内部威胁分析与防御综述

孙德刚[1,2], 刘美辰[1,2], 李梅梅[1,2,3], 王 旭[1,2], 石志鑫[1,2],
刘鹏程[1,2], 李 楠[1,2]

[1]中国科学院大学 网络空间安全学院 北京 中国 100049
[2]中国科学院 信息工程研究所 北京 中国 100093
[3]北京交通大学 计算机与信息技术学院 北京 中国 100044

摘要 内部威胁攻击是由可信的内部人员发起的, 相比较外部威胁更具有透明性、隐蔽性和高危性, 是当今最具有挑战的网络安全问题之一, 因此需要十分重视且关注该领域的研究成果和发展趋势。本文对内部威胁研究范畴内的成果进行了概述, 并使用扎根理论的方法进行严格的文献归纳和分析, 通过全景视图下的内部威胁系统性研究, 帮助组织减轻和消除内部威胁事件并根据自身实际情况快速制定防御方案。本文的研究对内部威胁领域有重要意义, 因为它(1) 概括了内部威胁的研究范畴, 包含定义与分类、数据集分析、事件分析、威慑、缓解和预防、检测、响应七个方面, 旨在建立内部威胁的研究框架, 该框架遵循从事件到解决方案的方向描绘了内部威胁研究的工作流; (2) 从定义与分类、数据集以及事件的角度对内部威胁进行了全面的分析, 提出了针对内部威胁的结构化分析与分类方法, 将威胁事件的重要特征维持一个易于维护和清晰的状态, 便于扩展、整合以及修改; (3) 基于内部威胁分析提出一个包含威慑、预防/缓解、检测和响应的分步防御框架, 该框架概括了用户行为、心理和犯罪学对于事件的影响, 并对防御框架内每一步包含的方法进行归纳分析; (4) 通过分析内部威胁案例和当前研究进展, 讨论现有研究的不足并从数据集、事件分析、防御三个方面展望进一步的研究方向。

关键词 网络安全; 内部威胁; 分析与防御; 文献归纳; 结构化分类; 综述
中图法分类号 TP309.2 DOI 号 10.19363/J.cnki.cn10-1380/tn.2023.06.02

A Survey of Insider Threat Analysis and Defense Solutions

SUN Degang[1,2], LIU Meichen[1,2], LI Meimei[1,2,3], WANG Xu[1,2], SHI Zhixin[1,2],
LIU Pengcheng[1,2], LI Nan[1,2]

[1] School of Cyberspace Security, University of Chinese Academy of Sciences, Beijing 100049, China
[2] Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
[3] School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Abstract Insider threat is initiated by trusted internal personnel. Which is more transparent, covert, and high-risk than external threat. It is a challenging cyber security issue, therefore we should pay more attention to the insider threat's current research findings and evolution trends. In this paper, we study the research category of insider threat and use grounded theory for rigorous literature review and analysis. Through the systematic study of insider threats in the panoramic view, we aim to help organizations obtain a panoptic view on this disparate topic and thereby quickly develop solutions according to their actual situation. This paper presents a novel insider threat survey of great significance to the field of insider threat. The main contributions of this survey can be summarized as follows. (1) It summarizes the research scope of insider threat, aiming at establishing the framework of this research. The research scope includes seven aspects: definition and classification, data set analysis, event analysis, deterrence, mitigation and prevention, detection and response. The framework describes the workflow of insider threat research, following the direction from event to solution. (2) It makes a comprehensive analysis of insider threats from the definition and classification, data sets and events, and proposes a practical and unified taxonomy. This method makes the important characteristics of threat events easy to maintain and keep a clear state, and makes it easy to expand, integrate and modify. (3) It proposes a step-by-step defense framework including deterrence, prevention/mitigation, detection, and response, it summarizes the impact of user behavior, psychology, and criminology on events, and then summarizes and analyzes the research results. (4) It analyzes the insider threat cases and current research progress, then discusses the deficiency of existing research and proposes further research directions from three aspects: data set, event analysis, and defense.

Key words cyber security; insider threat; analysis and defense solutions; grounded theory for rigorous literature review; practical and unified taxonomy; survey

通讯作者: 李梅梅, 硕士, 高级工程师, Email:limeimei@iie.ac.cn。
本课题得到国家重点研发计划课题(No. 2018YFF01014303)、中国科学院 C 类战略性先导科技专项(No. XDC02040300)资助。
收稿日期: 2020-09-24; 修改日期: 2021-01-19; 定稿日期: 2023-02-20

1 引言

Cybersecurity Insiders 与 Crowd Research 在 2018 年对网络安全专业人士进行了调查, 该调查显示 90%的被调查者认为他们的组织很容易受到内部威胁的影响, 主要的因素包括拥有不合理访问权的用户过多(37%)、敏感数据的访问设备过多(36%)、信息技术越来越复杂(35%)[1]。在过去一年里, 有 64%的企业越来越关注防范内部威胁、58%的组织关注威慑方法、49%的组织注重分析与发布违反规定取证。2019 年的一项调查显示[2], 30%的组织认为内部威胁造成的损害比外部攻击更为严重。虽然内部威胁事件的数量远远小于外部攻击, 但是其造成的损失却是巨大的。美国计算机紧急事件响应小组协调中心(Computer Emergency Response Team, CERT)的报告显示[3], 平均一个内部攻击损失约为 170 万美元, 最多的一次内部威胁事件造成了 870 万美元的巨额损失。

内部威胁不同于外部威胁, 攻击是由可信的内部人员发起的, 具有以下特征: 1) 透明性: 攻击者来自安全边界内部, 因此攻击者可以躲避防火墙等外部安全设备的检测, 导致多数内部攻击对于外部安全设备具有透明性; 2) 隐蔽性: 内部攻击者的恶意行为往往发生在正常工作的间隙, 导致恶意行为嵌入在大量正常行为数据中, 提高了数据挖掘分析的难度, 同时内部攻击者具有组织安全防御的相关知识, 因此攻击者可以采取措施逃避安全检测; 3) 高危性: 攻击者自身具有组织的相关知识, 可以接触到组织的核心资产(如知识产权等), 从而对组织的经济资产、业务运行以及组织信誉进行破坏, 对组织造成巨大损失[4]。内部威胁的攻击者具备内部知识, 因此可以直接访问核心信息资产, 对企业造成严重危害; 同时, 透明性与隐蔽性却使得这种威胁难以检测, 难以防范。任何有恶意行为倾向的内部人员都可能给组织造成严重的财产和名誉损害, 内部威胁形势严峻, 需要引起组织和个人的高度重视。因此, 本文认为十分有必要对当前内部威胁的研究进展进行系统的分析, 帮助研究者基于自身实际快速制定安全策略, 从而应对日益严峻的威胁挑战。

许多研究者对这一领域进行了归纳分析, 在研究了这些综述[5-10]之后, 我们发现一些问题。例如, 文献[5]仅仅关注伪装者检测; 文献[6-8]仅涉及检测系统或者只关注检测方法; 文献[9]仅关注内部威胁中的行为分析而没有注意到心理和社交分析; 文献[10]在排名最高的顶级期刊中选取了 90 篇文章, 采用扎根理论的统计学方法将内部威胁研究领域分为 6 类: 1) 内部威胁缓解; 2) 理论观点; 3) 内部威胁管理; 4) 内部威胁行为; 5) 内部威胁概述; 6) 其他, 但是这项工作只是对这 90 篇文章进行统计学分类, 旨在描述内部威胁研究的范畴, 而没有具体的对文献进行阐述和解释。因此, 对内部威胁的相关研究进行更为全面的回顾和分析是很有必要的。我们的综述不仅包含内部威胁的检测方法, 还在全景视图的下对内部威胁的研究领域进行分类, 并对每个类别进行拓展和细粒度分类。我们的研究对于防御方案的设计者和寻求内部威胁解决方案的研究者是十分重要的, 为他们选择和实施适当的防御方案提供了参考。

本文旨在解决现有综述研究中包含的局限性并进行完善, 提供一个更为详尽和新颖的文献集并对这一领域的知识和研究进行系统化。针对广泛的内部威胁问题, 我们使用扎根理论的文献分类方法, 将引用量高且新颖的文献的研究内容进行分类, 从而确定内部威胁的研究范畴。

本文的贡献如下:

(1) 使用严格的扎根理论的文献分析方法, 归纳出内部威胁的研究范畴, 并遵循从事件到解决方案的方向, 将研究范畴描述为内部威胁的工作流, 旨在系统化内部威胁研究, 有助于研究者在全景视图下快速了解内部威胁并制定安全策略。

(2) 对内部威胁进行了全面的分析并提出一种结构化的分析与分类方法, 将现有的内部威胁分析与分类方法整理为统一视图, 旨在帮助研究者简化信息收集的过程, 并将内部威胁核心特征维持在一个便于扩展和修改的状态。

(3) 在内部威胁分析的过程中, 提出以威慑、缓解/预防、检测和响应为主的分步防御框架, 并对每一步的研究进行拓展和细粒度的分类, 有助于研究者从不同层次不同类别了解内部威胁的防御, 帮助研究者基于自身实际建立更为专业化的安全策略。

文章组织结构如下: 第二章提出了内部威胁的研究范畴。第三章讨论内部威胁的定义与分类, 提出了一种结构化且易于维护的分析与分类方法; 对现有公开可用数据集进行介绍; 并从案例、行为、心理和犯罪学四个角度对事件进行分析。第四章提出了一种内部威胁分步防御框架, 并以此框架为基础细粒度的分类了当前的防御方法。第五章对本文进行总结并展望未来的研究方向。

2 内部威胁研究范畴

本文的目的在于系统化内部威胁研究。因此, 我们参考了文献[10]使用的统计学方法, 将引用量较高且新颖的 124 篇文献的研究内容进行统计和归纳, 从而提出内部威胁的研究范畴。还将此范围内的文献进行分门别类的概述, 帮助研究者在全景视图下快速了解该领域并制定符合自身实际的防御方案。这些参考文献的发布年份如图 1 所示。

图 1 参考文献年份分布

Figure 1 The number of references which distributed according to the year

本文调查了这些参考文献的研究内容, 并将它们分为 6 类: 内部威胁概述、事件分析、内部威胁管理、缓解/预防、检测以及其他。各个类别中文献的数量分布如图 2 所示。一些参考文献不仅对于某一类研究内容有贡献, 例如, 文献[3]不仅明确了内部威胁的定义还概述了预防和检测方法, 因此概述、缓解/预防和检测这几类都包含了该文献。

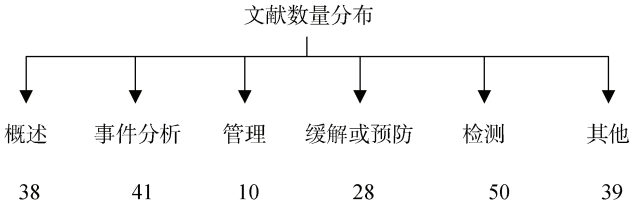

图 2 文献分类

Figure 2 Classification of references

本文由此确定了内部威胁的两大研究范畴, 随后的章节也将对每一类别进行扩展和细粒度的分类。我们首先对这两个主要类别进行概述:

(1) **内部威胁分析。**本文从定义与分类、数据集和事件几个方面对内部威胁进行分析。1) 定义与分类: 研究者在开始内部威胁领域研究之前, 最重要的工作就是明确内部威胁的定义, 而定义又随着所关注的威胁类型的不同有细微差别。为了帮助研究者简化这一信息收集的过程并快速了解内部威胁的重要信息, 我们参考了一种用于信息收集的科学分析方法, 提出了结构化的内部威胁分析和分类方法。2) 数据集分析: 数据集不仅可以描述真实世界的内部威胁事件案例, 还可以用于评估防御方案。本文基于数据集的用途将现有公开可用的数据集分为四类: 伪装数据集、叛徒数据集、混合数据集和身份验证/识别数据集。3) 事件分析: 事件分析对于理解攻击过程, 理解恶意内部人员的动机和行为具有重要意义, 是设计防御方案的基础。我们首先介绍案例分析的研究, 然后概括恶意内部人员在事件发生之前、正中和之后的行为、心理以及其他相关方面的研究。通过案例、行为、心理和犯罪学这四个方面的研究, 来对事件进行更为全面的分析。

(2) **内部威胁防御。**本文基于内部威胁分析提出一个包含威慑、预防/缓解、检测的三步防御框架, 在缓解/预防或检测捕捉到威胁后, 进行响应。1) 威慑: 内部威胁的根本问题是如何避免用户在社会环境下的恶意行为, 那么如何在内部威胁中应用社会科学方法将是一个决定性问题。为此, 我们的防御框架考虑了犯罪学和社会学, 试图通过改变组织结构和文化以及威慑犯罪来减少动机。我们考虑了三大社会科学理论, 分别是威慑理论(the general deterrence theory, GDT)、情景预防犯罪理论(Situational Crime Prevention, SCP)和计划行为理论(Theory of Planned Behavior, TPB)。2) 缓解和预防: 对信息和人员缺少安全属性的标识是造成内部威胁的一个主要原因, 如果能够识别资产和人员, 就能在一定程度上预防内部威胁事件的发生。我们的防御框架将这类预防机制作为第二道防线, 使用包括访问检查、访问控制、防数据泄露和诱饵在内的各类方法。3) 检测: 由于内部人员可以通过盗用权限、伪装等方法来进行恶意攻击, 前两道防线无法进行有效的防护, 我们将检测作为第三道防线。现有的检测方法可以分为三类: 异常检测、误用检测和混合检测。异常检测是指建立用户正常行为模型, 与之进行对比检测出偏移该模型的异常行为; 误用检测是指预先定义异常行为, 然后使用反向的异常检测方法; 混合检测方法结合了上述两种方法。4) 响应: 在缓解、预防以及检测给出警报之后, 如果能快速进行警报分析以及响应, 就能在可能造成任何损失之前及时阻止威胁。

除了分类概述外, 我们的另一个目的就是描绘内部威胁研究的工作流。遵循从事件到解决方案的方向, 首先应该从内部威胁的定义与分类出发; 在结合自身实际明确了内部威胁的定义和类别之后, 要寻求可用的数据集用于分析/评估; 随后结合案例理解攻击过

程和攻击意图; 最后确定符合自身实际的防御方案。而在组织部署了防御方案之后, 仍然需要对发生的内部威胁事件进行分析, 从而进行反馈更新以期望提高防御方案的精度。换句话说, 所有防御方案的设计都依赖于对内部威胁的定义以及数据集和事件分析, 而防御方案的部署能帮助组织获取真实的数据和事件, 从而进行更为准确的数据集分析和事件分析, 帮助组织更新一个更为精确和全面的防御方案。由此确定了本文的研究范畴及其工作流, 如图 3 所示。从左到右的方向表示事件到解决方案的方向, 这是面向目标的工作流, 而自右到左的方向表示从解决方案到事件的方向, 这些模块是有序而又相互影响的。

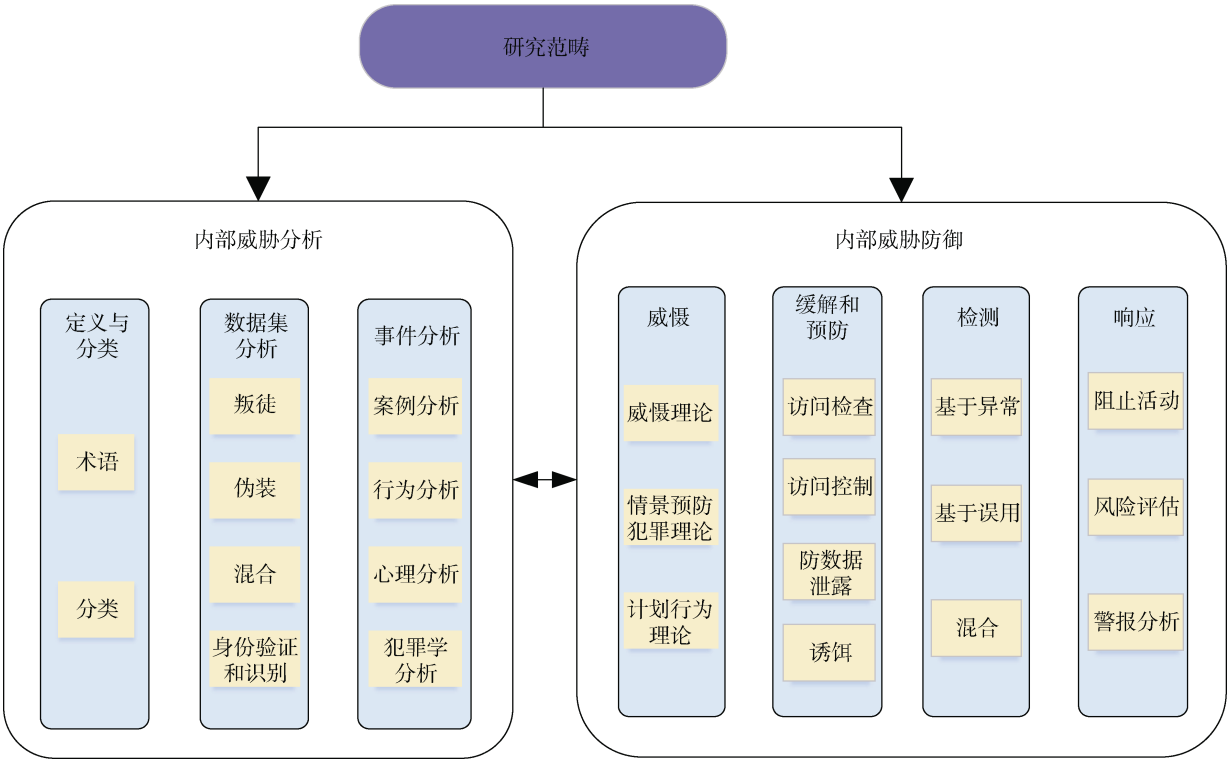

图 3　内部威胁研究范畴

Figure 3　Research scope of insider threat

3　内部威胁分析

本章从三个方面对内部威胁进行分析, 包括定义与分类、数据集分析以及事件分析。

3.1　定义与分类

本节重点介绍内部威胁的定义与分类。当组织的内部员工和外部人员同时在内部网络中活动时, 很难界定内部威胁的边界范围, 甚至还有内部人员(刚离职的员工)从外部对组织进行攻击。因此, 内部威胁很难被定义。通常研究者选择的定义取决于其关注的威胁类型。我们首先调查了不同研究对于内部威胁的定义, 然后借鉴了六何分析法(任何问题都可以从原因、对象、地点、时间、人员、方法六个方面进行分析), 将内部威胁从人员-Who、目标-Where、原因-Why 和方式-How 四个方面进行分析并分类。

3.1.1　内部威胁定义

大多数定义涉及到两个术语: 内部人和内部威胁。内部人是对个人的静态描述, 使用例如访问权限、信任、安全策略等术语进行描述; 而内部威胁是对应的行为, 使用例如滥用访问、违反安全策略等术语进行描述。

1) 内部人

内部人的定义最早出现在文献[11]中, 他们认为具有计算机和网络授权使用的是内部人。文献[12]将内部人定义为一个合法进入组织计算机和网络的人。文献[6]将内部人定义为对内部资产具有某种合法特权的人, 内部人有权改变组织的计算机设置、数据或者程序。文献[13]从数据访问的角度定义内部人, 提出用安全边界(例如, 防火墙或者局域网)来区分内部人和外部人。文献[4]从信任的角度将内部人定义为一个被信任的人, 内部人可以访问敏感信息和信息系统。文献[14]认为内部人能对资产进行合法访

问、使用和更改。文献[15]从安全策略的角度出发, 定义了语言策略、可行策略、配置策略以及运行策略四个等级, 从不同的等级对应不同的行为来定义内部人。文献[16]将内部人的定义进一步扩展为配偶、朋友或者商业伙伴。文献[3]也认为内部人也可以由外部实体来表示, 因此将内部人扩展为企业或组织员工(在职或离职)、承包商以及商业伙伴等, 并且内部人应该具有系统、网络以及数据的访问权。

2) 内部威胁

文献[12]将内部威胁定义为一个内部人的行为, 以不当的方式将组织的数据、程序或其他资产置于风险之中。文献[16]指出内部威胁不仅包括恶意内部人员的恶意行为, 还包括正常内部人无意的犯错。文献[17]定义内部威胁为来自被授权访问的人的威胁, 例如, 被授权的用户滥用其特权违反组织的信息安全政策。文献[4]定义内部威胁为受信客体违背对授信主体的承诺, 做出不利于授信主体合法利益的行为, 通过背信行为的具体化来定义内部威胁。文献[3]定义内部威胁为内部人利用合法获得的访问权对信息系统的机密性、完整性以及可用性造成负面影响。

3.1.2 内部威胁分类

上一小节我们对内部人和内部威胁这两个术语进行了描述, 本节从分类的角度对内部威胁进行分析。通常研究者在开始内部威胁研究的时候, 需要对内部威胁相关研究进行信息收集, 为了帮助研究者简化这一过程, 我们将内部威胁拆分为谁、在何地、为什么、做了什么, 即攻击人-WHO、攻击目标-WHERE、攻击动机-WHY、攻击方式-HOW。这种结构化的分析与分类方法, 能将现有的信息整理为统一视图, 如图 4 所示, 有助于系统性了解内部威胁, 将威胁事件的重要特征维持一个易于维护和清晰的状态, 便于扩展、整合以及修改。

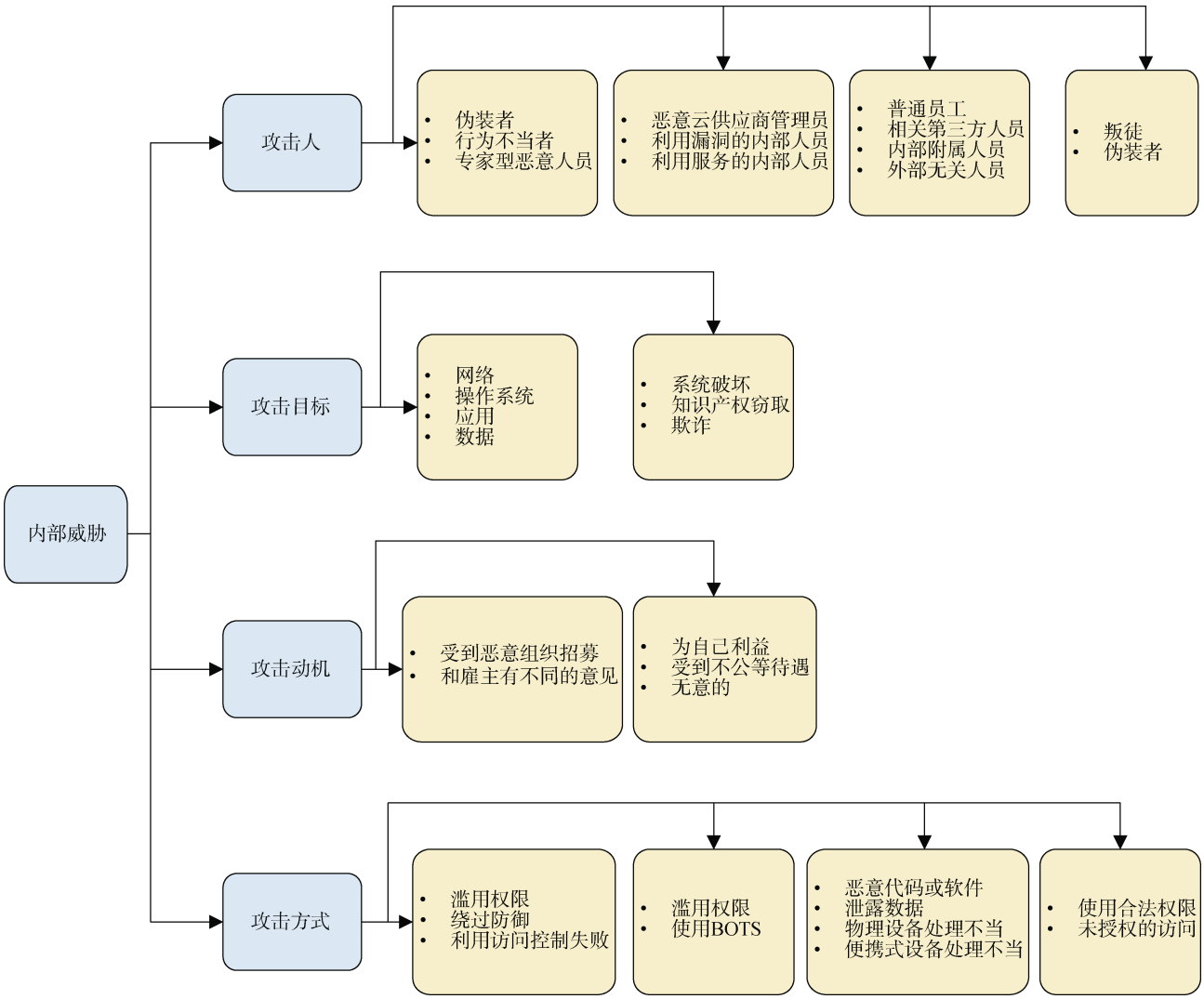

图 4 结构化分类方法

Figure 4 Structural taxonomy of insider threat

1) 攻击人

计算机系统内部威胁最早的分类由文献[18]提出, 根据非法用户将内部威胁分为三类: 1) 伪装者, 伪装者是指通过某种方式绕过安全机制, 并渗透到计算机系统内部的人, 或通过伪造/窃取来执行一些恶意行为的内部用户; 2) 行为不当者, 这类用户滥用自己权限; 3) 专家型恶意人员, 该类用户熟悉内部网络安全策略和防护机制, 能利用专业知识实施威胁行为, 比如知识产权窃取, 这类恶意人员最难察觉。文献[19]从云计算的角度将恶意内部人员分为三类: 1) 恶意云供应商管理员, 他可以通过访问潜在敏感数据或者其他租用的资源来进行泄密或者欺诈; 2) 利用漏洞的内部人员, 未经授权地窃取或者破坏内部数据; 3) 利用服务的内部人员, 例如破解密码文件, 进行分布式拒绝服务攻击等。文献[20]将内部人分为四类: 1) 只拥有必要权限的普通员工; 2) 相关第三方人员, 例如承包商或者供应商; 3) 内部附属人员, 比如员工的家庭成员、朋友, 这类人没有进入组织的权限, 但是可以通过窃取的方法获得权限, 并实施恶意行为; 4) 外部无关人员, 组织外部的不受信任人员可以通过网络攻击或者漏洞(钓鱼攻击)来获得访问权限。文献[6]基于用户的知识体量将恶意内部人分为叛徒和伪装者, 叛徒具有一定的知识体量, 对政策和安全机制有了解, 甚至有自己的权限和凭证; 而伪装者比叛徒的知识体量小很多, 他们通常窃取另一个用户的凭据来进行恶意活动。

2) 攻击目标

文献[21]将攻击目标分为四类: 网络、操作系统、应用和数据, 作者假设一个内部攻击可以在系统的特定级别上表现出来且一个内部攻击的痕迹可能存在于不同的级别, 比如, 违反数据完整性的行为可以在数据和应用上体现, 数据溢出在网络和操作系统上体现。文献[3]将内部威胁目标分为三类: 1) 系统破坏, 内部人员利用自己的权限来对组织或个人进行伤害; 2) 知识产权窃取, 内部人利用信息系统窃取组织的知识产权, 组织包括企业、政府以及各种机构, 信息窃取中信息的外延包括知识产权、组织信息(结构设置、工资福利数据、安全策略等)以及业务所需的客户信息; 3) 欺诈, 内部人出于个人利益, 利用信息系统非法修改、添加、删除组织数据, 或窃取信息进行身份欺诈, 内部欺诈有两类: 一是非法篡改组织数据, 从而直接牟利, 如金融数据、消费账单记录等; 二是与外部犯罪团伙苟合, 窃取内部数据用作身份欺诈, 如信用卡欺诈等。

3) 攻击动机

文献[20]基于攻击动机将内部威胁分为两类: 1) 内部人员受到恶意组织招募, 进行内部攻击, 通常恶意组织会针对那些有经济困难的人或者想要挣取外快的人进行招募; 2) 个人原因, 一些员工和雇主有不同的意见, 比如政治观点不和, 或者在雇主那里遭遇了不公待遇。招募者会使用诱饵或个人把柄来给员工制造陷阱。文献[6]根据动机将内部威胁分为三类: 1) 为自己利益进行恶意行动; 2) 因为不公待遇想要对组织进行破坏; 3) 由于操作不规范等原因无意中做出了恶意行为。

4) 攻击方式

文献[22]根据攻击方式将内部威胁分为三类: 1) 滥用权限, 内部人使用合法权限进行不当访问, 这是最难检测的; 2) 绕过防御, 内部人绕过防火墙或者绕过他人进行的攻击; 3) 访问控制失败, 利用访问控制存在的技术问题或者漏洞进行攻击。文献[23]监控网络事件, 将内部威胁的攻击方式分为两类: 1) 未经授权访问数据(与项目无关的数据)以及以不当的方式滥用权限(与无关的人共享数据); 2) 利用木马执行内部网络侦察或其他自动化程序识别内部系统漏洞。文献[24]定义了四种恶意攻击方法: 1) 恶意代码与恶意软件相结合的方式, 比如植入 USB 驱动或者网络钓鱼攻击和间谍软件的组合; 2) 泄露数据, 在网络上公开发布敏感信息或者通过邮件传送给无权限人员; 3) 物理设备不当处理, 非电子记录的丢失, 比如纸质文件; 4) 便携式设备处理不当, 比如丢失数据存储设备等。文献[25]将攻击方式分为两类: 1) 使用合法权限访问; 2) 未授权的访问。

3.2 数据集分析

数据集在每个应用领域中对于设计和评估新想法都至关重要。我们回顾了内部威胁的相关文献和防御方案, 将常用的公开可用数据集分为四类, 分类标准如图 5 所示。内部人员通过非授权访问来执行恶意行为, 检测伪装者的数据集将此类行为标记为恶意标签。伪装者的数据集将具有区别于正常用户的行为数据标记为恶意, 而检测叛徒的数据集将用户恶意违反策略的意图标记为恶意标签。这类内部人员通过合法的权限来执行恶意行为, 例如项目负责人拷贝其具有访问权限的知识产权。由伪装者和叛徒混合而成的数据集是内部威胁的通用测试数据集。基于身份验证/识别的数据集与用户的意图是否为恶意无关, 主要用于识别和认证

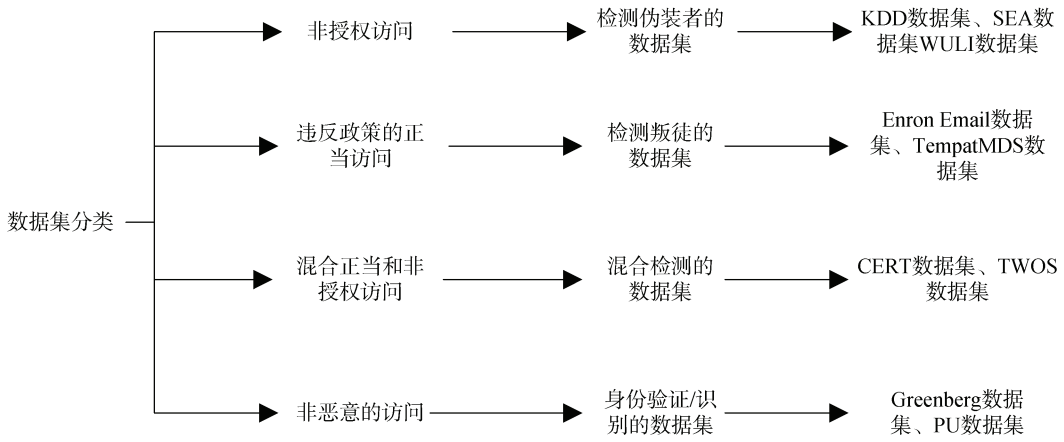

图 5　公开可用数据集分类

Figure 5　Categorization of public datasets

3.2.1　基于伪装者的数据集

KDD 数据集[26]源自于一个入侵检测项目, 美国国防部高级规划署分别从主机和网络收集了近 9 周的系统审计和网络连接数据。系统审计数据主要是用户信息、进程信息等。网络连接数据主要使用 TCP 记录。前 7 周收集到的数据用于训练集, 后 2 周的数据用于测试集。该数据集记录了 41 个特征的向量描述, 能够刻画探测攻击、拒绝服务攻击、远程连接和用户提权四种攻击。但是该数据集产生时间较早, 与实际内部威胁相差巨大。

SEA 数据集[27]被广泛应用于伪装者检测研究。该数据及记录了 UNXI 系统下的 70 个用户的约 15000 条命令, 数据集是由组织内的各种角色的 50 个正常用户生成的, 剩下的 20 个用户中会随机插入一些模拟攻击。每个用户的命令数据的前 50 个数据块为正常数据, 后 100 个数据块可能被插入了恶意行为数据。文献[28]提出了 SEA 数据集的变体, 被插入的模拟攻击考虑到了每个用户的命令频率, 试图将伪装者模仿的更像合法用户。

WULI 数据集[29]从 windows 系统用户的文件访问行为来刻画用户行为。该数据集记录了 20 个用户浏览文件和目录的行为, 每条信息包含 ID、访问事件、文件对象以及路径信息。与 SEA 数据集不同的是, 该数据集包含了用户知识背景, 20 个用户来自于不同的职业。虽然正常用户是从真实用户中收集的, 但是该数据集的攻击行为来自于模拟的伪装用户。这些伪装者的攻击能力也大有不同, 这些能力和用户的职业相关联。

3.2.2　基于叛徒数据集

Enron Email 数据集[30]是由 150 个用户的 50 万封电子邮件组成, 主要是安然公司的高级管理用户。虽然由于隐私问题, 附件和机密内容被删除, 但是仍然能够分析邮件中的文本和社交网络, 该数据集用于内部威胁中的叛徒检测。

TempatMDS 数据集[31]模拟计算机系统文件访问工程师的日常工作, 利用审计令牌来描述文件访问, 包括时间, 人员, 主题, 动作类型, 操作方向和对象令牌, 每个用户的文件访问记录按照时间跨度划分为多个文件子活动。该数据集包含良性工程师和恶意工程师, 恶意工程师的任务是基于良性分析师的任务来进行一些恶意操作, 比如, 浏览重要文件的内容, 然后将有价值的内容复制到 USB 中, 这使得检测更具有挑战性。

3.2.3　混合数据集

CERT 数据集[32]是来源于美国国家互联网内部威胁中心的数据, 该数据集模拟了各类涉及恶意用户的场景, 包含了这些场景下的用户行为数据和用户背景数据(如, 用户属性)。CERT 数据集中涉及了多个维度的文件访问数据、用户行为数据、设备使用数据、邮件收发数据、HTTP 访问数据以及系统登陆行为数据, 还包括用户的角色信息和心理测量信息。该数据集的问题在于攻击多来源于人工模拟, 无法反应真实的攻击行为。

TWOS 数据集[33]是一个多人游戏的数据集, 旨在体现真实组织的互动, 涉及 24 个用户, 这些用户组成了 6 个队。数据集由鼠标、键盘、网络和系统调用的主机监视器日志等数据组成。但是该数据集只能基于模板生成命令序列, 为了解决这个局限性, 文献[34]使用户可以手动创建/指定模板, 模板包括命令、元命令(一组功能相似的命令)、作业、元命令序列以及会话(一系列元命令序列)。

3.2.4　基于身份验证的数据集

Greenberg 数据集[35]是第一个基于身份验证的数

据集, 收集了 168 个 UNIX 用户的完整命令行, 包含会话开始/结束时间, 用户输入命令行, 当前工作目录, 上一个命令的别名扩展, 所输入的行是否有历史纪录扩展名以及在命令行中检测到的错误。该数据集根据知识和技能将用户分为四组, 分别由 52 名具有编程技能的科学家, 55 名新手程序员, 36 名高级程序员和 25 名非技术用户组成。和 SEA 数据集不同的是, 它不仅仅包含普通配置(命令和别名), 还包含完整参数。

PU 数据集[36]是预处理的 UNIX 命令数据, 这些数据是从普渡大学的八个计算机用户的 shell 中获取的, 包含 2 年的用户数据, 每个用户收集的命令数量从 7769 到 22530 不等, 平均每个用户 16500 个命令, 包含命令名称, 参数和选项。

3.3 事件分析

本节首先概述了各种类型的内部威胁案例研究, 然后概括恶意内部人员在事件发生之前、正中和之后的行为、心理以及其他相关方面的研究。通过行为意图研究、心理和社会理论研究, 来对内部威胁事件进行一个全面的分析。

3.3.1 案例分析

迄今为止, CERT 发布了许多丰富的内部威胁案例研究, 他们建立了 2001 年至今的 700 多例内部威胁案例数据库[37]。在此之前, 也有一些针对特定领域的案例总结, 例如, 文献[38]研究了 1996 年至 2002 年在金融领域发生的 23 起内部威胁事件, 其中包含 15 起欺诈事件、4 起知识产权窃取事件和 4 起系统破坏事件。文献[3]描述了 51 个内部威胁案例, 并把它们进行分类: 24 起破坏案例、3 起伴随着破坏的欺诈案例、6 起知识产权窃取案例、12 起欺诈案例以及 6 起其他案例。文献[20]将内部威胁组织分为政府和商业公司, 分析了 21 个政府的内部威胁案例和 15 个商业公司的案例。

3.3.2 行为分析

文献[39]分析了一个人的行为有哪些决定因素, 如图 6 所示, 包括反应效能(Response Efficacy)、社交影响(Social Influence)和自我效能(Self-Efficacy)。反应效能是指组织已有的措施能有效避免威胁的程度。员工对组织反应效能的认知将决定他们选择应对威胁的方式, 组织对威胁的反应效能越高, 越能减少威胁。自我效能也是决定因素之一, 员工是否会按照组织的建议或者规定来执行是至关重要的。一个人的行为也会受其社交圈内相关人员的影响, 例如, 一个人接受和使用组织的建议或者规定会受到其朋友/同事的影响。这些社会因素是指个人会参考群体主观文化, 以及个人在特定社会情况下与他人达成的具体人际协议。制裁严重性(Perceived Threat Severity)和制裁确定性(Perceived Threat Susceptibility)都会影响反应效能和自我效能。制裁确定性是指相信能够发现个人的不当行为, 制裁严重性是指明确不当行为将导致严重的惩罚。

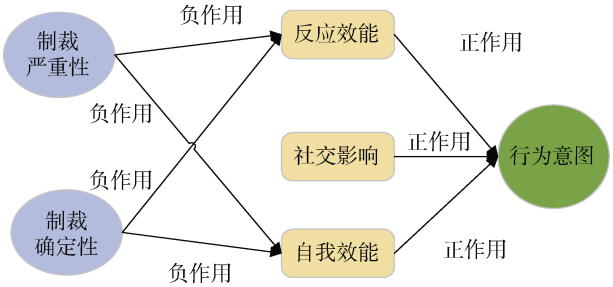

图 6 个人行为决定因素

Figure 6 Determinants of individual behavior

基于该行为意图的内部威胁分析有许多例子。他们通过阐述恶意内部人员的属性、知识、访问权限、特权、风险、策略、技能、动机、过程等来模拟事件。文献[12]基于组织、环境、系统、个人四个方面的相互作用, 提出了一种针对内部风险的概念框架。文献[40]确定在恶意内部威胁的整个过程, 从攻击发生到检测以及响应这一系列连续的共同事件中都有引爆点(第一次不满意的事件)和恶意行为(安装漏洞)。文献[41]认为用户特征、攻击特征和组织特征是内部事件是否发生的决定因素。文献[42]提出一种理论内部威胁模型, 该模型基于三种理论, 分别是基于动机、机会、能力三者结合的理论; 基于计划行为理论; 基于人格特征理论。文献[43]正式将恶意内部威胁建模在微观和宏观方面, 进行了意向性分析和行为分析, 通过贝叶斯网络对员工进行内部威胁的可能性分析, 当员工有攻击意图时, 利用马尔可夫决策过程进行行为分析。

3.3.3 心理分析

心理和社会研究发现, 内部威胁和人格因素、情绪、精神具有明确的联系[44]。文献[45]证实了自恋人格的用户更有可能造成内部威胁。下面的大量研究力图从心理学和社会学领域来研究内部威胁。文献[44]专注于系统管理员、程序员、网络专业人员并探索个人和文化的脆弱性, 以此为基础提出一系列指标, 包括内向、挫折、伦理、忠诚度、归属感、共情度等。文献[46]分析了影响员工安全行为的几个因素, 包括其他员工的行为、员工的安全感以及员工与公司的心理锲约, 潜在影响下的认知是影响内部人感知风险的最显著特征。文献[47]讨论了如何在工作环

境中形成公平感, 包括分配、纠纷处理程序、人际关系、尊严和信息公开化, 后来他们[39]又审查了现有的公平规则, 认为用户的内部工作环境合理化了他们的恶意行为。

文献[48]证明了一个假设, 一旦内部人有恶意意图, 群体传播的语言激励就发生了显著变化。因此, 心理学领域的研究一般是用侧写的方法从用户的语言和社交中推断其心理状态, 从而分析出其人格特征。例如, 文献[49]从游戏服务器中抓取用户的游戏数据, 通过其社交言行发现用户心理状态上的异常, 以此检测背叛游戏公会的用户。文献[50]从社交媒体应用的角度推断用户的人格特征以检测可疑的内部威胁用户。文献[51]抓取推特用户的状态数据, 包括用户 ID、昵称、简介、关注用户数和粉丝数等信息, 绘制用户的社交网络, 当用户与其所在的社交网络组适配度较低时, 该用户被判断为有自恋人格。文献[52]分析用户邮件内容, 提出了一个自动语言分析系统, 根据用户邮件的内容分析用户心理因素的变化, 进而检测出潜在的内部威胁用户。文献[53]对用户的网络浏览行为进行心理分析, 从网络浏览的网页中提出文本信息建立词汇计数和词向量之间的关系矩阵, 然后将关系矩阵和五大人格矩阵进行结合作为正常值, 通过计算用户浏览的新网页的结合矩阵与日常值的距离来判断异常。

3.3.4 犯罪学分析

内部威胁是很多问题的综合, 层出不穷的内部威胁定义以及各式各样的分类方法也证明内部攻击和威胁处在不断变化当中。内部威胁也可以认为是工作场所的不当甚至犯罪行为。因此, 犯罪学也可以被用来分析内部威胁和保护信息资产的安全 [54,55]。GDT 解释了人们如何避免在社交网络的影响下产生不良行为[56], 从最小化成本和最大化个人利益的角度描述了人类的行为和决策。比如, 罪犯如果意识到自己将受到严厉的惩罚, 他就不会进行恶意操作。文献[57]证实 GDT 能够有效的影响内部用户威胁的态度和意图, 从而有效防止内部威胁。SCP 解释了如何通过减少动机和机会来减少犯罪活动和行为[58], 通过帮助设计一种能让犯罪变得越来越困难的环境来减轻威胁行为。TPB 解释了当用户持积极态度, 认为其他人与其有同样的行为方式, 且自认为有能力进行行为控制的时候, 该用户的威胁意图会大大减少[59]。

3.4 小结

层出不穷的系统破坏、欺诈、知识产权窃取等事件使内部威胁成为了网络安全领域的重点话题, 国内外安全领域的研究者们针对内部威胁定义、分类、数据集以及事件分析进行了深入研究, 取得了一定的成果。1) 在定义与分类方面, 最早的研究伴随着用户访问控制研究而产生, 此时研究者们将内部人滥用权限定义为内部威胁, 之后随着内部威胁问题的逐步科学化与系统化, 定义也逐渐清晰。目前学界较为接受的是 CERT 的定义: 内部人利用合法获得的访问权对信息系统的机密性、完整性以及可用性造成负面影响, 内部人是指企业或组织员工(在职或离职)、承包商以及商业伙伴等。大量对内部威胁分类的研究也有助于研究者根据自身实际情况研究相应的解决方案。但是随着移动、云等新型设备融入内部网络, 内部网络的规模随之增大, 内部网络的安全架构也有了新的形态, 已经产生了许多无法归入到现有分类中的威胁。2) 在数据集方面, 研究者们为了提供更具有现实意义的防御方案测试集, 在创建数据采集环境时尽可能的描述真实世界内部威胁事件。不仅结合人口统计模型、心理测量模型来模拟恶意环境, 还将这些恶意环境中的恶意攻击和良性人为事件一同混入正常数据集中, 使得数据更贴近现实。但仍普遍存在不足, 合成了攻击的数据集缺乏验证, 无法证明与真实环境的相似性。此外, 考虑到用户行为数据的隐私性和安全性, 数据集发布之前必须要进行隐私处理, 但是匿名化过程会隐藏数据集中的一些重要信息。3) 在事件分析方面, 针对研究受到案例不足限制的问题, 不仅有包含 700 多例内部威胁案例的数据库, 还总结了针对特定领域的案例。研究者能在其中分析内部人的行为模式、心理特征等, 更好的研究应对内部威胁的方法。此外, 研究者已经注意到内部人的主观因素与内部威胁的关联, 并且从心理学、犯罪学、社会学的角度阐述内部人的动机和行为特征。但是本文认为事件分析方面的主要问题有两个。首先没有意识到心理/犯罪学因素结合行为意图的必要性, 心理/社会特征只能说明用户异常, 并不能判定其为恶意用户。并且由于心理、社会因素多涉及公司机密与个人隐私, 难以获取, 现有研究只能从邮件、网页中的数据进行侧写。其次, 内部威胁的刻画从动机到行为, 从仅刻画内部威胁某一方面到包含各类特征和属性的复杂模型, 给具体实现带来了巨大挑战, 导致目前的模型偏向形式化验证还未产生实际价值。

4 内部威胁防御

由此, 我们由事件分析得到了内部威胁的分步防御框架, 该框架概括了用户行为、心理和犯罪学对于事件的影响, 如图 7 所示。组织和员工, 员工和员

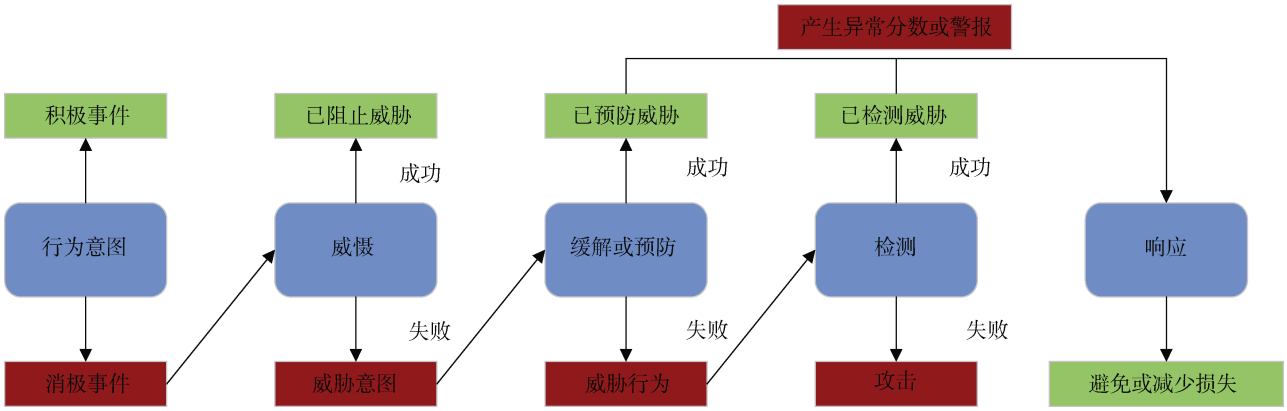

图 7　内部威胁分步防御框架

Figure 7　A step by step defense framework for insider threat

工之间的交互被建模，员工可以感知事件。友好的工作环境会导致积极反应/事件，而对组织或同事的不满、感受到不公待遇等会导致消极的事件。消极的事件形成消极态度甚至威胁意图，这可能导致威胁行为；但是这会受到政策安抚或者其他威慑手段的阻碍。如果威胁意图足够强烈，威胁行为将会被尝试执行，是否能够阻止取决于组织的预防和检测能力。如果预防不能够阻止威胁行为，那么该行为被成功执行后，组织只能通过检测等技术手段来阻止。预防和检测这两道防线可以直接阻止异常人员的活动，但是大多数情况下是通过发出警报/异常分数来对组织进行示警，此时安全分析员需要快速响应。如果该行为逃过了威慑、预防和检测三道防线，那么将造成不可挽回的损失。

我们基于上述框架，对内部威胁的防御方案进行分类和简要概述。首先对威慑方法进行总结，然后将缓解/预防方法进行分类，随后重点介绍现有的检测方法，最后总结了响应方法。必须明确的一点是内部威胁具有广泛性，通常和入侵、金融电信欺诈、网络营销诈骗等研究领域相关。因此，我们所讨论的方法不仅关注内部威胁领域还包含了与内部威胁有交集的相关领域。

4.1　威慑

许多研究者认为内部威胁是工作场所的一种不当甚至犯罪行为。因此，通过威慑犯罪来减少内部威胁的动机也是一步重要的防御解决方案。有一些重要的研究是利用犯罪学中的 GDT 预防用户在社会环境下的恶意行为[60,61]。GDT 理论中的制裁确定性和制裁严重性对于阻止内部威胁有重要作用。文献[61]将组织中的人员分成信息人员、安全员和恶意攻击者，从心理活动中学习出可疑程度和真实威胁概率的曲线，运用 GDT 理论降低内部威胁的频率。犯罪学理论中适用于内部威胁的还包括 SCP 和 TPB。文献[62]采用 TPB 框架对组织内的计算机滥用进行评估，并评估了其他犯罪学理论对 TPB 的影响。文献[63]展示了 SCP 如何应用于内部威胁的通用脚本框架，并展示了其在欺诈检测上的实际应用。文献[61]使用威慑和减少信息安全不良行为的机会来减少内部威胁，通过合成 GDT 和 SCP 来研究如何改变员工的态度和思维方式。文献[46]从组织的角度分析导致员工不满的原因，分别从程序公平、权力分配的角度来解释员工不满。文献[64]提供了企业的管理准则，包括评估优先级、循环审查以及补救等。文献[65]在物理、技术、逻辑和资产四个层次设计特定的行政政策，用来加强对内部人员的防御。

4.2　缓解与预防

本小节将现有的缓解/预防方法分为四类，分别是访问检查、访问控制、防数据泄露和诱饵。

4.2.1　访问检查

文献[66]认为安全策略机制是受保护系统不可分割的一部分，以此作为完整性检查的一部分来保护数据库免受内部威胁。文献[67]对机密文件和数据库记录进行指纹识别，指纹是一组数据的哈希值，将得到的哈希值和机密信息的指纹进行比较，如果匹配，那么阻止交易。每个指纹是从一系列单词中计算得到的，数据的一个字符的变化就会导致不同的哈希值，因此可以通过稍微改写内容来绕过指纹识别。文献[68]通过建立指纹核心机密内容，忽略非机密部分来解决上述问题。文献[69]通过检查每个网络模型分配给特定应用程序的顺序来阻止数据库被破坏。文献[70]针对数据库上的主体对特定数据进行访问的顺序建立威胁预测图，利用该图进行内部威胁

缓解。

4.2.2 访问控制

文献[71]拓展了基于角色的访问控制来评估风险和系统对用户的信任, 通过基于信任和风险的认知来支持访问控制。文献[72]提出了基于属性的群组控制, 结合安全策略用于缓解恶意内部威胁。文献[73]提出基于功能的访问控制, 对访问文档指定部分的操作进行控制。文献[74]提出基于 LINUX 容器的解决方案, 将系统管理员从与其无关的资源隔离, 同时允许他们在权限代理批准时获得额外的权限。还有一些研究是基于白名单/黑名单的访问控制方法, 定义用户对对象执行的操作策略, 例如, 文献[75]基于 XACML, 既可以作为白名单还可以作为黑名单, 允许指定正面和负面策略, 并且可以根据上下文的决策进行配置。

4.2.3 防止数据泄露

文献[76]使用信息检索、文献[77]创建敏感值模型, 来检测离开组织边界的敏感数据。在基于规则的解决方案中, 还有几种网络入侵检测系统, 文献[78]使用基于签名的 Snort 侦听网络流量并阻止其与内部规则匹配, 也适用于检测网络上未经授权的敏感数据泄露, 例如, 通过创建字符串匹配规则。

基于学习的方法是通过观察过去的活动自动学习正常行为模型, 并将模型的偏差标记为异常。文献[79]基于 SQL 命令和执行查询的顺序建立正常行为的配置文件。文献[80]使用贝叶斯分类器构建配置文件, 系统根据查询中的 SQL 命令、表和列来学习预测用户的标识和角色。文献[81]根据用户检索的数据来分析正常行为, 结合以结果为中心和以上下文为中心的混合方法。

文献[82]结合了基于规则和基于学习的方法, 但是他们在没有反馈的情况下结合了这两种技术, 因此无法定义新签名, 也无法从预定义的数据挖掘方案中推断出新的签名。文献[83]为了解决上述问题, 将领域知识用于反馈更新。

4.2.4 诱饵

还有些研究通过限制攻击的影响来缓解内部威胁, 这类方法允许攻击继续进行, 通过自动生成和分发高度可信的诱饵信息, 让攻击者无法区分真正的机密信息与诱饵信息, 并且能够检测和追踪诱饵信息的访问以及滥用行为。这些研究认为有些时候因为遭受到攻击而关闭系统比接受攻击事后做出回应更具有破坏性。文献[84]提出在邮箱中嵌入具有密标的伪邮件来对内部网络嗅探行为进行检测。文献[85]提出使用陷阱主机、灯塔诱饵和 HMAC 验证的方法, 在攻击者使用密标时获取信息。文献[86]从设置的诱饵主机中监测用户的行为轨迹, 分析其异常程度。文献[87]集成了蜜罐和用于流量分析的传感器来监视内部流量, 并使用攻击指示器来推断内部人员的恶意意图。文献[88]提出了几种针对内部人员的诱捕手段, 例如在文档中伪造内容、设置不可见的 HTML 链接、伪造 DNS 记录、在数据库中设置蜜罐、设置诱捕文档/账户等。

4.3 检测

本小节将现有的检测方法分为三类, 分别是基于异常的检测方法、基于误用的检测方法和混合检测方法。

4.3.1 基于异常的检测方法

1) 用户命令检测

文献[88-89]最早使用用户命令作为内部威胁研究的分析对象, 他们首先计算相邻命令模式出现的概率, 然后匹配新出现的命令与历史模式来计算异常。但是, 上述工作没有考虑到用户行为的隐含信息。文献[90]认为刻画用户行为的隐含关联数据也应该被考虑, 为此, 他们基于网络分层的方法将同时出现的非相邻命令提取出来补充模型。文献[91]是基于转换概率的方法, 基于贝叶斯模型来观察单步命令转移概率是否和历史转移矩阵一致。还有一些使用命令序列匹配的方法[92], 通常使用用户命令的统计特征来进行分析, 例如事件发生的频率、事件的持续时间等。但是基于命令的检测所依靠的数据源过于单一, 创建的用户行为模型过于简单, 导致多数方法的检测率不高。

2) 审计日志检测

审计日志检测主要涉及到的日志为: 系统登录/登出日志、文件访问日志、设备使用日志、HTTP 访问日志和邮件收发日志, 文献[32]认为以上五种日志可以刻画用户行为。文献[93]提出一种多源数据融合方法, 从用户的工作组属性出发定义域间一致性, 使用 Term frequency/Inverse document frequency (TF/IDF)思想融合用户在不同数据域上一致性的评分。文献[94]开发了一个名为 ELICIT 的检测系统, 收集了 13 个月的内部人员的数据, 使用 76 个检测器进行检测, 然后为每个用户计算一个威胁分数, 当分数高于阈值时, 该用户被定义为异常用户。文献[95]提出从不同层次来选择特征, 分别是指示器、异常模型和场景, 此外, 还基于异常检测语言建立针对场景的复杂内部威胁检测机制。

除了上述研究, 还有些重要工作是针对文件使用的内部威胁研究。文献[96]将文件目录作为用户

“任务”的抽象, 利用向量对信息系统的权限活动进行特征量化, 从而确定潜在的内部威胁, 这些检测方法主要是对用户的权限进行建模, 然后对相关的特征信息进行研究和分析, 从而实现异常的识别。他们都强调访问活动的外部特性。当两个文件具有不同的外部特征但包含相似的内容时, 很难识别所有潜在的内部威胁。文献[97]的系统用于伪装者攻击, 从用户遍历文件系统以及访问文件目录的角度建立行为模型。最终通过朴素贝叶斯与马尔可夫模型对比分析, 证明了其检测系统的有效性。文献[51]从文件访问行为的角度构建了用户和对象之间的二部图, 通过用户、用户组和对象之间的访问关系的偏差来识别异常。文献[98]使用文件系统层次结构来提取用户相关性信息, 但是过于依赖文件系统层次结构, 无法处理动态文件系统, 一个文件可以从一个目录移动到另一个目录。

3) 外设使用检测

IO 设备检测主要是研究用户使用计算机外设的行为模式, 主要是鼠标和键盘的使用。文献[99]使用生物认证的方法进行用户验证, 监控用户的鼠标移动, 提取基于角度的度量, 然后使用支持向量机(Support Vector Machines, SVMs)进行精确和快速的分类。文献[100]对鼠标定义了三类基本使用方式, 分别是移动、点击和推拉, 记录这些基本使用方式的坐标、移动距离、角度和速度等特征, 以此构建鼠标行为数据集。文献[101]从用户输入口令中分析用户输入键盘方式的变化, 从数据集中提取了 31 个不同的键盘输入特征。

4) 社交数据检测

这类方法主要是从社交数据中提取特征, 包括电子邮件通信模式和内容、网络在线活动数据等, 以此构建用户行为模式。文献[102]使用社交图发现社交网络活动中的异常, 核心是刻画图的输入、修改和删除等变化。不足是检测效率过低, 在 VAST 数据集上实验表明 1000 多个顶点的图计算耗费了 3 天时间。文献[93]创建用户的对等组, 然后对对等组的用户行为进行建模, 识别偏离社交组的用户。文献[103]通过每个时间实例将用户属性和其社交网络中的用户进行对比。文献[104]在文献[103]的基础上利用跨时间一致性的概念来识别用户属性随时间的异常变化, 有助于识别那些行为在每个时间点看起来都很正常, 但是单独查看却是异常的用户。文献[105]利用来自社交媒体的数据检测用户性能的偏差, 监控业务流程。还有一些研究从社交媒体数据中推断隐藏信息, 比如用户的情绪和心理变化, 以此评估用户的威胁程度。文献[106]使用基于 MOC 模型的心理测量来计算每个用户的威胁分数, NLP 通常用于推断各种心理和情绪指标。

4.3.2 基于误用的检测方法

误用检测和异常检测的不同在于误用检测事先定义了威胁行为, 之后再和威胁行为相匹配, 通常采用的匹配方式是相似性标识。文献[107]利用攻击树的概念进行内部威胁在线的检测和评估, 评估的方式是和系统的最小攻击树进行比较, 该攻击树是根据用户在使用系统前说明的使用意图所生成, 任何偏离其声明意图的行为都被看作攻击触发警报。但是无法枚举攻击方式全集而且存储分析用户意图集会大大地增加系统负担。文献[108]在攻击图中补充用户意图信息, 基于意图信息构建用户的合法元操作集合, 然后生成用户最小攻击树; 通过实时监控用户行为在最小攻击树中的进度判断用户的内部威胁等级。文献[109]在攻击树的基础上提出了用图顶点表示主机或服务器, 每个顶点存储资源信息, 如数据; 边关系表示实体间通信。用户访问行程形成关键序列, 以此计算内部攻击成本。文献[110]将校园环境中的内部威胁行为制定为基础规则, 并以此来检测威胁。文献[111]提出了防火墙语法, 能够捕获组织所采用的策略的抽象, 以及内部错误行为的签名, 当违反策略或匹配到内部不当行为的签名时, 将生成警报。

4.3.3 混合检测方法

也有许多研究同时使用异常检测和误用检测。文献[112]融合了攻击树与行为树提出了活动树模型, 记录用户的工作流模式; 从分支长度、对应节点相似性等方面判断新行为与已有工作流模式的相似性。文献[113]提出了敏感信息传播检测的概念, 用于检测流露到组织外部网络的敏感信息, 将敏感信息检测网络设备放置在网络边缘, 透明的执行三个任务: 从网络分组的有效载荷中识别应用、匹配内容签名以及隐蔽信道的检测。文献[114]提出了基于角色的综合监控方法, 在分离的会话-组织、操作系统、应用程序中监视用户行为。文献[115]提出了分层检测器的框架, 分成策略违反、阈值检测以及模式偏移三个层次, 每层检测到异常都会触发警报, 通过结果反馈实时更新检测模型。随后, 他们又基于(设备-操作-属性)三元组对用户及对应的角色行为进行树结构抽象, 更加全面地刻画用户行为[116]。

4.4 响应

当缓解/预防、检测触发警报后, 通常需要手动分析警报来确定是误报还是实际攻击。然而, 这需要

具备高级安全经验的专家来辨别异常是否为误报, 既加大了运营成本, 又会大大的增加安全事件的响应时间。大多数研究通过提高检测能力来降低误报率, 以此减少响应时间。还有一些研究通过风险评估来改善上述问题[117-119], 这些方法将异常检测模型结合威胁情报、外部系统生成告警的上下文信息, 对每个身份提供整体的风险评分, 辅助事件的进一步调查。文献[117]结合了用户行为检测和数据行为检测, 通过结合事件、行为和实体的风险来评估人员和项目的威胁等级, 并选取风险最高的前几名进行人为干预。文献[118]对用户设立多方面的监测指标, 包括键盘监控指标、屏幕监控指标、移动介质监控指标等, 在时间窗口内采取安全基线的方式进行响应。文献[119]利用行为特征构建异常行为分类器并进行威胁等级判定。另外一些研究通过帮助安全分析员简化警报分析的过程来减少警报分析的时间, 从而减少响应时间[120-124]。文献[121-122]通过计算异常分数中的特征贡献权重来帮助安全操作员简化警报分析的过程。文献[123]提供了用户在一段时间内的事件的统计概要, 一旦用户被标记为异常, 能迅速帮助安全研究员在各个维度对用户进行分析, 快速识别该异常用户是否为内部威胁人员。文献[124]提出将警报进行基于异常和基于无监督的可视化, 这帮助分析员区分用户异常是由于系统导致的还是由于用户自身行为导致的。

4.5 小结

内部威胁防御研究作为热点领域, 取得了显著成果, 大大提升了威胁的发现能力。1) 在威慑方面。研究者已经意识到内部威胁是工作场所的一种不当甚至犯罪行为, 从社会科学理论出发, 在动机上预防用户的恶意行为。2) 在缓解与预防方面。针对不同的信息系统与管理水平, 设置不同的安全策略。例如, 当因为遭受到攻击而关闭系统比接受攻击事后做出回应更具有破坏性的时候, 采用限制攻击影响的方法来缓解威胁。此外, 不仅仅依靠机器学习等数据分析方法, 还结合了基于规则/属性/角色的方法来更有效的应对威胁。3) 在检测方面。从最初的某个命令或访问的异常检测到基于场景或整体行为画像的检测, 检测方法也从基于贝叶斯、隔离森林、SVM 等异常检测算法到使用先进的深度神经网络、循环神经网络算法。此外, 从单一的检测指标到提出分层检测, 结合了安全策略违反、阈值检测和模式偏移这几个不同的层次, 并且通过结果反馈实时更新检测模型。但是本文认为内部威胁防御领域存在的问题主要有以下三点。1) 难以区分异常和恶意。恶意行为大多数表现为异常行为, 但是异常行为不一定是恶意行为。例如, 用户一次无意、偶然的过失被检测系统标记为异常, 但并不是威胁行为。这两者存在交集但不是完全重合。因此, 单纯将内部威胁问题转化为异常检测问题会导致较高的误报和漏报。2) 防御方案单一不全面。现有研究分别在威慑、缓解/预防或者检测方向进行了深入研究, 但是没有意识到分步防御是应对内部威胁的有效方法。3) 大多数研究仅依靠机器学习等数据分析方法来挖掘特征并建立分类器, 会造成分类器偏离实际行为模型。并且, 在选择算法之前缺乏算法选择理论研究, 实际应用中存在盲目性, 这制约了异常检测算法的实际应用。

5 结论

本文使用扎根理论的方法进行严格的文献回顾, 旨在系统化内部威胁研究领域。我们主要进行了两个方面的研究。(1) 内部威胁分析研究。本文首先给出了几种不同的内部人和内部威胁的定义, 还对内部威胁进行了全面的分析, 并基于现有的分类方法提供了一种新的结构化分析与分类法 3W1H。其次, 对现有的公开可用的数据集进行了分析并分类。最后, 不仅对案例分析进行研究, 还从三个方面对内部威胁进行分析, 分别是行为、心理和犯罪学分析, 旨在对内部威胁从动机到行为的整个生命周期进行概括。(2) 内部威胁防御研究。本文基于内部威胁分析确定了一种新的防御框架, 该方案有三道防线, 分别是威慑、预防和检测, 而每一道防线又可以分为几个不同的子类别, 在缓解/预防、检测发出警报后, 进行实时响应。总而言之, 我们的综述不仅注重全景视图下的内部威胁研究, 还遵循从事件到防御方案的方向将内部威胁进行细粒度的分类, 旨在提供系统化、全面化、层次化的内部威胁研究来降低组织的风险。

最后, 我们通过观察和总结, 认为内部威胁领域未来的关键研究方向有以下几点:

(1) 数据集方向: 内部威胁领域的主要挑战在于缺乏真实可用的数据集来评估防御方案, 只有少数合成了攻击的数据集, 而且这些数据集缺乏验证, 无法证明与真实环境的相似性。未来的研究方向主要有两个方面: 1) 数据的全面性, 在创建数据采集环境时, 尽可能的扩展数据采集的层次和范围, 例如, 囊括从系统层到应用层的多层次数据、包含新颖攻击(共谋攻击及其变体)、使用生成对抗网络丰富数据集等。2) 数据的隐私性, 数据集收集的过程中会不可避免的涉及到隐私问题, 设计隐私过滤保护机

制, 在保证数据全面性的基础上保护隐私。考虑到这两个方面的数据集能使防御方案的测试更具有现实性和挑战性。

(2) 事件分析方向: 事件分析的研究方向主要是特征关联。研究基于动机的特征(心理、社交、犯罪学影响)和基于行为的特征的结合, 例如从犯罪学和心理学出发, 建立用户的个体行为模式和社交网络行为模式, 在此基础上对用户的心理变化进行研究等。

(3) 防御方向: 防御方案研究方向主要是设计一个能为内部威胁提供全方位防护的解决方案, 我们认为一个强大的内部威胁防御方案应当是几种独立解决方案的组合: 1) 在第一道防线上将社会科学理论应用于内部威胁的通用脚本框架; 2) 在第二道防线上研究用于事前预防的防御机制或者研究限制攻击影响的方法来缓解恶意内部人员的攻击, 例如, 攻击图与攻击树结合构建用户意图、设置应用最小特权模式; 3) 在第三道防线上提高检测方法的精确度, 降低误报和漏报, 例如, 关联基于人格和基于行为的特征、使用先进的神经网络算法/克隆技术/人工智能算法; 4) 关联后两道防线的警报, 例如, 使用基于规则的预防方法产生签名, 随后使用基于学习的方法进行异常检测, 在这过程中使用异常检测的结果对预防方法进行及时的进行反馈更新; 5) 在响应阶段, 减少运行成本和响应时间, 例如, 研究异常检测分数的可解释性, 帮助安全操作员减少警报分析的时间。

参考文献

[1] Insider threat 2018. Cyber security Insiders and Crowd Research Partners. https://crowdresearchpartners.com/. 2018.

[2] SEI Cyber Minute: Insider Threats. R. F. Trzeciak. http:// resources.sei.cmu.edu/library/asset-view.cfm?assetid=496626. February 7, 2019.

[3] Cappelli D, Moore A, Trzeciak R. The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes (theft, sabotage, fraud)[M]. Upper Saddle River, NJ: Addison-Wesley, 2012

[4] Yang G, Ma J G, Yu A M, et al. Survey of Insider Threat Detection[J]. *Journal of Cyber Security*, 2016, 1(3): 21-36.
(杨光, 马建刚, 于爱民, 等. 内部威胁检测研究[J]. *信息安全学报*, 2016, 1(3): 21-36.)

[5] Bertacchini M, Fierens P. A Survey on Masquerader Detection Approaches[C]. *Proceedings of V Congreso Iberoamericano de Seguridad Informática, Universidad de la República de Uruguay*. 2008: 46-60.

[6] Ben Salem M, Hershkop S, Stolfo S J. A Survey of Insider Attack Detection Research[M]. Insider Attack and Cyber Security. Boston, MA: Springer, 2008: 69-90.

[7] Gheyas I A, Abdallah A E. Detection and Prediction of Insider Threats to Cyber Security: A Systematic Literature Review and Meta-Analysis[J]. *Big Data Analytics*, 2016, 1(1): 6.

[8] Sanzgiri A, Dasgupta D. Classification of Insider Threat Detection Techniques[C]. *The 11th Annual Cyber and Information Security Research Conference*, 2016: 1-4.

[9] Krawczyk B. Learning from Imbalanced Data: Open Challenges and Future Directions[J]. *Progress in Artificial Intelligence*, 2016, 5(4): 221-232.

[10] Ophoff J, Jensen A, Sanderson-Smith J, et al. A Descriptive Literature Review and Classification of Insider Threat Research[C]. *The 2014 InSITE Conference*, 2014: 211-223.

[11] Schultz E E. A Framework for Understanding and Predicting Insider Attacks[J]. *Computers & Security*, 2002, 21(6): 526-531.

[12] Pfleeger S L, Predd J B, Hunker J, et al. Insiders Behaving Badly: Addressing Bad Actors and Their Actions[J]. *IEEE Transactions on Information Forensics and Security*, 2010, 5(1): 169-179.

[13] Christian W. Probst, Rene Rydhof Hansen, Flemming Nielson. Where Can an Insider Attack?[C]. *Proceedings of USENIX Conference on File and Storage Technologies*, 2006: 127-142.

[14] Christian W. Probst, Jeffrey Hunker, Dieter Gollmann. Countering Insider Threats[C]. *Dagstuhl Seminar Proceedings 08302,* 2008: 1-18.

[15] Bishop M, Engle S, Frincke D A, et al. A Risk Management Approach to the "Insider Threat"[M]. Insider Threats in Cyber Security. Boston, MA: Springer, 2010: 115-137.

[16] Roy Sarkar K. Assessing Insider Threats to Information Security Using Technical, Behavioural and Organisational Measures[J]. *Information Security Technical Report*, 2010, 15(3): 112-133.

[17] Theoharidou M, Kokolakis S, Karyda M, et al. The Insider Threat to Information Systems and the Effectiveness of ISO17799[J]. *Computers & Security*, 2005, 24(6): 472-484.

[18] Anderson J P. Computer Security Threat Monitoring and Surveillance. Technical report, James P. Anderson Co, 2010.

[19] Wang G F, Liu C Y, Pan H Z, et al. Survey on Insider Threats to Cloud Computing[J]. *Chinese Journal of Computers*, 2017, 40(2): 296-316.
(王国峰, 刘川意, 潘鹤中, 等. 云计算模式内部威胁综述[J]. *计算机学报*, 2017, 40(2): 296-316.)

[20] Cole E, Ring S. Insider Threat Protecting the Enterprise From Sabotage, Spying, and Theft[M]. Rockland, Mass.: Syngress, 2006.

[21] Phyo A H, Furnell S M. A detection-oriented classification of insider IT misuse[C]. *The 3rd Security Conference*, 2009, 21(01).

[22] Hunker J, Probst C. Insiders and Insider Threats - An Overview of Definitions and Mitigation Techniques[J]. *J Wirel Mob Networks Ubiquitous Comput Dependable Appl*, 2011, 2: 4-27.

[23] Myers J, Grimaila M R, Mills R F. Towards Insider Threat Detection Using Web Server Logs[C]. *The 5th Annual Workshop on Cyber Security and Information Intelligence Research*: *Cyber Security and Information Intelligence Challenges and Strategies*, 2009: 1-4.

[24] Greitzer F L, Strozer J, Cohen S, et al. Unintentional Insider Threat: Contributing Factors, Observables, and Mitigation Strategies[C]. *2014 47th Hawaii International Conference on System Sciences*,

2014: 2025-2034.

[25] Bishop M, Gates C. Defining the Insider Threat[C]. *The 4th annual workshop on Cyber security and information intelligence research*: *developing strategies to meet the cyber security and information intelligence challenges ahead*, 2008: 1-3.

[26] Zhang X Y, Zeng H S, Jia L. Research of Intrusion Detection System Dataset-KDD CUP99[J]. *Computer Engineering and Design*, 2010, 31(22): 4809-4812, 4816.
(张新有, 曾华燊, 贾磊. 入侵检测数据集 KDD CUP99 研究[J]. *计算机工程与设计*, 2010, 31(22): 4809-4812, 4816.)

[27] DuMouchel W, Ju W H, Karr A F, et al. Computer Intrusion: Detecting Masquerades[J]. *Statistical Science*, 2001, 16(1): 1-17.

[28] Posadas R, Mex-Perera C, Monroy R, et al. Hybrid Method for Detecting Masqueraders Using Session Folding and Hidden Markov Models[C]. *The 5th Mexican international conference on Artificial Intelligence*, 2006: 622-631.

[29] Camiña J B, Hernández-Gracidas C, Monroy R, et al. The *Windows-Users and -Intruder Simulations Logs* Dataset (WUIL): An Experimental Framework for Masquerade Detection Mechanisms[J]. *Expert Systems With Applications*, 2014, 41(3): 919-930.

[30] Enron Email Dataset. CALO Project. http://www.cs.cmu.edu/ enron/, February 7, 2015

[31] Wang J R, Cai L J, Yu A M, et al. TempatMDS: A Masquerade Detection System Based on Temporal and Spatial Analysis of File Access Records[C]. *2018 17th IEEE International Conference on Trust*, *Security and Privacy In Computing and Communications/ 12th IEEE International Conference on Big Data Science and Engineering*, 2018: 360-371.

[32] Glasser J, Lindauer B. Bridging the Gap: A Pragmatic Approach to Generating Insider Threat Data[C]. *2013 IEEE Security and Privacy Workshops*, 2013: 98-104.

[33] Harilal A, Toffalini F, Castellanos J, et al. TWOS: A Dataset of Malicious Insider Threat Behavior Based on a Gamified Competition[C]. *The 2017 International Workshop on Managing Insider Security Threats*, 2017: 54-85.

[34] Chinchani R, Muthukrishnan A, Chandrasekaran M, et al. RACOON: Rapidly Generating User Command Data for Anomaly Detection from Customizable Template[C]. *20th Annual Computer Security Applications Conference*, 2005: 189-202.

[35] Saul Greenberg. Using UNIX: Collected Traces of 168 Users. Technical Report. Department of Computer Science, University of Calgary, 1988.

[36] Lane T, Brodley C E. An Application of Machine Learning to Anomaly Detection[C]. *In Proceedings of the National Information Systems Security Conference,* 1997: 366-380.

[37] Collins M L, Theis M C, Trzeciak R F, et al. Common Sense Guide to Prevention and Detection of Insider Threats[M]. CERT, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, 2016.

[38] Randazzo M R, Keeney M, Kowalski E, et al. Insider Threat Study: Illicit Cyber Activity in the Banking and Finance Sector. Technical Report. CERT Coordination Center, Software Engineering Institute, Carnegie, Mellon University, Pittsburgh, PA, 2005.

[39] Robert W, Merrill W. Beyond deterrence: an expanded view of employee computer abuse[J]. *Society for Information Management and the Management Information Systems Research Center*, 2013: 1-20.

[40] William R Claycomb, Carly L Huth, Lori Flynn, et al. Chrono-Logical Examination of Insider Threat Sabotage: Preliminary Observations[J]. *Journal of Wireless Mobile Networks*. Ubiquitous Computing, and Dependable Applications 3, 4 (2012), 4-20.

[41] Nurse J R C, Buckley O, Legg P A, et al. Understanding Insider Threat: A Framework for Characterising Attacks[C]. *2014 IEEE Security and Privacy Workshops*, 2014: 214-228.

[42] Maasberg M, Warren J, Beebe N L. The Dark Side of the Insider: Detecting the Insider Threat through Examination of Dark Triad Personality Traits[C]. *2015 48th Hawaii International Conference on System Sciences*, 2015: 3518-3526.

[43] Chen T L, Kammüller F, Nemli I, et al. A Probabilistic Analysis Framework for Malicious Insider Threats[C]. *The Third International Conference on Human Aspects of Information Security*, *Privacy*, *and Trust - Volume 9190*, 2015: 178-189.

[44] Eric S, Keli R, Post J. The insider threat to information systems: The psychology of the dangerous insider[J]. *Security Awareness Bulletin*, 1998, 2(98): 1-10.

[45] The insider threat, an introduction to detecting and deterring insider spy. US. Department of Justice and Federal Bureau of Investigation, https://www.fbi.gov/about-us/investigate/counterintelligence/the-insider-threat/, 2012.

[46] Leach J. Improving User Security Behaviour[J]. *Computers & Security*, 2003, 22(8): 685-692.

[47] Willison R, Warkentin M. Motivations for Employee Computer Crime: Understanding and Addressing Workplace Disgruntlement through the Application of Organizational Justice[C]. *In Proceedings of the International Workshop on Information Systems Security Research*, 2009: 127-144.

[48] Ho S M, Hancock J T, Booth C, et al. Demystifying Insider Threat: Language-Action Cues in Group Dynamics[C]. *2016 49th Hawaii International Conference on System Sciences*, 2016: 2729-2738.

[49] Brdiczka O, Liu J, Price B, et al. Proactive Insider Threat Detection through Graph Learning and Psychological Context[C]. *2012 IEEE Symposium on Security and Privacy Workshops*, 2012: 142-149.

[50] Chen Y, Nyemba S, Malin B. Detecting Anomalous Insiders in Collaborative Information Systems[J]. *IEEE Transactions on Dependable and Secure Computing*, 2012, 9(3): 332-344.

[51] Kandias M, Galbogini K, Mitrou L, et al. Insiders Trapped in the Mirror Reveal Themselves in Social Media[C]. *International Conference on Network and System Security*. Berlin, Heidelberg: Springer, 2013: 220-235.

[52] Brown C R, Watkins A, Greitzer F L. Predicting Insider Threat Risks through Linguistic Analysis of Electronic Communication[C]. *2013 46th Hawaii International Conference on System Sciences*, 2013: 1849-1858.

[53] Alahmadi B A, Legg P A, Nurse J R C. Using Internet Activity Profiling for Insider-Threat Detection[C]. *The 17th International Conference on Enterprise Information Systems - Volume 2*, 2015: 709-720.

[54] Shropshire J, Warkentin M, Sharma S. Personality, Attitudes, and Intentions: Predicting Initial Adoption of Information Security Behavior[J]. *Computers & Security*, 2015, 49: 177-191.

[55] Padayachee K. Taxonomy of Compliant Information Security Behavior[J]. *Computers & Security*, 2012, 31(5): 673-680.

[56] Goo J, Yim M S, Kim D J. A Path to Successful Management of Employee Security Compliance: An Empirical Study of Information Security Climate[J]. *IEEE Transactions on Professional Communication*, 2014, 57(4): 286-308.

[57] Son J Y. Out of Fear or Desire? Toward a Better Understanding of Employees' Motivation to Follow IS Security Policies[J]. *Information & Management*, 2011, 48(7): 296-302.

[58] Levan K, MacKey D A. Prevention of Crime and Delinquency[M]. International Encyclopedia of the Social & Behavioral Sciences. Amsterdam: Elsevier, 2015: 877-882.

[59] Ifinedo P. Understanding Information Systems Security Policy Compliance: An Integration of the Theory of Planned Behavior and the Protection Motivation Theory[J]. *Computers & Security*, 2012, 31(1): 83-95.

[60] Choi S, Martins J T, Bernik I. Information Security: Listening to the Perspective of Organisational Insiders[J]. *Journal of Information Science*, 2018, 44(6): 752-767.

[61] Safa N S, Maple C, Furnell S, et al. Deterrence and Prevention-Based Model to Mitigate Information Security Insider Threats in Organisations[J]. *Future Generation Computer Systems*, 2019, 97: 587-597.

[62] Lee J, Lee Y. A Holistic Model of Computer Abuse within Organizations[J]. *Information Management & Computer Security*, 2002, 10(2): 57-63.

[63] Willison R, Siponen M T. Overcoming the Insider: Reducing Employee Computer Crime through Situational Crime Prevention[J]. *Commun ACM*, 2009, 52(9): 133-137.

[64] Steele S, Wargo C. An Introduction to Insider Threat Management[J]. *Information Systems Security*, 2007, 16(03): 23-33.

[65] Viduto V, Maple C, Huang W. An Analytical Evaluation of Network Security Modelling Techniques Applied to Manage Threats[C]. *2010 International Conference on Broadband*, *Wireless Computing*, *Communication and Applications*, 2010: 117-123.

[66] Jabbour G, Menasce D A. The Insider Threat Security Architecture: A Framework for an Integrated, Inseparable, and Uninterrupted Self-Protection Mechanism[C]. *2009 International Conference on Computational Science and Engineering*, 2009: 244-251.

[67] Shabtai A, Elovici Y, Rokach L. Data Leakage Detection/Prevention Solutions[M]. A Survey of Data Leakage Detection and Prevention Solutions. Boston, MA: Springer, 2012: 17-37.

[68] Shapira Y, Shapira B, Shabtai A. Content-Based Data Leakage Detection Using Extended Fingerprinting[EB/OL]. 2013: arXiv: 1302.2028. https://arxiv.org/abs/1302.2028

[69] Chagarlamudi M, Panda B, Hu Y. Insider Threat in Database Systems: Preventing Malicious Users' Activities in Databases[C]. *The 2009 Sixth International Conference on Information Technology*: *New Generations*, 2009: 1616-1620.

[70] Yaseen Q, Panda B. Enhanced Insider Threat Detection Model that Increases Data Availability[C]. *The 7th international conference on Distributed computing and internet technology*, 2011: 267-277.

[71] Baracaldo N, Joshi J. A Trust-and-Risk Aware RBAC Framework: Tackling Insider Threat[C]. *The 17th ACM symposium on Access Control Models and Technologies*, 2012: 167-176.

[72] Bishop M, Engle S, Peisert S, et al. We Have Met the Enemy and He is us[C]. *The 2008 New Security Paradigms Workshop*, 2008: 1-12.

[73] Desmedt Y, Shaghaghi A. Function-Based Access Control (FBAC): From Access Control Matrix to Access Control Tensor[C]. *The 8th ACM CCS International Workshop on Managing Insider Security Threats*, 2016: 89-92.

[74] Shalev N, Keidar I, Moatti Y, et al. WatchIT: Who Watches your IT Guy? [C]. *The 8th ACM CCS International Workshop on Managing Insider Security Threats*, 2016: 93-96.

[75] OASIS XACML Technical Committee. eXtensible Access Control Markup Language (XACML) Version 3.0. Technical Report, OASIS Standard, 2013.

[76] Gessiou E, Vu Q H, Ioannidis S. IRILD: An Information Retrieval Based Method for Information Leak Detection[C]. *2011 Seventh European Conference on Computer Network Defense*, 2012: 33-40.

[77] Gómez-Hidalgo J M, Martín-Abreu J M, Nieves J, et al. Data Leak Prevention through Named Entity Recognition[C]. *2010 IEEE Second International Conference on Social Computing*, 2010: 1129-1134.

[78] Roesch M. Snort - Lightweight Intrusion Detection for Networks[C]. *LISA. USENIX Association*, 1999: 229-238.

[79] Fonseca J, Vieira M, Madeira H. Integrated Intrusion Detection in Databases[C]. *Latin-American Symposium on Dependable Computing*. Berlin, Heidelberg: Springer, 2007: 198-211.

[80] Kamra A, Terzi E, Bertino E. Detecting Anomalous Access Patterns in Relational Databases[J]. *The VLDB Journal*, 2008, 17(5): 1063-1077.

[81] Wu G Z, Osborn S L, Jin X. Database Intrusion Detection Using Role Profiling with Role Hierarchy[C]. *Workshop on Secure Data Management*. Berlin, Heidelberg: Springer, 2009: 33-48.

[82] Hwang K, Cai M, Chen Y, et al. Hybrid Intrusion Detection with Weighted Signature Generation over Anomalous Internet Episodes[J]. *IEEE Transactions on Dependable and Secure Computing*, 2007, 4(1): 41-55.

[83] Costante E, Fauri D, Etalle S, et al. A Hybrid Framework for Data Loss Prevention and Detection[C]. *2016 IEEE Security and Privacy Workshops*, 2016: 324-333.

[84] Spitzner L. Honeypots: Catching the Insider Threat[C]. *19th Annual Computer Security Applications Conference*, *2003*. *Proceedings*, 2004: 170-179.

[85] Bowen B, Ben Salem M, Hershkop S, et al. Designing Host and Network Sensors to Mitigate the Insider Threat[J]. *IEEE Security & Privacy*, 2009, 7(6): 22-29.

[86] Kandias M, Mylonas A, Virvilis N, et al. An Insider Threat Prediction Model[C]. *International Conference on Trust*, Privacy and Security in Digital Business. Berlin, Heidelberg: Springer, 2010: 26-37.

[87] Mark M, Penny C, Brant C, et al. Analysis and detection of malicious insiders[C]. *International Conference on Intelligence Analy-*

sis, 2005: 1-7.

[88] Brian D D, Haym H. Predicting sequences of user actions[C]. *AAAI/ICML 1998 Workshop on Predicting the Future AI Approaches to Time series Analysis*, 1998: 5-12.

[89] Ryan J, Lin M J, Miikkulainen R. Intrusion Detection with Neural Networks[C]. *The 10th International Conference on Neural Information Processing Systems*, 1997: 943-949.

[90] Mizuki O K, Yoshihiro O, Kazuhiko K. Eigen co-occurrence matrix method for masquerade detection[J]. *Proceedings of the 7th JSSST SIGSYS Workshop on Systems for Programming and Applications (SPA)*, 2004: 1-5.

[91] Dumouchel W. Computer intrusion detection based on Bayes factors for comparing command transition probabilities. Technical report 1999, National Institute of Statistical Sciences.

[92] Ju W H, Vardi Y. A Hybrid High-Order Markov Chain Model for Computer Intrusion Detection[J]. *Journal of Computational and Graphical Statistics*, 2001, 10(2): 277-295.

[93] Eldardiry H, Bart E, Liu J, et al. Multi-Domain Information Fusion for Insider Threat Detection[C]. *2013 IEEE Security and Privacy Workshops*, 2013: 45-51.

[94] Maloof M A, Stephens G D. elicit: A System for Detecting Insiders Who Violate Need-to-Know[C]. *International Workshop on Recent Advances in Intrusion Detection*. Berlin, Heidelberg: Springer, 2007: 146-166.

[95] Senator T E, Goldberg H G, Memory A, et al. Detecting Insider Threats in a Real Corporate Database of Computer Usage Activity[C]. *The 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013: 1393-1401.

[96] Zhang R, Chen X J, Shi J Q, et al. Detecting Insider Threat Based on Document Access Behavior Analysis[C]. *Asia-pacific Web Conference,* 2014: 376-387.

[97] Camiña J B, Rodríguez J, Monroy R. Towards a Masquerade Detection System Based on User's Tasks[C]. *International Workshop on Recent Advances in Intrusion Detection*. Cham: Springer, 2014: 447-465.

[98] Gates C, Li N H, Xu Z L, et al. Detecting Insider Information Theft Using Features from File Access Logs[C]. *Computer Security - ESORICS 2014*,: 383-400.

[99] Zheng N, Paloski A, Wang H N. An Efficient User Verification System via Mouse Movements[C]. *The 18th ACM conference on Computer and communications security*, 2011: 139-150.

[100] Adam W, Anil R, Pranav S, et al. Mouse Movements Biometric Identification: A Feasibility Study[C]. *Student/Faculty Research Day, CSIS*, 2007: 1-8.

[101] Killourhy K, Maxion R. Why Did My Detector Do That? ![C]. *International Workshop on Recent Advances in Intrusion Detection*. Berlin, Heidelberg: Springer, 2010: 256-276.

[102] Eberle W, Holder L. Insider Threat Detection Using Graph-Based Approaches[C]. *2009 Cybersecurity Applications & Technology Conference for Homeland Security*, 2009: 237-241.

[103] Young W T, Memory A, Goldberg H G, et al. Detecting Unknown Insider Threat Scenarios[C]. *2014 IEEE Security and Privacy Workshops*, 2014: 277-288.

[104] Gavai G, Sricharan K, Gunning D, et al. Detecting Insider Threat from Enterprise Social and Online Activity Data[C]. *The 7th ACM CCS International Workshop on Managing Insider Security Threats*, 2015: 13-20.

[105] Ritzalis D, Stavrou V, Kandias M, et al. Insider Threat: Enhancing BPM through Social Media[C]. *2014 6th International Conference on New Technologies*, *Mobility and Security*, 2014: 1-6.

[106] Raskin V, Taylor J M, Hempelmann C F. Ontological Semantic Technology for Detecting Insider Threat and Social Engineering[C]. *The 2010 New Security Paradigms Workshop*, 2010: 115-128.

[107] Ray I, Poolsapassit N. Using Attack Trees to Identify Malicious Attacks from Authorized Insiders[C]. *The 10th European conference on Research in Computer Security*, 2005: 231-246.

[108] Wang H, Liu S F. A Scalable Predicting Model for Insider Threat[J]. *Chinese Journal of Computers*, 2006, 29(8): 1346-1355.
(王辉, 刘淑芬. 一种可扩展的内部威胁预测模型[J]. *计算机学报*, 2006, 29(8): 1346-1355.)

[109] Chinchani R, Iyer A, Ngo H Q, et al. Towards a Theory of Insider Threat Assessment[C]. *2005 International Conference on Dependable Systems and Networks*, 2005: 108-117.

[110] Bhilare D S, Ramani A K, Tanwani S K. Protecting Intellectual Property and Sensitive Information in Academic Campuses from Trusted Insiders: Leveraging Active Directory[C]. *The 37th annual ACM SIGUCCS fall conference*: *communication and collaboration*, 2009: 99-104.

[111] Agrafiotis I, Erola A, Goldsmith M, et al. A Tripwire Grammar for Insider Threat Detection[C]. *The 8th ACM CCS International Workshop on Managing Insider Security Threats*, 2016: 105-108.

[112] Agrafiotis I, Legg P A, Goldsmith M, et al. Towards a User and Role-Based Sequential Behavioural Analysis Tool for Insider Threat Detection[J]. *J Internet Serv Inf Secur*, 2014, 4(4): 127-137.

[113] Liu Y L, Corbett C, Ken C A, et al. SIDD: A Framework for Detecting Sensitive Data Exfiltration by an Insider Attack[C]. *2009 42nd Hawaii International Conference on System Sciences*, 2009: 1-10.

[114] Park J S, Ho S M. Composite Role-Based Monitoring (CRBM) for Countering Insider Threats[C]. *International Conference on Intelligence and Security Informatics*. Berlin, Heidelberg: Springer, 2004: 201-213.

[115] Philip L, Nick M, Jason R C N. Towards a conceptual model and reasoning structure for insider threat detection[J]. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications,* 2013, 4: 20-37.

[116] Legg P A, Buckley O, Goldsmith M, et al. Automated Insider Threat Detection System Using User and Role-Based Profile Assessment[J]. *IEEE Systems Journal*, 2017, 11(2): 503-512.

[117] Warren M. Modern IP Theft and the Insider Threat[J]. *Computer Fraud & Security*, 2015, 2015(6): 5-10.

[118] Moore A P, Hanley M, Mundie D. A Pattern for Increased Monitoring for Intellectual Property Theft by Departing Insiders[C]. *The 18th Conference on Pattern Languages of Programs*, 2011: 1-10.

[119] Thompson, Salvatore Joseph, Keromytis, et al. Anomaly Detection at Multiple Scales (ADAMS) Broad Agency Announcement DARPA-BAA-11-04. General Services Administration. Retrieved

2011-12-05.

[120] Virvilis N, Vanautgaerden B, Serrano O S. Changing the Game: The Art of Deceiving Sophisticated Attackers[C]. *2014 6th International Conference on Cyber Conflict* (*CyCon 2014*), 2014: 87-97.

[121] Tuor A, Kaplan S, Hutchinson B, et al. Deep Learning for Unsupervised Insider Threat Detection in Structured Cybersecurity Data Streams[EB/OL]. 2017: arXiv: 1710.00811. https://arxiv.org/abs/1710.00811

[122] Brown A, Tuor A, Hutchinson B, et al. Recurrent Neural Network Attention Mechanisms for Interpretable System Log Anomaly Detection[C]. *The First Workshop on Machine Learning for Computing Systems*, 2018: 1-8.

[123] Chris B. Security Information and Event Management (SIEM) Implementation[M]. Mcgraw-Hill Publ.Comp, 2010.

[124] Colombe J B, Stephens G. Statistical Profiling and Visualization for Detection of Malicious Insider Attacks on Computer Networks[C]. *The 2004 ACM workshop on Visualization and data mining for computer security*, 2004: 138-142.

孙德刚 中国科学院信息工程研究所研究员、博士生导师。中国科学院大学网络空间安全学院教授。主要研究方向为电磁泄漏防护, 无线通信技术以及高安全等级的信息系统防护技术。Email: sundegang@iie.ac.cn

刘美辰 于 2017 年在中国地质大学(北京)电子信息工程专业获得学士学位。现在中国科学院信息工程研究所信号与信息处理专业攻读博士学位。研究领域为网络安全、内部威胁检测。Email: liumeichen@iie.ac.cn

李梅梅 于 2007 年在北京大学(北京)获得工学硕士学位。现任中国科学院信息工程研究所硕士生导师。研究领域为网络安全、内部威胁检测、信息保密技术、数据分析等。Email: limeimei@iie.ac.cn

刘鹏程 于 2012 年在华中农业大学计算机科学与技术专业获得学士学位。现在中国科学院信息工程研究所网络空间安全专业攻读博士学位。 研究领域为海量数据检索、异常检测。研究兴趣包括: 内部威胁检测、网络安全、大数据分析。Email: liupengcheng@iie.ac.cn

王旭 于 2017 年在兰州大学信息安全专业获得工学学士学位。现在中国科学院大学网络空间安全专业攻读工学博士学位。研究领域为内部威胁检测, 伪装者检测。研究兴趣包括: 异常检测、用户行为画像、机器学习。Email: wangxu1996@iie.ac.cn

石志鑫 1986 年生, 博士, 现任中国科学院信息工程研究所高级工程师, 研究领域为特定应用场景大数据挖掘分析、智能信息处理威胁检测等。Email: shizhixin@iie.ac.cn

李楠 于 2014 年在北京航空航天大学计算机科学与技术专业获得硕士学位。现任中国科学院信息工程研究所工程师, 研究领域为网络空间安全。研究兴趣包括安全大数据分析、入侵检测、虚拟化安全等。Email: linan@iie.ac.cn