数据出境安全风险监测预警关键技术综述

张凯[1]，时金桥[1]，马乐乐[1]，郭晓威[2]，刁毅刚[1,3]，李凤华[4]
（1.北京邮电大学网络空间安全学院，北京 102200；2.华中科技大学网络空间安全学院，湖北 武汉 430074；
3.中央网信办数据与技术保障中心，北京 100048；4.中国科学院信息工程研究所，北京 100085）

摘　要：数据出境场景存在主体多元、环境开放、业务动态、数据私密等特点，现有监测技术架构难以满足数据出境场景下的全域风险协同监测需求，传统风险监测技术在出境业务动态适配、监测隐私无干扰等方面尚存在较大差距。针对上述问题，提出一种数据出境安全风险监测预警研究框架，系统分析全链条监管需求与挑战，细化明确现有技术在数据出境场景下的适用性及差距，凝练支撑需求的关键技术体系，为国家数据出境安全风险监测提供系统化支撑。

关键词：数据出境；监测预警；全域协同；隐私无干扰
中图分类号：TP309.2
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2025221

Review of key technologies for data outbound security risk monitoring

ZHANG Kai[1], SHI Jinqiao[1], MA Lele[1], GUO Xiaowei[2], DIAO Yigang[1,3], LI Fenghua[4]
1. School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 102200, China
2. School of Cyberspace Security, Huazhong University of Science and Technology, Wuhan 430074, China
3. Data and Technology Support Center, Cyberspace Administration of China (CAC), Beijing 100048, China
4. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100085, China

Abstract: Data outbound scenarios are characterized by multiple stakeholders, open environments, dynamic operations, and sensitive data. Existing monitoring technology architectures struggle to meet the requirements for comprehensive, coordinated risk monitoring across these scenarios. Traditional risk monitoring techniques still exhibit significant shortcomings in dynamically adapting to outbound operations and ensuring privacy-preserving monitoring without interference. Addressing these challenges, a research framework for monitoring and early warning of data outbound security risks was proposed. The regulatory requirements and challenges across the entire chain were systematically analysed, the applicability and gaps of existing technologies in data outbound scenarios were clarified, and the key technological systems required to support demand were refined. This provides systematic support for national data outbound security risk monitoring.

Keywords: data outbound, monitoring and early warning, all-domain collaboration, private and undisturbed

0　引言

在我国法律语境下，“数据出境”是指数据处理者向境外提供在中华人民共和国境内运营中收集和产生的重要数据和个人信息的行为。当前，国内外高度关注数据出境安全，美国制定《CLOUD Act》[1]，发布《14117号总统令》等，构建数据出

收稿日期：2025–10–16；修回日期：2025–12–06
通信作者：时金桥，shijinqiao@bupt.edu.cn
基金项目：国家重点研发计划基金资助项目（No.2023YFB3106400）
Foundation Item：The National Key Research and Development Program of China (No.2023YFB3106400)

境规制体系。欧盟发布《通用数据保护条例》（GDPR, general data protection regulation）[2]，为个人数据从欧盟向其他国家或国际组织的出境传输提供明确的法律框架和规则。我国发布《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》和《网络数据安全管理条例》等法律法规，制定《促进和规范数据跨境流动规定》，明确数据出境安全评估、个人信息出境标准合同等制度，支持自由贸易试验区实施数据出境负面清单管理。国家大力推进国家数据基础设施建设，积极探索构建跨境可信数据空间，建立高效便利安全的数据跨境流动机制，构建数据跨境传递监控、存证备案、出境管控等能力体系。上述法律法规为数据出境活动提供明确的规范指导，促进数据出境有序合规流动。然而，数据出境活动业务场景多样，数据交互复杂，出境主体地域分散，合规意识和能力各异。监测识别数据出境业务运行过程中实际发生的风险对于支撑相关法律法规落地实施，支撑管理部门开展数据出境事前事中事后全过程监管，掌控全域出境安全风险态势具有重要意义。

数据出境风险监测的核心是数据流转全过程风险识别与管控，需要完备的技术架构以及适应业务特点的支撑技术。在监测架构方面，中国科学院面向分布式信息系统提出插件化风险采集与数据流转监测全周期架构[3]，支撑读写、保存、转发等细粒度操作层面的数据流转安全监控。亚马逊推出Amazon DataZone[4]，通过数据目录编制和数据血缘对数据进行识别与追踪，实现云环境下的可信数据流转管控。欧盟面向数字健康领域提出可互操作电子健康服务的安全与可信范式（KONFIDO）项目[5]，通过接入认证、区块链存证等实现医疗行业数据出境行为的风险监测。下一代互联网国家工程中心推出国际跨境可信数据空间，基于数据基础设施通过IPv6身份认证、区块链存证、数字合约控制等实现行业级数据出境行为的风险监测管控。美国在联邦政府、国防信息网络等受控内网及跨国情报机构中推行数据安全标识和数据控制安全标记，并在关键互联网数据汇聚出入口部署传感器进行通信数据的采集与监测[6]。实现国家级敏感数据流转行为监测管控。上述架构围绕受控协作的信息系统、分布式网络实现面向特定数据的多层次流转管控，但在数据出境场景下，出境主体地域分布跨度大、出境数据类型多样、流转网络环境复杂且非全面可控，上述架构难以全面有效地支撑开放网络环境下的敏感数据出境风险监测。

在监测技术体系方面，英国信息专员办公室依据法律法规提出结构化风险评估模板协助出境机构抽取风险要素和风险分级评估[7]，李金等[8-10]基于网络边界监测构建二分网络结构刻画数据出境流转路径，结合出境主体属性、实际出境传输行为统计特性对数据出境风险路径进行量化评估，上述研究可一定程度上支撑数据出境风险的泛化、动态评估，但在复杂出境业务场景的普适化适配、域内出境的覆盖性方面尚有不足，且缺乏数据广泛流转条件下的违规危害预估。Guamán等[11]提出隐私策略分析、应用程序行为分析和数据传输合规检查的协同分析框架，实现移动应用个人信息出境合规性判别。美国数据管理解决方案提供商BigID、人工智能与数据治理公司OneTrust及国内天融信等通过出境数据合规性监测、数据流转关系异常分析，实现企业级数据出境违规风险预警。上述研究一定程度上可以支撑数据出境风险的运行时识别，但在出境数据隐私保护、多主体协同出境的异常判别、违规出境溯源追责方面尚存在不足。在网络安全领域中，多点协同监测和处置方面已有大量研究成果，然而，在数据出境场景下，如何在复杂多跳的数据处理流转链条下，关联主体众多且业务处理角色多样条件下，实现出境异常事件的有效处置仍然是一个技术难题。基于可编程协议无关包处理（P4, programming protocol-independent packet processor）的网络信息流控制系统P4control[12]和带内网络遥测（INT, in-band network telemetry）[13-14]等已有研究基于新型可编程网络架构实现数据中心、企业广域网等环境下数据流转的实时跟踪和多点协同处置，但受限于可编程网络设备的依赖，尚不足以实现开放网络环境下的出境业务无干扰监测处置。

因此，数据出境安全风险监测与传统网络安全监测与信息内容安全监测有本质的不同，数据出境安全风险监测往往是在数据流转路径难以全面覆盖、出境业务架构复杂、出境数据隐私加密、出境业务无干扰等条件下开展的，构建国家级数据出境监管技术体系面临监测代价、效果、

覆盖性、准确性等技术难题。数据出境风险管控本质上是以数据安全为驱动的研究，出境业务在数据类型、主体关系、流转模式、合规逻辑等方面特征复杂各异，已有工作在风险辅助评估[7]、风险动态评估[8-10]、合规判别[11]等直接针对数据出境场景进行研究，但缺乏自动适配业务流程、在数据对象和具体操作层面精准识别细粒度风险行为、全面掌控风险态势的技术能力支撑。另外，传统网络安全与信息安全领域，在业务流程自适应管控[15-20]、细粒度风险识别[21-28]、风险态势感知[12-14]等方面已有大量研究，但这些技术在适配数据出境场景业务多样、风险动态隐蔽、数据敏感、业务无干扰等实际问题还存在差距，亟须解决数据出境业务场景适配、技术能力缺口系统识别、管理需求与工程桥接等问题。为此，本文首先对数据出境场景进行了系统化的描述与分类，以“全过程监管、细粒度监测、全区域覆盖”为主线，提出一种数据出境安全风险监测预警研究框架（以下简称研究框架），分析了数据出境事前、事中、事后全链条监管的挑战与技术能力需求。进一步，本文以数据出境安全风险监测预警各维度核心步骤为逻辑，对数据出境直接相关的关键技术进行了综述，分析了现有相关关键技术在实际数据出境场景下的支撑能力与适用性差距。基于上述研究总结与技术能力需求，本文总结并归纳了适配场景与需求的关键技术要点，凝练了一套业务无干扰、隐私安全保护、动态多阶段数据出境风险全面识别与管控的关键技术体系，介绍了关键技术体系的落地应用与未来研究展望。

1　数据出境场景与研究框架

为支撑数据出境风险监测预警关键技术的系统研究与综述，本节围绕场景刻画与框架构建2个方面展开，首先对数据出境场景进行描述与分类，在此基础上提出数据出境安全风险监测预警研究框架。

1.1　场景描述与分类

数据出境场景是指在特定业务目的驱动下，数据从境内向境外传输、处理和共享使用的全过程，涉及多元主体、多样数据类型、多重处理环节及差异化合规要求的复合性情境。其传输与流转主要依托网络环境及基础设施实现，通过出境通信链路、云平台或网络系统等完成数据交换与处理。现有工作虽然通过提出架构、工具和技术解决了部分场景的研究问题，但缺乏数据出境场景的规范化定义或适用范围界定，容易导致技术分析路径偏差，表现为评估对象与范围难以稳定界定、数据保护强度与管控粒度难以平衡等问题。因此，为更好地描述数据出境场景，本文通过考虑中国以及欧美的法律框架，并结合具体实践，从产业类、研究类、业务角色链条类、数据敏感等级类等核心差异化维度构建数据出境场景分类。

产业类场景，是指特定产业领域内，因业务、服务或协作需求而产生的数据出境流转活动，其主要特征由产业属性、数据类型及业务模式共同决定。现行监管体系虽无统一的产业类数据出境分类标准，但产业规范已呈现出差异化监管雏形，产业特性决定了其所在的“宏观环境”和“准入门槛”。不同产业的数据属性和业务模式的差异，导致数据出境合规与监管强度的巨大鸿沟。理解产业分类及特性，是精准定位合规路径、平衡发展与安全的关键。为更好地对产业分类进行描述，从数据出境实际监管需求出发，参考《国民经济行业分类标准》（GB/T 4754−2017）及《数据出境安全评估申报指南》的相关表述，面向数据出境场景的核心管控行业及其数据类型与业务应用场景，建立表1所示的产业分类体系。

研究类场景，是指为实现科研、创新或治理等研究目的，研究主体基于非营利或营利动机，在数据传输、处理与使用过程中产生的数据出境流转活动，具有多主体参与、数据价值导向和合规路径差异化的特征。在数据出境过程中，出于不同研究目的所处理的数据，其形态以及相关数据处理主体的角色分工并非固定，风险管控往往具有较强的不确定性和情境依赖性。如表2所示，本文基于不同的出境特征与研究动机对数据出境研究类进行了分类与描述。

业务角色链条类场景，是指在数据出境传输过程中，不同主体基于控制与处理职责形成的多层级业务流转关系，体现数据在网络环境中经由多个角色环节动态传输、处理与共享使用的合规管理情境。业务角色链条关注的是谁在什么条件下、以何种方式处理和传递数据，是实现场景化流转与权责

表1 数据出境场景产业类

产业领域	典型产业	典型数据类型	典型业务场景
信息通信与互联网类	电信、互联网、跨境电商、零售	注册信息、通信日志、订单数据	物流追踪、全球云服务部署、国际通信
金融与保险类	银行、证券、保险、支付机构	客户身份、账户数据、交易记录	跨国结算、反洗钱审查、集团风控
医疗与生命科学类	医疗、公共健康、制药与生命科学	医疗影像、基因数据、临床样本	国际科研合作、药物开发、疫情监测
制造与工业控制类	工业、科技、国防科工	运行日志、控制指令、实验数据	跨国生产协同、远程维护
能源与交通类	能源、电力、交通运输、自然资源	能源调度数据、地理空间与遥感影像	国际能源协作、远程监控、资源研究
教育类	高校、教育信息化机构	学籍信息、教学日志	国际合作办学、教育科研数据共享

表2 数据出境场景研究类

研究类别	出境特征与动机	典型研究主体	典型场景
非营利类	公益或科学驱动：以知识探索、学术共享、社会治理或公共利益提升为目标，强调开放合作与伦理合规	高校、科研院所、国际科研联盟、公共实验室、政策研究机构、统计部门、公共治理平台	向境外科研机构共享实验数据、上传至国际数据平台、向国际组织提交社会经济统计与分析数据
营利类	经济驱动：以技术创新、商业化应用和市场收益为核心，通过跨国联合研发和数据合作提升企业竞争力	企业研发中心、科技公司、产业联盟	向境外合作方传输用户或市场数据以支持商业化研发与技术迭代

划分的关键。如表3所示，本文以数据控制者和数据处理者为主要角色，对数据出境场景的角色链条类进行描述，其中数据控制者是指单独或与他人共同决定个人数据处理目的和方式的实体，数据处理者是指代表控制者处理个人数据的实体，数据控制者和数据处理者的身份直接影响可选择的合规路径，通过对数据控制者和数据处理者及其相互关系的细致刻画，可以支撑具体且具有操作性的数据出境场景构建，实现从抽象法规到具体实践的对接，且避免因忽略角色间的权责差异而导致出境风险管控方案的失效。

数据敏感等级类场景，是指依据数据的重要性及潜在影响程度，将不同敏感等级的数据在出境流转中映射到差异化合规要求和安全管控措施的出境情境，用以确定数据可出境性及其合规路径。数据敏感等级是合规义务的逻辑起点与分水岭，理解数据敏感等级是有效进行数据出境场景分类的基础。如表4所示，本文依据《网络数据安全管理条例》《数据安全技术 数据分类分级规则》（GB/T 43697—2024）等，结合不同数据的合规要求与影响程度等因素对数据出境敏感等级进行了归纳与分类。

本文对数据出境场景的分类旨在形成一种描述性、可复核的标签化表达体系。然而，所提出的4维分类并非对数据出境场景进行唯一划分，而是用于稳定呈现场景分析的关键前提与差异要素。当不同标签间存在交叠时，遵循“从严一致性合并原则”处理。此外，诸如网络与基础设施形态、数据发送方与接收方所在国家法律制度及监管政策等因素属于跨场景的公共约束条件，虽不单列为独立分类

表3 数据出境场景业务角色链条类

业务链条类别	角色构成	关键合规特征	责任主体	典型场景示例
直接传输模式	境内控制者→境外控制者	双方建立法律关系，适用标准合同条款	境内控制者承担全部技术控制责任	跨境电商向合作伙伴传输用户订单数据
链式传输模式	境内控制者→处理者A→处理者B→境外控制者	每个环节需进行独立评估，相关处理者需获得单独授权	每一环节均需部署技术控制，责任连带	境内制造企业将生产数据传输给ERP系统服务商，ERP系统服务商将数据传输至境外物流协调平台，最终数据由境外采购商用于库存管理和采购决策
联合传输模式	境内外由多方控制者与多个处理者共同参与的复杂处理链条模式	各方承担连带责任，需明确约定各自权利义务	各方根据约定比例分担技术投入	境内车企收集车辆运行数据，委托境内数据服务商进行初步清洗和脱敏，境外零部件供应商和保险公司分别通过境外云服务商接收这份数据

表4　数据出境场景数据敏感等级分类

数据类别	概念	合规要求	典型数据示例
核心数据	对领域、群体、区域具有较高覆盖度或达到较高精度、较大规模、一定深度的，一旦被非法使用或共享，可能直接影响政治安全的重要数据	绝对禁止数据出境	特种工程塑料、高性能膜材料、光刻胶等关键材料生产工艺数据
重要数据	特定领域、特定群体、特定区域或达到一定精度和规模的，一旦被泄露或篡改、损毁，可能直接危害国家安全、经济运行、社会稳定、公共健康和安全的数据	有条件数据出境，需进行目录管理、安全评估与目的限制	未公开的领陆、领水、领空数据和政府、军工单位等敏感客户清单等
一般数据	核心数据、重要数据之外的其他数据	自由流动原则	产业名称与产业分析等报告数据、活动名称与活动总结等营销推广数据

轴，但在评估与论证环节中被统一作为前置约束条件加以考虑。本文提出的场景分类体系旨在与整体研究框架形成协同作用，使后续的问题与需求刻画、关键技术综述与体系设计能够在统一的语义基础上展开，从而减少适用性偏差，提升研究结论的一致性与可复核性。

1.2　研究框架

在数据出境场景下，出境业务流程环节多样，面向多角色链条，涉及多产业、多过程流转，违规出境手段复杂隐蔽，多维风险交织。此外，出境行为存在动态性，出境主体多，分布广，监管代价敏感。目前，现有研究在风险量化评估[8,10]、出境合规检测[11]与出境风险监测[3,6]等方面提出了一些方法与观点，但整体仍缺乏系统化的研究框架。为了充分结合数据出境场景更全面地分析相关技术差距、梳理相关研究需求，本文面向数据出境安全风险识别与管控，从过程监管、风险预警和监测处置等视角提出一种数据出境安全风险监测预警研究框架，围绕数据出境业务场景，全面、系统地梳理了现有研究存在的问题与技术挑战，进一步归纳分析了各维度所需的技术能力需求，支撑数据出境安全风险监测预警关键技术体系构建。

如图1所示，为了能够更好地支撑覆盖事前、事中、事后全过程管控，提出过程监管，重点分析了现有研究应如何优化评估流程与响应能力、加强违规线索核查水平。为了充分结合实际业务甄别多维风险事件及其关联扩散风险，提出风险预警，讨论了现有研究应如何安全地识别动态风险、强化预警能力。为了全面支撑覆盖风险源头、传输路径与边界节点的协同监管，提出监测处置，分析了现有研究应如何强化源头管控与协同处置水平、实现业务无干扰的全域风险长效管控。

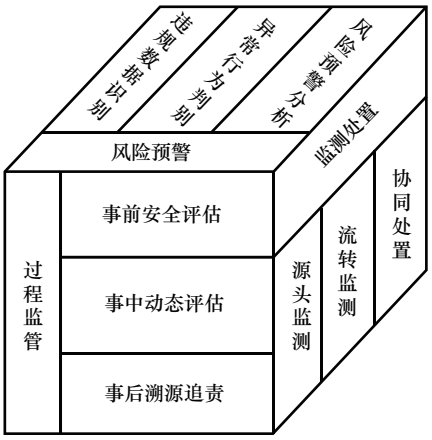

图1　数据出境安全风险监测预警研究框架

图2梳理了研究框架各维度的作用、信息交互机制及其工作流程。过程监管是整体框架的时序逻辑主线，通过将事前的数据出境合规扩展为动态、全过程的风险监管，实现事前、事中、事后全过程监管。在过程监管采集的出境业务信息如业务传输时间、数据量、传输范围等可为风险预警中的异常行为判别提供业务背景支持，所采集的业务架构、系统拓扑等信息为监测处置明确监测部署提供支撑。风险预警通过在数据出境过程中深入挖掘安全事件，所识别的违规数据、异常行为信息等可为过程监管的事中动态评估提供安全事件支撑，为监测处置中的风险监测布控范围、策略等提供支撑。监测处置从空间的视角，研究整体框架中的监测响应部分的部署策略与交互关系，通过完整覆盖数据出境源头与流转路径，为过程监管的事中动态评估与事后溯源追责提供基础数据支持，为风险预警的异常或违规事件发现及扩散分析提供风险线索支持。过程监管、监测处置、风险预警3个维度分别从数据出境监管的时间逻辑、空间部署以及深度感知开展研究，并通过相互支撑形成数据出境风险的发现、监测、处置闭环。

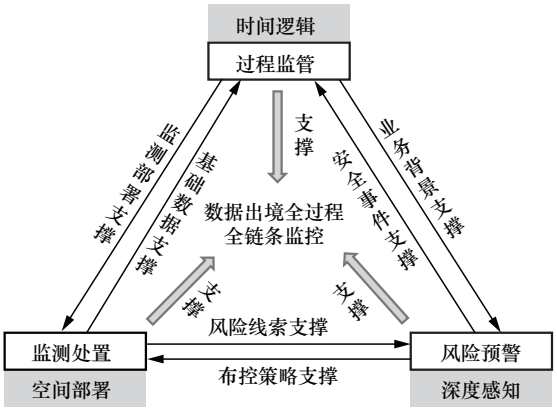

图2　研究框架各维度交互与工作流程示意

1.2.1　过程监管

当前出境场景多样，业务流程复杂，过程监管面临如下技术挑战。首先，出境机构普遍缺乏对不同产业场景出境业务的精细化描述，导致普适化适配的风险评估困难，难以实现风险源头的精细化管控。其次，随着技术演进、监管趋严及业务多元化，出境路径和合规要求不断演化，如何在出境风险要素变化下实现动态风险评估成为技术难点。此外，违规行为责任归属依赖于操作行为可溯性与主体责任边界的清晰划定，而实际出境事件时空跨度大，关联主体多，溯源信息碎片化，导致精准溯源与责任判定困难。出境数据敏感，并且隐私保护与溯源效力存在根本性矛盾，导致监管效能弱化，造成溯源追责难度进一步上升。

针对上述问题，过程监管主要包括事前安全评估、事中动态评估和事后溯源追责等部分。在事前安全评估方面，需强化数据出境安全风险辅助评估手段建设，提升对不同产业场景出境业务精细化建模和描述能力，加强对境外主体基本信息（如行业类型、单位性质、位置）、业务过程（如研究目的、数量、方式）和行为要素（如传输链路、规模、频次）的抽象分析能力。构建普适化风险智能评估机制，提升事前安全评估自动化水平，建立风险指标体系的自适应构建机制，弥补人工评估手段在效率与主观性方面的不足。

在事中动态评估方面，需要提升关键风险要素动态变化的感知与识别能力，针对实际出境过程中的风险要素，建立风险要素动态变化下风险评估响应机制，设计具备动态更新能力的风险评估体系架构，需强化评估指标与权重动态优化的自适应调整能力，实现对风险要素变化的实时响应与量化评估，提升动态出境风险的精准识别水平，支撑动态风险防范持续改进。

在事后溯源追责方面，需提升对碎片化、多源化异构数据的关联分析能力，强化关键线索缺失下的信息增强与补全能力，重构具备时空完整性的证据链条，为责任识别与归属提供数据支持。完善面向复杂业务下的违规责任认定机制，明确数据提供方、处理方及境外接收方等所属角色职责和责任层级，推进责任认定的标准化与精细化。此外，需统筹考虑隐私保护与监管溯源需求，设计具备隐私增强能力的溯源机制，实现对溯源深度与隐私保护强度的均衡优化，构建可验证、可审计的可信责任追溯体系。

1.2.2　风险预警

当前出境机构规模多样，出境数据敏感，风险预警面临如下技术挑战。首先，出境流量呈加密特征，受加密策略和数据敏感性影响，违规数据识别准确性与隐私安全面临挑战。其次，出境业务动态变化，而现有方法风险识别能力不足，缺乏静态备案信息与动态业务运行间的差异感知。违规出境方式隐蔽多样、路径不确定、行为特征弱化，导致异常行为难以界定与捕获。此外，现有方法缺乏扩散风险感知能力，而主体间存在规范标准与适用边界差异，易引发风险扩散与链式关联，造成合规压力向外溢出。风险态势动态变化，而现有预警策略以静态规则为主，缺乏自适应调整能力，难以支撑敏捷响应需求。

针对上述问题，风险预警主要包括违规数据识别、异常行为判别和风险分析预警等部分。在违规数据识别方面，需要协同提升管理部门与数据处理主体在加密环境下的风险识别能力，构建融合行政监管与自监测的双轨治理体系。加强数据在加密前和传输后的日志审计能力，应要求数据处理主体对传输内容进行结构化描述与记录，确保可感知信息的有效留痕与审查。其次需建立持续运行的自监测机制，开发面向加密场景的合规自检工具，提升企业对风险事件的自我发现能力，支撑解决加密环境下违规数据识别难题。

在异常行为判别方面，需提升动态业务运行状态的识别与核验能力，构建支持静态备案信息与动态业务行为间的一致性核验机制，应具备关联分析、关键要素抽象和差异化比对能力，支撑对数据

出境运行时风险的常态化管控。其次应构建面向多维特征的异常行为模式识别手段，实现对访问端口异常、传输时间异常和通信IP异常等非常规行为的识别。增强流转数据、出境行为和主体在时间、空间和逻辑上的关联建模能力，构建多视角协同的异常行为判别体系，支撑解决复杂流转条件下的异常行为发现难题。

在风险分析预警方面，需提升对扩散风险的识别与响应能力，强化主体出境约束（如合规要求、管控强度等）差异化分析手段建设，增强关键风险传播点识别能力。建立覆盖多主体、多业务的风险传播与扩散感知机制，实现对风险传导路径与影响域的协同识别。构建自适应动态预警体系，支持基于风险事件等级驱动的精细化预警配置，提升预警系统在动态风险态势下的决策能力，增强数据出境风险防控水平。

1.2.3　监测处置

当前，出境主体地域分散，业务架构快速演变，监测处置面临如下技术挑战。数据出境通联模式多样，通联数据多源异构，采集规模与类型显著扩张，导致风险指征抽象分解难度增加，难以实现精细化风险监测，采集开销同样面临挑战。风险特征与感知重点随风险态势变化，固定采集策略难以支撑精准持续的监测需求。其次，开放的流转网络环境中流量规模海量，出境业务混杂，通信特征弱化，环境异构，难以快速精准识别出境数据。数据出境涉及多节点、多阶段流转，业务角色链条不透明，流转轨迹还原困难，且监测过程易造成业务干扰，甚至导致隐私泄露风险。此外，违规风险传播链条分散、影响范围动态，处置对象多源，现有机制联动响应、优化调整等能力不足，难以支撑动态风险的精细化协同处置。

针对上述问题，监测处置包括源头监测、流转监测和协同处置等部分，在源头监测方面，需提升多源风险指征体系化梳理能力，建立覆盖网络会话、应用协议、服务交互等多维度的风险指征分类框架，构建面向通信会话和关键行为要素的细粒度源头风险采集机制。需充分考虑通信数据的行为特征与结构属性，完善风险指征的要素级解构能力（如将会话类指征分解为会话ID、起止和持续时间等要素）。构建差异化低开销采集策略，支持依据通信行为模式异质性进行粒度可调的策略适配，强化动态风险水平与采集配置的映射能力建设，推动静态规则驱动向动态采集范式演进。

在流转监测方面，需提升海量流量下出境数据流的识别与分析能力，构建跨网络、跨边界与跨平台环境的流转监测能力体系，强化具备高鲁棒性、高安全性和低干扰性的出境识别能力建设，加强流转网络中业务角色链条、关键环节与链路结构的分析能力，提升开放分布式网络中监测点优化部署水平，构建覆盖全域的流转路径刻画机制，增强数据出境流转可见性，支撑面向数据出境的全生命周期监管体系构建。

在协同处置方面，需提升面向动态风险演化的处置范围识别与联动处置能力，强化面向多源处置对象（如多个风险源流量、设备等）的精细化处置手段建设。加强风险源头和扩散链条及潜在影响域的分析预判能力，支撑处置范围精准刻画。构建支持跨主体、跨网络、跨域环境的分布式多点协同处置机制，同步完善处置反馈的效果评估能力，推动反馈机制闭环优化，实现协同处置的高效响应与策略优化。

2　相关关键技术

基于上述出境场景描述与研究框架，本节在过程监管、风险预警和监测处置等方面，以数据出境安全风险监测预警各维度的核心步骤为逻辑，对数据出境直接相关的关键技术作出系统化梳理与综述，提出并分析数据出境场景下关键技术的典型可量化对标指标。然后总结并归纳现有相关的关键技术研究现状，在此基础上，重点分析现有相关关键技术在数据出境场景下能否有效支撑所需的关键技术能力，总结现有相关关键技术的不足，最后指出了应重点突破的技术难题。

2.1　过程监管相关关键技术

2.1.1　事前安全评估

事前安全评估其本质是数据出境前对潜在风险进行系统识别、量化和管控的治理手段，主要思路是将模糊未知的风险转化为可量化、可追溯的决策依据。目前，各国逐步建立起以风险预防为导向的监管体系。主要呈现3种差异化监管模式，分别为欧盟、美国和中国模式，其中，欧盟以GDPR为核心，通过充分性认定、标准合同条款（SCC, standard contractual clauses）、约束性企业规则（BCR,

binding corporate rules）等机制构建“高保护标准与灵活工具”的体系。美国奉行数据自由流动原则，限制最小化，主要通过行业自律与事后执法实现监管。而中国主张建立“风险自评估和安全评估申报”的双层机制，采取“分类分级与许可审查”的混合模式，结合安全评估申报指南、标准合同和个人信息保护认证等实现灵活的事前评估。上述3种模式并非完全独立，而是共同塑造了复杂的评估标准，例如Meta公司因未能满足GDPR关于数据转移影响评估（TIA, transfer impact assessment）的要求而受到巨额罚款[29]，所以即便是在美国注册，也无法绕开欧盟的监管效力。

然而，面对日益增长的评估需求和技术挑战，监管力量和专业化水平存在区域差异，人工评估方法存在效率低下与主观性问题，导致数据处理者选择在监督力度较弱的区域落户以规避管制风险的管辖竞择现象。事前安全评估技术实际是以数据为驱动、以业务为核心的研究，业务建模是事前安全评估的首要工作，文献[15]提出对业务流程的佩特里网（Petri Net）进行结构化改造，使其符合某类可验证的标准形态。文献[16]提出面向工业互联网的业务流程模型和标记（BPMN, business process model and notation）建模方法，能够在不破坏BPMN兼容性的前提下，表达工业互联网场景的特有概念与语义，有效帮助理解数据出境业务流程及其相互关系。为了丰富业务表达与建模能力，文献[30]针对同一业务的多源信息进行系统建模，引入归因与视角概念，构建了具备视角兼容性的知识图谱模型。在数据出境场景下，为客观评估不同建模技术的适用性与优劣，可从模型运行效率和业务建模质量等维度进行量化评估，前者包括业务流程建模延迟（Latency）、吞吐量（TPS）等，用于衡量不同方法在实际运行中的效率和资源消耗；后者包括业务流程覆盖率（Coverage）和建模精确率（Precision）等，用于衡量不同方法进行目标业务建模的完整性和正确性。数据出境场景下的业务建模主要侧重于业务流程的完整刻画能力，通常作为离线过程进行操作，所以要求较高的业务流程覆盖率与建模精确率，以便支撑风险要素识别与量化分析。对于大多数出境产业场景，在一定的业务流程建模延迟区间内，若能保持较高的业务流程覆盖率与建模精确率，一般能够满足需求。对于互联网和信息通信产业，业务数量庞大，其建模方法需要面向大规模流量场景专门设计兼顾吞吐量与业务流程建模延迟的算法，以应对实时处理需求。对于金融和医疗等合规要求较高的行业，主要考量则是保持较高的业务流程覆盖率与建模精确率，支撑监管决策和风险评估。

风险要素识别是建立在业务流程建模基础上的关键环节，主要是针对合同条款、申报与备案信息等实现关键风险要素（如数据敏感等级、研究目的）发现与挖掘。文献[31]测试并验证了自然语言处理（NLP, natural language processing）能够自动化识别复杂合同文本中的风险要素。文献[32]为解决NLP在可解释性、数据标注成本与跨任务迁移方面的痛点，采用自动生成规则来匹配识别合同中的风险要素（如风险术语与句式），适用于数据稀缺或对审计可追溯性要求高的出境场景。为提升文本的上下文理解与泛化任务适配能力，文献[33]提出基于大语言模型（LLM, large language model）的端到端框架网络威胁情报知识图谱AutoCTI2KG，在不需要大规模标注的情况下，针对不同场景的非结构化文本实现关键要素的自动化检索与抽取。针对上述关键技术，可引入F1分数（F1 score）与召回率（Recall）等指标衡量模型在该任务下的整体识别性能。引入数据标注效率（DAE, data annotation efficiency）指标，表示为模型达到目标性能（如F1 score≥0.8）所需的最小人工标注样本数量，用于衡量不同方法在标注成本维度的效率。识别延迟指标可用于评价不同方法的响应速度与可用性。在数据出境场景中，待审查文档中的风险要素多样且语义复杂。因此，如何自动化且准确地识别待评估的风险要素，直接影响事前风险评估的准确性和有效性。F1 score与召回率可视为主要性能要求。高召回率意味着最大化地识别所有潜在的风险要素，而F1 score则兼顾了准确性与召回率，为评估模型的综合性能提供依据。金融与医疗等产业由于涉及个人隐私数据和高度合规要求，对覆盖率和精确率有更高的性能标准。而对于互联网、通信、教育等产业，由于业务规模庞大、数据敏感性较低，通常对风险要素识别的时效和数据标注效率的要求更为看重，倾向于在保证一定精确度的前提下，优化识别的速度和标注的效率。

风险评估是事前安全评估的核心技术，需综合

考虑数据出境业务场景，针对多维风险要素实现高效的、普适的评估。早期手段主要采用NIST SP-800-30、ISO 27005等通用评估方法[34]辅助评估数据出境。在此基础上，英国信息专员办公室通过发布数据出境传输风险评估工具[7]，实现风险要素与风险级别关联分析，支撑数据出境安全风险分级评估。为提升风险量化评估能力，文献[35]提出一种多过程、多周期数据安全风险评估模型，通过整合国内外数据出境流程，采用层次分析法抽取数据出境安全风险共性评估指标并确定权重，结合小波神经网络实现数据出境风险量化。为扩大评估范围，文献[36]提出综合数据出境方和数据接收方两大机构主体数据保护现状和能力、数据境内存储和出境使用双阶段的多周期评估框架，动态模拟风险因素评价过程，实现数据出境风险量化计算。国际隐私专业协会则基于概率风险分析理论，提出欧盟SCC数据转让事项的风险评估框架[37]，通过分析数据出境多维风险要素联合发生概率，估算安全事件发生概率及损失，量化数据出境风险。针对上述关键技术，可引入F1分数与评估准确率（Accuracy）等指标衡量不同方法的风险评估性能。在数据出境场景中，对于各产业而言，准确率为核心要求，高数值的准确率意味着模型在大多数情况下能够准确判断潜在的数据出境风险。而F1分数则能够在数据样本不平衡的情况下，例如风险事件通常会远少于正常事件，帮助评估数据出境风险评估性能，高F1分数的方法能够显著提高风险评估的实用性。

从上述研究的综述可以看出，数据出境事前安全评估不是一项简单的法律合规任务，而是呈现法律文本到工程实现、单一技术到技术生态、静态合规到动态治理三大特征。本文在表5中总结并归纳了事前安全评估的关键问题与需求、核心要点描述、代表性文献与典型评价指标。总体而言，在欧盟、美国和中国3类监管背景下，现有研究能够通过形式化业务建模实现数据出境过程的结构化刻画，利用NLP与LLM技术开展风险要素的自动化识别与语义抽取，通过多指标量化模型为核心构建多维度、可追溯的风险评估框架，有效支撑事前安全评估。但是上述研究在跨产业场景业务的精细化建模方面还存在差距，难以在数据出境场景下实现高精确率的业务建模。此外，由于数据出境风险描述语义具有多样性，现有方法在高召回率的风险要素识别任务上仍存在不足，并可能导致过高的数据标注成本。而且现有评估指标体系多以静态风险要素为依据进行启发式构建与评估，尚未充分刻画不同数据出境产业场景中风险要素的差异性，难以满足多样化产业场景的普适性评估需求。因此，未来研究应聚焦于如何在多样化数据出境产业场景下设计具备高精确率的事前安全评估方法，以提升事前风险识别与前瞻性预判能力。

2.1.2　事中动态评估

由于事前安全评估存在监管与促进数据流通利用的矛盾，仅通过事前检查难以发现出境业务实际开展中的违规问题，因此，事中动态评估是必要的技术保障，体现了《数据出境安全评估办法》中“事前评估和持续监督相结合、风险自评估与安全评估相结合”的基本原则。事中动态评估是指对数据出境行为进行实时、持续、自适应的安全合规评估，技术基础是动态风险要素感知能力，文献[38]针对工业控制领域的安全监测场景以滑动窗口的方式对每个时间窗内业务数据的风险特征做归因分析并计算风险要素评分，实现动态风险要素感知。为分析实时行为或安全事件引发的风险要素变化，文献[39]通过场景化仿真，分析医疗物联网安全事件对风险要素的影响，识别不同时间阶段的动态风险要素。为识别由外部规制对风险要素的影响，文献[40]利用NLP模型识别财经新闻中的企业涉及的多类风险事件，如监管、财务、政策等，并通过风险评分实现企业风险状态的动态更新，从而识别新增的风险要素。针对上述关键技术，可引入F1分数评估不同方法的动态风险要素感知能力，引入检测延迟评估识别任务的速度表现。动态风险要素的准确感知驱动了事中动态评估的有效性，间接影响风险预警与处置的效率。在数据出境各产业场景中，通常需要在保证低检测延迟的情况下，使得F1分数能够达到相应的性能要求。

为了更准确地支撑事中动态风险评估，必须对评估指标做出优化调整。文献[41]从优化原始风险评估指标贡献调整视角出发，提出一种结合进化算法与决策树的模型结构优化方法，对树形评估结构的叶节点应用最小绝对收缩和选择算子（LASSO）实现指标稀疏化选择，利用进化算法搜索最优树结构，达到对风险评估指标优化调整的目的。为提升动态性，文献[42]以专家咨询的方式针对安全事件

等对评估指标进行筛选、修正和补充，并利用熵权法对指标赋予权重，实现评估体系动态优化，支撑风险动态评估。区别于专家评分，文献[43]面向物联网场景以一种数据驱动的方式对动态风险利用引力搜索、灰狼优化等启发式策略筛选并重构风险评估指标。文献[44]则提出一种基于条件互信息的动态指标选择方法，学习单样本序列化指标的获取策略，基于信息贡献的贪心思想实现指标集合的动态重构。针对上述关键技术，可引入曲线下面积（AUC, area under curve）、F1分数等指标，对不同方法在动态调整前后风险评估任务预测性能的提升进行验证。引入调整响应延迟评估不同方法的更新效率。在数据出境场景中，低延迟、准确的评估指标体系调整能够显著增强事中动态风险的量化能力，使系统能够更及时地识别与管控潜在风险事件。在制造与工业控制、信息通信与互联网等业务规模大、风险要素动态变化的产业中，通常需要在更低调整响应延迟下仍能维持较高的F1分数。而金融与保险、医疗与生命科学等产业由于风险样本稀缺、类别分布不均衡且合规压力更高，更强调在可接受的响应延迟范围内达到更高的AUC要求，从而支撑更稳健的风险体系优化调整。

在上述技术研究支撑下，事中动态评估的关键在于通过量化手段将感知的风险要素与优化的评估指标转化为可操作的风险决策。早期探索主要依赖静态评估工具，核心特征是预定义规则库与周期性审计，但无法适应动态性。伴随GDPR的生效，催生出具备动态分析能力的自动化数据出境合规评估工具链[11]，标志着评估从声明式合规转向行为式验证。为提升动态量化评估能力，文献[10]提出将数据出境构造为包含境内传出机构和境外接收机构的二分网络，根据网络结构、传输时间特征等动态因素，通过设置权重计算动态风险，实现了基于业务日志的数据出境事中风险量化评估。为了更灵活、更精准地量化风险，文献[45]提出一种基于层次分析法与熵权法混合权重的量化评估框架，对各阶段、各时期的关键风险指标计算权重实现风险动态度量。而文献[46]则提出一种基于模拟风险演化过程的动态风险量化框架，构建风险要素之间的演化关系网络，利用灰色关联分析（GRA, grey relation analysis）对风险指标重要性进行排名，并根据综合影响度计算风险指标权重，实现风险动态量化评估。针对上述关键技术，可引入评估模型校准度（ECE）用于衡量结果的预测概率和真实的经验概率是否保持一致，验证在动态风险环境下模型是否能够真实地反映风险状况。引入F1分数验证不同方法地风险评估效能。引入评估延迟以评估响应效率。在数据出境场景中，评估量化结果的可解释性与可决策性为各产业场景的首要标准，直接影响风险预警与处置的执行正确性。因此通常以更低的模型校准度为优先考量，并尽可能压缩评估延迟。此外，在信息通信与互联网、制造与工业控制等高动态产业场景中，对F1分数通常设定更高要求，以在快速演化的行为模式下保持稳健识别能力，而在金融与保险、医疗与生命科学等强合规与高敏感度场景，则往往设定更低的校准度目标以确保能可靠支撑审计与决策。

本文在表5中总结并归纳了事中动态评估的关键问题与需求、核心要点描述、代表性文献与典型评价指标。总体而言，事中动态评估关键技术正从静态、粗粒度的工具转向精细化实时动态评估，现有研究通过滑动窗口分析、场景化仿真及外部信息挖掘等方法，实现了对业务行为变化、监管环境变化以及安全事件影响的实时捕捉。通过进化算法、启发式优化与熵权法等手段，逐步构建了可自适应调整的指标体系，使风险评估输入能够随环境变化持续优化。这些研究共同奠定了事中动态评估的技术基础，但由于数据出境风险要素非平稳变化，不同产业场景业务流程多样且动态运行，高风险要素并非显著、频繁出现，导致现有研究在隐性风险挖掘与感知、评估体系低延迟迭代更新能力上存在差距，动态风险要素识别、指标体系调整以及风险动态量化评估的F1分数难以突破瓶颈。因此，未来研究应聚焦于解决如何能够面向动态变化的多元风险实现低延迟且高预测性能的事中动态评估难题。

2.1.3 事后溯源追责

事后溯源追责通常是违规行为发生后，通过技术手段对数据链条、违规源进行追踪并确定责任主体的过程。现有监管体系明确了其核心地位，欧盟GDPR强调“问责制原则”，要求企业有可证明的合规性。美国积极推动的全球跨境隐私规则（CBPR, cross-border privacy rule）倡议，强调自我评估与问责代理认证。中国《促进和规范数据跨境流动规定》明确强化事前、事中、事后全链条全领

域监管。在技术研究方面，事后违规行为可追溯的实现基础是完备的数据重建工作，文献[47]针对主机侧审计日志构建了主机级的数据流图，并根据异常事件识别与规则推理实现了端到端的数据流转链条还原。为实现全局数据流转图构建，文献[48]拓扑、资产、安全事件等多源异构数据统一表示为能够反映数据流转关系的攻击图结构，并引入吸收马尔可夫链对状态转移过程进行建模，对数据流转路径进行概率推演，实现被攻击数据在网络中的完整流转过程重建。而在真实的监测环境下，重建的数据难免存在缺失、噪声、甚至错误值的情况，为解决此问题，文献[49]提出一种图结构的数据重建策略，在存在数据缺失、噪声错误值的情况下，通过图卷积等算子综合邻居节点的历史观测信息对当前节点的数据进行插补、去噪与重建。还有部分技术研究是将缺失补全建模为链路预测问题，例如文献[50]针对每条待预测边自动选择其最相关的个性化邻域子图，针对子图利用图神经网络（GNN, graph neural network）进行结构表示学习，从而预测该缺失的边是否真实存在。针对上述关键技术，可引入F1分数、召回率、平均绝对误差（MAE, mean absolute error）等典型指标评估不同方法的数据重建性能。在数据出境场景下，数据重建的首要目标是不遗漏关键路径与关键主体，确保溯源结果的全面性与准确性。在此准则下，召回率通常被视为最重要的评估指标，召回率越高表示模型越能避免遗漏关键违规路径。在确保较高召回率的前提下，再以F1分数评估不同方法在抑制伪链路与维持整体判定可靠性方面的综合表现。而对于包含时间戳、数据量、计数等数值型风险要素的重建过程，则引入MAE衡量插补与去噪后的量化偏差，以评价不同方法在时间线与量值判断方面的可信度。

违规行为的溯源追责可追溯至20世纪90年代，学术界将其定义为“数据日志、数据存档或数据来源[51]”，为实现上述目标，传统方法多采取集中式日志审计，但是人力成本高，且抗篡改能力弱。数字取证作为一种事后审计技术被引入，但在多角色、复杂链条的数据出境场景下，不同国家受法律约束影响对于电子证据的采录差异巨大且成本高，造成溯源困难。在此基础上，区块链技术因其去中心化、不可篡改、时序固化特性，被视为解决数据出境溯源审计问题的有效方案，但由于数据共享困难且真实性难以判断，限制了其广泛应用。因此，碎片化证据支撑下的违规溯源追责成为关键技术要点，文献[52]引入反事实因果解释框架，从数据流转交互图中学习因果关系，指导反事实样本生成，针对违规行为通过干预特定特征节点来识别导致结果变化的直接因果因素，实现节点级的溯源推断。而文献[53]提出一种根据观测状态反推初始源以及传播路径的扩散模型，其建模过程采用易感者–感染者–康复者模型（SIR, susceptible infected recovered model）结合马尔可夫链形式构造状态转移矩阵，并设计了一个可逆残差图卷积网络作为反向生成器，从观测状态重建出可能的扩散路径和源节点分布。在此基础上，为增强溯源方法的可解释性，文献[54]将数据传输路径、邻居节点行为和间接信任证据进行融合，并通过路径回溯对可疑行为的跨节点流转过程进行重建，从而追溯数据违规传输路径并识别关键违规节点。对于数据出境而言，部分场景涉及多角色复杂链条，违规事件责任判定是必要的工作，也是新的技术难题，文献[55]提出一种矩阵组织结构，涉及8种“单位角色类型”的分类方法，明确了矩阵结构中各单位角色“应做什么”“不能做什么”“与谁协作”的责任边界，从而减少冲突和责任模糊。文献[56]针对医疗健康产业领域提出数据监护者的概念，用于承担数据获取、管理、存储、使用和共享等环节的信托责任，明确业务角色链条中不同参与方的职责分工，形成面向数据全生命周期的责任判定机制。针对上述关键技术，可引入准确率指标评估不同方法的违规溯源与责任判定能力。在数据出境场景中，各产业场景对于违规链路重建与责任主体识别的核心要求是保持较高的溯源追责结果判定的准确性。较高的准确率能够确保溯源过程真实反映违规行为链条，为违规事件后续的风险处置、合规治理和责任追究提供可信依据。

此外，受法律法规影响，安全与隐私问题是事后溯源追责另一重要的研究方向。文献[57]提出轻量级安全搜索方案以支持医疗数据共享中的加密检索，使隐私保护检索在保持效率的同时具备可审计性和责任可追踪性。文献[58]则提出一种基于电子健康记录分类的分布式数据完整性审计方案，利用分类标签的思想并配合布隆过滤器提升审计的效率，并降低隐私泄露的风险。在此基础上，为提升

可解释性和溯源效力，文献[59]基于压缩感知和数字水印技术构建了多层级可逆隐私保护模型，一方面，该模型通过嵌入式水印实现对数据来源及其传播路径的追踪，另一方面，不同权限的用户可获取不同精度的数据视图，并在满足授权条件时恢复原始数据，从而为事后违规行为的审计与取证提供了有效的技术支撑。针对上述关键技术，可引入MAE与误报率（FPR, false positive rate）等指标评价不同方法抵御非授权方对敏感数据的识别能力，支撑判断隐私保护强度。可引入响应延迟、存储开销（Storage overhead）用于评估不同方法应用后造成的开销情况。在数据出境场景中，各产业均涉及不同程度的敏感数据，因此在保证可审计性与可追责性的同时，应尽量降低敏感信息在溯源过程中的暴露风险。溯源追责通常作为离线过程执行，可以通过索引技术减少响应延迟，但需要考察存储开销。基于此，一般首先要求保持较低的MAE，用于衡量在提供可用溯源证据的同时，引入的扰动是否处于可控范围，从而避免隐私泄露风险。对于金融与保险、医疗与生命科学以及教育等高隐私敏感产业，由于隐私违规会直接引发监管风险，不仅要求更严格的MAE，还需降低FPR，以确保非授权主体在溯源过程中无法通过误判推断出敏感信息。在信息通信与互联网产业，由于数据量巨大且交换频繁，溯源证据的存储成本随系统规模迅速扩大，因此需在不削弱隐私保护能力的前提下将存储开销控制在较低水平，以保障工程化部署的可行性。

本文在表5中总结并归纳了事后溯源追责的关键问题与需求、核心要点描述、代表性文献与典型评价指标。总体来看，研究从重建数据流、跨网络攻击图建模、图结构数据补全与缺失链路预测等方向推动了数据流转链条的可追溯性，同时引入因果推断、扩散反推、路径回溯等方法增强违规溯源的可解释性。在责任划分方面，则从组织矩阵结构与产业数据监护者机制上明确业务链条中各角色的职责边界。在安全与隐私约束下，通过可审计检索、完整性审计、多层级可逆隐私保护等技术兼顾了隐私保护与可追责性。然而，这些研究仍多依赖出境主体角色边界清晰、上下文稳定的出境链条来实现违规行为的溯源追责，但是数据出境场景产业多样、出境业务动态，链条长短不一，现有研究在复杂的碎片化链条下的违规溯源追责准确率方面仍存在差距。而且受法律约束，数据出境违规溯源与隐私保护存在根本性矛盾，导致现有方法在隐私保护强度与溯源开销方面的平衡能力不足。在未来的研究中，应聚焦于研究包含多重角色主体（例如既是控制者又是处理者）的动态复杂处理链条的精准违规溯源与责任清晰判定难题。

2.2　风险预警相关关键技术

2.2.1　违规数据识别

违规数据识别的目标是在保障数据自由流动的同时，有效防范因数据不合规出境而引发的安全风险。在数据出境监管初期，技术手段主要源自传统的安全领域，其核心思想是在企业的网络出口或关键链路上部署监控设备，采用深度包检测技术通过深入检查数据包的有效载荷并根据预设规则来识别和拦截可疑的数据传输[60]。而随着加密协议的普及和法律对处理用户隐私数据的限制，违规数据识

表5　过程监管相关关键技术分析

过程监管维度	关键问题与需求	核心要点描述	代表性文献	典型评价指标
事前安全评估	数据出境精细化描述	跨产业出境业务流程精细化建模	文献[15-16, 30]	Latency、TPS、Coverage、Precision
	高效率事前安全评估	海量评估文档中风险要素自动化抽取	文献[31-33]	F1 score、Recall、DAE、Latency
	差异化风险普适评估	动态多样化产业场景评估体系自适应构建	文献[7, 35-36]	F1 score、Accuracy
事中动态评估	动态风险要素识别	动态风险变化下的风险要素感知辨识	文献[38-40]	F1 score、Latency
	风险评估指标调整	风险要素变化下的评估指标优化调整	文献[41-43]	AUC、F1 score、Latency
	动态风险量化评估	面向多维风险要素的动态量化评估	文献[10, 45-46]	ECE、F1 score、Latency
事后溯源追责	碎片化数据证据重建	多源碎片化数据下重构时空线索视图	文献[48-50]	F1 score、Recall、MAE
	精准溯源与责任判定	复杂角色交互链条下违规溯源与清晰定责	文献[52,54,56]	Accuracy
	违规溯源的隐私保护	不暴露原始敏感信息的精准违规溯源	文献[57-59]	MAE、FPR、Latency、Storage overhead

别在监管侧更倾向于采用侧面的辅助手段来识别违规数据出境行为。文献[61]采用NLP技术在契约协议文本层面，对数据类别、处理目的、敏感等级等摘要信息进行内容分析，实现了基于GDPR提炼规则的自动化合规检查。文献[62]则针对流转事件日志，精确识别流程执行过程中违反业务或监管规则的不合规行为。针对上述关键技术，可引入F1分数、召回率、FPR等指标用于评估不同方法在识别任务方面的能力。违规数据的识别能力直接决定了风险预警与处置的有效性。在数据出境各产业场景中，通常以较高的召回率为首要考量，以评估方法在识别违规数据与违规行为时的漏检能力。在此基础上，以F1分数衡量在既定召回率下“查全–查准”的综合效能，并通过约束FPR，以减少误报带来的不必要审查与引发的合规和运营成本。

此外，在个人隐私逐渐被重视的情况下，风险自评估则成为违规数据识别的另一重要手段，既能够配合监管方提升合规水平，也能够在与监管方互动中积累信任资本。文献[63]采用文本指纹与相似性比对技术实现重复文本检测，为相似违规内容的自动识别提供了有效支撑。在此基础上，文献[64]面向文件内容、网络数据包有效载荷、日志行、邮件正文等任意大规模字符串或文本流，基于高效的多模式匹配实现了规则式风险扫描。虽然企业风险自评估能够帮助监管侧提升风险管控水平，但其自评估操作及报告可信度是另一个需要关注的问题，为解决此问题，文献[65]借助受管的关键基础设施，对各类事件进行集中采集与标准化处理，从而实现对安全操作的统一管理与可信存证。为缓解基础设施的存储压力，文献[66]引入高效的级联纠删码提出区块链分片存储方案，即使在节点失效、网络抖动时仍能恢复全链数据，从可用性和完整性维度增强链上操作记录的长期可信留存能力。为增强监管方对日志的审计能力，文献[67]设计了一种可证明不可篡改、可追责的证据日志通用框架，通过引入日志服务器、监控者与审计者等角色，以及基于梅克尔树（Merkle Tree）的动态承诺结构，该方案将事件记录组织为可公开验证并具备防抵赖、防诬陷特性的透明日志体系，实现更高等级的可信存证保障。针对上述关键技术，可引入准确率指标评估不同方法的违规数据识别能力，引入识别延迟指标用于评估不同方法的识别效率。由于风险自评估通常发生在数据出境源头侧，因此对于大部分数据出境产业场景而言，自评估系统应在较低的违规数据识别延迟下保持较高的准确率，以确保风险识别结果能够及时且可靠地支撑后续治理流程。在金融与保险、医疗与生命科学以及教育等安全敏感产业，自评估结果往往直接影响合规审计质量，因此对识别准确率的要求更为严格，以避免因误判导致合规处罚、隐私风险或安全责任扩大。而在信息通信与互联网产业，由于数据规模庞大，更强调对识别延迟的控制，避免因识别延迟过高影响业务流转效率。

本文在表6中总结并归纳了违规数据识别的关键问题与需求、核心要点描述、代表性文献与典型评价指标。综上所述，现有违规数据识别相关技术能够在可观测平面上较为系统地发现显性违规模式、对合规文本与操作日志进行自动化规则匹配与审计，并借助透明日志与链上存证机制为后续合规取证和责任划分提供一定的抗篡改技术基础。然而，在数据出境场景中，强加密网络环境导致违规流量与业务场景语义高度脱节，现有研究难以将各类流量行为与具体的数据敏感等级、研究目的以及境外接收方职责等精确对齐，导致数据出境的违规数据识别准确率受损，此外，企业侧自评估机制虽能提升整体合规水平，但其自我申报的可信度和可验证性仍然是难以彻底解决的结构性问题。未来研究亟需突破的关键难题在于如何在不过度侵入用户隐私的前提下，依据侧信号（流量模式、日志操作记录等）精准识别出境违规数据。

2.2.2　异常行为判别

异常行为判别的主要目标是从行为侧识别与正常业务模式不符、可能违反安全策略或法规的出境行为。早期依赖于传统网络安全手段，主要采取由专家预定义规则集[68]，例如阈值、目标、协议等规则，用于描述不合规或者恶意的出境行为，但是其规则维护成本较高。随着《数据出境安全评估办法》《促进和规范数据跨境流动规定》等法规的持续颁布，事前备案机制成为强有力的规则判别依据，在一定程度上缓解了人力成本，有效识别备案信息与动态业务运行之间的差异是支撑异常行为判别的高效手段。在早期的解决方法中，文献[69]提出一种基于树编辑距离的算法，通过将运行态数据

和规则文件分别抽象为两棵具有层级结构的有序树，通过计算当前运行状态与规则树之间的树编辑距离来刻画其结构差异，并根据得到的最小编辑序列精确定位差异位置。为增强差异化识别的可解释性，文献[70]通过构建运行态节点与规则节点的二分图，并以节点结构特征与属性信息计算边权作为相似度度量，再通过最小代价匹配获得全局一致的节点映射关系，基于该映射生成的变更脚本可直观展示节点新增、删除、修改及移动等差异轨迹，从而显著增强差异化识别的可解释性与精确性。文献[71]则将规则化流程与运行态数据建模为带属性的图，并将差异化识别转化为图编辑距离估计，该方法基于格罗莫夫–沃瑟斯坦距离（GWD, Gromov-Wasserstein distance），利用图的结构关系与节点属性进行对齐，能够识别网络级的复杂结构差异。针对上述关键技术，可引入准确率、召回率、FPR等指标评估不同方法在异常或违规行为的判别能力，引入检测延迟验证不同方法的异常识别效率。在数据出境场景中，能否高效、准确地识别动态业务运行与备案信息之间的偏离情况，直接影响整体数据出境风险管控的质量。在此基础上，还需保持较低的检测延迟，以确保异常行为能够在可接受的时效窗口内触发风险预警与处置流程。因此，在维持较低检测延迟的前提下，应优先提升准确率，并保持FPR符合要求，以避免误报带来的审计成本或不必要的监管触发。按照数据出境产业实践与指导意见，建议准确率不低于95%，FPR不超过25%。此外，在制造与工业控制以及能源与交通等领域，异常行为可能对生产连续性或公共安全造成直接影响，因此强调在检测延迟要求下保持更高的召回率。在金融与保险及医疗与生命科学产业，由于业务链条复杂且合规压力较高，还需在确保召回率的同时进一步降低FPR。另外，在信息通信与互联网产业，规则偏离检测也需要考虑数据规模对算法整体效率的影响，通常在确保一定准确率的前提下，降低检测延迟以保障业务连续性。

上述基于规则的手段能够直接识别违规或异常出境行为，然而，一旦出境企业不配合或者假意配合，上述手段难以应对。为克服静态规则的局限性，研究者们提出基于行为模式的识别方法，核心思想是首先为“正常”的数据出境行为建立一个基线模型，然后将实际行为与该模型进行比较，若偏差超过某个统计显著性水平，则判定为异常。在早期研究中，文献[24]采用自回归滑动平均模型（ARIMA, auto-regressive moving average model）对行为序列进行线性时间序列建模，以学习其正常变化轨迹，并通过预测值与观测值之间的残差是否超出置信区间来判定异常时刻。针对异常样本数量不足导致的识别能力下降问题，文献[23]在隐马尔可夫模型（HMM, hidden Markov model）基础上设计了对抗数据增强机制，通过生成对模型具有挑战性的样本来强化其判别能力。为了在数据中学习到更复杂的行为模式表征，文献[25]提出基于序列到序列（Seq2Seq, sequence to sequence）模型的深度神经网络方法，并结合滑动窗口的序列建模策略，以捕捉序列中的强非线性关系、长程依赖及多维耦合特性，为异常行为模式识别提供了更具表达力的特征基础。针对上述关键技术，可引入F1分数指标评估不同方法的异常行为识别能力，引入检测延迟验证不同方法的异常识别效率。引入受试者工作特征曲线下面积（ROC-AUC, area under the receiver operating characteristic curve）、精确率–召回率曲线下面积（PR-AUC, area under the precision-recall curve）等指标评估不同方法能否有效区分正常与异常模式。在数据出境场景中，基于行为模式的异常识别方法通常需要更为复杂的算法设计，但也提供了一种具有实际工程落地价值的异常识别辅助手段。该类方法在各产业场景下对检测延迟的敏感度相对较低，因此在可接受的延迟范围内，保持较高的F1分数成为当前应重点优化的性能目标，用以评估不同模型对异常行为的整体判别能力。此外，在金融与保险以及医疗与生命科学等高敏感产业中，异常行为样本比例极低，但其潜在影响高度严重，因此还需依赖较高的PR-AUC与ROC-AUC，以确保模型能够在样本不平衡的条件下有效区分少量异常样本。

异常行为模式识别虽然在一定程度上克服了规则匹配的刚性缺陷，但是忽视了数据出境过程中多主体间复杂的交互拓扑与协同关系。当面对分布式协同异常行为（如多节点协同数据渗漏、跨主体责任规避等），因其缺乏全局视野而难以识别跨主体的行为关联性与意图一致性。针对上述问题，文献[72]通过建立基于历史窗口的时间序列基线实现统计式异常预警，将多主体协同异常视为多个监测序列在

同一时间段内的同步异常升高，从而为可能的协同行为提供初步的预警依据。进一步，文献[26]将监测节点及其拓扑关系建模为具有空间约束的多元时间序列，通过无监督的时空状态估计识别偏离正常时空依赖的节点或链路，并据此揭示在时空维度上呈协同偏移的群体性异常行为。为了更精准地捕获多主体协同异常行为，文献[73]提出基于时空轨迹建模的活动时空动力学（ActSTD, activity spatio-temporal dynamics）模型，通过学习多主体正常的时空演化模式构建行为基线，进而将显著偏离该基线的联合轨迹视为潜在的协同异常行为。文献[74]则提出在时间网络中识别预定义的子图结构，通过统计这些带有时间顺序的局部交互模式，为多主体的协同行为异常检测提供了微观网络结构视角。针对上述关键技术，可引入F1分数、召回率等指标评估对不同方法的异常行为识别能力，引入ROC-AUC、PR-AUC等指标评估不同方法能否有效区分协同异常与正常模式。在数据出境场景中，多主体协同异常往往具有更强的隐蔽性与破坏性，可能引发更高等级的数据安全风险。因此，召回率为首要性能标准，用于评估模型对协同异常行为的不漏检能力。在此基础上，F1分数用于衡量模型在保证召回率的同时对误判的抑制效果，从而反映整体判别质量。ROC-AUC与PR-AUC可用于评估模型在不同阈值下区分正常模式和协同异常模式的判别稳定性，其中PR-AUC在协同异常样本比例极低的场景中具有更高代表性，能够反映模型在样本不平衡场景下的识别能力。

本文在表6中总结并归纳了异常行为判别的关键问题与需求、核心要点描述、代表性文献与典型评价指标。综上所述，现有异常识别相关技术研究总体形成从规则差异识别、行为基线建模到多主体协同分析的技术演进路径，规则差异类方法通过树与图结构对齐实现备案与运行态的一致性校验，具备良好的可解释性。行为模式类方法利用统计与深度序列模型刻画正常出境行为，从而识别偏离基线的异常模式。多主体协同方法将时空依赖与网络结构纳入建模，为识别跨节点、跨主体的协同行为提供了初步能力。然而，由于数据出境违规手段多样，现有方法尚无法在同一语义空间内统一刻画法律规制目标、备案要求与实际出境行为，难以直接支撑对“超范围传输”“目的外使用”“未备案共享”等问题的明确判定。同时，现有技术在“低速慢流”式的长期渗漏出境与新型未知违规出境手段识别方面仍存在差距，容易被掩盖在长期正常业务噪声中。此外，多主体协同识别普遍依赖固定拓扑或同步异常等强假设，难以覆盖跨云、多域、异构协议环境下真实的数据流动路径。导致现有技术在异常行为识别能力方面普遍存在不足。未来亟须解决的核心难题在于如何在不完全观测且具有强对抗性的多主体协同违规出境行为。

2.2.3　风险预警分析

风险预警分析的核心在于针对异常或违规事件进行分析并触发相应的预警响应，帮助数据出境安全风险提前发现和主动防控。一旦发生异常或违规的数据出境行为，其影响可能迅速扩散并波及更大范围。因此，风险扩散趋势预测技术十分必要，可用于评估潜在风险的传播方向与影响程度，从而支持更及时、更精准的干预措施。现有研究方法多依赖于信息扩散理论支撑风险建模，可以描述风险如何在节点（如企业、服务器）之间传播。文献[75-76]将流行病学机理模型与深度神经网络相结合，能够在复杂网络上刻画和预测“类似传染过程”的扩散轨迹与趋势，为风险在网络中的时序传播建模提供支撑。文献[77]在多元时间序列上构建图结构，通过图注意力机制和扩散建模刻画不同系统变量之间的依赖关系与信息传导过程，能够有效分析异常或风险在多维指标间的扩散路径与传播模式。文献[78]针对属性图提出深度自适应生成（DAG, deep adaptive and generative）方法，利用编码器对节点属性与拓扑结构进行联合表示学习，并设计可学习的社区隶属度推断模块，在无监督条件下自动识别图中高度耦合的节点子群体与潜在结构社区，定位风险在网络中可能快速聚集与扩散的关键区域。针对上述关键技术，可引入F1分数、AUC等评估指标衡量不同方法在风险传播预测成功率方面的能力，引入MAE评估不同方法对扩散范围的预测误差。在数据出境场景中，较高成功率的风险扩散趋势预测能够有效阻断违规风险的进一步蔓延，为风险预警与监测处置响应提供关键的策略依据。对于各类数据出境产业场景而言，AUC往往处于更高的优先级，其较高数值意味着模型在区分“是否可能被风险波及”方面具有可靠的判别能力，是风险扩散预测是否具备落地实用性的核心指标。而对于制造与

工业控制、能源与交通等安全关键产业，由于资源投放与隔离窗口有限，更强调较低的MAE，以确保对扩散范围与强度的预测足够精确，从而支撑快速且准确的防控决策。在金融与保险、医疗与生命科学等高敏感产业中，潜在风险外溢可能引发严重的合规后果，因此需要进一步提升AUC以稳健区分可能被波及的对象，并在此基础上提高F1分数，以降低漏判带来的合规与安全风险。相对而言，在教育等风险强度整体较低的产业场景中，通常以AUC与F1分数作为主要评价组合，并将MAE控制在运维可接受水平，以保障总体识别质量与资源投入之间的平衡。

风险预警的早期研究主要借鉴信息安全监控技术，通过将领域专家的知识和经验固化为预设规则，以辅助风险预警响应。随着技术的发展，简单的规则引擎逐渐与统计模型相结合，例如利用回归分析对历史数据进行建模，从而识别异常行为并支持预警决策[79]。然而，这类方法往往需要对数据分布做出较强假设，在处理高维度、非线性等复杂特征的数据时表现受限，而这些特征正是数据出境流转过程中普遍存在的。随着人工智能技术的成熟，数据处理能力逐渐增强，文献[20]基于历史预警事件提取多种风险预警特征，通过K-means聚类结合专家评分构建多级预警标签，并据此训练多分类随机森林模型，实现了工业系统场景中风险事件预警等级判别。而文献[80]面向工业互联网场景构建了由预测模块与预警模块组成的早期预警框架，采用GNN与时间卷积网络（TCN, temporal convolutional network）的双通道结构对多维时间序列进行联合建模，实现未来风险状态的预测，并基于统计特性与专家评分设计了时间敏感的异常评分与分级预警机制。为避免依赖专家评分带来的主观性偏差，文献[81]基于历史风险事件数据构建了随机森林等监督学习回归模型，通过同时建模事件发生频率与影响程度，对未来特定时间点的风险等级或强度进行预测与排序，有效支撑数据出境智能化风险预警。针对上述关键技术，可引入预警延迟指标比较不同方法的预警效率，引入F1分数指标评估风险预警能力，引入ECE指标用于验证是否能够真实反映风险状况。在数据出境场景中，通常以预警延迟、ECE与F1分数3项指标共同衡量整体效果。其中，预警延迟反映系统对风险信号的响应时效，是工程部署的基础约束。F1分数可以刻画在不同风险等级下的综合命中效果。ECE则用于检验风险评分或预警概率与真实风险水平的一致性，确保分级阈值、资源调度与联动处置具备可解释性与可决策性。总体上三者需协同优化，但不同产业侧重存在差异，制造与工业控制、能源与交通等安全关键产业受限于短时效窗口，通常要求更低的预警延迟。金融与保险、医疗与生命科学等高敏感场景风险预警直接影响合规动作，更强调更低的ECE，并同步维持较高的F1分数以降低漏判与误判。

本文在表6中总结并归纳了风险预警分析的关键问题与需求、核心要点、代表性文献与典型评价指标。综上所述，现有风险预警分析相关技术研究已能在复杂网络和多元时间序列上刻画“类似传染过程”的风险传播，利用机理模型与深度神经网络、图注意力及DAG等方法识别关键节点和风险传播群体，并刻画时序风险演化，并通过聚类结合随机森林、GNN和TCN等模型实现风险状态的分

表6 风险预警相关关键技术分析

风险预警维度	关键问题与需求	核心要点	代表性文献	典型评价指标
违规数据识别	违规数据内容识别	面向多源异构日志的违规数据审计	文献[60-62]	Recall、F1 score、FPR
	违规风险自我发现	动态多样化风险的自评估与可信存证	文献[63-65]	Accuracy、Latency
异常行为判别	备案与动态业务间差异感知	度量宏观备案信息与微观行为间差异	文献[69-71]	Recall、Accuracy、FPR、Latency
	异常行为模式识别	动态出境业务的非常规行为模式识别	文献[23-25]	F1 score、Latency、ROC-AUC、PR-AUC
	复杂协同行为识别	多主体协同参与的异常出境行为识别	文献[26,73-74]	F1 score、Recall、ROC-AUC、PR-AUC
风险预警分析	风险扩散趋势预测	差异化出境主体约束下风险扩散分析	文献[75, 77-78]	F1 score、AUC、MAE
	风险智能预警	多维动态风险事件的精细化分级预警	文献[20, 80-81]	Latency、F1 score、ECE

级预警与强度预测，具备较高的自动化与精细化水平。然而，上述研究仍主要延续传统网络安全视角下的技术导向，缺乏对数据安全场景的适配能力，在建模方法上缺乏合规能力、监管力度等约束特征，对数据敏感等级、业务及产业语境的描述能力仍然不足，难以实现高准确率的风险扩散预测及预警。未来应明确以数据安全为核心，聚焦于解决适配出境场景多维风险特征的风险扩散分析建模与预警难题。

2.3　监测处置相关关键技术

2.3.1　源头监测

源头监测主要是指在数据出境企业的近端借助自动化工具实时监控并采集数据出境传输行为特征，支撑异常判断、违规处置、风险评估等下游任务，同时也能够在一定程度上防止敏感数据未经授权泄露或被恶意窃取。早期方法主要依赖于网络边界防火墙实现对IP地址、端口和少数协议等有限特征信息的采集。随后，数据防泄露（DLP, data loss prevention）技术被提出[82]，其中网络DLP通过在网络出口旁路部署，实时捕获文件操作、网络传输等行为，识别数据层面的风险活动。受数据加密与敏感等级影响，传统的内容安全手段并不适用于当前用户隐私友好的数据出境监管环境。在违规风险日益隐蔽且数据出境活动日趋频繁的背景下，精细化的数据出境风险信息采集成为异常与违规识别的关键基础。文献[83]在网络流量记录（NetFlow）基础上提出流量雷达（FlowRadar），通过将流五元组及包、字节计数写入交换机侧的可解码编码表，按短时间窗导出并在收集端统一解码实现全流量级特征采集，但是其可采集指标仍然有限。文献[84]则依托采样流量技术（sFlow）在网络监控设备上实现了部分包级样本及部分载荷信息的采集，扩展了基于流的监测能力，但其核心思想是随机采样，在采样率较低或流量突发的情况下，部分关键流量可能未被捕获，存在采集完整性不足的风险。文献[85]提出基于规则的网络入侵检测工具Snort，对监控链路上全包数据进行捕获与逐字节分析，实现极为精细化的流量采集。但是在大规模出境链路环境下可能导致交换设备带宽或存储瓶颈。此外，文献[12]在受控环境下提出P4control，在链路采集信息的基础上通过扩展的伯克利包过滤器（eBPF, extended Berkeley packet filter）补充了主机侧的进程及上下文信息，从而实现了跨主机信息流采集，然而，此类采集方式可能涉及用户隐私，具有一定的入侵性。针对上述关键技术，可引入监测覆盖率（Coverage）用于评估给定网络负载下，实际被捕获的数据流或包所占比例，用于评估采集监测能力。引入采集延迟、TPS等验证监测效率，引入可部署性（Deployability）评估指标，综合部署位置灵活性、部署难度、部署成本、能否轻量化部署、业务无干扰能力、隐私合规风险等多个与当前场景适配的维度，可通过专家评估或实证案例确定各维度权重，构建加权综合评分，比较不同方法在实际监管环境中的落地可行性。在数据出境场景中，高评分的可部署性通常被视为源头采集技术的首要要求，是判断方案能否在真实监管环境中落地的基本准则。相较而言，监测覆盖率、采集延迟与TPS构成并列的评价指标，用于共同刻画采集面的充分性、时效性以及在高并发负载下的稳定性表现。在实践中，不同数据出境产业侧重存在差异。在金融与保险、医疗与生命科学等高敏感产业中，可部署性的评分通常更高，尤其强调业务低干扰性与隐私合规性。在信息通信与互联网产业，由于数据规模和并发度极高，需要在可部署性约束下同时保障高TPS与可接受的采集延迟。在制造与工业控制、能源与交通等安全关键产业，受制于短时效处置窗口与业务连续性要求，更强调较低采集延迟与可接受的TPS，并在满足实时性的前提下确保监测覆盖率达到支撑下游分析任务的必要阈值。

随着数据出境规模的持续扩大，源头监测面临显著的采集开销压力，如何实现低开销的监测机制成为亟须解决的关键问题。在早期研究中，文献[13]提出将采集逻辑嵌入可编程数据平面，并将采集信息附加到业务流量的数据包自定义头部中。该设计将采集、存储与传输开销从交换设备转移到业务数据包中，但增加了数据包头部负担和整体传输开销。文献[83]进一步设计了一种基于压缩编码的采集方法，能够以较低的存储和传输开销实现业务流量信息的高效采集。文献[86]则将采集决策建模为马尔可夫决策过程，利用深度强化学习根据上下文环境自适应调整采样频率或强度。该方法能够有效根据当前网络状态动态控制采样率，并实现持续的在线策略更新。针对上述关键技术，应引入CPU利用率（CPU utilization）、采集延迟、TPS、带宽开销

（Bandwidth overhead）、存储开销、内存开销（Memory overhead）等指标用于比较不同方法的监测开销性能。不同指标在各类数据出境产业场景中的重要性存在显著差异。在信息通信与互联网场景，由于数据规模庞大且并发度极高，通常优先保障高TPS与可接受的采集延迟，并将带宽开销与存储开销控制在容量规划阈值内，以支撑持续的高流量业务运行。在制造与工业控制、能源与交通等安全关键产业中，处置窗口短且链路拓扑相对固定，更强调低采集延迟与可控的CPU利用率，以避免对控制面的干扰，同时将内存开销与存储开销维持在采集设备的可承载范围内。在金融与保险、医疗与生命科学等涉及敏感信息的产业，为满足长期留存与审计需求，需要在确保采集延迟可接受的前提下，进一步降低存储开销并合理平衡带宽开销。相对而言，教育等场景的整体负载较为温和，更关注在较低CPU利用率与内存开销下维持可接受的TPS，并通过合理控制带宽与存储开销来降低运维成本。

本文在表7中总结并归纳了源头监测的关键问题与需求、核心要点描述、代表性文献与典型评价指标。目前，源头监测的相关技术研究已从早期依赖防火墙采集IP、端口与协议等有限特征，发展到结合DLP内容识别、基于NetFlow、FlowRadar和sFlow的流与包级采样，以及Snort全包捕获和跨主机信息流融合等精细化方案，同时也探索了可编程数据平面嵌入采集逻辑、压缩编码降低存储与传输开销，以及利用深度强化学习自适应控制采样率等低开销智能采集机制。总体来看，上述工作在提升流量可视性和采集效率方面取得了一定进展。然而，在数据出境场景中，企业间通信模式多样、各产业风险指证效用存在显著差异且出境流量规模庞大，现有研究尚未系统考虑面向不同通信模式和产业特征的差异化采集策略，在一定程度上制约了在数据出境场景下的采集覆盖率与开销。因此，未来研究应聚焦于在多样化通信模式和产业风险差异化场景下，研究兼顾采集效率、覆盖率、开销与可部署性的数据出境源头监测技术体系。

2.3.2 流转监测

流转监测是指运用技术手段或管理措施针对数据从境内向境外传输、存储或处理的全过程流转链条进行监测，能够支撑包含多主体角色、复杂流转链条场景下的异常识别与溯源追责下游任务。传统研究并非直接对数据流转路径进行采集，而是利用网络连通性测试工具ping和逐跳路径探测诊断工具traceroute测试流转路径的连通性并进行故障诊断。随着IP流信息导出协议（IPFIX）、NetFlow、sFlow等网络流量监控技术的出现，为数据出境流转监测提供了基础支撑。在实际的流转监测场景下，由于通信网络环境复杂，背景流量类型多样且规模庞大，如何从中精准识别出需要监测的业务流量成为流转监测的首要任务。文献[13, 87-88]通过在数据包头中构造具有可识别性的专用字段，实现了对目标业务流量的显式标记，有效支撑业务流量识别。文献[89]通过利用相邻包间隔、包数目等模式特征构造逻辑上可检测的隐蔽水印，实现对目标业务流量的隐蔽识别。为提升网络噪声下的鲁棒性，文献[90]提出一种时间抖动不敏感、协议无关的流量识别方法，以IP分组序列为主要载体，并借助IP-ID字段进行辅助编码与检测，在复杂网络干扰条件下表现出更好的适用性。文献[91]则通过在指定时间窗口内对相邻报文进行局部可控的乱序与间隔调制，构造融合报文序列与时序特征的混合水印载体，显著增强了在时延抖动和丢包环境下对目标网络流的识别与溯源能力。针对上述关键技术，可引入F1分数、召回率等指标用于评估不同方法的业务流量识别成功率。引入抗检测性（Undetectability）指标表示标记或水印机制是否难以被第三方察觉或主动过滤，可以量化体现为攻击者采用技术工具对于标记或水印的识别成功率。引入识别延迟指标用于量化评估识别效率。在数据出境场景中，业务流量识别的可靠性直接决定后续数据流转路径采集与分析的可行性。通常情况下，以较高的召回率作为首要目标，按照数据出境产业实践与指导意见，对于一般数据管理和重要数据集中管控2种出境场景及其交互，2种场景的召回率值要求不低于80%与90%，以确保目标业务流在复杂背景流量中不被漏检。在存在对抗威胁或可控性不足的环境中，可靠的抗检测性优先级可上调至与召回率并列，以避免显式标记或水印机制被第三方识别、过滤或规避。而识别延迟作为运行层面的时效约束，应维持在可接受范围内以支撑后续监测与处置流程。不同产业场景侧重点则存在差异。在制造与工业控制、能源与交通等时效敏感产业中，低识别

延迟与高召回率并列为首要标准。在金融与保险、医疗与生命科学等高敏感且对抗风险较高的产业中，可靠的抗检测性与高召回率共同构成最核心要求，以确保识别机制的隐蔽性与稳定性。在信息通信与互联网的大规模高噪声场景中，需优先保障高召回率与F1分数，并将识别延迟控制在可用阈值内，以保证识别能力与业务连续性的平衡。

在实现对目标业务流量进行可靠识别的前提下，如何刻画其数据出境传输过程中所经过的流转链条，构建端到端的可观测链路，是流转监测闭环中的另一关键环节。文献[92]提出一种跨自治系统（AS, autonomous system）的转发路径重构方法，基于边界网关协议（BGP, border gateway protocol）路由表与更新、域名系统（DNS, domain name system）以及多测量点数据的IP-AS映射，实现AS粒度的端到端路径重构。文献[93]则把关注点从AS级拓扑推进到单域内部、数据包级别的流转路径构建，提出网络可见性系统NetSight，为每个数据包设计一个“明信片”，并通过在控制平面对所有明信片进行聚合与排序，实现数据包流转路径的追踪与重建。但会带来巨大的控制面开销，随着可编程数据平面的发展，文献[13]将流转路径等遥测信息在数据平面内直接写入业务数据包，有效降低了对控制面的依赖及其处理开销，但也显著增加了报文头部空间占用和链路传输开销。在此基础上，文献[12]通过为主机、进程、文件等实体创建分布式信息流控制（DIFC , decentralized information flow control）标签，在受控域内实现了主机侧与网络侧一体化的信息流转路径管控与追踪。针对上述关键技术，可引入F1分数、精确率等指标评估不同方法流转路径捕获能力，引入可部署性用于衡量不同方法的落地部署能力。引入流转路径捕获粒度（Granularity）用于表示路径监测数据的细化程度，细化程度越高，例如进程级路径，粒度越小，细化程度越低。在数据出境流转监测场景中，高评分的可部署性通常被视为整体落地的首要要求，决定了技术方案能否在真实网络环境中稳定运行。在满足可部署性的前提下，F1分数与捕获精确率共同衡量路径监测在链路完整性与捕获准确性方面的核心效能。流转路径捕获粒度则作为补充性指标，用于反映不同方法在路径细化程度上的表现。在数据出境产业差异方面，金融与保险、医疗与生命科学等高敏感产业由于审计要求严苛，通常需要更细化的路径捕获粒度，以支持复杂链路条件下的责任链条重建。在信息通信与互联网场景中，链路规模大且拓扑变化频繁，需在保持高可部署性前提下优先保证较高的F1分数与捕获精确率，并根据业务需求对粒度进行适度权衡。在教育产业场景中，则侧重在成本可控前提下维持可接受的捕获粒度与稳定的捕获精确率，以支撑基础流转监测能力。

围绕数据出境场景下的流转监测，现有相关技术研究能够在复杂网络背景下稳健识别目标业务流、还原业务流量流转路径，并打通主机侧与网络侧的信息流关联。通过引入多样化的流量标记与识别机制、结合网络路径观测与跨层追踪手段，逐步增强了数据流从业务层到网络层的可观测性和可追踪性，为数据出境流转链条级别的风险管控提供了技术支撑。然而，数据出境所依托的开放流转网络环境中，数据出境业务高度混杂、通信特征不断弱化且背景噪声显著增加，导致部分非侵入式标记流量识别技术在此环境下稳定性不足，而侵入式标记方案又易引发业务干扰并放大隐私安全风险，从而削弱数据出境流转监测的准确性与抗检测性。在未来研究中，应聚焦于如何在开放的流转网络环境下解决具备业务关联能力的隐私安全、业务无干扰的流转监测难题，加强受控域外数据流转的监测管控能力。

2.3.3　协同处置

协同处置是在满足法律法规要求的前提下，利用安全可靠的响应技术手段，对异常或违规行为进行协同化联动处置。处置技术的早期研究主要受重大安全事件驱动，多倾向于设计完备的入侵防御系统，在检测到攻击时对恶意流量实施拦截。随着各国数据与网络安全法律法规的完善，处置技术逐渐受到重视。在数据出境场景中，高效开展各参与方之间的处置活动，离不开处置指令的有效设计。文献[94]通过进程内安全监视器Endokernel构建监控与隔离机制，针对越权或违规访问定义了一套面向进程的系统级处置指令，如拒绝系统调用、终止子域执行等。文献[95]设计了一种快速、可扩展且精确的速率限制机制，提供了面向流量级的底层处置动作，如对特定流实施限速、调度等。文献[96]在交换机平面扩展了包级处置能力，设计了如丢弃、转发、重标记等包级动作原语。在此基础上，对上

述处置指令进行统一抽象与编排，设计高效的协同处置策略，是协同处置的关键任务。早期研究依赖于离线编排思想，文献[97]通过预计算攻击图和响应策略，将复杂的安全评估前移到离线阶段，使在线阶段可以在多设备组件之间快速下发联合防御动作，实现可解释的实时协同处置。随着网络与人工智能技术的发展，部分研究提出在线的协同处置策略，文献[17]提出了一种基于可信执行环境的自动化处置策略编排框架，通过设计在线执行协议针对特定安全事件安全地编排并执行跨系统的联动响应行为。文献[19]提出一种基于智能体化大语言模型（Agentic-LLM）的安全编排、自动化与响应架构，通过多智能代理协作建模不同角色的职责与行动选项，自动生成多主体协同的安全处置策略。针对上述关键技术，应引入召回率、精确率等指标用于评估不同方法的处置成功率，引入处置延迟指标用于评估不同方法的处置响应效率。引入可部署性衡量落地部署能力。引入带宽开销指标衡量资源开销。在数据出境场景中，高评分的可部署性通常被视为在多系统、多主体的分布式环境中实现落地的首要条件，是构建大规模协同处置体系的基础。在满足可部署性要求后，通常以较高的召回率与处置准确率协同衡量处置动作的成功率，前者体现对违规行为的处置覆盖能力，后者反映处置动作的准确性。处置延迟与带宽开销通常作为效率与资源消耗的次要指标，在满足前述要求的前提下予以优化，但在时效敏感行业场景（如工业控制、能源等）中，处置延迟的优先级可相应上调以确保处置时效。

除处置策略本身的设计与编排之外，对其执行效果进行系统性的评估与分析，也是协同处置体系中不可或缺的关键环节。文献[98]在5G/6G物联网（IoT, Internet of things）场景下提出智能感知平面（ISP, intelligent sensing plane）与分布式安全平面（DSP, distributed security plane）双层自治闭环的分布式拒绝服务（DDoS, distributed denial of service）攻击协同处置架构，通过多节点流量控制代理实现分布式联动阻断，并在通用开放研究仿真器（CORE, common open research emulator）中就执行延迟、攻击阻断比例和组件负载进行了量化评估。文献[99]提出智能车辆的自主入侵响应系统REACT，通过在车内多个控制单元上执行多种安全响应动作并对响应延迟和资源开销进行实验评估，从执行效果上验证了多实体协同响应的可行性。文献[100]提出云环境中资源感知的DDoS攻击缓解框架（RAM, resource-aware DDoS attack mitigation framework），并在多种攻击强度和资源配置场景下，从请求分类准确率、服务延迟和服务可用性等维度系统评估其协同处置效果。针对上述关键技术，应引入评估维度覆盖率用于评价不同方法所设计的评估维度数量占预期目标的比例。在数据出境产业实践中，评估维度覆盖率通常被视为协同处置效能评估的基础性要求。达到较高水平的覆盖率意味着评估体系能够充分反映处置策略在时效性、准确性、资源消耗及业务影响等多方面的实际表现。所以各产业的监管应以“尽可能接近完整的评估维度覆盖率”为目标，通过覆盖预期关键维度来提升协同处置体系的整体效能与可靠性。

本文在表7中总结归纳了协同处置的关键问题

表7 监测处置相关关键技术分析

监测处置维度	关键问题与需求	核心要点	代表性文献	典型评价指标
源头监测	源头风险指证采集	多源风险指证精细化采集	文献[83-85]	Coverage、Latency、TPS、Deployability
	低开销采集	自适应低开销动态采集	文献[13,83,86]	CPU Utilization、Latency、TPS、Storage overhead、Bandwidth overhead、Memory overhead
流转监测	出境业务流量识别	开放网络环境下出境流量精准识别	文献[13,89,91]	F1 score、Recall、Undetectability、Latency
	数据出境流转监测	开放网络环境下全域流转路径采集	文献[13,92-93]	F1 score、Precision、Deployability、Granularity
协同处置	协同联动处置	动态编排多点处置路径与策略组合	文献[17,19,97]	Recall、Precision、Latency、Deployability、Bandwidth overhead
	处置策略优化调整	分析与评估处置动作执行的有效性	文献[98-100]	Coverage

与需求、核心要点、代表性文献与典型评价指标。目前，现有协同处置相关研究已能够在多主体、多层次的安全体系中实现对异常行为的联动响应，打通进程级、流级到数据平面等不同粒度的处置能力，支撑从主机到网络的一体化闭环防御。通过不断丰富可组合的处置原语，引入从离线预编排到在线自动编排的多种策略框架，并在物联网、车联网和云环境等典型场景下对时延、阻断效果与服务稳定性开展量化评估，逐步形成了具有工程可行性的协同处置技术体系。然而，数据出境处置对象多源，风险扩散链条难以预测，既有策略多围绕网络攻击事件构建，在面向动态出境风险开展协同处置时，尚难有效适配数据敏感等级、产业背景与研究用途等复杂出境场景，导致现有研究的处置成功率仍存在较大差距。此外，尽管现有研究在处置有效性与效率等维度开展了一定程度的量化评估与分析，但评估维度覆盖度方面存在局限性，难以为处置策略的持续优化提供系统支撑。未来研究应聚焦于解决如何利用有限数目的处置设备实现高处置成功率与响应效率的协同处置难题。

3　关键技术体系与应用

通过前述研究工作综述与分析，可见现有相关关键技术应用于数据出境安全风险监测预警场景中解决问题仍存在局限性。为此，本文围绕研究框架的技术能力需求与现有相关工作面临的挑战，总结并归纳了适配场景与需求的关键技术要点，概述了其实现路径，并凝练了一套数据出境安全风险监测预警关键技术体系。在此基础上，本文介绍了关键技术体系的应用情况，并对未来研究进行了展望。

3.1　关键技术体系

本文以数据所有权人、出境企业、管理部门等相关方的需求为牵引，以各产业业务场景为主要驱动力，以国家数据基础设施为应用布局，在上文提出的研究框架和相关工作系统分析的基础上，凝练了一套包含 3 个维度、9 个方面的数据出境安全风险监测预警关键技术体系，支撑数据出境事前、事中、事后全过程监管、隐私保护的风险识别、全域覆盖的无干扰监测处置，如图 3 所示。为后续数据出境风险管控技术的应用、深入研究以及优化提供支撑。下面将对关键技术体系的各维度分别作出详细介绍。

3.1.1　过程监管

事前安全评估是数据出境合规管理的关键基础，为解决数据出境精细化描述能力不足问题，应构建元信息级数据出境业务描述机制，突破数据出境元信息表征技术，细化表示数据类型、敏感等级、摘要等数据元信息，突破出境业务信息表征技术，细化描述业务类型、路径、主体信息等业务元信息，支撑精细化源头管控。为提升事前安全评估效率，构建业务抽象过程模型自动化分析数据出境申报文档，突破风险要素抽取技术，结合元信息级数据出境业务描述机制实现跨产业领域风险要素自动化抽取。为满足差异化产业场景出境业务的普适性风险评估需求，应构建通用化智能风险评估模型，突破共性风险要素挖掘技术，提取不同业务的动静态共性要素，突破指标体系自适应构建技术，结合数据出境共性与差异化风险要素自动化关联匹配评估指标，支撑普适性风险评估。

事中动态评估是保障数据出境安全有序流动的重要阶段，为满足出境风险要素动态变化的识别需求，应构建风险要素变化辨识机制，突破风险要素感知技术，识别实际出境过程中由异常行为、安全事件及规制变化等引发的动态风险要素。围绕风险要素变化驱动的评估体系调整需求，应构建评估指标体系响应机制，突破风险评估指标遴选技术，识别风险要素变化对评估指标的影响路径与联动模式，筛选关键评估指标。突破指标权值自适应校准技术，设定权重调整安全边界，针对风险状态完成评估指标权重的自适应校正。为解决风险要素演变下的风险量化评估问题，构建数据出境风险动态评估模型，突破多风险要素融合量化分析技术，基于风险要素变化下的多维评估指标融合计算风险，实现出境风险的实时识别与动态评估。

事后溯源追责是提升数据出境治理效能的必要措施，为解决碎片化风险证据重建问题，应构建多源违规线索增强机制，突破碎片数据聚合重建技术，通过关联流量、系统日志、数据库等数据源，构建统一的时空语义线索共现视图。突破违规线索感知补全技术，推断缺失流转链路与行为片段，重建具备时空一致性与行为连续性的证据链条，支撑违规溯源追责。为解决复杂业务下的精准违规溯源和责任判定难题，应构建违规责任认定推演模型，突破流转序列挖掘技术，推演具有时间依赖性与操

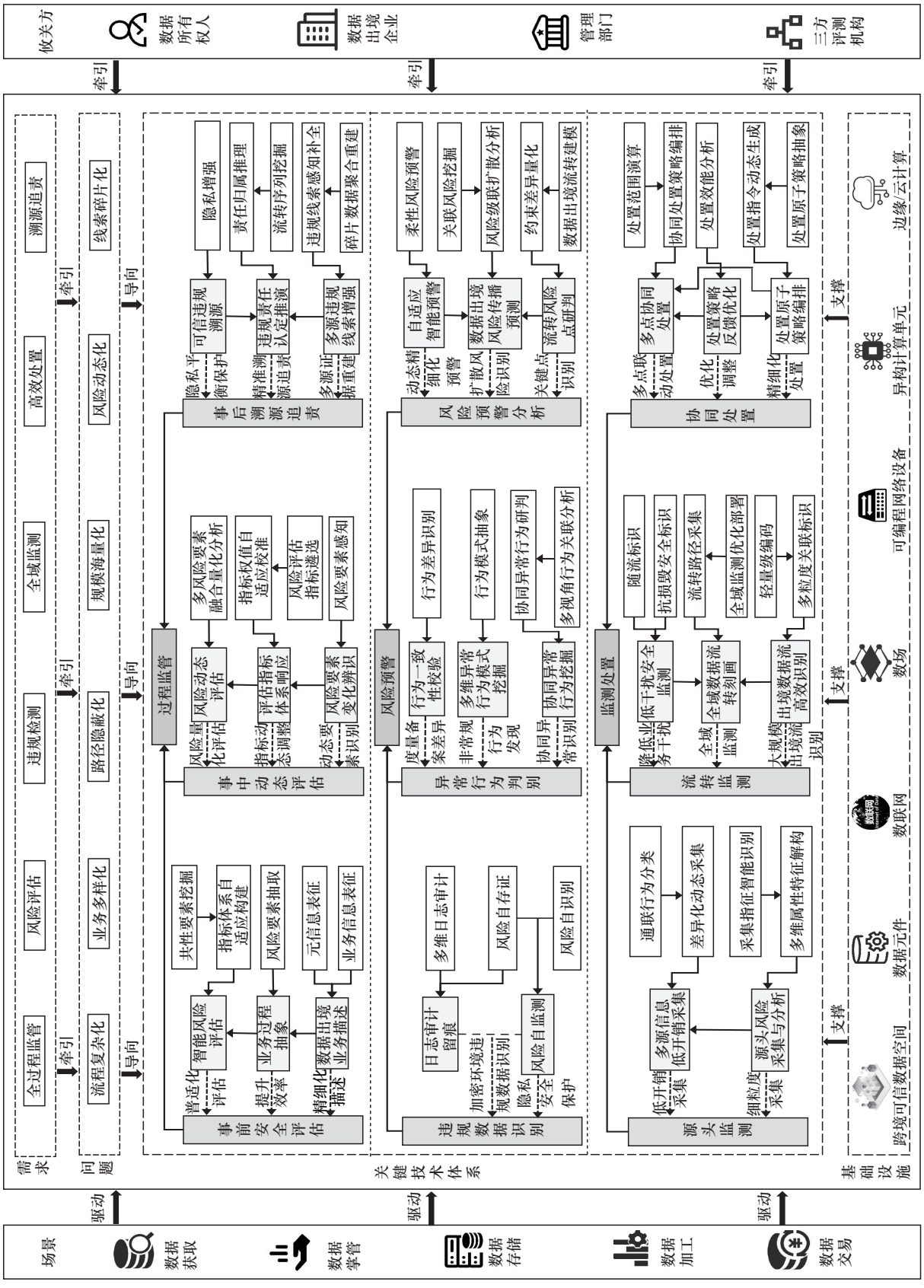

图 3 数据出境安全风险监测预警关键技术体系

作逻辑性的违规出境流转轨迹，精准追踪违规源头及责任归属路径。突破责任归属推理技术，针对违规事件结合流转轨迹对违规行为链条各实体进行责任归属推演，实现责任主体分级判定。为满足违规溯源追责过程的隐私保护需求，应构建可信违规溯源机制，突破隐私增强技术，设计可控身份保护策略以平衡溯源追责与隐私保护，以数据最小可用性为基础，保障在不暴露原始敏感信息的情况下实现精准违规溯源。

3.1.2 风险预警

违规数据识别是数据出境安全管控的重要防线，为解决加密环境下违规数据识别难题，应构建日志记录审计机制，突破多维日志审计技术，结构化记录出境业务关键内容摘要，记录通信、访问控制和安全告警等日志信息，提取数据、业务和操作行为特征，通过内容合规性校验与异常操作识别实现违规数据推演。为满足企业违规风险自我发现需求，构建出境风险自监测机制，突破风险自识别技术，数据处理者在源头侧针对数据出境活动对异常业务、违规内容特征持续自检，在保障隐私安全的同时，及时发现风险并处置违规数据。突破风险自存证技术，对出境业务活动运行风险情况进行标准化存证，重要风险事项及相关证据信息按相关要求实时推送管理部门，实现原始信息不暴露下的违规数据有效识别。

异常行为判别是实现数据出境风险主动监管的核心支撑，为解决静态备案信息与动态业务运行间的差异感知问题，应构建行为一致性校验机制，突破行为差异识别技术，针对实际出境业务运行情况，识别与备案信息不一致的时间、协议、范围与体量等字段，度量实际行为与备案信息的差异程度。围绕非常规异常行为的识别需求，应构建多维行为异常模式挖掘模型，突破行为模式抽象技术，融合多维行为特征构建数据流转行为基线，识别基线外非常用端口、非常规时间流动、非备案及高危IP通信等异常行为模式。为解决复杂流转模式下的异常协同行为识别难题，应构建协同异常行为挖掘模型，突破多视角行为关联分析技术，抽象主体间业务交互流程，解析其时间演化、空间轨迹与逻辑操作行为间的耦合关系，识别具有高度关联性的异常群体。突破协同异常行为研判技术，挖掘关联主体协同行为，量化关联强度与协同异常程度，支撑识别多路径分拆聚合出境、二次转移出境和多主体合谋出境等隐蔽协同异常行为。

风险预警分析是提升数据出境动态防控能力的重要手段，为满足潜在关键风险传播点识别需求，构建流转风险点研判机制，突破数据出境流转建模技术，针对业务流程进行处理环节拆分，基于业务角色链条与数据交互关系构建数据跨主体流转风险传播图。突破约束差异量化技术，通过数据流转沿途主体出境约束量化标注，分析主体间约束强度偏差，支撑关键风险传播点识别。为解决机构主体出境约束差异化条件下风险扩散传播趋势预测难题，构建数据出境风险传播预测模型，突破关联风险挖掘和风险级联扩散分析技术，分析出境主体间依赖关系与交互特征，抽象风险传播与扩散机理，实现潜在扩散节点、路径及影响域的精准预测。为提升风险预警响应水平，构建自适应智能预警机制，突破柔性风险预警技术，根据风险事件严重与紧急程度及扩散范围等特征，动态量化风险等级并匹配预警策略，实现精细化风险分级预警。

3.1.3 监测处置

源头监测是实现数据出境治理前移的关键环节，为解决面向多源风险指征的细粒度（会话级、要素级）采集难题，应构建数据出境源头风险采集与分析机制，突破采集指征智能识别和多维属性特征解构技术，通过通信会话映射与协议识别梳理多源风险指征，分析风险指征信息本质特征，对多源风险指征进行要素级解构与标准化抽象，建立多维度、分层级、精细化的风险指征要素分类体系。为满足复杂通信行为模式下的低开销采集需求，应构建多源风险信息低开销采集机制，突破通信模式分类技术，基于通信基础元信息，依据通信频率、连接稳定性、端口活跃度等特征对通信行为进行模式划分。突破差异化动态采集技术，针对通信模式特征量化其风险等级，并结合当前风险状态智能调整采集指标配置与频度，支撑动态环境下低开销细粒度多源风险指标采集。

流转监测是数据出境精细化监管的重要举措，为解决大规模流量下出境数据流快速识别问题，应构建出境数据流高效识别机制，突破多粒度关联标识和轻量级编码技术，针对业务与流量特征等多维属性，轻量化压缩生成能识别出境业务流量的唯一标识，并实现业务与标识双向绑定，标识可携带业

务与行为等多粒度信息，配合硬件加速实现出境数据流高效识别。为满足覆盖全域的数据出境流转监测需求，构建全域数据流转刻画机制，突破全域监测优化部署技术，针对分布式流转环境，衡量覆盖能力与监测成本，生成资源最优的全域监测部署方案。突破流转路径采集技术，借助标识识别能力，针对流转行为生成唯一的路径标识，通过重构路径标识片段实现全域流转路径刻画。为解决监测过程中的隐私安全与业务干扰问题，构建低干扰安全监测机制，突破随流标识技术，构造独立于原始业务流量的随流标识，高速注入数据出境业务流中，保障跟随出境数据传输且不对业务产生干扰与入侵。突破抗损毁安全标识技术，通过在数据出境流量中多层柔性注入随流标识，降低标识数据包丢失影响，通过对标识信息随机化密钥加密并进行多层冗余编码，提升标识抗篡改能力，支撑数据出境隐私安全保护。

协同处置是实现数据出境风险最大化可控的有效保障，为解决多源对象下的精细化处置问题，应构建处置原子策略编排机制，突破处置原子策略抽象技术，将复杂处置策略拆解为原子级、可组合的处置动作。突破处置指令动态生成技术，针对主体类型、通联模式、应用模式等要素，建立处置对象与原子处置动作间的映射关系，优化生成算法减少计算成本，支撑处置指令快速响应。为满足违规行为影响范围动态变化下的联动处置响应需求，构建多点协同处置机制，突破处置范围演算技术，针对违规出境行为，分析风险传播扩散趋势和演化路径，演算违规风险影响域边界，确定处置范围。突破协同处置策略编排技术，分析处置范围与资源约束，动态编排跨主体、跨链路的处置路径与策略组合，构建执行优先级与处理状态同步策略，规避并发冲突，支撑多点联动处置。为提升处置策略优化调整能力，构建处置策略反馈优化机制，突破策略执行效能分析技术，识别处置指令执行偏差并评估干预有效性，支撑处置指令与策略动态调优。

3.2 应用与未来展望

上述关键技术体系的落地实施需要充分考虑业务运行逻辑与空间分布逻辑，为此，本文面向数据出境前、中、后等过程，结合数据出境各阶段风险管控需求与实际业务场景，根据本文所提出的各关键技术的逻辑衔接与功能协同，梳理了如图4所示的关键技术体系的运行流程，有效支撑关键技术融合实际应用。进一步，如图5所示，本文以全域协同监管为主线，通过抵近式安全无干扰低开销监测、多维风险动态识别、安全事件柔性预警与协同精细化处置、违规精准溯源与责任判定等需求与策略驱动，根据关键技术能力支撑，考虑实际管理部门、数据所有者、企业、服务平台等数据出境参与主体，结合互联网、运营商网络、卫星网络等基础设施支撑，形成关键技术与出境场景的融合应用图。

综合现有研究成果来看，当前工作尚难以全面满足实际需求，仍面临诸多挑战。未来可重点突破如下几个方面的研究。在技术研究方面，为支撑大规模流量下出境风险快速隐私化识别，提升数据采集与异常识别能力，研究软硬件融合一体化技术，引入可编程设备作为数据处理加速单元，将标识识别、特征提取、异常识别等关键逻辑通过硬件逻辑固化、流水线构建，形成具备高吞吐、低延迟特性的硬件加速引擎，保障在不侵入出境业务的前提下实现线速度并发处理。软件层则负责策略配置与动态调度，通过软硬融合接口驱动底层硬件资源，实现灵活性与可维护性统一。软硬件协同一体化技术能够提供数据出境安全事件的动态监测、便捷部署能力，支撑隐私无干扰的监测需求。此外，数据出境线索多元，监管对象语义模糊、规则动态，为支撑实现监管过程中非结构化数据出境信息高效解析与合规语义识别匹配，提升风险评估效率与异常识别能力，研究具备深层语义理解与知识泛化能力的数据出境大语言模型技术，通过对非结构化内容（如申报文档、API调用语句、用户操作日志、合规条款等）的上下文语义建模，实现对数据出境目的、内容与合规风险的高精度识别。数据出境大语言模型应具备适配多种行业语义、接口类型与文档风格的多样化能力，支持插件化部署。数据出境大语言模型以知识驱动的方式不断强化自身能力，通过引入监管知识图谱、数据出境规则库与历史评估样本等知识源，增强对合规边界与政策适用场景的理解能力。在此基础上，结合检索增强生成（RAG, retrieval-augmented generation）、指令微调与多轮对话机制，实现对复杂合规评估任务的迭代优化。数据出境大语言模型技术能够提供插件化、多维度的安全评估与异常分析能力，支撑跨境可信

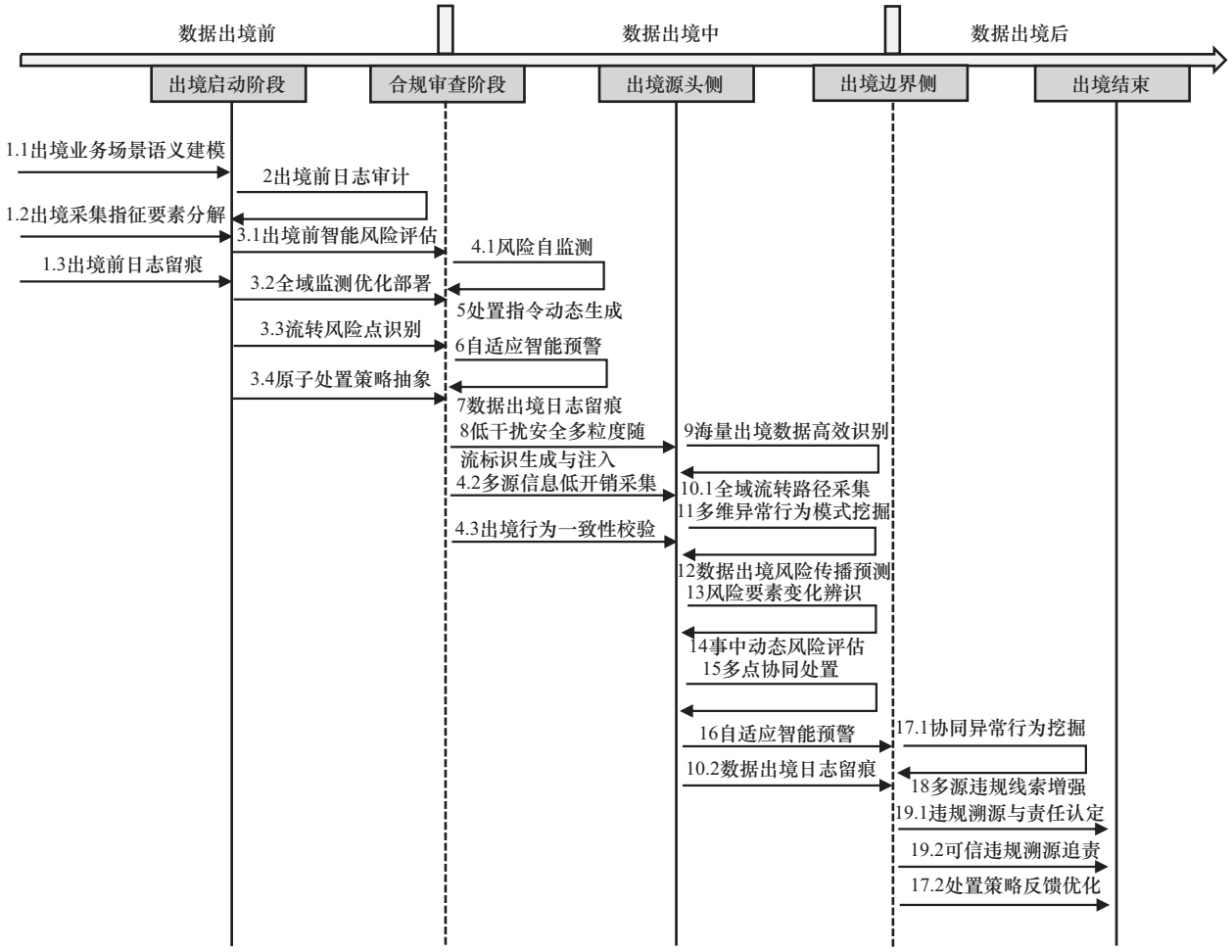

图4　关键技术体系的运行流程

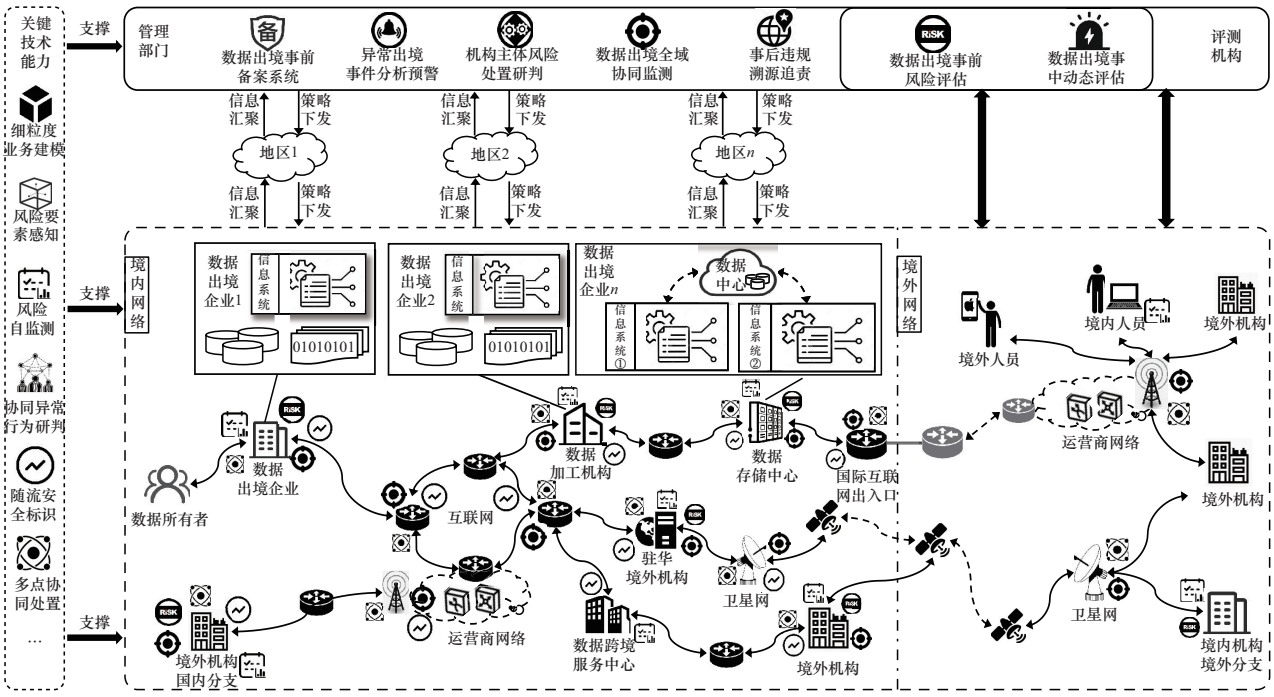

图5　关键技术与出境场景融合应用

数据空间构建，推动风险识别从静态规则向语义驱动的智能检测转型。

在应用场景研究方面，现有数据出境大量依赖第三方云场景、托管场景等开展业务。由于云上环境复杂，业务数据交叉融合，为支撑实现云环境下出境数据流转路径追踪与行为约束，提升云环境下的监测覆盖与违规处置能力，研究云上数据出境监管技术，通过在云原生基础设施中部署可观测探针与轻量化策略执行组件，通过埋点、云日志、网络协议等方式，采集数据在云节点和传输通道上的流转信息。基于数据指纹和多粒度数据安全标识等数据识别技术，配合数据流转记录构建完整的流转链路，确认出境数据的来源、目标与路径，并借助可视化工具将数据的流转链路、分布拓扑、血缘关系等信息进行展示，实现云上数据出境流转态势感知。依托细粒度（对象级、操作级）访问控制、标识感知与运行时策略引擎，将监管策略动态下沉至虚拟机、容器、无服务器函数等多样计算单元，实现与运行环境解耦的数据使用权限管控与实时阻断。云上数据出境监管技术能够提供云环境下数据出境行为的全面感知与合规管控能力，夯实数据出境全域可控的云端基础。此外，针对数据出境场景，可依托重要城市与自贸区构建可管可控的数据出境特区，建设出境数据安全托管与安全监测平台，部署数据出境业务托管设施和数据出境监管技术设施，可高速连接境外互联网，对数据出境业务交互网络行为实行全方位的监控，为在平台上托管出境业务的数据出境企业提供一站式标准化的监管合规服务，促进数据出境有序流动，维护重要数据和个人信息安全。

综上所述，数据出境安全风险监测预警关键技术未来仍需要在以下几个方面寻求重点突破。

1)如何在海量流转的广域网下，建立针对出境流量的高效追踪策略，实现流通痕迹存证。

2)如何在复杂多变的业务场景中，挖掘特征与出境行为合规性的隐式关联性，降低部署难度。

3)如何在数据出境监管过程中，设计多源信息的长距离融合方法，减少技术带来的误报与漏报。

4 结束语

针对数据出境过程中的安全风险与挑战，遵循业务无干扰、隐私安全保护等原则，本文在对数据出境场景进行系统化描述与分类的基础上，从过程监管、风险预警与监测处置等维度提出一种数据出境安全风险监测预警研究框架，重点分析了关键问题与技术能力需求，系统化梳理了数据出境直接相关的现有研究工作，并讨论了其在数据出境场景下的适用性。在此基础上，总结并归纳了适配场景与需求的关键技术要点，凝练了一套覆盖风险评估、全域监测、异常识别、协同处置与溯源追责等方面的关键技术体系，旨在强化数据出境事前事中事后全过程的监管能力，支撑全域风险态势有效掌控。最后围绕数据出境业务场景，梳理了关键技术体系的运行逻辑流程，呈现了关键技术与出境场景的融合应用，并介绍了未来研究展望与需重点突破的核心技术难题。

参考文献：

[1] RUTHERFORD M. The CLOUD act[J]. Berkeley Technology Law Journal, 2019, 34(4): 1177-1204.

[2] VOIGT P, VON DEM BUSSCHE A. The EU general data protection regulation (GDPR)[M]. Berlin: Springer, 2017.

[3] 谢绒娜, 郭云川, 李凤华, 等. 面向数据跨域流转的延伸访问控制机制[J]. 通信学报, 2019, 40(7): 67-76.
XIE R N, GUO Y C, LI F H, et al. Extended access control mechanism for cross-domain data exchange[J]. Journal on Communications, 2019, 40(7): 67-76.

[4] FRANCIS R, GUPTA R, OKE M. Amazon redshift: the definitive guide: jump-start analytics using cloud data warehousing[M]. California: O'Reilly Media, Inc., 2023.

[5] DIAMANTOPOULOS S, KARAMITROS D, ROMANO L, et al. Secure cross-border exchange of health related data: the KONFIDO approach[C]//Proceedings of the 2019 15th European Dependable Computing Conference (EDCC). Piscataway: IEEE Press, 2019: 73-74.

[6] 赵阳光, 黄海波. 美国“爱因斯坦计划”研究[J]. 信息安全研究, 2020, 6(11): 1013-1016.
ZHAO Y G, HUANG H B. American “Einstein plan” research[J]. Journal of Information Security Research, 2020, 6(11): 1013-1016.

[7] ESPEEL T, COLSON E, CRUQUENAIRE A. International data transfers under GDPR: applicable requirements and practical implementation[J]. Annals of Operations Research, 2022 (141): 19-41.

[8] 李金, 张黎明, 李建平, 等. 跨境数据传出机构的风险分类管控和影响因素分析[J]. 系统科学与数学, 2022, 42(9): 2347-2366.
LI J, ZHANG L M, LI J P, et al. Classified control and influencing factors for risks management in institutions with cross-border data flow[J]. Journal of Systems Science and Mathematical Sciences, 2022, 42(9): 2347-2366.

[9] 李金, 申苏浩, 孙晓蕾, 等. 重要数据跨境流动背景下风险路径的识别与分级[J]. 中国管理科学, 2021, 29(3): 90-99.
LI J, SHEN S H, SUN X L, et al. Identification and classification for

risk paths in the context of cross-border important data flow[J]. Chinese Journal of Management Science, 2021, 29(3): 90-99.

[10] 李金, 徐姗, 卓子寒, 等. 数据跨境流转的风险测度与分析: 基于数据出境统计信息的实证研究[J]. 管理世界, 2023, 39(7): 180-201.
LI J, XU S, ZHUO Z H, et al. Risk measurement and analysis for cross-border data flow: an empirical study based on statistics of outbound data[J]. Journal of Management World, 2023, 39(7): 180-201.

[11] GUAMÁN D S, RODRIGUEZ D, DEL ALAMO J M, et al. Automated GDPR compliance assessment for cross-border personal data transfers in Android applications[J]. Computers & Security, 2023, 130: 103262.

[12] BAJABER O, JI B, GAO P. P4Control: line-rate cross-host attack prevention via in-network information flow control enabled by programmable switches and eBPF[C]//Proceedings of the 2024 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE Press, 2024: 4610-4628.

[13] KIM C, SIVARAMAN A, KATTA N P K, et al. In-band network telemetry via programmable dataplanes[R]. 2015.

[14] LANDAU-FEIBISH S, LIU Z X, REXFORD J. Compact data structures for network telemetry[J]. ACM Computing Surveys, 2025, 57(8): 1-31.

[15] SU Y, ZHOU M C, QI L, et al. A reachability-decidable Petri net modeling method for discrete event systems[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2024, 55(1): 453-464.

[16] KANG G S, CHENG H Y, LIU J X, et al. Business process modeling for industrial Internet application via BPMN extension[J]. IEEE Transactions on Automation Science and Engineering, 2024, 22: 813-829.

[17] JEGAN D S, SWIFT M, FERNANDES E. Architecting trigger-action platforms for security, performance and functionality[C]//Proceedings 2024 Network and Distributed System Security (NDSS 2024) Symposium. Piscataway: IEEE Press, 2024: 1-8.

[18] BEMTHUIS R, WANG W, IACOB M E, et al. Business rule extraction using decision tree machine learning techniques: a case study into smart returnable transport items[J]. Procedia Computer Science, 2023, 220(C): 446-455.

[19] ISMAIL, KURNIA R, BRATA Z A, et al. Toward robust security orchestration and automated response in security operations centers with a hyper-automation approach using agentic artificial intelligence[J]. Information, 2025, 16(5): 365.

[20] TAN Q, FU M, WANG Z X, et al. A real-time early warning classification method for natural gas leakage based on random forest[J]. Reliability Engineering & System Safety, 2024, 251: 110372.

[21] SUI L K, JIANG Y G. Argo data anomaly detection based on transformer and Fourier transform[J]. Journal of Sea Research, 2024, 198: 102483.

[22] TAMBUWAL A I, NEAGU D. Deep quantile regression for unsupervised anomaly detection in time-series[J]. SN Computer Science, 2021, 2(6): 475.

[23] CASTELLINI A, MASILLO F, AZZALINI D, et al. Adversarial data augmentation for HMM-based anomaly detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(12): 14131-14143.

[24] NEWBOLD P. ARIMA model building and the time series analysis approach to forecasting[J]. Journal of Forecasting, 1983, 2(1): 23-35.

[25] YU J, LU K D, JING M H, et al. Sliding window Seq2seq modeling for engagement estimation[C]//Proceedings of the 31st ACM International Conference on Multimedia. New York: ACM Press, 2023: 9496-9500.

[26] 孙海丽, 黄炎, 韩兰胜, 等. 基于无监督时空状态估计的信息物理系统细粒度异常诊断[J]. 通信学报, 2025, 46(7): 45-59.
SUN H L, HUANG Y, HAN L S, et al. Unsupervised spatio-temporal state estimation for fine-grained anomaly diagnosis of cyber-physical systems[J]. Journal on Communications, 2025, 46(7): 45-59.

[27] POOLSAPPASIT N, DEWRI R, RAY I. Dynamic security risk management using bayesian attack graphs[J]. IEEE Transactions on Dependable and Secure Computing, 2011, 9(1): 61-74.

[28] ZHOU X K, WU J Y, LIANG W, et al. Reconstructed graph neural network with knowledge distillation for lightweight anomaly detection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(9): 11817-11828.

[29] HAVRANEK T, IRSOVA Z. Do borders really slash trade? a meta-analysis[J]. IMF Economic Review, 2017, 65(2): 365-396.

[30] PLÖTZKY F, BRITZ K, BALKE W T. A conceptual model for attributions in event-centric knowledge graphs[J]. Data & Knowledge Engineering, 2025, 159: 102449.

[31] KAZEMI M H, ALVANCHI A. Application of NLP-based models in automated detection of risky contract statements written in complex script system[J]. Expert Systems with Applications, 2025, 259: 125296.

[32] KIM J, KWON B, LEE J, et al. Inherent risks identification in a contract document through automated rule generation[J]. Automation in Construction, 2025, 172: 106044.

[33] 赖清楠, 金建栋, 周昌令. 基于大语言模型的网络威胁情报知识图谱构建技术研究[J]. 通信学报, 2024, 45(S2): 33-43.
LAI Q N, JIN J D, ZHOU C L. Research on the construction technology of network threat intelligence knowledge map based on large language model[J]. Journal on Communications, 2024, 45(S2): 33-43.

[34] BANINEMEH E, JANSEN S, LABUNETS K. A security risk assessment method for distributed ledger technology (DLT) based applications: three industry case studies[J]. arXiv Preprint, arXiv: 2401.12358, 2024.

[35] WANG N, WU G F, YUE Q L, et al. Research on security assessment of cross border data flow[C]//Frontiers in Cyber Security. Berlin: Springer, 2022: 327-341.

[36] WANG N, WU G F, RONG J F, et al. Cross-border data security from the perspective of risk assessment[C]//Information Security Practice and Experience.Berlin: Springer, 2023: 91-104.

[37] CHAPPLE M, SHELLEY J. IAPP CIPM certified information privacy manager study guide[M]. New Jersey: John Wiley & Sons, 2023.

[38] FUNG C, ZENG E, BAUER L. Attributions for ML-based ICS anomaly detection: from theory to practice[C]//Proceedings 2024 Network and Distributed System Security Symposium, 2024: 23216.

[39] CZEKSTER R M, WEBBER T, FURSTENAU L B, et al. Dynamic risk assessment approach for analysing cyber security events in medical IoT networks[J]. Internet of Things, 2025, 29: 101437.

[40] PEI J X, VADLAMANNATI S, HUANG L-K, et al. Modeling and detecting company risks from news[C]//Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 6: Industry Track). Stroudsburg: ACL, 2024: 63-72.

[41] CZAJKOWSKI M, JURCZUK K, KRETOWSKI M. Steering the interpretability of decision trees using lasso regression - an evolutionary perspective[J]. Information Sciences, 2023, 638: 118944.

[42] XIANG P C, CHONGQING C U, et al. Integrated measurement of public safety risks in international construction projects in the Belt and Road initiative[J]. Engineering, Construction and Architectural Management, 2025, 32(7): 4673-4699.

[43] DEY A K, GUPTA G P, SAHU S P. A metaheuristic-based ensemble feature selection framework for cyber threat detection in IoT-enabled networks[J]. Decision Analytics Journal, 2023, 7: 100206.

[44] COVERT I, QIU W, LU M Y, et al. Learning to maximize mutual information for dynamic feature selection[J]. arXiv Preprint, arXiv: 2301.00557, 2023.

[45] KALYAN S, BANSAL P, KUMAR S. A novel hybrid AHP-entropy weighted dynamic network DEA framework for comprehensive efficiency assessment[J]. Expert Systems with Applications, 2026, 296: 129137.

[46] SHI J H, LIU Z J, FENG Y W, et al. Evolutionary model and risk analysis of ship collision accidents based on complex networks and DEMATEL[J]. Ocean Engineering, 2024, 305: 117965.

[47] WANG S, SUN H B, WANG Z L, et al. End-to-end attack scene reconstruction in a host with rules and anomaly-based detection models[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 7317-7332.

[48] 康海燕, 龙墨澜. 基于吸收马尔可夫链攻击图的网络攻击分析方法研究[J]. 通信学报, 2023, 44(2): 122-135.
KANG H Y, LONG M L. Research on network attack analysis method based on attack graph of absorbing Markov chain[J]. Journal on Communications, 2023, 44(2): 122-135.

[49] WANG K, TANG X Y, ZHAO S M. Robust multi-step wind speed forecasting based on a graph-based data reconstruction deep learning method[J]. Expert Systems with Applications, 2024, 238: 121886.

[50] TAN Q Y, ZHANG X, LIU N H, et al. Bring your own view: graph neural networks for link prediction with personalized subgraph selection[C]//Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2023: 625-633.

[51] LIU C Y, CHEN X H, LI J, et al. A novel data traceability model based on blockchain and digital watermarking in edge computing[J]. Journal of Physics: Conference Series, 2020, 1682(1): 012041.

[52] SHAO X Y, WANG H Z, CHEN X, et al. CUBE: causal intervention-based counterfactual explanation for prediction models[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(6): 2416-2429.

[53] YAN X, FANG H, HE Q. Diffusion model for graph inverse problems: towards effective source localization on complex networks[J]. Advances in Neural Information Processing Systems, 2023, 36: 22326-22350.

[54] CHEONG C, SONG Y J, CAO Y, et al. Multidimensional trust evidence fusion and path-backtracking mechanism for trust management in VANETs[J]. IEEE Internet of Things Journal, 2024, 11(10): 18619-18634.

[55] GOOLD M, CAMPBELL A. Making matrix structures work: creating clarity on unit roles and responsibility[J]. European Management Journal, 2003, 21(3): 351-363.

[56] ROSENBAUM S. Data governance and stewardship: designing data stewardship entities and advancing data access[J]. Health Services Research, 2010, 45(5p2): 1442-1455.

[57] 谢晴晴, 宋亮晴, 冯霞. 面向医疗数据分享的轻量级且安全的搜索方案[J]. 通信学报, 2024, 45(11): 206-222.
XIE Q Q, SONG L Q, FENG X. Lightweight and secure search scheme for medical data sharing[J]. Journal on Communications, 2024, 45(11): 206-222.

[58] 张晓旭, 陈宇辰, 哈冠雄, 等. 基于分布式存储的外包EHR隐私保护分类审计方案[J]. 通信学报, 2024, 45(9): 26-39.
ZHANG X X, CHEN Y C, HA G X, et al. Classification auditting scheme for privacy protection of outsourced EHR based on distributed storage[J]. Journal on Communications, 2024, 45(9): 26-39.

[59] SUO Z F, XIA C, JIANG D H, et al. Multitiered reversible data privacy protection scheme for IoT based on compression sensing and digital watermarking[J]. IEEE Internet of Things Journal, 2023, 11(7): 11524-11539.

[60] STALLA-BOURDILLON S, PAPADAKI E, CHOWN T. From porn to cybersecurity passing by copyright: How mass surveillance technologies are Gaining legitimacy ⋯ The case of deep packet inspection technologies[J]. Computer Law & Security Review, 2014, 30(6): 670-686.

[61] CEJAS O A, AZEEM M I, ABUALHAIJA S, et al. NLP-based automated compliance checking of data processing agreements against GDPR[J]. IEEE Transactions on Software Engineering, 2023, 49(9): 4282-4303.

[62] VAN BEEST N, GROEFSEMA H, CRYER A, et al. Cross-instance regulatory compliance checking of business process event logs[J]. IEEE Transactions on Software Engineering, 2023, 49(11): 4917-4931.

[63] GONG C C, HUANG Y L, CHENG X Q, et al. Detecting near-duplicates in large-scale short text databases[C]//Advances in Knowledge Discovery and Data Mining. Berlin: Springer, 2008: 877-883.

[64] WU S, MANBER U. A fast algorithm for multi-pattern searching[M]. Tucson: University of Arizona, 1994.

[65] GONZÁLEZ-GRANADILLO G, GONZÁLEZ-ZARZOSA S, DIAZ R. Security information and event management (SIEM): analysis, trends, and usage in critical infrastructures[J]. Sensors, 2021, 21(14): 4759.

[66] 田有亮, 黄钰清, 王帅. 基于级联编码的区块链分片存储方案[J]. 通信学报, 2024, 45(7): 159-170.
TIAN Y L, HUANG Y Q, WANG S. Blockchain sharding storage scheme based on concatenated coding[J]. Journal on Communications, 2024, 45(7): 159-170.

[67] CHASE M, MEIKLEJOHN S. Transparency overlays and applications[C]//

Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2016: 168-179.

[68] KAZEMI S, ABGHARI S, LAVESSON N, et al. Open data for anomaly detection in maritime surveillance[J]. Expert Systems with Applications, 2013, 40(14): 5719-5729.

[69] ZHANG K, SHASHA D. Simple fast algorithms for the editing distance between trees and related problems[J]. SIAM Journal on Computing, 1989, 18(6): 1245-1262.

[70] WANG Y, DEWITT D J, CAI J Y. X-Diff: an effective change detection algorithm for XML documents[C]//Proceedings of the 19th International Conference on Data Engineering. Piscataway: IEEE Press, 2003: 519-530.

[71] TANG J H, ZHAO X, KONG L M, et al. Fused Gromov-Wasserstein alignment for graph edit distance computation and beyond[J]. Proceedings of the VLDB Endowment, 2025, 18(10): 3641-3654.

[72] HUTWAGNER L, THOMPSON W, SEEMAN G M, et al. The bioterrorism preparedness and response Early Aberration Reporting System (EARS)[J]. Journal of Urban Health, 2003, 80(1): i89-i96.

[73] YUAN Y, DING J T, WANG H D, et al. Activity trajectory generation via modeling spatiotemporal dynamics[C]//Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2022: 4752-4762.

[74] PARANJAPE A, BENSON A R, LESKOVEC J. Motifs in temporal networks[C]//Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2017: 601-610.

[75] LA GATTA V, MOSCATO V, POSTIGLIONE M, et al. An epidemiological neural network exploiting dynamic graph structured data applied to the COVID-19 outbreak[J]. IEEE Transactions on Big Data, 2020, 7(1): 45-55.

[76] RODRÍGUEZ A, CUI J M, RAMAKRISHNAN N, et al. EINNs: epidemiologically-informed neural networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(12): 14453-14460.

[77] LANKO V, MAKAROV I. Graph-attention diffusion for enhanced multivariate time-series anomaly detection[J]. IEEE Open Journal of the Industrial Electronics Society, 2024, 5: 1353-1364.

[78] LIU C, YANG Y W, DING Y, et al. DAG: deep adaptive and generative K-free community detection on attributed graphs[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2024: 5454-5465.

[79] KARAOGLAN A D, BAYHAN G M. A regression control chart for autocorrelated processes[J]. International Journal of Industrial and Systems Engineering, 2014, 16(2): 238-256.

[80] GUO C, PI D C, CAO J J, et al. Early warning model for industrial Internet platform based on graph neural network and time convolution network[J]. Journal of Ambient Intelligence and Humanized Computing, 2023, 14(3): 2399-2412.

[81] KIA A N, MURPHY F, SHEEHAN B, et al. A cyber risk prediction model using common vulnerabilities and exposures[J]. Expert Systems with Applications, 2024, 237: 121599.

[82] LIU S, KUHN R. Data loss prevention[J]. IT professional, 2010, 12(2): 10-13.

[83] LI Y L, MIAO R, KIM C, et al. FlowRadar: a better NetFlow for data centers[C]//Proceedings of the Symposium on Networked Systems Design and Implementation. Berkeley: USENIX Association, 2016.

[84] GIOTIS K, ARGYROPOULOS C, ANDROULIDAKIS G, et al. Combining OpenFlow and sFlow for an effective and scalable anomaly detection and mitigation mechanism on SDN environments[J]. Computer Networks, 2014, 62: 122-136.

[85] ROESCH M. Snort: Lightweight intrusion detection for networks[C]//13th LISA Conference. Berkeley: USENIX Association, 1999: 229-238.

[86] HUANG W, ZHAN J, SUN Y, et al. Context-aware adaptive sampling for intelligent data acquisition systems using DQN[J]. arXiv Preprint, arXiv: 2504.09344, 2025.

[87] CUI J, HAN K Y, SHA L, et al. An efficient hexadecimal network flow watermark method for tracking attack traffic[J]. Scientific Reports, 2023, 13: 21111.

[88] ATTEBURY G, BABIK M, CARDER D, et al. Identifying and understanding scientific network flows[J]. EPJ Web of Conferences, 2024, 295: 01036.

[89] MIRNAJAFIZADEH S M M, SETHURAM A R, MOHAISEN D, et al. Enhancing network attack detection with distributed and {in-network} data collection system[C]//33rd USENIX Security Symposium (USENIX Security 24). Berkeley: USENIX Association, 2024: 5161-5178.

[90] FENG W X, LUO X Y, LI T Y, et al. IP-pealing: a robust network flow watermarking method based on IP packet sequence[J]. Chinese Journal of Electronics, 2024, 33(3): 694-707.

[91] FENG W X, LUO X Y, LI T Y, et al. HSTW: a robust network flow watermarking method based on hybrid packet sequence-timing[J]. Computers & Security, 2024, 139: 103701.

[92] MAO Z M, REXFORD J, WANG J, et al. Towards an accurate AS-level traceroute tool[C]//Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. New York: ACM Press, 2003: 365-378.

[93] HANDIGO N, HELLER B, JEYAKUMAR V, et al. I know what your packet did last hop: using packet histories to troubleshoot networks[C]//Proceedings of the Symposium on Networked Systems Design and Implementation. Berkeley: USENIX Association, 2014.

[94] YANG F, IM B, HUANG W, et al. Endokernel: a thread safe monitor for lightweight subprocess isolation[C]//33rd USENIX Security Symposium (USENIX Security 24). Berkeley: USENIX Association, 2024: 145-162.

[95] WANG Z L, WAN X C, LI L Y, et al. Fast, scalable, and accurate rate limiter for RDMA NICs[C]//Proceedings of the ACM SIGCOMM 2024 Conference. New York: ACM Press, 2024: 568-580.

[96] BIANCHI G, BONOLA M, CAPONE A, et al. OpenState: programming platform-independent stateful openflow applications inside the switch[J]. ACM SIGCOMM Computer Communication Review, 2014, 44(2): 44-51.

[97] EOM T, HONG J B, AN S, et al. A framework for real-time intrusion response in software defined networking using precomputed graphical security models[J]. Security and Communication Networks, 2020(1): 7235043.

[98] BENLLOCH-CABALLERO P, WANG Q, ALCARAZ CALERO J M. Distributed dual-layer autonomous closed loops for self-protection of 5G/6G IoT networks from distributed denial of service attacks[J]. Computer Networks, 2023, 222: 109526.

[99] HAMAD M, FINKENZELLER A, KÜHR M, et al. REACT: Autonomous intrusion response system for intelligent vehicles[J]. Computers & Security, 2024, 145: 104008.

[100] XING F, TONG F, YANG J, et al. RAM: a resource-aware DDoS attack mitigation framework in clouds[J]. IEEE Transactions on Cloud Computing, 2024, 12: 1387-1400.

[作者简介]

张凯（1998– ），男，山东济南人，北京邮电大学博士生，主要研究方向为数据安全、网络测量、图异常检测等。

时金桥（1978– ），男，黑龙江哈尔滨人，博士，北京邮电大学教授、博士生导师，主要研究方向为隐私保护、人工智能安全、匿名通信技术、数据安全等。

马乐乐（1999– ），女，河南驻马店人，北京邮电大学硕士生，主要研究方向为网络安全、数据安全等。

郭晓威（1996– ），男，广西贺州人，华中科技大学博士生，主要研究方向为数据安全标识、源代码作者溯源、软件供应链分析等。

刁毅刚（1977– ），男，四川成都人，北京邮电大学博士生，中央网信办数据与技术保障中心正高级工程师，主要研究方向为自然语言处理、网络安全、数据安全、个人信息保护等。

李凤华（1966– ），男，湖北浠水人，博士，中国科学院信息工程研究所研究员、博士生导师，主要研究方向为网络与系统安全、隐私计算、数据安全等。