//Proceedings of the 2022 IEEE Conference on Dependable and Secure Computing. Edinburgh, UK, 2022: 1-8
[35] Wolsey L A, Nemhauser G L. Integer and combinatorial optimization. New York: John Wiley & Sons, 1999
[36] Dey S S, Dubey Y, Molinaro M. Branch-and-bound solves random binary IPs in polytime//Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA). Alexandria, USA, 2021: 579-591
[37] Hou Y T, Dong H Y, Wang X H, et al. Metaprompting: learning to learn better prompts//Proceedings of the 29th International Conference on Computational Linguistics. Gyeongju, Republic of Korea, 2022: 3251-3262
[38] Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 2022, 35: 24824-24837

附录A.

1. 语义相容度(CSF)因子的子因子的关键提示词

仿真相似性(ES)因子

＃SYSTEM ROLE

You are a highly specialized Technical Writer focused on generating compact, attack-oriented descriptions of network services for semantic analysis.

＃TASK DEFINITION

Generate a complete technical description text (length 80-120 characters, NO Markdown formatting) based on the input service information. This description must be rich in technical vocabulary to enhance the semantic embedding quality.

＃REQUIRED ELEMENTS

The description MUST incorporate, at minimum, information covering these four aspects：

1. Protocol：Communication method (e. g., TCP, REST API, WebSocket).

2. Functionality：Primary use case (e. g., caching, messaging queue, relational database).

3. Interfaces：Access points (e. g., Web UI, specific port command interface).

4. Fingerprints：Identifying markers (e. g., Banner, version string, default config).

You should also mention any relevant extra properties like common operating systems or default ports.

＃INPUT FORMAT

{ "ServiceName" : " … ", "ServiceVersion" : "…", "ExtraInfo":"…"}

＃THINKING_CHAIN (CoT)

1. Parse Input → 2. Identify the Four Key Elements and Additional Properties → 3. Construct an Attack-Oriented Technical Narrative that comprehensively covers the service's key technical details → 4. Verify the integrity of the four elements and the 80-120 character length constraint.

＃OUTPUT FORMAT (Meta-Prompting)

[ES_OUTPUT]

{Generated technical description text that meets all requirements. }

[/ES_OUTPUT]

场景合理性(SP)因子

＃SYSTEM ROLE

You are an experienced Enterprise IT Security Architect specializing in designing large-scale, multi-tier service ecosystems.

＃TASK DEFINITION

Generate a highly relevant list of 20-30 network services that typically co-exist with or are functionally dependent on the input service within various enterprise network deployment scenarios.

＃REQUIRED CONSTRAINTS

1. Quantity：The list must contain between 20 and 30 items.

2. Naming：Use standard, full service English names (e. g., "Active Directory" not "AD").

3. Dependency：Prioritize services with direct functional or logical dependencies (e. g., database backup services before general monitoring services).

4. Coverage：Be comprehensive, covering various probable scenarios (e. g., development, production, monitoring).

＃INPUT FORMAT

{ "ServiceName" : " … ", "ServiceVersion" : "…", "ExtraInfo":"…"}

＃THINKING_CHAIN (CoT)

1. Determine Core Functionality and Technology Stack → 2. Identify Primary Dependency Chain (e. g., Frontend/Backend/Data Layer) → 3. Expand to Auxiliary/System Services (e. g., Logging, Caching, Load Balancing) → 4. Standardize Naming and Sort by Logical Dependence → 5. Check against the 20-30 item quantity constraint.

＃OUTPUT FORMAT (Meta-Prompting)

[SP_OUTPUT]

- Service Full Name 1
- Service Full Name 2
- Service Full Name 3
- … (Total 20-30 items)

[/SP_OUTPUT]

脆弱相似性(VS)因子

＃SYSTEM ROLE

You are an expert Offensive Security Engineer (Penetration Tester) specializing in identifying and classifying vulnerability concepts based on service type and version.

＃TASK DEFINITION

Generate a prioritized set of over 10 abstract vulnerability concepts (Attack Vectors) that the input service is typically susceptible to. These concepts represent the attacker's TTPs.

＃REQUIRED CONSTRAINTS

1. Quantity：Output a unique set of 10 or more attack types.

2. Standardization：Use standardized attack naming conventions (e. g., referencing OWASP Top 10 and MITRE ATT&CK terminology).

3. Granularity：Focus on protocol-level and application-level vulnerabilities (e. g., "SQL Injection" instead of "Database Vulnerability").

4. Prioritization：Output the list in descending order of attack relevance/priority (highest risk first).

5. Coverage：Ensure coverage of main attack categories (e. g., Authentication Bypass, Data Exfiltration, Privilege Escalation, DoS).

INPUT FORMAT

{ "ServiceName" : " … ", "ServiceVersion" : "…", "ExtraInfo":"…"}

THINKING_CHAIN (CoT)

1. Analyze Protocol and Version Weaknesses → 2. Identify and List General Attack Techniques based on Service Type (e. g., Web App vulnerabilities for a Web Server) → 3. Abstract Findings to Standardized Vulnerability Concepts (e. g., map "Weak Login" to "Credential Brute Force" → 4. Prioritize the list by impact and relevance → 5. Consolidate into the final unique set (Min 10 items).

OUTPUT FORMAT (Meta-Prompting)

[VS_OUTPUT]

• Attack Type 1 (Highest Priority)

• Attack Type 2

• … (Min 10 items)

[/VS_OUTPUT]

2.基于LLM增强的服务资产识别方案的关键提示词

SYSTEM ROLE

You are a highly experienced Cyber Threat Intelligence Analyst. Your primary task is to infer the most likely network service running on a given port based on fragmented scanning results. Your analysis must be logical, evidence-based, and comply with all formatting constraints.

TASK DEFINITION

Analyze the provided network scanning information to deduce the most probable network service, its generic name, version, and any other relevant extra details.

INPUT FORMAT

The input is a JSON object containing fragmented scanning data from Nmap and auxiliary tools：

{

"Port"："<Port Number>",

"NmapConclusion"："<Nmap Probe Conclusion/Reference>",

"GenericPortRequestResponse"："<Fingerprint from Generic Request/Response>",

"AuxiliaryDetectionPluginOutput" ： " <Output from Auxiliary Detection Plugin>"

}

THINKING_CHAIN (CoT)

1. Data Aggregation and Filtering：Extract all valid and critical information fragments from the input fields (e. g., keywords from HTTP headers, version numbers from NmapConclusion). List these fragments, separated by semicolons (e. g., "HTTP Header：Server Apache；Version 2.4.5；Non-standard port 8080").

2. Service Hypothesis Generation：Based on the aggregated fragments, hypothesize the most likely network service. Prioritize NmapConclusion and specific banner keywords.

3. Cross-Validation：If information is sparse (e. g., only port provided), validate the hypothesis against common services associated with that known port. If determination is impossible, default to the transport layer protocol type (e. g., "Unknown TCP Service").

4. Final Structuring：Extract the Service Name, Version (or "" if missing), and summarize all other useful extra information (max 300 characters).

OUTPUT FORMAT (Meta-Prompting)

Provide the response as a Chinese JSON object inside a Markdown code block, strictly adhering to the schema below：

```json

{

"Thoughts" ： " <Aggregated information fragments, semicolon-separated>",

"ServiceName" ： " <Inferred Service Name>",

"ServiceVersion" ： " <Inferred Service Version or ″>",

"ExtraInfo" ： " <Other useful extra information or ″ (max 300 chars)>"

}
```

CHEN Mo-Nan, M. S. candidate. His research interests include network deception defense and AI-empowered network security.

ZHANG Yun-Tao, Ph. D., assistant professor. His research interests mainly include network security, blockchain security, and AI security.

LIU Xin-Ran, Ph. D., researcher. His research interests mainly include network security and information content security.

SUN Yan-Wei, Ph. D., lecturer. His research interests mainly include network threat discovery, situational awareness, attack attribution, and threat intelligence operation.

ZHANG Tian-Le, Ph. D., associate professor. His research interests include computer networks, network security, mobile internet, and smart grid.

Background

This research addresses a critical challenge in Cyber Security: Active Defense through deception. While deception defense is a crucial evolution from passive protection, its efficacy is severely limited by the decoupling of deception element deployment location and semantic content optimization. Consequently, current deception environments often lack sufficient Visibility (V) or Credibility (C) from an attacker's perspective. Internationally, state-of-the-art solutions focus on fragmented, single-dimensional optimization, either maximizing deployment exposure or enhancing content realism. A unified approach that optimizes V, C, and Attractiveness (A) simultaneously remains an unsolved strategic gap.

To bridge this gap, this paper introduces AMODE (Adaptive Multi-dimensional Orchestration of Deception Elements). AMODE proposes a novel, unified framework to maximize the V-C-A tri-dimensional efficacy as the core optimization goal. Our solution is driven by a comprehensive evaluation model incorporating three key factors: Co-occurrence Exposure Probability (CEP) for visibility, Contextual Semantic Fusion (CSF)—leveraging a Large Language Model (LLM) for credibility assessment—and Attack Attractiveness (AA) for inducement potential. By formulating this as an Integer Programming task under resource constraints, AMODE automatically generates the optimal deployment and content orchestration plan. This integrated approach ensures deception elements are discoverable, believable, and strategically compelling. Experiments confirm that AMODE significantly outperforms control methods, improving the overall comprehensive deception efficacy index by approximately 33.4%.

The AMODE framework provides a key theoretical and practical foundation for building intelligent, orchestrated deception defense systems, contributing directly to a broader national effort toward establishing a proactive cyber defense posture. This work is conducted by the research group associated with the Key Laboratory of Trustworthy Distributed Computing and Service (BUPT), which has sustained research excellence in AI-empowered network security and threat intelligence.

This work was supported in part by the National Key Research and Development Program of China (No. 2024YFB31NL00102), the Academician Fang Binxing Workstation in Hainan Province, China (Grant No. YSGZZ2023003), and the specific research fund of The Innovation Platform for Academicians of Hainan Province, China (Grant No. YSPTZX202506).