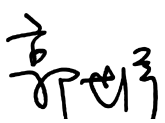
“网络空间抗测绘理论与技术”丛书序

关键信息基础设施(简称“关基设施”)是信息时代经济社会高效运行的神经中枢、网络空间安全的重中之重，是关乎国家安全的命门所在。近年来，随着网络攻防对抗的体系化升级与威胁手段的智能化演变，网络空间安全博弈已步入深水区，我国关基设施正面临日益复杂的隐蔽性探测与高频次扰动。习近平总书记多次强调，要加快构建关键信息基础设施安全保障体系。没有抗测绘，便无真安全；看不住家门，防线必被穿透。

网络空间抗测绘技术正是化被动为主动的战略抓手，旨在保护各类网络空间资源免遭非法测绘，使受保护网络对非法测绘者而言“测不了、测不准、绘不对”。开展网络空间抗测绘技术研究，掌握抗测绘主动权，让敌手成为盲人瞎马，陷于“迷阵”之中，为我国网络疆域筑起“铜墙铁壁”，对于捍卫国家关基设施安全具有重要战略意义。

在上述背景下，“网络空间抗测绘理论与技术”丛书应运而生，这是我国网络空间安全科技工作者践行网络强国战略掷地有声之作。丛书由中国人民解放军网络空间部队信息工程大学和科学出版社联合发起，依托国家重点研发计划项目支持，由我国网络空间抗测绘领域知名专家学者联合撰写。丛书从基础理论破题，到抗扫描探测、抗拓扑测绘、跨空间关联映射对抗等核心关键技术攻坚，再到平台应用落地，贯通了抗测绘技术体系的全链条。丛书起点高，内容新，导向性、原创性强，体现了科学出版社“高层次、高水平、高质量”的特色和“严肃、严密、严格”的优良作风。

该丛书为国内首部网络空间抗测绘专著丛书，我相信这套丛书的出版，必将推动我国网络空间抗测绘技术的跃升与突破，为筑牢关基设施防护的坚实防线提供战略支撑。筑墙御敌，久久为功，丛书付梓亦是一个新起点，期盼广大读者与同仁不吝赐教，共臻完善，携手将我国网络空间安全屏障锻造得坚不可摧！



郭世泽

中国科学院院士

前　言

网络空间作为与现实空间深度交融的第五大战略疆域，承载着国家安全、经济发展和民生福祉的核心利益，其战略价值与日俱增。制网权已成为继制陆权、制海权、制空权、制天权之后，全球各国博弈的关键制高点。当前，我国网络与信息化建设进入高质量发展阶段，经济社会的运行对网络依赖度持续攀升，但网络空间的复杂性、动态性与开放性也使其面临的安全威胁日趋多元化。特别是，跨空间关联映射技术的滥用可能导致关键信息基础设施的部署位置和社会归属等敏感信息，以及广大网民的常驻位置和社会身份等隐私信息存在暴露风险，给国家安全和社会稳定带来严重隐患。

习近平总书记多次强调，要加快构建关键信息基础设施安全保障体系，为网络空间安全防护工作指明了方向。建设网络强国，既要具备对网络空间资源的动态感知与精准画像能力，也要构建有效的防御体系抵御非法测绘与信息窃取行为。跨空间关联映射技术作为网络空间测绘的核心支撑，为网络空间态势感知提供了技术路径，但也可能被非法利用，对关键信息基础设施安全和用户隐私保护构成严峻挑战。因此，发展跨空间关联映射对抗技术，筑牢网络空间安全防线，已成为保障国家网络空间主权的必然要求。

跨空间关联映射技术是将网络空间中的实体资源和虚拟资源映射到地理空间和社会空间的重要技术。该技术打通网络空间、地理空间与社会空间的固有限制，让网络中的各类虚拟资源与现实中的地理坐标、社会属性建立起精准关联，使管理者对网络空间的认知不再局限于抽象的逻辑拓扑和零散的设备列表，而是能够将虚拟网络与物理环境、社会场景深度融合，构建出直观可感的网络地图，为网络空间态势感知、风险预判和防御部署提供了坚实支撑。然而，一旦跨空间关联映射技术被非法测绘者滥用，该技术就会成为窃取重保单位网络设施架构、挖掘重要用户隐私数据的工具，直接威胁我国关键信息基础设施的物理安全与网络安全，以及重要用户的信息权益。为抵御这一风险、筑牢网络空间安全防线，跨空间关联映射对抗技术的研究应运而生。

跨空间关联映射对抗技术是指通过干扰、欺骗、阻断或变幻等手段，使得测绘者无法根据探测或开源的数据准确映射出网络空间实体资源和虚拟资源在地理空间和社会空间的属性信息，或者映射出地理空间范围和社会空间组织机构的网络空间实体的属性信息与结构信息。当前，跨空间关联映射对抗技术已成为保障国家网络空间主权与安全的重要研究方向，但相关理论体系与技术方法仍需进一步完善。

此前，本书作者和有关课题组围绕网络空间测绘，特别是跨空间关联映射技术持续开展了多年相关研究。2000 年左右即开展了网络拓扑分析方面的探索，2011 年开始网络目标定位技术研究，2013 年提出了网络空间测绘概念并开展了一系列相关理论和方法研究。网络空间测绘概念提出后受到了学术界、产业界和国家科技主管部门的关注和认可，国家自然科学基金委员会和科技部在“十三五”期间都对网络空间测绘相关项目给予了专项资助，为课题组开展跨空间关联映射对抗技术研究奠定了坚实基础。自 2022 年起，作者课题组在国家重点研发计划“网络空间安全治理”重点专项的支持下，系统性开展了一系列跨空间关联映射对抗技术和方法研究，涉及网络设备地理定位对抗、网络设备社会归属分析对抗、网络用户地理定位对抗、网络用户社会身份识别对抗、应用服务信息泄露检测等多个方面，形成了一系列具有自主知识产权的研究成果。

本书系统梳理跨空间关联映射对抗技术的研究进展，构建涵盖网络设备、网络用户、应用服务三个维度的技术体系。全书共 6 章：第 1 章概述跨空间关联映射及其对抗技术的核心概念、发展现状与关键技术；第 2 章为网络设备地理定位对抗技术，主要介绍三种 IP 位置泄露风险评估方法、三种网络 IP 非协作地理定位对抗方法；第 3 章为网络设备社会归属分析对抗技术，主要介绍基于大语言模型和对比学习的网络资产所有者识别方法、基于知识图谱的网络资产归属泄露风险评估方法，以及基于单节点数据微扰动的网络设备抗社会归属识别方法；第 4 章为网络用户地理定位对抗技术，主要介绍基于文本、图像和社交关系的用户位置泄露风险评估方法和三类用户地理定位对抗方法；第 5 章为网络用户社会身份识别对抗技术，主要介绍跨平台网络用户社会身份识别与社区检测技术的研究现状及对抗方法；第 6 章为应用服务信息泄露检测技术，主要介绍基于权限频率统计的数据越权访问检测方法、基于两阶段污点分析与数据流链接的多元应用共谋检测方法，以及针对

SharedUserId 机制的应用共谋检测方法。

本书由中国人民解放军网络空间部队信息工程大学网络空间安全学院杨春芳副教授、罗向阳教授、丁世昌讲师、刘琰教授、杜少勇副教授和国防科技大学李阳副教授等撰写。各章撰写与整理分工如下：第 1 章由杨春芳副教授和罗向阳教授撰写与整理；第 2 章由丁世昌讲师、罗向阳教授和卢恩尚等撰写与整理；第 3 章由李阳副教授、屠铖、窦开洋撰写与整理；第 4 章由杨春芳副教授、于龙、吴南、孙思悦等撰写与整理；第 5 章由刘琰教授、龙子琦、王潞瑶等撰写与整理；第 6 章由杜少勇副教授、关钦臣、唐昊等撰写与整理。本书的撰写还得到实验室刘粉林教授、朱玛副教授、裴阳讲师、张萌讲师、王旭讲师、袁福祥讲师，以及王泽康、刘毅敏、张远东、祖烁迪、姬少杰等研究生的大力支持，同时得到相关领域专家学者的指导与帮助。在此，对他们长期以来的支持和帮助表示衷心感谢。

跨空间关联映射对抗技术是一个动态发展的研究领域，随着网络空间技术的不断演进，新的安全威胁与技术挑战将持续涌现。本书的出版旨在系统呈现作者课题组的研究成果，为相关领域的研究者提供参考，推动跨空间关联映射对抗技术的创新发展。

由于作者水平有限，书中难免存在疏漏之处，恳请各位专家和读者批评指正。

杨春芳、罗向阳等

2025 年 11 月于郑州

目　　录

第 1 章　概　　述

1.1　跨空间关联映射及其对抗技术

1.1.1　网络空间测绘和跨空间关联映射技术

自 1969 年 ARPANET 诞生以来，网络已经渗透人们日常生活和工作的各个方面。以此为基础形成的网络空间已经成为继陆、海、空、天之后的第五大战略疆域。以往的测绘理论与技术主要对陆、海、空、天中人们物理上能够直接感受到的地形、地貌、海床、行星等进行测绘。然而，网络空间存于斯，又超于斯，既存在于传统空间，又是在传统空间中以众多信息技术基础设施网络为纽带而构成的一个超空间。它不仅包括互联网、电信网以及各种工业控制网络所组成的关键信息基础设施网络，还包括通过这些基础设施网络进行通信的自然人、组织机构，以及社交平台、业务系统等信息系统中注册的账号，甚至包括传输的信号和数据内容。传统的依赖光、电、磁手段的测绘理论与技术显然无法适用于这种超空间测绘。因此，为了能够摸清网络空间的底数，以更好地对其进行治理，网络空间测绘概念应运而生。

网络空间测绘这一概念最早于 2013 年，由信息工程大学罗向阳教授等提出。2016 年，赵帆等[1]在发表于《网络与信息安全学报》的论文《网络空间测绘技术研究》中，首次公开明确阐述了网络空间测绘的内涵：网络空间测绘技术主要是指在互联网环境下，利用网络探测、采集或挖掘等技术，获取网络设备等实体资源、用户和服务等虚拟资源的网络属性，通过设计有效的定位方法和关联分析方法，将实体资源映射到地理空间，将虚拟资源映射到社会空间，并将探测结果和映射结果绘制出来。赵帆等[1]还给出了网络空间测绘的一般技术框架，从探测层、映射层和绘制层分别阐述了网络空间测绘相关技术的研究进展。其中，映射层是网络空间测绘技术的核心层，主要研究实体资源向地理空间的映射技术和网络虚拟人物、虚拟社区等虚拟资源向社会空间的映射技术，并将映射结果提供给绘制层[2]。

网络空间测绘技术框架中映射层的任务实际上就是利用网络探测或开源数据，关联推断出网络空间中的实体资源(如网络设备)和虚拟资源(如社交账号)的地理位置和社会归属，以实现网络空间资源向地理空间和社会空间属性的映射，即跨空间关联映射。“十四五”期间，本书作者将跨空间关联映射的概念进一步完

善，形成如下定义：跨空间关联映射是指利用网络探测或开源数据，关联推断出网络空间中的实体资源(如网络设备)和虚拟资源(如社交账号)在地理空间和社会空间的属性信息，如地理位置、归属单位和社会身份等，或者地理空间范围和社会空间组织机构的网络空间资源属性和结构信息，如某个城市的 IP 地址范围、某个单位的网络拓扑、某个用户的社交账号 ID 等。跨空间关联映射技术打破了网络空间、地理空间和社会空间之间的壁垒，将网络空间中的各种资源与人们普遍接受的地理空间、社会空间建立联系。这使得管理者和决策者对网络空间的认识不再是抽象的逻辑拓扑和零散的各种网络设备，而能够与现实接触的物理环境和社会环境融合形成更为直观的网络地图，从而为网络空间态势感知和防御提供支撑。但是，跨空间关联映射技术也很容易被非法测绘者利用，以窃取我国重保单位网络信息和重要用户隐私数据，给我国关键信息基础设施和重要用户的物理安全和网络安全构成威胁。为了应对这一风险，跨空间关联映射对抗技术应运而生。

1.1.2 跨空间关联映射对抗技术

跨空间关联映射对抗技术是指通过干扰、欺骗、阻断或变幻等手段，使得测绘者无法根据探测或开源的数据准确映射出网络空间资源在地理空间和社会空间的属性信息或者地理空间范围和社会空间组织机构的网络空间资源属性和结构信息。当前，不仅互联网上有各种开源免费的工具与平台可用于实施跨空间关联映射，而且国际上各学术期刊和会议上也发表了大量跨空间关联映射技术的研究成果。这些工具和技术方法很容易被不法分子用于从事威胁网络空间安全的活动。因此，开展跨空间关联映射对抗技术研究意义重大，至少能够在以下 4 个方面发挥重要作用。

1. 完善网络空间测绘学科技术体系

网络空间安全具有很强的对抗性，作为该学科的一个重要发展方向，网络空间测绘只有经过反向对抗技术的持续对抗检验，才能逐步完善成熟。以此类推，作为网络空间测绘的核心内容——跨空间关联映射要发展，也必然需要大力发展反向对抗技术，以推动其不断升级换代，而且成熟的网络空间测绘学科方向也必然包含对其自身反向技术问题的系统研究，正如密码分析对于密码编码，网络攻击对于网络防御。因此，跨空间关联映射对抗技术研究将驱动网络空间测绘学科方向突破现有理论、技术和认知局限，深化与完善网络空间测绘学科方向技术体系，是其走向成熟不可或缺的催化剂。

2. 提升关键信息基础设施的安全性

在网络空间安全态势日益复杂的形势下，电信网络、金融系统、能源网络

等关键信息基础设施的设备和管理者已经成为敌对我实施攻击的重要目标，如近期曝光的美国国家安全局网络攻击入侵中国科学院国家授时中心事件。在该事件中，美国国家安全局利用某境外品牌手机短信服务漏洞，控制中国科学院国家授时中心工作人员手机并窃取敏感信息，再通过窃取的信息多次入侵内部网络系统，测绘出网络架构，并在此基础上部署了数十款网络武器，试图瘫痪我国高精度授时系统。由此可见，攻研跨空间关联映射对抗技术，以防止我国关键信息基础设施的设备和管理用户属性或身份信息被对手关联得到，已经成为保护我国关键信息基础设施安全的必然要求，对于保障国家经济社会稳定运行具有不可替代的战略意义。

3. 保护广大网民的隐私与财产安全

网络空间测绘中的跨空间关联映射技术不仅会被境外国家级组织用于攻击我国关键信息基础设施，还可能被黑客和不法分子用于窃取广大网民的位置和身份信息。例如，近年来频繁曝光的电信诈骗案件中，犯罪分子利用广大网民在网络上泄露的位置和身份信息编制对应话术，从而骗取大量钱财；也有一些犯罪分子利用窃取的用户位置和身份信息，识别其价值和脆弱点，从而发起精准的网络钓鱼、勒索软件攻击或账号盗取，以窃取更多的隐私信息。因此，开展跨空间关联映射对抗技术研究，能够提高网络犯罪的技术与成本门槛，从而为广大网民构筑起一道坚实而隐蔽的防线，保护其隐私与财产安全。

4. 抢占网络空间战略的全球制高点

网络空间的竞争，归根结底是规则与主导权的竞争。当前，全球网络空间治理规则仍处于激烈博弈与形成阶段。率先对跨空间关联映射对抗技术进行深入研究，取得胜敌一筹的技术领先优势，才能在界定网络空间测绘行为“正常”与“非法”的国际规则制定上掌握话语权。例如，在国际网络空间标准制定和治理谈判中，基于我国技术水平，提出领先于对手的技术或法律主张，防止对手利用技术优势窥探我方敏感信息。从我国建设“网络强国”战略的角度考虑，拥有领先的对抗技术能力，才能抢占网络空间战略的全球制高点，为维护国家网络空间主权奠定坚实基础，从而塑造全球共同发展的未来网络空间秩序。

正是由于跨空间关联映射对抗技术在学科发展、关键信息基础设施安全、民众信息保护和国家战略中的重要地位，在 2022 年立项的“十四五”国家重点研发计划“网络空间安全治理”重点专项共性关键技术项目“网络空间抗测绘关键技术”中，将跨空间关联映射对抗技术列为网络空间抗测绘关键技术的主要研究内容之一。

由于当前跨空间关联映射技术主要聚焦于网络设备和社交用户的地理位置和

社会属性分析(如单位归属映射、社会身份识别)，相对应地，现有跨空间关联映射对抗技术研究也主要关注网络设备和社交用户的地理位置和社会属性信息泄露发现与对抗。如无特殊说明，本书所称网络用户均指社交平台用户。下面将简要阐述当前跨空间关联映射技术的发展现状。

1.2 跨空间关联映射技术发展现状

根据关联映射的对象和目的空间的不同，现有跨空间关联映射技术大致可以分为网络设备地理定位技术、网络设备社会归属分析技术、网络用户地理定位技术、网络用户社会身份识别技术和应用服务信息泄露技术。上述技术主要根据探测获取的信息和开源数据，对网络设备或网络用户的地理位置、社会归属或身份进行推断。除此之外，随着智能移动终端在日常生活和工作中的普及，尤其是基于位置服务和实名认证登录的广泛推广，智能移动终端不仅可以获取和存储用户的位置信息，而且存储着大量与用户社会身份密切相关的数据。因此，在智能移动终端中下载安装的应用服务已经成为不法分子测绘用户及其接入网络的位置、社会归属和身份信息的一个重要数据来源。鉴于上述原因，本节主要从网络设备地理定位技术、网络设备社会归属分析技术、网络用户地理定位技术、网络用户社会身份识别技术、应用服务信息泄露技术等 5 个方面，简要介绍跨空间关联映射技术的发展现状。

1.2.1 网络设备地理定位技术

由于网络中主要利用 IP 地址来标识网络设备，现有的网络设备地理定位技术大致可以分为基于 IP 定位数据库的 IP 定位和基于网络测量的 IP 定位[3]。

早期的网络设备地理定位主要通过查询 IP 定位数据库中存储的 IP 和位置对应关系来实现。在构建 IP 定位数据库时，主要从多种开源和商用数据渠道获得 IP 位置数据，如 WHOIS[4]、IP2Location[5]、MaxMind[6]等。基于 IP 定位数据库的 IP 定位方法不需要进行实时测量，并且定位速度快，可以满足普通用户的需求。但 IP 定位数据库的精度不高，往往只能保证国家级、省级地理位置的可靠性，而且在城市一级，主流位置数据库的查询结果还经常出现互相冲突的情况[7]。而对于更精确的街道级地理位置信息，IP 定位数据库往往仅能覆盖极少部分 IP 地址。此外，大部分 IP 定位数据库的构建过程未公开，无法衡量数据库中地理位置信息的来源是否可靠，因此无法保证 IP 定位数据库返回的地理位置的可靠性，并且在实际互联网中，存在部分 IP 地址由运营商动态分配的情况，如果 IP 定位数据库更新不及时，将导致错误的定位结果。

为提供具有较高可靠性的城市级和街道级地理定位结果，近年来研究人员高度关注基于网络测量的 IP 定位方法[8]。此类方法对探测源、地标 IP(已知地理位置的 IP 地址)、目标 IP 之间的时延和拓扑等进行测量，分析地标的时延、拓扑与其地理位置之间的关系，从而实现对目标 IP 位置的估计。基于网络测量的 IP 定位方法大致可以分为基于时延测量的定位方法和基于拓扑分析的定位方法。前者主要通过探测包最小时延来判断目标 IP 与其周边地标 IP 的相对位置关系，进而推测、映射目标 IP 的地理位置，如街道级地理定位(street-level geolocation, SLG)[9]方法等，一般以与目标 IP 时延最低或者高度相似的地标的位置为目标 IP 的位置估计；后者试图分析目标 IP 的路由路径与地理距离之间的关系，从而根据路由路径相似度来实现定位，如 Router-Geo[10]方法等。最近的一些 IP 定位方法倾向于同时依赖时延和拓扑两类信息实现更高精度的定位，此类方法也常称为基于紧密结构的 IP 定位方法，如 GNN-Geo 方法[8]、IP2Vec[11]方法等。此类方法也高度依赖获取的准确地标以及相应时延和拓扑信息才能开展正确定位。

基于网络测量的 IP 定位方法比基于 IP 定位数据库的 IP 定位方法具有更高的定位精度，其中最新、最先进的方法甚至可以在部分网络环境中达到街道级，而且在理论上，只要目标 IP 能够被网络测量，就可以进行基于网络测量的 IP 定位，无须目标 IP/目标网络设备所属单位、管理者、网络运营商等配合，具有高精度、非协作的特点。

1.2.2　网络设备社会归属分析技术

网络设备社会归属分析技术是近年来网络空间测绘领域的研究热点——攻击面管理(attack surface management, ASM)所采用的核心技术，用于解决网络资源具体隶属于哪个组织的关键问题，进而协助企业全面准确地掌握当前所有在互联网上暴露的攻击面。

1. 基于 WHOIS 查询的网络设备社会归属分析

作为网络设备社会归属分析体系中的基础且常用技术分支，基于 WHOIS 查询的网络设备社会归属分析技术的核心是 WHOIS 协议及配套工具，从域名关联的公共注册信息库中提取关键数据，通过多维度数据整合与关联分析，推断网络设备(以域名为核心关联载体)所隶属的组织或主体。

基于 WHOIS 查询的网络设备社会归属分析技术的实现围绕“数据获取-数据处理-关联分析”三个核心环节展开，其中数据获取层借助标准化 WHOIS 协议及工具实现信息采集，包括本地命令行工具(如 Windows、Linux 系统内置的 WHOIS 命令，适用于单域名快速查询)和在线 WHOIS 服务平台或第三方应用程序接口；数据处理与关联分析层则对采集到的 WHOIS 原始数据进行结构化提取，核心字

段包括注册人信息(个人/组织名称/联系方式)、注册机构(域名注册商)、注册时间与到期时间、域名系统(domain name system, DNS)服务器地址、注册地点(国家/地区)等，通过建立多维度关联规则，挖掘域名间的归属关联。

基于 WHOIS 查询的网络设备社会归属分析技术存在以下优点：一是无须复杂技术部署，依托公开工具即可实现查询，且多数基础 WHOIS 服务免费；二是数据时效性较强，域名注册、续费、DNS 变更等信息会实时更新至 WHOIS 库，可快速捕捉资产动态；三是覆盖范围广，全球绝大多数域名(包括.com、.org、.cn 等主流后缀)均需要提交 WHOIS 注册信息，可实现跨地域、跨后缀的资产归属排查；四是具备直接性，对于未启用隐私保护服务的域名，能直接获取注册组织名称、联系方式等关键归属线索，无须间接推断，大幅提升了分析效率。

但是基于 WHOIS 查询的网络设备社会归属分析技术也存在局限性：一是隐私保护技术导致的信息缺失，随着《通用数据保护条例》等隐私保护法规的施行，多数域名注册商默认提供隐私保护服务(如将注册人信息替换为注册商代理信息)，导致核心归属信息被隐藏，无法直接获取；二是信息真实性风险，注册人可通过伪造名称、填写虚假联系方式等方式规避溯源，或者通过域名转让、注册商迁移等操作使 WHOIS 信息与实际归属主体脱节，易引发误判；三是数据碎片化，不同域名注册商的 WHOIS 数据格式不统一，部分小众后缀域名的 WHOIS 信息字段不完整(如缺少注册地点、DNS 信息)，增加数据整合与分析难度；四是历史信息查询受限，多数免费 WHOIS 服务仅提供当前最新信息，无法回溯历史注册记录，难以追踪域名的归属变更轨迹，对长期资产溯源支持不足。

2. 基于网页内容推断的网络设备社会归属分析

作为网络设备社会归属分析体系中的进阶技术分支，基于网页内容推断的网络设备社会归属分析技术的核心是突破单一域名注册信息的局限，通过多模态信息融合与智能分析，从网页文本与视觉元素中挖掘隐性归属线索，实现对网络设备(以网站为核心关联载体)所属组织或主体的精准推断，为攻击面管理中的“深度资产确权”提供高阶技术支撑，尤其适用于 WHOIS 信息缺失或失真场景。

基于网页内容推断的网络设备社会归属分析技术以“多源数据采集-跨模态特征提取-融合分析与建模”为核心路径，依托人工智能技术实现归属推断。多源数据采集层采集网页全量信息，包括站点公告、新闻报道、备案信息等结构化文本类数据和网页 Logo、品牌标识图像、产品图片、横幅广告等视觉类数据，同时通过图像预处理保障分析质量，采集过程中结合网页爬虫框架与浏览器渲染技术确保内容完整捕获。跨模态特征提取层采用专业化模型工具实现特征结构化，文本特征提取基于 Transformer 架构的预训练语言模型，通过上下文语义建模提取组织名称、品牌关键词等实体信息，提升识别精度；视觉特征提取则采用视觉 Transformer 融合

架构对 Logo 等视觉元素进行深层特征解码，通过学习图形、色彩等特征实现品牌标识精准识别。融合分析与建模层通过多模态注意力机制，如文本引导的跨模态映射构建文本语义-视觉特征关联矩阵，训练阶段基于带标签的多模态数据集优化模型参数，推理阶段通过余弦相似度计算匹配特征向量，输出归属概率与置信度，同时结合网页资源引用关系识别关联资产集群。此类方法中的典型有：李亚超[12]在 2022 年提出了一种多模态数据融合的网络设备社会归属分析识别方法，使用了网站的文本数据和图像数据来训练识别模型，取得了较好的效果，在构建的数据集中识别准确率达到了 82%；Ren 等[13]在 2023 年提出了所有者命名实体识别(owner name entity recognition, ONER)的概念，并且构建了一个融合多尺度图像信息和多模态信息的命名实体识别模型来识别网站归属实体，识别准确率也达到了 82%。

基于网页内容推断的网络设备社会归属分析技术存在以下优点：一是抗信息缺失能力强，无须依赖 WHOIS 注册信息，通过网页固有内容即可实现归属推断，有效规避隐私保护带来的溯源障碍；二是识别精度与鲁棒性高，多模态信息互补性显著，文本提供明确语义线索，视觉元素提供客观标识佐证，大幅降低了单一信息源失真导致的误判风险；三是覆盖场景更广泛，既能处理正常运营站点，也能对钓鱼网站、仿冒站点等异常资产进行归属溯源，且不受域名后缀与注册地域限制；四是隐性资产识别能力突出，可以通过品牌关键词与 Logo 的跨站关联，发现企业未公开的“影子网站”，完善资产图谱。

但基于网页内容推断的网络设备社会归属分析技术也存在局限性：一是技术门槛与成本高，需要部署人工智能模型并进行定制化训练，带标签的多模态数据集获取难度大、标注成本高；二是数据质量依赖性强，图像模糊、Logo 变形会影响视觉特征提取精度，网页文本杂乱会降低实体识别准确性；三是模型泛化挑战大，不同行业(如金融、电商)的网页设计风格与信息呈现方式差异显著，易导致模型跨领域应用时精度下降；四是分析效率相对较低，相比 WHOIS 查询的即时性，多模态模型推理与特征计算需要消耗更多的算力资源。

3. 基于知识图谱关联分析的网络设备社会归属分析

作为网络设备社会归属分析体系中的深度智能化技术分支，基于知识图谱关联分析的网络设备社会归属分析技术的核心是突破单模态信息的局限，通过多源实体关系建模与图结构推理，将分散的网络资源与组织实体构建成结构化知识网络，实现对网络设备(以域名、IP 地址等为核心关联载体)所属主体的系统性推断，为攻击面管理中的全域资产关联确权提供底层知识支撑，尤其适用于跨域资产归属追溯与隐性关联挖掘场景。

基于知识图谱关联分析的网络设备社会归属分析技术以“多源数据融合-知识图谱构建-图结构推理”为核心路径，依托知识工程与图智能技术实现归属分析。

其中，多源数据融合层实现全域数据采集与预处理，覆盖网络层数据(如网页内容、DNS 解析记录、IP 与服务关联信息)、公开数据库数据(如 WHOIS 注册库、企业工商信息库、域名备案数据库)，以及行业报告、学术文献等非结构化文本数据，采集后进行数据清洗、格式标准化处理，为知识抽取奠定基础。知识图谱构建层通过三重核心技术实现数据到知识的转化，实体识别采用基于 Transformer 架构的命名实体识别模型提取网络资产实体与组织实体，并补充属性，然后利用远程监督与 Few-Shot 学习技术提取归属类、运营类、合作类等关联关系，形成实体-关系-实体三元组，最后基于领域知识定义实体类型层级与关系约束规则，采用图数据库存储结构化知识，以完成本体构建与图谱存储。图结构推理层通过两类推理方式实现归属关系推断。基于规则的路径分析利用图遍历算法挖掘间接关联路径确定归属主体，基于机器学习的关系预测采用知识图谱嵌入模型或图神经网络预测缺失关系，结合属性图污点传播技术量化资产与组织关联强度，提升推断可信度。其中，典型方法有 Simeonovski 等[14]提出的基于属性图的互联网基础设施建模方法，该方法构建了包含 IP 地址、域名、DNS 区域、自治系统、国家和组织的属性图，并提出了一种属性图上的污点传播技术，量化了组织在互联网中的影响力，但并没有探索具体哪些资产属于同一个组织。

基于知识图谱关联分析的网络设备社会归属分析技术存在以下优点：一是多源信息整合能力强，可以打破数据孤岛，将分散的网络资产数据与组织信息系统化关联，相比单一技术更能还原资产归属全貌；二是语义关系丰富且可视化，通过图谱清晰呈现资产-资产、资产-组织的多层级关系，便于人工核查与管理决策；三是复杂推理与隐性关联挖掘能力突出，通过路径分析可以发现间接控股、“影子运营”等隐性归属关系，解决传统技术难以察觉的跨域资产关联问题；四是支持动态更新与迭代，可以通过增量抽取技术补充新实体与关系，适配网络资产动态变化特性。

基于知识图谱关联分析的网络设备社会归属分析技术也存在局限性：一是构建成本高，需要投入大量人力进行领域本体设计与数据标注，高质量标注数据集获取难度大；二是关系抽取准确性依赖数据质量，当原始数据存在信息冲突或缺失时，易导致关系误抽，影响后续推理结果；三是推理复杂度高，随着图谱规模扩大，路径分析与模型推理的计算开销呈指数级增长，需要依赖分布式计算框架优化；四是部分场景精度有待提升，对于无明确关联线索的孤立资产，仅能通过相似性匹配推测归属，存在误判风险。

1.2.3　网络用户地理定位技术

根据定位时所依据的数据，现有的网络用户地理定位技术主要分为两大类：基于发布内容的网络用户地理定位技术和基于社交关系的网络用户地理定位技术。

1. 基于发布内容的网络用户地理定位技术

当前，基于发布内容的网络用户地理定位技术主要通过分析用户发布的文本或图像内容，从中抽取出具有地理辨识性的内容或特征，以判定用户驻留的地理位置。根据发布内容类型的不同，现有基于发布内容的网络用户地理定位技术主要有基于文本的网络用户地理定位技术和基于图像的网络用户地理定位技术。

1) 基于文本的网络用户地理定位技术

用户使用社交媒体的一个主要行为是发布博文，而且可能在博文中提及其周边的地理位置名词，不同地区的用户在博文内容和用词习惯等方面也存在较大差异。因此，基于文本的网络用户地理定位技术主要分析用户发布的文本内容，寻找与地理位置相关的词语[15]、话题[16]、偏好[17]等特征，并根据特征的地理分布情况将用户与地理位置关联，从而推断用户位置。

早期方法大多使用地名词典来匹配用户发布内容中提到的地理位置名词[18-20]。例如，Gelernter 等[18]使用命名实体识别软件进行发布内容解析，以发现文本中与特定位置关联的单词；Quercini 等[19]引入来自维基百科的空间证据，从语义关联文献中提取维基百科中与给定位置空间相关的概念，从而根据局部词汇在空间上确定文本的读者范围；Middleton 等[20]将地名词典和命名实体识别方法相结合，实现了位置特征的良好识别效果。

目前，基于文本的网络用户地理定位技术更多的是从用户发布的文本中识别位置指示词，以推断用户位置。此类方法根据用户使用语言的地理偏好来推断用户位置，例如，Han 等[15]将用户位置推断问题视为用户文本分类任务，认为仅使用位置指示词作为用户特征比使用文本中所有词语的效果更优，并提出基于城市频率和词频的方法、基于信息增益率的方法和基于最大熵的方法三种位置指示词选择方法，以提取位置指示词，然后使用朴素贝叶斯模型来推断用户位置；Chi 等[16]首先对文本进行预处理，根据空格将文本标记为单个单词，然后提取位置指示词、城市/国家/地区名等文本特征作为特征词，最后基于朴素贝叶斯分类器进行文本分类，实现用户位置推断；Tian 等[21]改进了 Han 等[15]的方法，将词嵌入与基于文本的用户位置推断相结合，提出一种基于词聚类和包裹式特征选择的位置指示词提取方法。

随着表示学习技术的成熟，近年来也出现了一些基于词嵌入和神经网络的位置推断方法[22, 23]，例如：Rahimi 等[22]提出了基于二元高斯混合模型的位置表示方法，并结合词汇方言学模型准确预测地理方言词；Roller 等[24]使用机器学习算法从文本中选取特征，训练分类器，并提出基于 K-D 树的自适应网格划分方法，提升定位成功率；Lourentzou 等[25]对基于深度学习的社交网络用户位置推断方法进行了分析与综述；Huang 等[26]使用字符感知的词嵌入层处理发布内容中的噪声数

据，捕获语料库之外的词语。

2) 基于图像的网络用户地理定位技术

基于图像的网络用户地理定位技术是指在有地理位置标记的参考图像库中找出与给定待查询图像对应的图像，从而推断出用户的地理位置[27-29]。根据参考图像拍摄视角是否与用户发布的图像相同，现有基于图像的网络用户地理定位技术大致可以分为单视角图像地理定位技术与跨视角图像地理定位技术。

(1) 单视角图像地理定位技术主要针对待查询图像与参考图像均为地面视角图像的场景。现有此类主流方法大多利用已有深度神经网络的骨干部分提取图像局部特征，然后利用设计的特征聚合模块获得更具有辨识度和鲁棒性的全局描述符，通过匹配待查询图像与参考图像的全局描述符实现定位。2016 年 Arandjelovic 等[30]提出了 NetVLAD，将 VGG16 的输出作为局部特征，并根据 VLAD[31]原理构建了热插拔的 NetVLAD 聚合层以提取描述符，然后用其进行图像地理定位；王懋[32]提取了图像多个区域的卷积区域最大化激活(regional maximum activation of convolutions, R-MAC)特征，并将其与全局 R-MAC 特征融合形成了同时包含局部信息和全局信息的描述符，实现图像地理定位；刘强[33]利用多尺度融合卷积特征建立了相似度矩阵作为图像描述符，以进行图像地理定位。

(2) 跨视角图像地理定位技术主要针对待查询图像为地面视角图像而参考图像为空中视角图像(如卫星图像)的场景，例如，Regmi 等[34]将生成对抗网络引入跨视角图像地理定位领域，利用条件生成对抗网络[35]合成待查询地面图像对应的卫星图像，并将该合成卫星图像作为辅助信息对待查询地面图像与卫星图像进行匹配；Shi 等[36]将极坐标变换应用于卫星图像，生成卫星图像中心点对应坐标的伪地面全景图像，以弥合空间布局差异；Shi 等[37]为了使地面图像与卫星图像空间布局对齐，提出了一个基于圆周卷积的动态相似匹配模块，通过对伪地面全景图像特征图进行循环滑动卷积运算，实现了地面图像特征与卫星图像特征的对齐；Yang 等[38]采用 Transformer 中的位置编码技术[39]对图像每个像素进行相对位置编码，帮助地理定位模型更好地理解图像；Zhu 等[40]设计了一种去除卫星图像中建筑物顶部等地理位置无关内容的训练方法；Zeng 等[41]引入了无人机图像作为桥接，并利用交叉扩散将地面-无人机和卫星-无人机两个向量空间的特征进行融合。

2. 基于社交关系的网络用户地理定位技术

除了发布博文外，用户在社交媒体平台上另一主要行为是与其他用户进行交互，包括关注、提及、点赞、评论、转发等。现有基于社交关系的网络用户地理定位技术通常假设在社交媒体上关系更紧密的朋友在现实世界中地理位置更接近，从而根据用户朋友的位置来推断用户的位置。