中国科学: 信息科学 2024 年 第 54 卷 第 4 期: 817–852
SCIENTIA SINICA Informationis
《中国科学》杂志社
SCIENCE CHINA PRESS
评述

虚实融合网络空间安全综述

赵沁平, 周忠*, 梁晓辉, 李帅, 汪淼, 王焱

中关村实验室, 北京 100094
* 通信作者. E-mail: zz@buaa.edu.cn

收稿日期: 2023–06–22; 修回日期: 2023–08–20; 接受日期: 2023–12–15; 网络出版日期: 2024–04–09

国家自然科学基金 (批准号: 62272018, 62272019, 62372025) 资助项目

摘要 在计算机与网络基础设施不断发展的推动下, 越来越多的人类活动从物理世界向数字世界迁移, 产生了构建新型虚实融合网络空间的动因和思想, 增强现实、数字孪生、元宇宙等相继成为国际关注热点. 虚实融合网络以互联网、物联网为基础, 进一步将具有独立身份的计算机、各种物理对象及其数字孪生, 以及计算机生成的数字原生对象进行互联, 将物理世界和人类世界与数字世界贯通, 成为“泛联网”, 形成人、机、物泛联互通的虚实融合网络空间, 带来全新的大众体验、社交形态、生产模式和数字经济发展路径. 这种新型网络空间极大地拓展了互联网、物联网的空间边界和应用领域, 同时也带来了新的安全与隐私保护问题. 本文首先介绍了泛联网与虚实融合网络空间的概念及架构, 分析其存在的安全与隐私风险, 然后从用户认证与权限控制、数据安全、隐私保护、感知与交互安全、关键基础设施与软硬件安全、应用安全与网络空间治理等方面的国际研究现状和发展趋势进行综述, 最后给出需要解决的十个问题.

关键词 虚实融合网络空间, 泛联网, 数字孪生, 安全, 隐私

1 引言

随着计算机与互联网的不断升级换代, 人类社会在过去的 30 多年, 通过信息化和网络化变革传统运行模式, 重塑社会各行业和人类活动, 不断创新发展路径, 实现持续高速发展. 当前, 世界主要经济体的信息化已经较为普及, 数字世界日益膨胀壮大, 数字经济形成规模, 成为改变全球竞争格局的关键力量.

虚拟现实 (virtual reality, VR) 概念 [1] 的出现, 使得在物理世界与数字世界共享生存并自由穿梭成为人类的追求目标. 众多研究者在虚拟现实、增强现实 (augmented reality, AR)、混合现实 (mixed

引用格式: 赵沁平, 周忠, 梁晓辉, 等. 虚实融合网络空间安全综述. 中国科学: 信息科学, 2024, 54: 817–852, doi: 10.1360/SSI-2023-0188
Zhao Q P, Zhou Z, Liang X H, et al. Security in virtual-real mixing cyberspaces: a survey (in Chinese). Sci Sin Inform, 2024, 54: 817–852, doi: 10.1360/SSI-2023-0188

reality, MR)、远程呈现 (telepresence)、信息空间 (cyberspace, 也称赛博空间)、信息 – 物理系统 (cyber-physical system, CPS) 等技术领域 [2] 进行了长期探索[1)], 取得显著的技术进步和应用成效, 并开始形成相关新兴产业, 但距离构建可以普遍使用、高度社会化的数字世界还有很大差距. 近年来, 随着虚拟现实、人工智能、物联网、5G/6G、区块链、新型网络等技术的快速发展, 出现了构建虚实融合网络, 将物理世界和人类世界与数字世界贯通融合, 形成一种人、机、物三元混合的虚实融合泛联空间的思想, 为人类构建去中心、可构想、易组装、高度社会化的数字世界提供了可行的技术路径, 带来了全新的大众体验、社交形态、生产模式和数字经济发展路径, 极大地拓展了当前互联网、物联网支撑人类社会和生产力发展的边界, 被普遍寄予期望.

虚实融合网络, 我们称其为泛联网, 是以互联网为基础, 通过多种传感和信息交互设备, 按照有关协议, 将具有独立身份的计算机、各种真实物理对象及其数字孪生 (digital twins) 对象, 以及计算机生成的数字原生 (digital natives) 对象进行互联, 并能够独立寻址的网络系统, 是贯通融合物理、人类、数字三界, 形成人、机、物泛联互通的虚实融合空间的基础设施, 可称为继计算机网 (连接对象是计算机)、物联网 (连接对象是计算机和物理对象) 之后的第三代互联网. 新型网络、5G/6G、虚拟现实/增强现实/混合现实、人工智能、虚拟化身、数字孪生、区块链等技术是虚实融合网络空间的支撑技术, 近年来火爆的元宇宙可认为是虚实融合网络空间终极应用的愿景. 关于元宇宙已有大量讨论, 包含许多社会、金融、人文、认知方面的内容, 目前对其内涵构成尚未形成共识. 本文论述的虚实融合网络空间主要集中于信息技术范畴.

虚实融合网络空间作为物理、人类、数字三界贯通融合的泛联空间, 也会具有社会空间的关系属性, 因此, 除了传统物理层、网络层的安全问题之外, 必然会带来新的空间安全问题, 其安全风险对象将不限于数字世界, 也将涉及物理世界和人类世界, 例如人体、实物和场景的物理安全、人类社会关系的重塑等. 考虑到未来虚实融合网络可能在云端或边缘侧的数字化基础设施进行集中运维, 大量攻击将很自然地集中到网络应用层, 而且攻击类型将多样化; 另一方面, 由于虚实融合网络空间中有身份、可寻址、能交互的对象包括人及其化身、物理对象及其数字孪生对象, 以及计算机生成的数字原生对象等, 交互粒度更细、更多样, 因此虚实融合网络空间先天要求获取比现有互联网应用更多样化的隐私数据、虚拟场景等附带的地理信息数据, 还可能涉及国家安全等问题, 更容易产生严重的隐私泄露风险.

虚实融合网络空间是一个具有前瞻性的交叉研究领域, 国际上, 包括作者团队, 在虚拟现实、数字孪生、互联网、元宇宙等方向发表了一些文章 [1~5], 2021~2022 年左右元宇宙成为热点还引发出了大量的在线开放访问文章, 从不同角度提出观点, 促进了相关领域的思考. 本文的主要贡献包括: 在分析数字孪生、元宇宙等新型互联网应用的基础上, 进行归纳概括, 提出了更能体现核心技术和具有普遍意义的 “泛联网” 与虚实融合网络空间的概念及其技术架构, 分析了其带来的新的安全与隐私风险, 讨论了相关风险产生的原因; 从用户认证与权限控制、数据安全、隐私保护、感知与交互安全、关键基础设施与软硬件安全、应用安全与网络空间治理等方面, 综述相关的安全与隐私风险, 以及防御策略和技术的最新研究进展, 最后给出虚实融合网络空间中 10 个需要解决的开放性安全技术问题或未来研究方向.

1) 赵沁平. 发展数字孪生技术, 推动 VR 深度应用. 中国电子报. 第 87 期第 1 版, 2022.12.1.

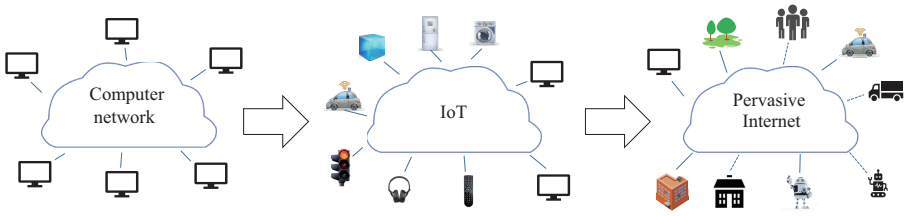

图 1 (网络版彩图) 计算机网 – 物联网 – 泛联网的连接对象

Figure 1 (Color online) Interconnection nodes of computer network, IoT, and Pervasive Internet

2 “泛联网”与虚实融合网络空间及其安全问题

2.1 “泛联网”与虚实融合网络空间

虚实融合网络空间是一种新型事物, 是数字世界不断扩张, 并与物理世界和人类世界逐步融合过程中形成的. 数字世界与物理世界融合, 历经完全与物理世界平行而独立存在的虚拟现实和数字原生, 虚拟对象在物理对象或现实环境上叠加显示的增强现实, 虚拟对象和其物理对象连通并同步镜像演化的数字孪生 [3], 发展到数字孪生、数字原生二者完全融合的数实共生等几个阶段 [4,5].

在数字世界和物理世界逐步融合发展过程中, 以社会关系网为主要特征的人类世界一直渗透其中. 人的本体存在于物理世界, 但由于人类及其活动具有不同于物理世界中其他对象的独特性和社会性, 人类创造出来的物理对象和数字对象也都具有社会空间属性, 从而构成了与物理世界不同的人类世界. 伴随着数字世界与物理世界的融合, 在数字空间中形成了现实世界中的物理环境和物理对象的镜像. 同时, 人类通过三维内容制作、人工智能内容生成等技术和工具构造与物理对象不存在映射或关联的数字原生对象, 并以虚拟化身方式融入数字空间, 形成了物理世界、人类世界、数字世界贯通的虚实融合网络空间. 人的化身、数字孪生、数字原生等对象都可以通过人机或其他信息交互手段在虚实融合空间进行交互和响应. 所谓元宇宙就是要建立一个最大化的、相互关联的虚实融合体验世界, 人们可以通过比现有社交平台中更加逼真的化身登录其中进行娱乐、游戏、购物, 从事各种社交活动和生产、工作.

从互联网连接对象, 或可独立寻址对象来说, 第一代互联网是计算机网, 实现了计算机设备的互联, 以 IP (Internet protocol) 地址来唯一标记身份和进行路由转发, 其应用导致了社交网络、电子商务等现代生活消费模式的产生; 第二代互联网是物联网 (Internet of Things, IoT), 连接对象扩展至物理对象, 通过各种传感器和信息交互方式, 将需要连接的物理对象接入互联网, 成为具有独立身份的可寻址对象, 其应用扩展至工业生产和生活服务. 物联网以通用唯一识别码 UUID (universally unique identifier) 等编码规范来标识计算设备、物和人的身份 [6], 我国 “国家物联网标识管理公共服务平台”[2] 作为全国的物联网标识管理节点, 提出以 128 位的 IPv6 专用编码来为基于 IP 的物联网系统提供网络直达的物联网对象标识. 随着数字孪生、元宇宙等新的互联网应用出现, 以及数字经济的兴起, 出现了大量的数字内容, 越来越多的数字对象被创建, 包括物理对象对应的数字孪生对象、计算机生成的数字原生对象、人的虚拟化身等. 这些对象可能分散在不同规模的异质平台上, 也可在不同用户和平台之间进行交易流转, 通过多中心或去中心化的结构进行互联, 将现有的互联网发展成为虚实融合网络, 我们称其为 “泛联网” (Pervasive Internet), 如图 1 所示.

泛联网是虚实融合网络空间的基础设施, 由互联网主干网、无线通信设施, 以及相关协议和基础

2) http://cniotroot.cn.

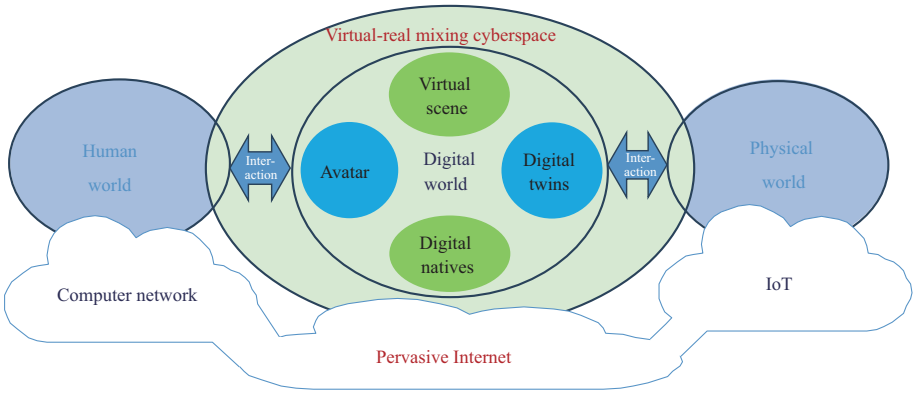

图 2　(网络版彩图) 虚实融合网络空间

Figure 2　(Color online) Virtual-real mixing cyberspace

软件系统构成, 包括了互联网发展所形成并持久运行的各种基础性支撑, 如目前集中式、分布式、层次式的各类互联网服务. 泛联网空间, 即虚实融合网络空间, 则由泛联网及其所有互联对象构成, 包括具有独立身份并能够独立寻址和进行交互的计算机、人及其虚拟化身、各种真实物理对象及其数字孪生对象、数字原生对象、虚拟场景, 以及物理世界中的场景, 如图 2 所示.

虚实融合网络空间的形式化静态模型如下.

设 W 为物理世界物体状态的集合, 将 W 划分为不交子集的集合 T, 以使不同的建模方法可以模型化对应划分中的物理世界状态. M 是人的状态集合, C 是计算机状态的集合, D 是网络地址标识的集合.

首先定义一个选择操作 sel, 它把 W 中的状态映射到它所选择的划分:

$\mathrm{sel} : W \to T$.

$\mathrm{env} : T \to \rho(C)$, 其中 $\rho(C)$ 为 C 的幂集, env 函数把现实世界状态划分映射到计算机状态序列集, 生成数字环境.

$\mathrm{dt} : W \to \rho(C)$, 其中 $\rho^{\mathrm{dt}}(C)$ 表示 dt 函数生成的 C 的子集的集合, 也就是数字孪生对象的集合.

$\mathrm{dn} : C \to \rho(C)$, 其中 $\rho^{\mathrm{dn}}(C)$ 表示 dn 函数生成的 C 的子集的集合, 也就是数字原生对象的集合.

$\mathrm{ava} : M \to \rho(C)$, 其中 $\rho^{\mathrm{ava}}(C)$ 表示 ava 函数生成的 C 的子集的集合, 也就是虚拟化身的集合.

$\mathrm{addr} : \langle \rho^{\mathrm{dt}}(C), \rho^{\mathrm{dn}}(C), \rho^{\mathrm{ava}}(C) \rangle \to D$. addr 函数给每一个独立对象分配地址标识.

定义虚实融合网络空间为十元组: $\langle W, M, C, D, \mathrm{sel}, \mathrm{env}, \mathrm{dt}, \mathrm{dn}, \mathrm{ava}, \mathrm{addr} \rangle$.

虚实融合网络空间中具有独立身份的对象通过泛联网基础设施和传感交互设备进行各种数据、控制信息交互, 实现各类对象的状态和形态变化等动态行为.

2.2　新型网络空间的安全与隐私保护

虚实融合网络空间是在泛联网基础设施之上, 实现物理世界、人类世界和数字世界的贯通融合, 从而支撑数字孪生、元宇宙等新的互联网应用. 虚实融合网络空间的基础技术架构由 XR 引擎/终端、人工智能、数字孪生、区块链 4 种基础技术引擎构成, 如图 3 所示. 其中人工智能支持数字世界中的人类化身和其他数字原生智能体的行为模型构建与运行, 以及高效内容生成、智能检索与资源定位; XR 引擎/终端支撑各种数字原生和数字孪生对象的几何、物理模型构建, 以及各类对象之间, 跨越物理世界和数字世界的自然交互; 数字孪生实现物理世界和数字世界对象的同步演化; 区块链支撑自组

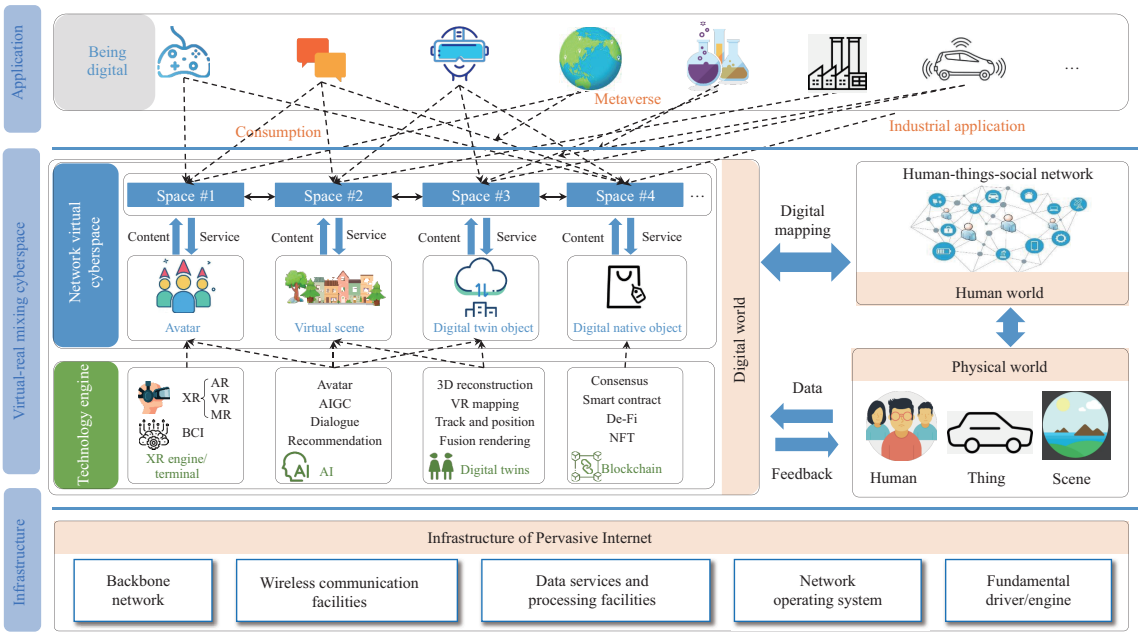

图 3 (网络版彩图) 虚实融合网络空间技术架构

Figure 3 (Color online) Technology architecture of virtual-real mixing cyberspace

织、无中心的可信安全交易. 在技术引擎的基础上, 网络空间将出现虚拟化身、虚拟场景、数字孪生对象、数字原生对象等四类数字主体, 其中虚拟化身是用户在虚拟世界中的数字替身 (avatar), 虚拟场景是计算机生成的、一般不能被用户改变的数字化空间环境, 数字孪生对象是与物理世界存在镜像映射的数字实体, 数字原生对象包括计算机生成的无物理世界镜像对象的智能体.

如上所述, 泛联网将互联网和物联网的连接对象扩展至人的虚拟化身、物理对象的数字孪生对象, 以及各种数字原生对象, 形成虚实融合网络空间, 再次突破传统网络空间的边界, 带来了新的互联网应用领域, 相关理念日益受到发达国家各信息龙头企业和各种组织的重视, 如 Meta、微软、苹果等, 都纷纷部署研发力量, 投入巨资, 开发相关平台、交互装置, 以及进行应用. 与此同时, 由于虚实融合网络空间拥有了多种新的独立身份的对象, 因此不可避免地带来了新的网络空间安全和隐私保护问题. 除了泛联网基础设施层的安全问题之外, 其安全风险对象将从数字世界扩大到物理世界. 在数字世界有人的化身、数字对象、场景的安全风险, 在物理世界还存在人体、物理对象和场景的物理安全等, 攻击类型也将多样化, 出现一些新型攻击. 由于交互粒度更细、更多样, 虚实融合网络空间先天要求获取比现有互联网应用更多的隐私数据, 更容易产生严重的隐私泄露风险.

其次, 虚实融合网络空间模糊了物理和数字世界的边界, 特别是 XR 引擎和终端提供了沉浸式的自然交互环境甚至脑机接口的低消耗、迅敏交互方式, 极度降低了用户的学习门槛, 将很容易提升大众的渗透率与接受度, 推动数字世界中的社会关系发展. 但在去中心化、初期以年轻人为主要参与者的网络空间中进行高度沉浸感的数字化交流、生活和工作, 缺乏物理世界中的实体监管手段和规则, 将冲击现有的文化传统和国家治理模式, 必然会产生新的伦理道德及法律问题.

虚实融合空间生命周期的各个阶段与组成中都存在安全问题, 本文后续部分将从用户认证与权限控制、数据安全、隐私保护、感知与交互安全、关键基础设施与软硬件安全、应用安全与网络空间的治理等 6 个方面对国际相关研究现状和发展趋势进行综述, 最后提出有待解决的 10 个安全技术问题

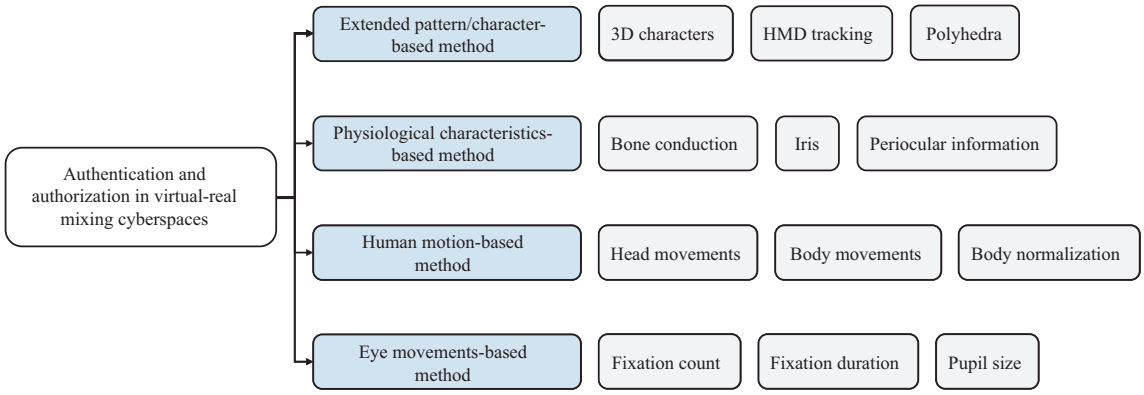

图 4 (网络版彩图) 虚实融合网络空间用户认证与权限控制方法分类

Figure 4 (Color online) Methods of authentication and authorization in virtual-real mixing cyberspaces

或未来研究方向.

3 用户认证与权限控制

虚实融合网络空间的用户认证与权限控制, 是指在一个贯通数字和真实的网络空间环境中, 用户按照一定方式验证用户身份并获得系统授权, 以访问该用户身份下的各类资源, 这里用户是指参与到虚实融合网络空间中的人. 用户获得信息系统授权以访问资源, 是参与系统活动的基本条件, 常规上通过输入字符口令或滑动手势进行用户认证与权限控制, 虚实融合网络空间新型的交互方式和三维呈现方式, 使其在用户认证与权限控制方面具有新的特点, 进而产生了基于扩展字符/图案、基于生理特征、基于人体运动、基于眼动数据等新的方法, 如图 4 所示.

3.1 基于扩展字符/图案的用户认证与权限控制

此类方法探讨利用 VR/AR 的新型交互设备, 扩展熟知的字符/滑动手势方式进行虚实融合网络空间的用户认证与身份识别. George 等 [7] 对采用头戴式显示器及交互设备的认证机制进行了评估, 设计了 8 种 PIN 和图案模式的实验, 证明了常规的口令和 2D 图案锁密码在虚实融合网络空间的可用性; Yu 等 [8] 在常规口令和图案锁密码的基础上, 引入了组合空间更加复杂的 3D 口令, 并证明了其对虚实融合网络空间的有效性、易用性和安全性; Olade 等 [9] 探讨了利用移动设备在数字空间输入图案锁口令的方式, 比较了使用手柄、头戴式显示器、LeapMotion 和视线跟踪 4 种输入模式, 发现在数字空间输入图案锁口令比在移动终端具有更高的安全性; Mathis 等 [10, 11] 给出一种利用 3D 几何体的用户认证方法, 该方法使用了包含颜色和 9 个数字的六面体, 用户通过指向设备对六面体操作完成认证, 同时也分析了这种认证方法的安全性; 针对头戴式显示器, Funk 等 [12] 提出了一种基于头注视跟踪和空间映射的图形身份验证机制, 口令由空间中的数字对象组成的一组连续序列构成, 用户需按照正确的顺序选取预先设定的数字对象.

3.2 基于生理特征的用户认证与权限控制

此类方法通过探索分析用户生理特征, 识别用户并进行权限控制, 主要思想是利用在一段时间内用户的生理状态具有唯一性的特点, 通过分析在每个用户不同的生理特征进行用户认证和权限控制. 这类方法的问题是易受到时间/生理变化等因素的影响, 如利用毛发生长或体重增加等特征. Schneegass

等 [13] 根据头骨解剖结构的个体差异导致的频率反应不同, 基于谷歌眼镜的骨传导能力通过用户头骨所传播的声音, 分析其传播差异得到头骨差异, 并进行用户识别; Raja 等 [14] 使用多模态融合方法, 通过将面部、虹膜和眼周等信息结合用于用户认证, Venkatasubramanian 等 [15] 提出基于生理特征的密钥协议 (physiological-signal based key agreement, PSKA), 利用生理信号作为认证手段, 实现对称口令密钥的认证, 且无需密钥材料的先验分布, 从而用于人体区域网络的安全传感器间通信和认证.

3.3 基于人体运动的用户认证与权限控制

为了减少口令被窃取或被转移带来的安全风险, 研究者还探索了分析人体运动的用户认证与权限控制方法, 该类方法的核心思路是分析虚实融合网络空间中, 不同用户的头部运动、身体运动所产生的位移和时间等方面的差异, 从而进行用户认证和权限控制. 这类方法的主要问题在于其方式容易被旁观者窃取. Mustafa 等 [16] 探索了通过头部运动进行用户认证的可行性, 用户通过头部运动操控数字物体, 并对齐数字空间随机产生的若干点, 通过分析运动的时间、偏差等信息获得用户认证信息; Sivasamy 等 [17] 设计了一个认证系统 VRCAuth, 该系统使用多种 VR 头盔内置传感器的信息表征头部运动, 并通过多种机器学习模型对不同种类数据进行分析, 从而实现用户认证; Pfeuffer 等 [18] 利用 HTC VIVE 和眼动仪研究了可用于用户识别的虚实空间交互行为, 其工作涵盖了 4 种基本动作, 通过分析不同用户完成基本动作的数据特征进行用户认证, 但该方法准确率不高; Kupin 等 [19] 通过追踪用户在执行面向目标任务时的具体行为, 比如观察用户向某一指定目标投掷球的动作, 来进行用户身份认证, 极大地提升了准确率; Ajit 等 [20] 在上述工作基础上, 通过结合头戴式显示器、右手控制器和左手控制器的位置和方向特征, 并使用感知分类器学习归集的匹配权重, 实现无缝持续的用户认证, 最大准确性可达 93%; Liebers 等 [21] 提出了将用户身高和手臂长度归一化的概念, 通过标准化参与者的身体比例, 提高了通过分析运动数据进行用户识别的准确性; Miller 等 [22,23] 分析了不同交互设备 (oculus quest, HTC vive 和 HTC vive cosmos) 在相同运动时的差异, 以及头部、左右手控制器参数对用户认证的重要性, 并使用神经网络构建了认证系统; Miller 等 [24] 在之前工作的基础上, 通过引入同一用户上不同交互设备间的空间位置关系和平滑约束, 利用深度神经网络进行身份认证, 提升了认证的准确性; Shen 等 [25] 结合 VR/AR 耳机中的车载惯性测量单元, 使用步态识别模型 Dynamic-SRC 进行用户认证, 通过几步行走来验证用户的合法性, 无需使用额外的硬件. 除了人体运动外, 还存在相关工作 [16] 通过跟踪手部运动实现虚实空间中的用户认证.

3.4 基于眼动数据的用户认证与权限控制

目前, 很多头戴式显示器配备了眼动跟踪设备, 通过这一设备可以获取用户的眼动信息, 该信息可与头部运动、控制器操作等结合进行用户认证. 与基于人体运动的方法相比, 这类方法不易被旁观者察觉, 有更好的安全性. Rogers 等 [26] 让用户观看快速变化的数字和字母图像, 使用头戴式显示器捕捉无意识动作, 如眨眼和头部运动, 并结合红外、加速度计和陀螺仪传感器进行用户识别; Holland 等 [27] 利用眼球运动过程中涉及的复杂神经系统相互作用而具有的独特性, 提出了多种基于眼动的生物特征, 如注视次数、平均注视时间等, 并将其融合来进行用户认证; Luo 等 [28] 从人类视觉系统特点出发, 在眼动信息基础上, 结合其他设备获得的眼睑、眼外肌等眼电信息, 进行用户认证. 在基于眼动数据的研究中, 也有结合眼动生物力学的工作, Lohr 等 [29,30] 实现了一个基于复杂眼球运动行为的生物识别框架, 通过提取眼部静态和动态有关的 12 类特征, 进行用户认证; Olade 等 [31] 提出了利用物理运动和虚实融合网络空间中头部和眼睛凝视数据的用户识别, 分析了眼动的生理特征模式, 比较了头部、手部、眼部不同生理特征下的用户识别与认证的差异; Liang 等 [32] 提出了基于眼球追踪技术的

生物识别模型, 从眼动记录中提取视觉注意特征, 并将其作为生物特征来识别用户, 其研究表明基于视频的注视分析是生物特征应用的可行解决方案; Zhu 等 [33] 分析了用户登录数字世界的安全性问题, 提出了基于眨眼和瞳孔大小变化模式的两因素用户身份认证方案, 通过广泛的实验验证和评估, 证实了该方案在用户认证场景中的可行性和有效性. 除了上述工作, 一些研究还探索了多种方式结合的用户认证以及存在的风险. Mathis 等 [34] 提出了一种知识驱动的行为生物识别认证方案 RubikBiom, 将口令与用户行为结合, 给出了知识驱动的用户认证与权限管理方法; Wang 等 [35] 验证了采用多属性进行用户认证的方案, 可以有效防范人在室内的攻击, 该方法给出了物体的属性, 用户通过对属性的识别进行认证; Von Willich 等 [36] 认为在虚实融合网络空间, 由于用户不完全感知周围环境, 旁观者会获取其行为并窃取身份, 从而带来安全问题, 为此, 其设计了有效的感知机制以进行用户认证; Liebers 等 [37] 将人体视为一组函数, 该函数可以对用户认证系统产生的 “刺激” 作出 “反应”, 二者随后将共同用于用户认证.

3.5 小结

用户认证与权限控制是虚实融合网络空间与用户衔接的重要环节, 与常规认证方式相比, 结合用户生理特征、运动等方式具有新的特点, 也更符合用户便捷交互的要求. 其中, 基于扩展字符/图案的方法扩展了传统认证方式, 用户可在新型交互设备上利用熟知方式认证, 具有较好的便利性; 基于生理特征的方法利用了具有唯一性的特征, 与用户身份便于结合; 基于人体运动的方法结合用户在虚实融合环境操作的特点, 分析用户在运动模式上的差异, 能够与使用过程结合; 基于眼动数据的方法利用头戴式显示设备将眼部和头部等运动数据结合, 使用户认证便利化. 与其他信息系统的用户认证与权限控制方法类似, 在虚实融合网络空间中, 这些用户认证与权限控制方法仍有可能被攻击, 使用户的身份被非法窃取和假冒, 从而带来互操作中的风险.

4 数据安全

在虚实融合网络空间, 数字对象是对数据、操作、服务、系统的数字化表示, 并可以根据实时驱动信息或人机交互进行更新. 虚实融合网络空间中的数据安全是指在相关数据全生命周期中, 例如数据创建、数据传输、数据存储、数据处理等过程中, 保证数据完整性、真实性、可用性、机密性的预防、检测、预警等处理手段. 其中, 数据完整性指在处理过程中不发生人为或非人为的非法篡改. 数据真实性是指数据内容是真实且无错误的. 数据可用性指授权方可以正常访问数据. 数据机密性是指在处理过程中不发生泄露. 传统网络安全中的数据保护仅影响到数字空间中的信息或资产, 然而虚实融合网络空间一般对物理实体的多维度、多粒度、多时空属性数据进行更为完备的映射, 如果在相关数据传输、存储和管理过程中出现非法访问或数据损坏, 那么会对现实物理空间直接造成影响. 例如, Alcaraz 等 [38] 在一篇关于数字孪生安全风险的综述中指出, 孪生系统面临着从数字空间攻击孪生体而达到伤害物理层面目的的风险, 同时, 以物理世界资产为蓝本而创建的数字孪生体也存在知识产权被侵犯的风险. 例如, 攻击者可以操纵和伪造相关信息, 破坏孪生体的属性数据的保真度和映射粒度; 攻击者也可通过人为控制孪生体宿主计算机的计算开销以限制孪生对象的仿真过程, 进而间接影响对物理实体运行状态的即时控制; 攻击者还可能从数字空间控制实体资产, 以窃取敏感信息等. 此外, 如果关键系统主要依赖虚拟仿真服务来进行运维、优化和恢复, 那么最终攻击的后果将是毁灭性的, 可能导致物理系统的中断甚至瘫痪. 传统网络空间数据安全领域已经在数据保护、数据丢失预防、数据维护更新和数据泄露反馈几个方面进行了广泛研究. 如图 5 所示, 本文着重考虑虚实融合网络空间的数据安全

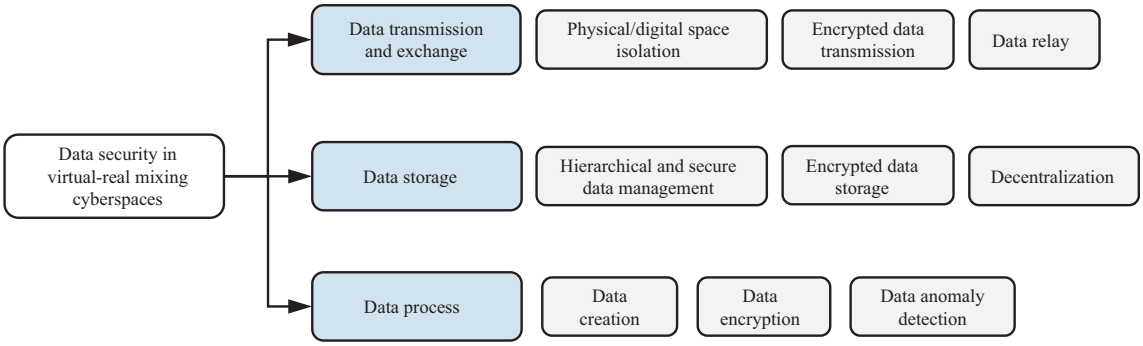

图 5 (网络版彩图) 虚实融合网络空间数据安全中的风险分类

Figure 5 (Color online) Classification of data security threats in virtual-real mixing cyberspaces

传输、交换、存储、处理等生命周期过程中的安全问题.

4.1 数据传输和交换安全

在虚实融合网络空间的数据传输和交换过程中, 数字对象和物理世界之间会形成完整的映射链路. 通过数据在传输过程中的干扰和篡改, 可以对物理实体产生直接或间接的影响. 因此, 在传统网络空间数据传输安全问题的基础上, 虚实融合网络空间需要着重研究数字空间与物理空间的数据安全交换, 构建物理 – 数字空间数据的隔离带和防火墙. 针对虚拟对象可能受到的数据传输入侵, 主要可通过数据加密 [39]、孪生体隔离保护 [40] 和构建数据安全中继防御点 [41] 等方式来进行数据安全防护, 从而阻隔由于虚拟对象被攻击诱发的物理实体安全问题. 例如, 针对数字孪生对象在安全测试、入侵检测等方面可能遇到的问题. Eckhart 等 [39] 提出了一种基于信息加密的数字孪生对象构建方法, 并阐述了如何有效且安全地创建、维护和运行数字孪生对象, 降低其对应物理实体及信息 – 物理系统带来的安全隐患. 同时, 该团队提出了一种基于数字孪生的网络环境安全测试平台, 建立了一种虚拟孪生体对真实网络环境的模拟和复制. 该孪生体支持网络拓扑、可编程逻辑控制器、人机接口和物理设备的虚拟复制, 使其可以仿真、重播、可视化对真实网络的攻击及其影响, 并同时隔绝入侵攻击仿真过程对真实物理系统稳定的影响 [40]. 针对基于数字化物联网的海上运输系统 (maritime transportation system, MTS) 的安全性能, Liu 等 [41] 提出了基于中继协作物联网的海上运输数字孪生模型构建方法. 他们将 “人 – 船舶 – 环境” 对象上的射频识别标签等传感器作为标识, 建立各节点无线传输接收能量和信息的衰减关系, 并利用海上丝绸之路的历史运输数据对其性能进行了分析. 当模型中天线数为 6 个时, 信息安全曲线的增长逐渐平缓, 中断概率低, 节点安全率高. 同时, 在 100% 的成功传输概率下, 该模型可以将数据传输延迟保持在 700 ms 以下, 保障信息传输的安全性和时效性.

4.2 数据存储安全

由于虚实融合网络空间表示和交互物理实体丰富多样, 其产生和需要存储的数据会更加复杂且数量庞大. 再加上边缘计算等技术的应用, 虚实融合网络空间在运行过程中产生的大量数据一般主要以云端方式存储. 因此, 确保数据在云端存储时信息的真实性和完整性是虚实融合网络空间需要面临的问题. 因此, 需重点考虑数据加密管理 [42] 和去中心化加密方法 [43], 以保证数据在云端存储和调用时数据的真实性和机密性.

针对云环境数字孪生对象数据的安全问题, Susila 等 [42] 分析了在云端环境中原始数据加密的必要性. 存储数据经过运营商提供的加密算法后可以形成多层加密状态, 保证数据在处理过程中数据的

真实内容不会泄露于未经授权的一方, 以此增强数据的机密性. 此外, 他们还提出了一种哈希 (Hash) 文件加密算法配合 RSA 签名 (rivest-shamir-adleman signature) 身份验证算法, 进行数据保护和访问权限控制, 以保证存储数据安全. Seth 等 [44] 提出了一种基于外包数据库的安全存储系统, 通过 3 个步骤来增强外包信息的安全. 首先对原始数据进行加密, 确保在传输和存储过程中不被窃取或篡改. 其次使用安全存储服务提供商来存储加密后的数据. 最后通过访问控制和审计机制来保护数据, 防止未经授权的访问和恶意攻击. 这种系统可以有效地提高外包信息的机密性, 同时降低企业的安全风险和管理成本. Sajay 等 [45] 提出了一种结合同态加密和 Blowfish 加密的算法, 提高云端存储数据机密性的同时减少了数据存储的开销. 针对数字孪生对象的数据管理和安全问题, Suhail 等 [43] 提出了一种基于区块链技术的数字孪生网络框架, 可有效防止恶意攻击对虚拟对象信息的篡改和访问. 他们将可靠来源的数据利用区块链的防篡改和总账不可变特性, 跨多个参与实体进行分布式、分散存储, 避免数据遭受外界篡改而导致存储内容失真, 从而保证了数据的真实性. 基于区块链的透明度和隐私机制的强制执行特性, 保障存储数据只在授权情况下才能接受访问. Aks 等 [46] 通过智能合约技术, 可以将元宇宙中的数字资产映射到区块链上, 实现数字资产的安全交易. 智能合约可以防止数据真实内容被篡改或删除, 同时也可以防止欺诈行为.

4.3 数据处理过程安全

虚实融合网络空间中存在大量的用户交互、数据创建、数据更新等操作. 数据处理过程中, 恶意第三方可能通过非法访问和网络攻击, 进行窃取、监听、替换数据的操作, 破坏数据的真实性和完整性, 或者以伪装的方式发布失真信息. 为此, 目前学术界对数字孪生体的安全创建 [47]、数据加密方法 [48] 以及数据异常检测 [49] 等方面进行了研究, 讨论了数据在处理过程中的安全问题.

针对数字孪生网络对物理实体的加密和保护中可能遇到的问题, Bitton 等 [50] 提出了一种特定于网络、具有成本效益、高度可靠且面向安全测试的数字孪生体创建方法. 该方法由两个模块组成: 问题构建器获取有关被测系统的真实数据, 并将其转换为反映系统拓扑和数字孪生实施约束的规则集; 问题求解器接受输入, 并使用非线性规划来找到满足所有约束的安全的解决方案. Wu 等 [48] 探讨了基于深度学习的无人机信息系统的安全问题. 针对无人机系统受到攻击时的安全问题, 他们采用改进的长短期记忆 (long short-term memory, LSTM) 网络对信息 – 物理系统数据进行分析, 从攻击发生前预测系统控制信号数据的角度进行预测. 同时, 他们引入差分隐私频繁子图 (differential privacy frequent subgraph, DPFS) 保护数据隐私, 利用数字孪生技术在物理空间映射无人机的操作环境, 基于差分隐私改进的 LSTM 构建了无人机数字孪生攻击预测模型后, 并以田纳西伊斯曼 (Tennessee Eastman) 为仿真平台, 对所构建的模型进行仿真, 确定了性能的可靠性. 针对信息 – 物理系统中数据处理安全问题, Benedictis 等 [49] 利用数字孪生体可以提供的先进异常检测能力, 在不影响物理网络的前提下为真实数据交换提供保障. 该方法构建了孪生体服务层, 使用联邦学习对非法匿名访问进行检测, 并构建了由物理层、存储层、虚拟层组成的安全孪生体. 其中, 物理层采集真实空间的大量传感数据; 存储层利用区块链对必要的实时数据提供数据授权访问控制, 并隔离虚拟层和物理层的直接交互; 虚拟层除了存储模型信息外, 还配备了大量的人工智能检测算法, 用于数据异常和入侵检测.

4.4 小结

数据安全是虚实融合网络空间安全的重要组成部分. 虚实融合网络空间中的虚拟资产与物理世界资产之间存在深度联系, 因此, 数据安全不仅需要维护数据真实正确、保证正常访问、拒绝非法篡改, 也需要建立虚拟数据和物理数据之间的异常交互检测和应对措施. 数据安全问题主要存在于数据传输

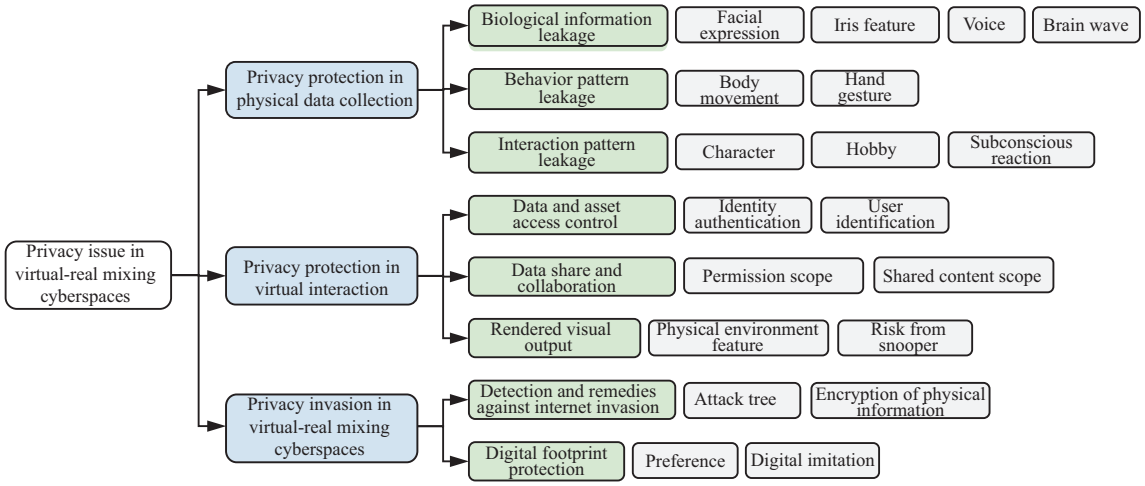

图 6 (网络版彩图) 虚实融合空间隐私保护的风险分类

Figure 6 (Color online) Classification of privacy threats in virtual-real mixing cyberspaces

交换、数据存储和数据处理阶段. 其中, 数据传输交换过程中可以通过数据加密、虚实隔离保护和数据中继等保证数据的机密性、真实性和稳定传输; 数据存储阶段主要考虑海量数据的安全分级管理和云端去中心化加密问题; 数据处理阶段需要解决如何安全创建虚拟 – 物理空间的完备映射, 以及如何建立针对数据异常操作的检测和应对措施.

5 隐私保护

虚实融合网络空间带来新的数字革命的同时, 也带来了新的隐私安全问题. 虚实融合网络空间的隐私安全与保护是指在物 – 数 – 人三界贯通的空间中, 防止通过主动暴露或被动入侵等方式, 造成物理对象和数字对象的所属控制权、内容利用权、公开范围决定权等基本权利侵害的预防和处理措施. 相较于传统人机交互和社交平台, 虚实融合网络空间采集的隐私信息数量更加庞大, 而且种类繁多、更加敏感、具有较高的识别性, 同时在新的交互模式中也出现了更多形式的隐私泄露风险. 数字空间中所对应的物理实体数据覆盖范围十分广泛, 许多数据背后的物理实体具有可观的应用价值或者直接涉及社会空间中公民隐私信息, 所承受的网络入侵攻击程度也会更加严重. 因而, 隐私数据的保护已成为虚实融合网络空间中的一个重要问题. 考虑到传统数字空间的隐私保护与安全已经在通讯 [51]、数据库 [52]、物联网 [53] 等方面有了广泛的研究. 如图 6 所示, 我们将着重讨论在虚实融合网络空间中, 从物理世界数据采集、数字世界交互、隐私入侵攻击 3 个方面对虚实融合网络空间的隐私保护问题进行梳理分析.

5.1 物理世界数据采集的隐私保护

物理世界的数据采集贯穿于人类对可穿戴设备、体感设备、脑机接口等物理设备的操作的全过程, 也是人类通过物理世界与数字世界产生交互联系的信息入口. 虚实融合网络空间交互中需要对人或物理实体采集多维度的数据, 包括生物信息、行为信息、语言交互信息等. 相较于传统的个人隐私信息, 这几类信息蕴含着更为敏感和高识别度的匹配特征. 交互硬件设备的加密程度不足和交互过程中对隐私信息的管理不当和肆意使用, 将会导致以人类用户为主体的虚实融合网络空间的用户的隐私

泄露.

(1) 个人基本生物信息泄露: 虚实融合空间为了达到沉浸式的交互体验, 需要在多个维度上进行用户数据采集. 除了传统的个人信息之外, 还可能包括但不限于面部表情、虹膜特征、语音和生物特征, 甚至脑电波. 这些新形式的个人隐私数据增加了隐私泄露的风险. Martinovic 等 [54] 使用消费级的脑机接口设备与脑电图测试了旁路攻击的可行性, 其实验结果证明: 捕获的脑电信号可以揭示用户的银行账户、PIN 码 (personal identification number)、居住区域、人际关系等私人信息. Bozkir 等 [55] 使用多种分类器混合模型对人眼跟踪所必需的最小时间片段进行判定, 在驾驶模拟器中成功使用截取的时间片段进行用户认知测试, 并通过交叉验证的方式证明了其有效性. 这种方法仅使用部分时间片段的传感器信息进行数据分析, 因此, 合理运用此方法可在用户信息采集过程中最大程度地保护用户隐私. John 等 [56] 使用高斯滤波器对虹膜信息进行模糊处理, 可以有效地防止用户注视信息的泄露. 同时, 该团队还通过在眼球跟踪器上增加一个透镜伸缩臂, 让用户自主调节虹膜的清晰程度, 来主动防御虹膜信息泄露 [57].

(2) 用户物理行为模式泄露: 围绕可穿戴设备与体感设备展开的旁路攻击已成为隐私安全保护领域的关注重点与研究热点. 为了设计安全的虚实融合系统并建立用户信任, 必须首先了解该类系统中用户物理行为模式所面临的安全威胁. 现有的攻击方法或潜在的威胁大都源自人体不同部位泄露的信息, 包括肢体动作 [58,59]、手势 [60]、语义交互行为 [61] 等. Chen 等 [58] 提出了一种视频辅助 PIN 码推断系统 ArmSpy, 该系统通过从用户背后观察击键时引起的手臂姿势细微变化, 包括肘部弯曲角度的变化和不同手臂关节之间的空间关系, 来推断 PIN 码输入. 实验结果表明, ArmSpy 可以通过 3 次尝试获得超过 67% 的 PIN 码推断平均准确率, 对于部分用户甚至可以达到 80% 以上的准确率. 由此可见, 肢体动作攻击可对虚实融合系统造成严重威胁. 在虚实融合环境中, 用户通常使用虚拟键盘键入关键字、浏览网页或键入密码来访问账户, 从而与数字世界进行交互. Al Arafat 等 [60] 提出了一种基于 WiFi 信号信道状态信息的虚拟按键识别方法. 该方法的核心思想是: 与每个虚拟按键相关的精细手部运动的侧通道信息, 在信道状态信息波形中具有独特的模式. 因此, 该方法通过利用信号处理技术从信道状态信息的变化中提取相应的模式, 可以实现 69.75% 的虚拟击键识别准确率. Luo 等 [59] 研究发现, 部分虚实融合系统并不对运动行为数据进行访问权限控制, 这个漏洞加大了恶意软件的攻击风险. 恶意软件的击键推断攻击可通过 6DoF (degree-of-freedom) 头部运动驱动的按键跟踪方法和基于空气敲击模式的按键推断方法来实现. 除此之外, Li 等 [61] 提出了一种利用人体关键节点信息的智能图推理模型, 可以通过视频片段准确地预测双人交互行为类别, 最高可以达到 92.1% 的准确程度. 第三方可以通过对用户交互行为的语义模式进行分析, 识别和锁定特定用户的身份信息.

(3) 用户交互沟通模式泄露: 通过分析用户在不同环境和条件下的心理活动特征, 可以得出用户的人格特征、兴趣爱好和潜意识反应, 从而被用于骚扰行为甚至间谍策反活动. Buck 等 [62] 指出, 用户与周围环境的互动方式本身就是一条关键信息, 因为个人空间是可塑的, 可以根据实时的互动而变化. 当用户接近喜欢的对象时, 个人空间会被压缩; 反之, 当用户不欣赏当前的互动或者事物时, 个人空间会因为用户的远离而被拉伸. 这些因素都有可能会被运营公司收集并分析出有关个人偏好的敏感信息.

个人基本生物信息、用户物理行为模式和用户交互沟通模式是 3 种核心的用户隐私数据类型. 相比于传统互联网涉及的用户隐私, 虚实融合网络空间中的隐私数据类型更加丰富且敏感. 基本生物信息, 如脑电波、虹膜和眼动信息等, 可能暴露用户的相关账户密码、居住区域以及人际关系, 可以通过模糊处理和规避采集过程暴露风险等方式进行保护. 用户物理行为模式包括全身关节的运动模式, 攻击者可以从手臂和头部的运动状态推理出用户 PIN 码等关键隐私数据, 甚至推断用户所进行的多人

物理交互类别. 用户交互沟通模式隐私攻击可以通过用户的对话和与周围环境的交互模式, 推断出该用户的喜爱偏好和心理活动特征, 并暴露出个人的人格特征和兴趣爱好.

5.2 数字世界交互的隐私保护

数字世界的隐私信息会在身份认证、协作交互、终端显示等用户与数字世界交互过程中暴露在恶意第三方的感知范围之内. 人与数字空间交互的过程需要开放部分用户的数字资产权限和环境信息, 而在沉浸感为主的虚实融合网络空间中, 对于 "隐私" 权限以及可以共享的范围界定尚不明确, 因此, 可能造成用户数字资产丢失和物理空间环境暴露等问题, 最终会导致物理世界中的资产权限完整性被破坏以及泄露用户在人类社会中的位置、环境, 甚至偏好等敏感隐私信息.

(1) 个人数字资产访问隐私保护: 隐私数据和个人数字资产的安全性涉及身份验证和授权决策. 在虚实融合空间中, 数字资产的数量庞大且可能含有个人敏感信息, 因此, 保证身份验证和用户身份识别过程的安全性十分重要, 因为身份授权可决定这个标识的用户可以访问哪些资源. 目前数字认证的方式主要有基于知识信息的验证、基于生物特征的验证, 以及混合方式的验证. Yu 等 [8] 开发出了 3D 模式验证、模式锁定和 PIN 码验证系统, 并验证了其保护元宇宙访问权限的有效性. 生物识别认证主要取决于生物识别的数据类型, 常用的有脑电图 (electro-encephalogram, EEG)、身体运动和眼电图 (electro-oculogram, EOG) 读数. 脑电图数据因为其独特性, 是目前比较可靠的方案之一. Li 等 [63] 在元宇宙环境中研发了一个基于大脑信号的身份认证系统, 其身份关联准确率可以达到 80.91%. 然而, 因为脑电信号本身信息足够敏感, 它本身也需要信息加密和保护措施. 此外, 目前相对成熟的认证方式还包括 RubikAuth 和 OcuLock 等. RubikAuth 的名字来源于魔方的发明者, 其本质上是覆盖在一个彩色立方体的 9 位数 PIN 码, Mathis 等 [34] 基于 RubikAuth 以及对用户头部 6DoF 方位和眼部跟踪信息, 设计了一种全连接卷积神经网络分类器, 使得身份关联准确程度可以达到 98.91%. OcuLock 主要利用人类视觉系统 (human vision system, HVS) 信息, 包括眼球、眼睑、眼睛周围的神经、眼外肌、细胞等信息 [28]. 通过在预定义的时间刺激眼睛来激活 HVS 的行为, 其身份关联错误率仅在 3.55%∼4.97% 之间变化. Groshev 等 [64] 提出了一个基于人工智能的数字孪生构建方法, 通过限定数据的可信度、完成性、获取权限等方面特征, 实现了孪生对象的安全验证, 提升了虚实融合系统的内容访问安全性和抗攻击性.

(2) 数据共享与多人协作安全: 支持数据共享和多人协作是虚实融合空间应用的重要创新和特色. 类似于现实空间中的多人合作, 在虚拟空间中也期望用户能够共享相同的物理空间或相同的虚拟内容. 无论是现实物理空间还是虚拟的数据资产, 在多人交互协作和共享过程中, 都存在被第三方恶意窃取或泄露其他用户隐私数据的风险. Ruth 等 [65] 探讨了多用户 AR 内容共享和访问控制的实际需求, 设计了一个 AR 应用多用户共享控制模块, 支持用户控制如何与他人分享现实内容, 并形成了基于 HoloLens 的可信应用系统 ShareAR. 针对虚拟环境下多个用户在云平台共享数据的用户安全及信息感知问题, Ritzdorf 等 [66] 提出了一种分布式数据分布框架, 通过对不同种类数据及不同用户设定特定的访问权限并实时更新, 保证了数据共享、数据协同操作的安全性.

(3) 虚实融合交互呈现安全: 虽然虚实融合可将信息以更为透明的方式增强到物理世界和人类社会, 但也增加了新的隐私和安全风险. 其中, 虚拟内容可视输出带来的安全风险, 或者是渲染应用程序篡改用户对现实世界的观察能力所产生的风险尤其需要重视 [67]. 虚实融合空间的渲染输出一般需要对现实物理世界的空间结构进行分析. 例如, 通过当前拍摄的场景图像结合传感器信息, 计算出设备在三维世界中的准确位姿, 从而在当前视图上叠加虚拟场景中的对象进行渲染. 因此, 基于渲染输出来对物理世界信息进行解析可能会引起用户所处环境信息的泄露. 例如, Zhang 等 [68] 提出了场景几何

及语义信息推理模型, 可以通过局部点云片段推理预测场景的完整几何状结构及其语义分布信息. 同时, 该团队 [69] 提出了一种三维场景图推理方法, 可进一步预测场景实体的语义关系, 并据此提取用户所处环境的语义信息. 由于真实世界的场景很大, 所以目前商业的增强现实服务一般预先建立一定范围的三维地图放到云端或边缘侧, 将终端的查询图像提交给服务进行定位 [70], 终端根据返回的全局定位信息在局部范围利用即时定位与地图构建 (simultaneous localization and mapping, SLAM) 技术实现位姿计算. 其中, 设备中的定位模块会定期拍摄图像以及采集数据进行上传, 然而用户并不知道这些设备在多大程度上记录数据并存储和发送. 现有方法通常采用运动结构恢复 (structure from motion, SfM) 生成的稀疏点云来建立三维地图, 同时丢弃 SfM 原图实现隐私保护 [71]. 然而 Pittaluga 等 [72] 提出了一种针对三维地图的隐私攻击方法, 利用级联的可见性估计网络、重建网络和对抗网络从包含深度信息和描述符的特定视点, 输出场景的合成图像. 合成图像可恢复出大量细节, 在视觉上与真实图像很接近. 此外, 可视化渲染输出的可观看范围因设备不同而大相径庭, 显示范围较大的设备还可能出现被窥视的风险. Vilk 等 [73] 倡导限制虚实融合环境中可视交互设备的检测范围, 仅允许检测人体骨骼信息、房间内表面信息、设备位置和方向信息. Lin 等 [74] 提出利用人眼感知和成像设备之间的差异, 通过时间心理 – 视觉调节 (temporal psycho-visual modulation, TPVM) 来在图像中嵌入一种特殊的数字 AR 标签, 该标签不能被人眼发现但可以被摄像机检测到, 增加了渲染内容被窥视的风险. 因此, 虚实融合交互呈现过程中的场景构造、渲染输出对真实场景信息具有严重的隐私威胁, 需要研究新的隐私保护方案.

数字世界交互的隐私保护包括 3 个方面的内容: 个人数字资产访问、数据共享与多人协作、交互呈现安全. 个人数字资产访问的隐私保护, 主要通过大脑信号和眼部生物特征等与用户身份高度绑定的信息, 对数字世界身份进行认证, 技术要点为生物信息的特异性模式识别. 数据共享与多人协作是虚实融合网络空间的重要特性, 而共享数据和环境信息暴露程度控制是在多人协作过程中维护个人隐私的核心, 通常可以通过用户自定义决定共享内容、对环境信息细节隐藏等方式进行保护. 交互呈现安全问题包括: 呈现内容对用户环境感知能力带来的风险, 渲染过程对环境几何细节和语义分布的泄露, 以及渲染内容被窥视的风险. 目前, 此类安全问题的研究尚在初级阶段, 需要针对虚实融合呈现过程中的环境构造和渲染内容等提出新的隐私保护方案.

5.3 虚实融合空间中的隐私入侵攻击

与传统的网络空间安全类似, 虚实融合网络空间也不可避免地面临被恶意第三方主动入侵以获取个人或者平台隐私信息的风险, 这些隐私信息包括存储在云端平台的个人数据、物理世界位置信息、数字足迹等. 这些隐私信息被获取后, 恶意第三方可能利用远程操控、覆写、数字伪装等手段对用户的数字资产可信度、完整性、可用性进行破坏.

(1) 隐私入侵攻击形式和应对策略: 在虚实融合空间中, 攻击者可以通过网络入侵、窥探、篡改、泄露用户化身隐私数据. 针对虚实融合网络空间中的隐私入侵攻击, Valluripally 等 [75] 提出了导致隐私泄露的攻击树. 针对拒绝访问、虚假数据注入、数据伪造等常见攻击手段, Suhail 等 [76] 提出了一种基于区块链的数字孪生框架, 该框架可根据多维度数据进行交叉验证, 并使用区块链对数据进行存储, 同时通过操作过程监控、诊断和优化控制, 保证了登录、身份识别、数据维护等操作的可信度和完整性. 针对远程入侵和恶意操控等问题, Donno 等 [77] 提出了一种保护云端数字孪生对象安全性的方法, 通过对不同攻击的发生条件、潜在影响及应对方法进行分析评估, 可有效应对物联条件对各类物理对象的入侵攻击. 此外, 在使用 AR 定位服务时, 用户需要上传查询图像在云端服务器上实现相机定位和辅助位姿估计. 当消费者在家中、工业、商业以及保密场所使用此类服务时, 可能会引发隐

私泄露问题. Speciale 等 [78] 提出对查询图像进行处理, 用随机定向的 2D 线替换图像中的 2D 图像特征点, 在去除特征点信息隐藏图像内容的同时, 仍然提供足够的几何约束来支持准确的相机位姿估计. 针对三维地图中的隐私问题, 他们还提出将地图中的每个 3D 点转换成具有随机方向的 3D 线, 生成的三维地图只需仅存储 3D 线和特征描述符, 丢弃了携带隐私信息的原始点云特征. 这种新颖的表示不仅仍支持现有相机定位和位姿估计算法, 还混淆了基础场景几何结构, 保护了真实场景隐私 [79]. Do 等 [80] 提出了一种新的相机定位方法, 可以基于场景关键点实现相机位姿估计. 为了避免隐私泄露, 他们通过训练场景关键点检测器获取查询图像中的场景点, 计算相机位姿时不再采用传统方法学习和计算场景中图像特征点与点云的对应关系, 只学习图像 2D 特征与关键点的对应关系, 无需存储详细的点云信息和视觉特征.

(2) 数字脚印安全保护: 由于虚实融合空间中化身的行为模式、偏好、习惯和活动可以反映他们的真实生活状态, 攻击者可以收集化身的数字足迹, 并利用与真实用户相关的相似性, 进行准确的用户分析, 甚至伪装成数字用户进行非法活动. Falchuk 等 [81] 给出了虚实融合空间中虚拟形象的数字足迹隐私的 3 种类型: 个人信息 (例如, 虚拟形象分析)、虚拟行为、虚拟形象与 NPC (non-player character) 之间的互动或交流. 同时, 该团队还提出了一种可用于数字隐身、私人飞地等内容的信息封锁机制. 该机制允许虚实融合空间内的某些位置在一定时间段内被私人占据, 并通过用户指定权限对该飞地进行有限制的访问, 从而达到隐藏数字行踪的目的. 针对数字脚印隐藏问题, Ning 等 [82] 提出了一种化身克隆方法. 该方法可以创建多个虚拟克隆体, 并同时创建多个与之对应的数字脚印来迷惑攻击者. 然而, 如何管理每个用户数量众多的克隆体, 也是亟待解决的问题之一.

虚实融合网络空间的攻击者可以通过入侵、窥探、篡改和泄露等方式攻击用户化身的隐私数据. 面对拒绝访问、虚假数据注入、数据伪造等常见攻击, 可以使用区块链技术对数据进行安全存储, 并配合网络攻击的监控、诊断、潜在影响评估与响应等方法进行隐私数据保护. 此外, 虚实融合网络空间的数字脚印是反映用户行为模式和个人喜好的重要数据, 可以通过虚拟化身克隆和私人飞地等技术对数字脚印进行伪造和隐身, 从而达到对用户在数字空间内隐私数据的保护.

5.4 小结

虚实融合网络空间隐私数据的安全隐患主要来源于 3 个方面: 物理世界的隐私数据采集、数字世界交互的隐私数据保护、虚实融合空间的主动隐私入侵攻击. 虚实融合网络对物理世界的观察和利用相较于传统网络更加深入, 隐私数据会涉及敏感且多样的生物信息、行为模式和交互模式等. 同时, 对于个人数字资产访问、数据共享多人交互和虚拟内容呈现过程中涉及的个人生物信息、所处环境信息、渲染内容信息, 通过对其模糊和加密, 有助于在虚实融合网络空间交互过程中减少隐私的泄露. 此外, 隐私入侵攻击防范和数字脚印保护也是虚实融合网络空间重要的隐私保护措施. 对于隐私入侵攻击, 需要对攻击手段进行监控和诊断, 并做出对应级别的相应措施, 保证用户在登录、身份识别等过程的隐私安全, 可以利用数字隐身、私人飞地、化身克隆等方式保护用户在数字世界的脚印痕迹, 保证数字化身不被盗用并处于安全的状态. 综上, 相较于传统互联网, 虚实融合网络空间中涉及的数据源和隐私信息类别更多, 需要在传统互联网隐私保护技术的基础上, 进行系统性深入研究.

6 感知与交互安全

感知与交互安全是指虚实融合网络空间中具有独立身份的对象在通过泛联网基础设施和传感交互设备进行各种数据、控制信息交互, 实现所需要的状态改变等动态行为过程中, 保障沉浸感知、数

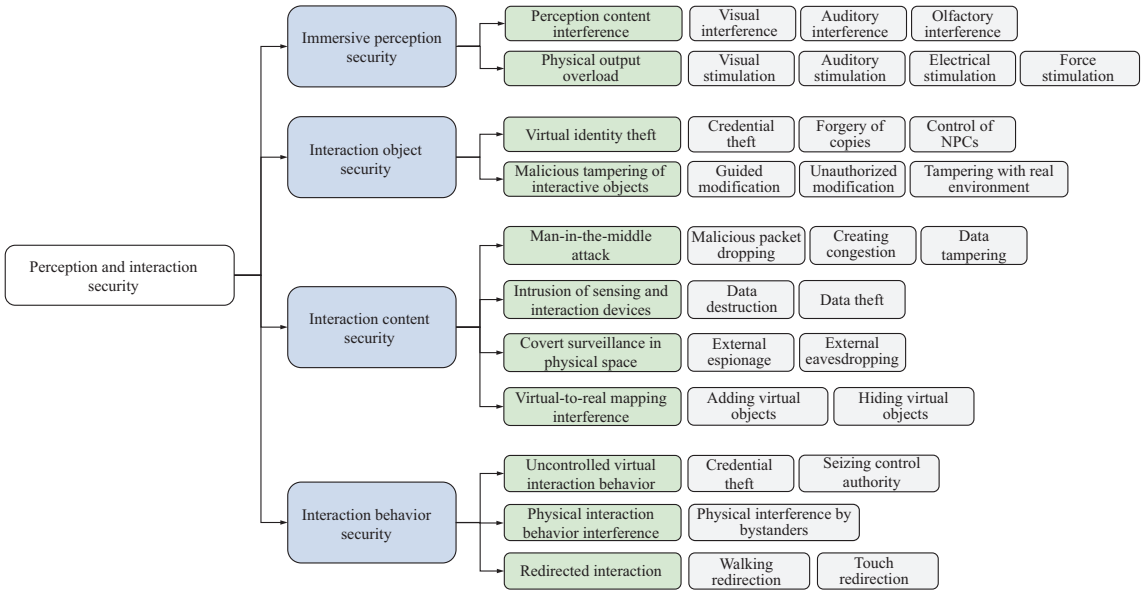

图 7 (网络版彩图) 虚实融合空间感知与交互安全的风险分类

Figure 7 (Color online) Risk classification of augmented reality (AR) and mixed reality (MR) space perception and interactive security

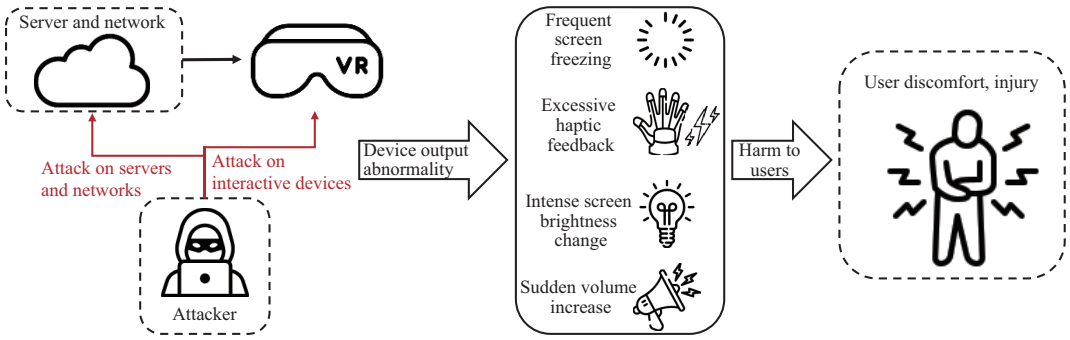

图 8 (网络版彩图) 沉浸感知安全问题概览

Figure 8 (Color online) Overview of immersive sensory security issues

字对象、交互内容以及行为控制的完整性、机密性和一致性的安全防护手段和机制. 感知与交互安全问题产生的原因包括感知与交互的数字对象被篡改、传感交互设备被攻击、交互过程被干扰、物理环境被入侵控制等, 被侵害的主体包括物理世界的用户以及数字世界中的虚拟化身、数字原生等独立对象. 如图 7 所示, 下面从沉浸感知安全、交互对象安全、交互内容安全、交互行为安全 4 个方面对典型感知与交互安全问题进行分析, 并探讨相应风险的安全防范建议.

6.1 沉浸感知安全

沉浸感知安全是指虚实融合网络空间中的用户借助感知交互设备对数字对象和物理环境沉浸感知过程中面临的威胁和防御措施. 沉浸感知安全问题产生的原因包括感知内容干扰和物理输出过载, 例如攻击者通过降低画面帧率或引入画面抖动的方式破坏用户视觉体验, 造成眩晕等模拟器病; 或者篡改设备物理输出使其超出用户的承受极限, 干扰和破坏用户感知, 如图 8 所示.

(1) 感知内容干扰: 攻击者可通过修改或破坏用户接收的视觉信号来干扰用户感知. Odeleye 等 [83] 发现, 通过攻击交互设备的网络连接可以降低头盔所渲染的画面帧率, 从而造成用户身心疲劳, 引起头晕、恶心等模拟器病症的发生. 在 Gulhane 等 [84] 的研究中, 实验人员利用增加网络拥塞的方式针对多个用户同时体验的虚拟网络课堂场景发起攻击, 导致网络延迟升高、头盔渲染画面卡顿, 甚至部分虚拟场景无法完整渲染. 实验结论表明, 该攻击方法会增加用户的感知负担和疲劳感, 破坏用户间的协作流畅程度. 相比这种引入轻微扰动的攻击方式, 另一种更直接的攻击方式是攻击者直接中断用户与原定数字对象的交互, 或在用户观察视野前固定放置攻击者想要呈现的画面阻碍用户视线, 从而达到破坏用户感知、恶意广告推销或勒索用户财产的目的 [75,85].

除了在视觉上对用户的沉浸感知进行干扰, 攻击者还可以通过听觉、嗅觉等感官通道影响用户准确感知数字原生内容, 混淆内容接收者的判断并引导其做出错误的行为. 例如, 在多人虚拟现实竞技游戏中, 用户的视线可能被建筑物遮挡, 这时往往需要用户通过环境中的声音源判断其他用户所在位置. 而攻击者可以篡改用户佩戴耳机的声道音频信息, 影响用户利用双耳效应对其他用户的位置进行准确判断和决策, 破坏公平性和游戏体验. 除此以外, 改变用户的嗅觉感知也是破坏用户间的交互和协作的一种潜在攻击手段. 例如, 攻击者可以通过程序恶意操控嗅觉仪释放令人不悦的气味, 影响用户原有的嗅觉灵敏度或使其产生负面情绪, 从而影响用户交互过程 [86,87].

(2) 物理输出过载: 当用户通过头戴显示器、力触觉反馈装置等感知交互设备与其他对象进行交互时, 部分感官和躯体控制能力被将设备接管, 因此物理输出过载会危害用户的身体安全. 攻击者可通过直接入侵并修改感知交互设备参数的方式使其输出的物理刺激超过用户承受上限, 从而直接对用户身体造成严重威胁. Hamed 等 [88] 证实攻击者可通过入侵头戴显示器设备, 突然调大音频的音量, 损伤用户的听力. Roesner 等 [89] 提出, 除了瞬时调高的音量, 虚实融合空间安全还需警惕剧烈屏幕闪光和强度过高的触觉反馈, 这些物理输出过载的现象可能造成用户视力受损、手指受伤, 特定频率的剧烈闪光甚至会诱发少数用户的癫痫发作, 造成生命危险. Pfeiffer 等 [90] 提出了一种新的应用可变电肌肉刺激工具包, 能够通过粘贴在皮肤上的电极, 以微弱的电流驱动特定肌肉收缩. 攻击者可能会针对该工具包发起攻击, 控制电极发出过大的电流, 导致用户疼痛、受伤, 甚至干扰心脏起搏器功能, 影响大脑的正常工作, 诱发心脏病和癫痫.

感知内容干扰和物理输出过载都可破坏用户虚拟环境中的正常感知和身体安全. 前者主要通过修改用户视听觉等感官接收的内容, 而后者主要通过操控交互设备物理输出, 使其超过身体承受上限. 为了防止沉浸感知安全问题的发生, 虚实融合网络空间的传感交互设备应提高安全等级, 从底层限制设备输出最大值, 从而加强对恶意攻击的防御能力; 同时, 对系统输出内容进行管理, 改善用户交互逻辑, 保证渲染画面的稳定性, 减轻用户在网络阻塞、攻击发生时的不适感; 数字内容提供者需对内容进行严格核查, 对可能面临的危险进行完整标注和公开; 用户需要严格遵守安全规范和使用流程, 使用前确保所处环境的安全性, 尽可能降低事故发生的可能性.

6.2 交互对象安全

交互对象安全是指虚实融合网络空间中的具有独立身份的对象在与其他对象交互过程中, 保障虚拟化身、孪生体、虚拟资产和虚拟场景等可交互数字对象的主体不被入侵或篡改的安全防护手段和机制. 交互对象安全问题的攻击方式包括虚拟化身盗用攻击、交互对象恶意篡改. 例如, 攻击者冒用虚拟化身实施违法行为, 或篡改数字对象外观从而影响或干扰用户交互结果, 如图 9 所示.

(1) 虚拟化身盗用: 当用户登录虚实融合网络空间的认证方式遭到泄露或破解时, 其虚拟资产可能会被侵犯或盗用. 其中, 虚拟化身是虚实融合网络空间中一类特殊的虚拟资产, 其背后登录的用户

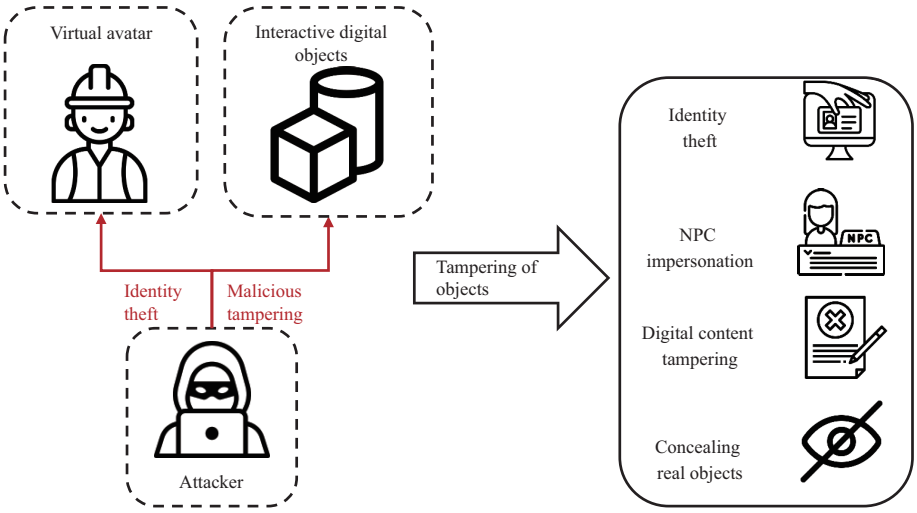

图 9 (网络版彩图) 交互对象安全问题概览

Figure 9 (Color online) Overview of interactive object security issues

身份难以通过虚拟化身形象或行为准确分辨, 一旦被攻击者入侵或控制, 将可能对其他相关用户的隐私或财产进行侵犯, 产生不良后果. 虚拟化身盗用的手段主要包括攻击者盗取用户登录凭证后操控其虚拟化身, 或获取虚拟化身的关键参数信息, 仿制生成虚拟化身副本. Heartfield 等 [91] 指出, 攻击者可盗用用户的虚拟化身, 冒用其身份进行社会活动, 在获取亲友信任后可在难以察觉的情况下骗取私人信息或个人财产. 虚实融合网络空间中的虚拟化身除了由用户直接控制, 也可以是由通过程序或人工智能算法自动控制, 该类虚拟角色统称为非玩家角色 (non-player character, NPC). NPC 一方面能够减少服务业的人力劳动成本, 另一方面在高效算力的加持下, 可以实时分析用户的行为并进行更加合理的决策, 以改善用户体验 [92], 部分情形下更容易取得用户的信任 [93]. 然而, NPC 由后台代码控制逻辑, 攻击者可以通过注入恶意代码的方式获得 NPC 的控制权, 进行非法活动. 例如, 攻击者通过操控 NPC 在用户不知情的情况下监视用户在虚拟环境中的行为, 引导用户泄露隐私或做出危害社会的行为. 该攻击方式难以察觉且通用性强, 极有可能在虚拟旅游向导、虚拟课堂教育、多人安全培训和虚拟远程协助等应用中产生安全隐患.

(2) 交互对象恶意篡改: 攻击者篡改虚拟融合网络空间中交互对象的方式包括恶意引导用户对数字对象进行不适当的修改、通过直接攻击服务器非法获取权限后篡改数字对象、通过 AR 投影篡改真实物理对象外观.

当用户在虚拟环境中与其他用户交流时, 如果用户受到攻击者的恶意引导, 使用了违背其他用户价值观的虚拟形象, 可能造成不良社会影响, 甚至遭受网络暴力或人身威胁 [94,95]. 攻击者还可能使用注入攻击等方式非法取得服务器权限 [96], 直接修改其中存储的数字对象, 通过擦除署名、数字水印等信息, 实现数字对象篡改和盗用等目的 [97]. 除了篡改 VR 场景中的数字内容, 攻击者还在 AR 场景中生成虚假的数字原生内容或篡改数字孪生内容, 干扰用户对真实场景的认知, 恶意引导用户. Lebeck 等 [67] 指出, 攻击者可通过遮挡真实环境中的关键信息或篡改真实物体外观来修改 AR 设备输出的内容, 造成安全隐患. 例如, 在汽车挡风玻璃的特定位置使用虚假交通标志覆盖真实交通标志, 或者在用户行走时遮挡用户即将面临的障碍物. 此外, 攻击者还可以利用投影技术修改用户正在交互的真实物体外观, 影响用户做出判断. Ueda 等 [98] 验证了使用 AR 技术改变食物外观, 使用户接受原本不愿食

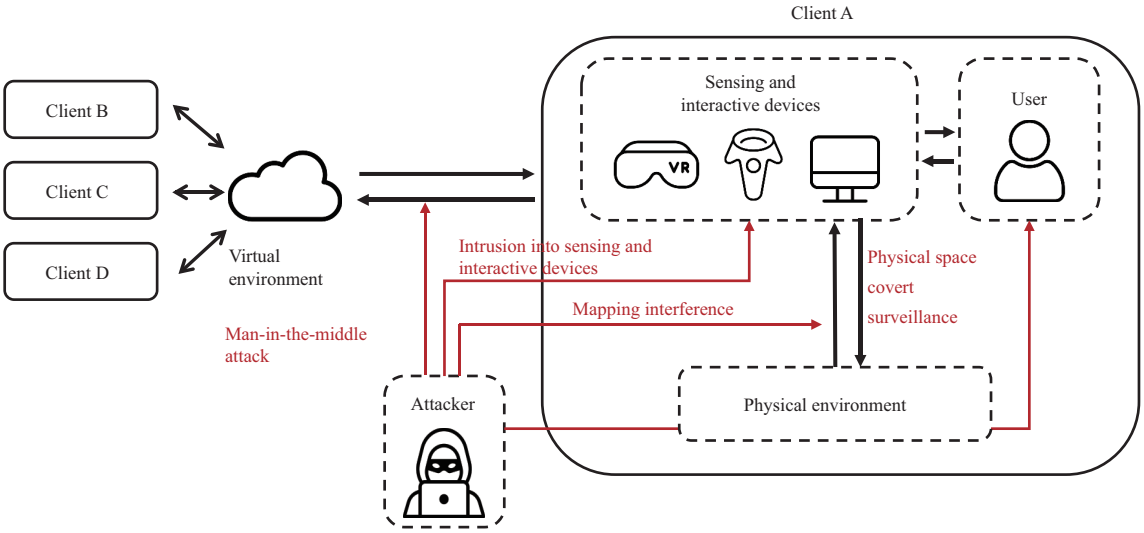

图 10 (网络版彩图) 交互内容安全问题概览

Figure 10 (Color online) Overview of interactive content security issues

用的食物的可行性; Nakano 等 [99] 进一步将该思路应用到带有前置摄像头的 VR 设备中, 提高用户使用 VR 设备时的进食体验. 以上应用场景很有可能被攻击者恶意使用, 使用户做出错误判断, 从而主动食用对身体有害的食品. 避免此类风险需要从 AR 平台和操作系统层面入手, 制定相应的规则, 限制 AR 应用对数字对象的修改. Lebeck 等 [67] 人工定义了 10 条 AR 输出管理策略, Ahn 等 [100] 使用分布广泛的雾计算和深度强化学习生成了适用范围更广的 AR 输出管理策略, 该类方法希望通过限制 AR 应用的行为和功能, 阻止对交互对象的恶意篡改.

虚拟化身盗用和交互对象恶意篡改都能误导用户判断, 产生不良影响. 前者主要通过非法获取用户或服务商虚拟形象进行欺骗, 而后者主要通过修改虚拟对象外观或属性. 为了防止交互对象安全问题发生, 虚实融合网络空间安全系统应能自动检测并阻止攻击者的身份冒用行为, 例如在创建账户时对用户的真实身份进行审核, 增强登录身份验证流程安全, 检测到可能暴露用户财产隐私时及时提醒用户注意防范. 虚拟现实传感设备的交互过程需要进行加密保护, 而虚拟现实系统需要加强对数字内容的管理和保护, 防止其数字内容被来历不明的攻击源进行恶意篡改. 此外, 虚拟现实设备要提升自检能力, 当侦测到可疑篡改行为时及时提醒用户注意防范. 同时也应该向用户定时普及安全风险教育, 让他们认识到虚拟网络空间安全的重要性.

6.3 交互内容安全

交互内容安全是指虚实融合网络空间中的具有独立身份的对象在通过传感交互设备与其他对象交互过程中, 保障传感数据、交互信息等内容在传递过程中的完整性、机密性和一致性的安全防护手段和机制. 交互内容安全问题产生的原因包括中间人攻击、传感交互设备入侵、物理空间隐蔽监视、虚实映射干扰等, 如图 10 所示.

(1) 中间人攻击: 当交互信息在虚实融合网络传递过程中, 攻击者可窃取网络数据包, 并经过篡改后再发送, 使用户收到虚假数字内容. Gulhane 等 [84] 使用 Clumsy 等软件对一个 VR 远程课堂场景进行了中间人攻击, 通过恶意丢包、制造网络拥塞, 导致用户与服务器的连接断开, Valluripally 等 [101] 进一步使用该方法篡改了用户接收到的教学内容.

(2) 传感交互设备入侵: 用户在虚实融合网络空间中与其他对象交互需要传感交互设备的支持, 包括在真实物理环境中用于记录用户物理空间动作和行为数据的相机或跟踪器, 以及记录用户声音的麦克风等. 泛联网通过传感交互设备采集交互数据, 并对其进行相应处理, 最后传递至其他具有独立身份的对象. 运动传感器准确采集并完整传递数据是保障用户在虚实融合空间中稳定运动的基础. 攻击者可干扰或入侵运动传感器来破坏用户的正常运动. Rafique 等 [102] 的研究表明, 对运动传感器进行强电信号脉冲攻击或连续的虚假脉冲信号干扰可以影响甚至操纵虚拟现实空间定位和姿态追踪系统, 从而实现对用户行走和运动的干扰和控制. 此类攻击不仅会威胁用户安全, 还会影响虚实融合空间的稳定性, 例如在未来虚实融合智慧城市中, 攻击者会通过替换硬件、修改电路和重新写入操作系统等方式干扰传感器的正常工作 [103], 使其上传错误的数据, 对系统的数据一致性造成损害, 而数据不一致的影响会很快扩散到整个虚实融合空间, 造成巨大的安全隐患 [104]. 此外, 交互过程中产生的信息蕴含用户的真实身份、心理状态、交互对象间的社交关系等, 与用户隐私与安全息息相关 [94]. 攻击者可以在用户察觉不到的情况下入侵传感交互设备, 窃取用户数据. 例如, 利用头戴式显示器中的内置耳机和麦克风监听用户间合作交流时的对话内容, 或者捕获用户头戴显示器渲染画面, 掌控用户交互情境. 这些设备甚至可以在用户未登录虚实融合网络空间时被攻击者擅自开启以捕获图像或声音信息, 窃取用户隐私 [105].

(3) 物理空间隐蔽监视: 当用户以完全沉浸的方式体验虚拟环境时, 往往难以察觉其周围现实世界中的环境, 此时攻击者可以潜入物理环境或通过安装隐蔽摄像头、窃听器的方式获取用户的行为、语音信息, 从而达到窃取用户隐私的目的. 例如, Ling 等 [106] 研究了一种基于计算机视觉和运动感知的用户交互侧信道攻击 (side-channel attack) 方式, 利用安装在用户房间内的立体摄像机捕获用户运动数据, 从而推断出用户在虚拟环境中正在输入的个人隐私信息. 通过用户运动数据可以还原出用户输入的密码, 从而威胁用户的虚拟资产.

(4) 虚实映射干扰: 攻击者通过在虚实融合网络空间中改变虚拟对象与真实对象间的虚实映射关系, 从而对用户交互造成干扰的攻击方式称为虚实映射干扰 [107]. 虚实映射干扰攻击形式包含以下三类: 在虚拟环境中创建真实场景中不存在的虚拟对象、在虚拟环境中通过图像覆盖的形式隐藏虚拟对象、在虚拟环境中渲染与真实对象不一致的虚拟对象. 不同的攻击方式对用户交互的真实性和有效性造成不同的影响, 甚至可能危害用户的人身安全. 例如, Casey 等 [85] 通过实验证实可通过入侵虚拟现实渲染系统, 在用户的视野中叠加或覆盖虚假的环境图像, 使用户迷失方向并迫使其撞击物理物体或墙壁, 对用户造成威胁.

中间人攻击、传感交互设备入侵、物理空间隐蔽监视和虚实映射干扰是 4 种主要的破坏交互内容安全的攻击手段. 中间人攻击通过篡改数字内容欺骗用户, 传感交互设备入侵通过非法获取操纵用户数据, 物理空间隐蔽监视通过获取用户隐私信息, 而虚实映射干扰通过修改虚实对应关系造成用户误判. 为了降低攻击者窃取用户交互信息以及干扰用户判断带来的影响, 可以采取如下措施: 提高 VR 设备的使用权限和虚拟环境的操控权限, 阻止来历不明的攻击者远程调用其功能对用户进行信息泄露; 加密用户之间的通信, 降低攻击者窃听并直接破译用户隐私信息的可能; 对用户的物理空间进行定期安全隐患排除, 防止攻击者安装的隐蔽监控装置窃取用户隐私信息; 此外, 为防止虚实映射干扰, 虚实融合网络空间安全系统应反复检验虚实对象的对应情况, 发现异常应及时提醒用户注意人身安全.

6.4 交互行为安全

交互行为安全是指虚实融合网络空间中的具有独立身份的对象通过传感交互设备与其他对象交互过程中, 保障其产生的交互行为或对应的交互信息不被拦截或篡改的安全防护手段和机制. 交互行

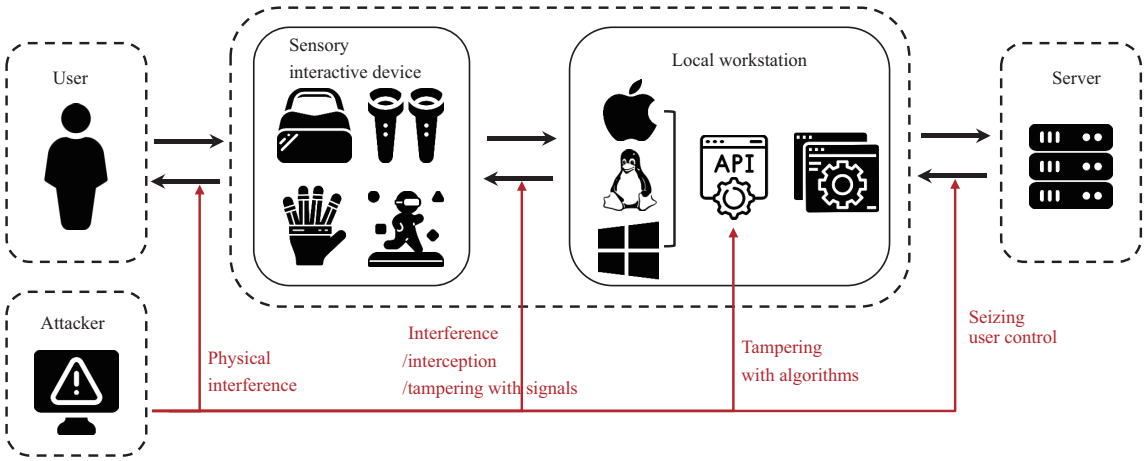

图 11 (网络版彩图) 交互行为安全问题概览

Figure 11 (Color online) Overview of interactive behavior security issues

为安全问题产生的原因包括物理干扰、拦截和篡改交互信号、篡改交互算法、截获用户控制权等, 体现在虚拟交互行为失控、物理交互行为干扰、重定向交互 3 个方面, 如图 11 所示.

(1) 虚拟交互行为失控: 虚拟智能体是指驻留在虚实融合网络空间中, 能持续自主地发挥作用, 具备驻留性、反应性、社会性、主动性等特征的数字对象, 包括虚拟代理、虚拟宠物等. 攻击者可能通过程序控制虚拟智能体产生恶意交互行为. 例如, 在多人社交应用中, 攻击者可通过恶意入侵用户的虚拟代理, 对其他用户或数字对象实施欺骗、恐吓等恶意交互行为. 此外, 攻击者可以通过拦截用户控制信号或篡改用户控制算法来截获用户控制权限, 使其无法正常控制智能体或代替用户进行控制. 例如, 在涉及控制器交互的多人合作任务中, 攻击者可以通过阻断控制信号的传输禁用用户的手柄控制, 或通过隐藏控制器的渲染模型实现在虚拟环境中隐藏用户手柄的方式欺骗用户, 使用户认为该时间段无法与其他用户进行交互, 从而取代该用户与其他用户交互的控制权 [85].

(2) 物理交互行为干扰: 除了对虚拟交互行为进行攻击, 攻击者也可以在物理空间中实时干扰用户的交互行为. 虽然头戴显示器可以隔绝用户对物理环境的视觉和听觉感知, 但是攻击者在物理世界发起的干扰与攻击仍然会对用户造成影响, 甚至导致意外发生 [107]. 为了让用户在体验虚拟内容时感知到物理空间中旁观者的存在并提前作出应对, Kudo 等 [108] 提出了一种物理空间中的旁观者可视化方法, 即当旁观者靠近用户时, 在头显中显示特定的虚拟形象让 VR 用户感知到旁观者实时空间位置和状态变化. 用户实验结果表明这种可视化方法虽然影响虚拟内容的完整呈现, 但能有效防止意外发生.

(3) 重定向交互: 人类感知系统具有视觉感知比其他感官通道感知能力更强的特点, 即当视觉通道和其他感官通道接收的刺激信号不一致时, 人脑更倾向于相信视觉刺激 [109]. 利用这一特点, 当用户在虚实融合空间进行主动交互和状态改变时, 攻击者可在用户难以察觉的情况下, 通过篡改交互过程中的关键参数和控制信息, 修改相应的视觉呈现效果, 以干扰用户对交互状态的正确感知, 影响用户对数字对象的交互操作, 此类攻击方式称为重定向交互攻击. 其中, 重定向行走技术是一种虚拟现实中的移动交互技术, 它可以实时引导用户在真实世界中的行走路径偏离其在虚拟环境中的行走路径, 从而实现在空间尺寸较小的真实物理空间中漫游大型的虚拟空间. 该技术的主要原理是通过对用户在虚拟环境中的行走路径叠加微小的增益来影响用户对虚拟环境中对方向和位置的感知, 从而引导用户偏离原定的行走路径. 当用户虚拟环境中漫游时, 攻击者可以使用重定向行走技术恶意引导用户的在物理

空间中的行走路径, 从而对用户造成安全威胁. Tseng 等 [107] 指出, 攻击者通过在虚拟环境中施加用户行走位移和方向的增益, 实现对用户行走路径的恶意操控, 引导用户行走到可能发生坠落、碰撞等意外的危险区域. 此外, Kohli [110] 通过实验证实, 当用户在虚拟空间利用接触式交互触摸不同虚拟对象时, 算法可以微调和扭曲虚拟空间的视觉呈现, 引导用户触摸同一物理对象, 实现重定向交互. 攻击者可以利用该机制, 引导用户偏离原触摸目标物, 甚至重定向至危险物体, 从而对用户造成伤害.

虚拟交互行为失控、物理交互行为干扰和重定向交互是 3 种主要的破坏用户交互行为安全的攻击手段. 虚拟交互行为失控通过控制虚拟智能体产生恶意交互, 物理交互行为干扰通过在物理空间对用户的交互产生干扰, 而重定向交互通过修改关键参数误导用户交互判断. 虚实融合网络空间安全系统可采取多种措施来应对这类问题, 例如使用加密技术来保护交互信息不被拦截或篡改, 使用身份验证和访问控制来防止未经授权的用户访问虚拟智能体的控制权与敏感信息等, 以及使用安全协议来确保交互过程的安全性. 此外, 系统还可以通过监控物理空间中的异常交互行为, 防止物理交互行为干扰和重定向交互等问题的发生.

6.5 小结

感知与交互安全是保障虚实融合网络空间用户体验的关键, 其中沉浸感知安全需要保证感知数据的正常传输以及物理设备在安全范围内输出; 交互对象安全主要涉及虚拟场景中的虚拟化身和交互对象, 攻击者仅通过获得这些虚拟资产的操控权就能利用其外观进行用户欺骗和恶意引导; 交互内容和交互行为安全则需重点防止交互数据被非法监听和篡改, 以及对用户交互行为的故意干扰导致的风险. 具体安全问题包括感知内容被篡改、数字对象被盗用、交互信息被泄露等. 为了防范各种攻击手段可能带来的安全隐患, 可以从不同层面采取相应对策: 硬件层面需提高传感交互设备的安全设计, 设置输出限制, 增强抗干扰能力; 系统层面需实现访问控制和身份验证, 监控异常情况, 限制非授权用户的操作权限; 软件层面需加密用户通信, 认证数字对象的可靠来源, 检测可能的篡改; 数据传输层面需使用安全传输协议, 避免在网络传输过程中的泄露; 算法层面需设计更智能的防御模型, 实现主动防范和攻击检测; 管理层面需制定安全规范, 对用户进行安全教育, 定期检查物理环境安全. 采取以上对策, 可以为用户感知与交互构建一个安全可靠的虚实融合环境.

7 关键基础设施与软硬件安全

虚实融合网络空间是构建在泛联网的信息基础设施之上的, 所以传统的网络安全风险仍然存在, 只是随着云服务的广泛部署, 运营商将可以集中监测和处理安全风险并通知租户及时采取对应措施, 上层的应用系统和服务商需要关注的传统安全风险将越来越少. 另外, 虚实融合网络空间的构建与运行依赖于 XR 引擎/终端、人工智能、数字孪生、区块链等基础技术引擎, 存在关键软硬件的安全风险. 关键基础设施与软硬件安全是指主干网、无线通信设施、数据服务与处理设施 (含云计算与数据存储中心) 等关键基础设施被破坏或入侵的风险, 以及操作系统、XR 引擎/终端、人工智能、数字孪生、区块链等基础软硬件被更改、伪造、拒绝服务等的风险, 如图 12 所示.

7.1 关键基础设施的安全

美国网络空间和基础设施安全委员会 (cybersecurity & infrastructure security agency, CISA) 将关键基础设施定义为 “现代社会保持国防安全、经济活力和公共卫生与安全所必需的物理和虚拟的资

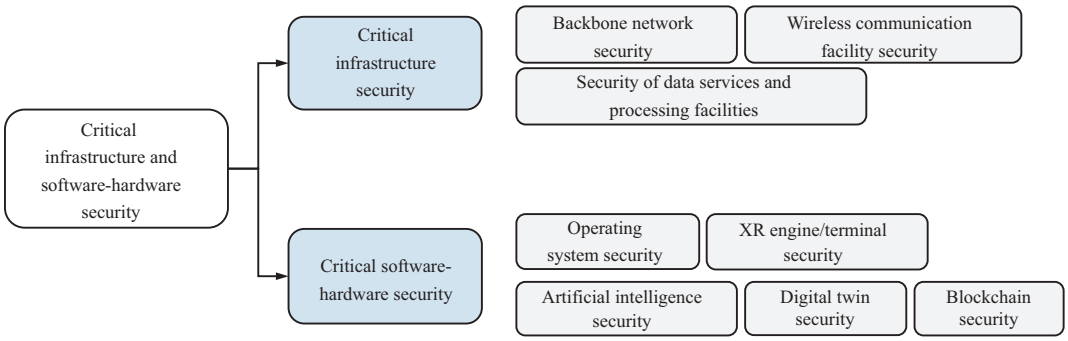

图 12 (网络版彩图) 虚实融合网络空间关键基础设施与软硬件安全的风险分类

Figure 12 (Color online) Risk classification of key infrastructure and software-hardware security of virtual-real mixing cyberspace

产、系统和网络", 包括国家安全、经济、生产、能源、网络等方面的 16 类设施[3]. 2023 年 2 月美国两党政策中心 (Bipartisan Policy Center) 发布了《网络空间安全中最大的风险 2023》[111], 提出美国关键基础设施中存在操作系统和软硬件未及时打补丁或不再维护等在内的重大风险.

虚实融合网络空间实现了物理世界、数字世界、人类世界的贯通, 其底层泛联网关键基础设施安全主要包括物理世界基础设施安全和数字化基础设施安全. 物理世界的基础设施安全一般是通过安防系统来进行保障, 网络、云服务、边缘服务等通用数字化基础设施的安全是发展较为充分的专业领域, 本文不作讨论. 目前物联网设施一般做私有化部署来提供服务, 随着数字世界中数字孪生部分的持续发展, 一些具有公用服务能力的物联网设施也可能开放和持久存在, 成为泛联网基础设施的一部分, 所以相应的传感/控制设备、通信设备、计算和存储设备也需要作为基础设施来运维以保障公共服务. IEEE 标准协会 2888 工作组[4] 正在制定信息世界和物理世界的传感器/关节点交互接口、同步性、结构、评估、全息可视化接口等方面的标准, 但尚未关注到相关安全问题.

7.2 关键软硬件的安全

关键软硬件的安全主要包括操作系统、XR 引擎/终端、人工智能、数字孪生、区块链等虚实融合网络空间运行依赖的基础软硬件的安全风险. 例如人工智能模型在数字世界中的应用将越来越多, 广泛使用的模型将可能成为基础服务被海量调用, 这将导致其攻击和防御也变得很重要. 报告《网络空间安全中最大的风险 2023》中指出广泛使用的开源软件不能及时修复漏洞以及被劫持、恶意软件的商业化分发等都是目前网络空间安全中操作层面的具体风险. 鉴于去中心化系统可能将大量出现, 美国国会已经起草了在线平台的互操作性要求 [111], 但如果终端系统的底层技术引擎被恶意破坏, 将可能影响特定群体接入泛联网空间以及正常访问. 鉴于目前泛联网操作系统、XR 引擎/终端、数字孪生、区块链相关的软硬件还在发展前期, 尚未形成垄断, 下面重点讨论更容易部署推广的人工智能模型的安全.

作为未来数字世界的支柱性基础技术, 人工智能模型的安全性已得到了很多关注和研究, 在虚实融合网络空间中, 对真实场景的感知与识别是数字世界处理的首要步骤, 人们借助于 VR/AR 设备可以在虚实融合空间中互动, 也引入了真实场景中虚假信息的攻击与防御问题. 例如用户戴着 AR 眼镜不仅可以看到物理世界, 还能看到通过后台深度学习算法识别并虚拟标注的对象或场景信息, 然而, 这

3) http://cisa.gov.
4) https://sagroups.ieee.org/2888/.

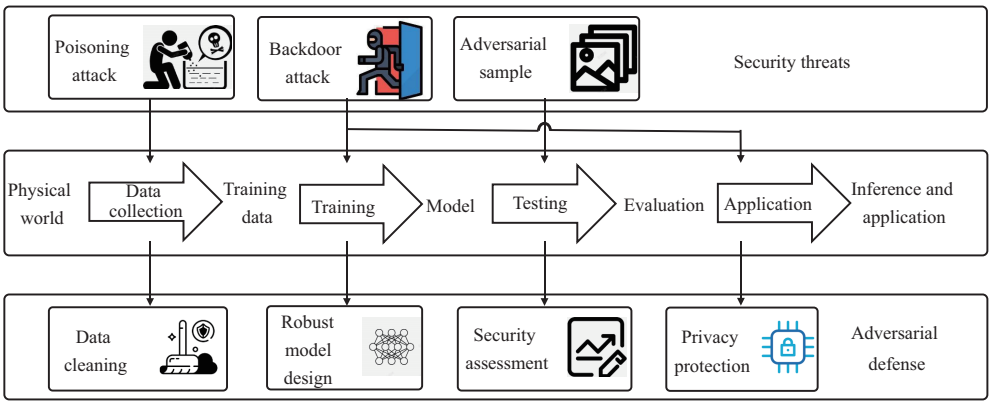

图 13 (网络版彩图) 虚实融合网络空间中人工智能模型的安全威胁与对抗防御

Figure 13 (Color online) Security threats and adversarial defenses of artificial intelligence models in virtual-real mixing cyberspaces

些深度神经网络模型可能受到虚假信息的攻击, 影响用户的信息获取正确性甚至危害其人身和财产安全, 如黑客在场景图像中注入一些难以察觉的特殊噪声或图像补丁, 以及在物理世界中部署一些经过特殊设计的纹理图像, 都可能使模型做出错误的识别, 从而损害系统功能. 虚假信息攻击可以在多个阶段产生不同的安全威胁, 如图 13 所示, 需要针对不同的攻击方式采取相应的防御策略.

(1) 虚假信息攻击: 基于神经网络的主流人工智能模型具有黑盒不可解释, 容易受到多种虚假信息的对抗性攻击, 包括对抗样本 (adversarial examples)、数据投毒 (data poisoning) 以及后门攻击 (backdoor attack) [112].

对抗样本攻击是指攻击者在模型测试和推理阶段, 通过在输入中人为注入一些人类无法察觉的特殊噪声 (如虚假信息) 来生成对抗样本, 使深度神经网络在这些特殊的输入样本上做出错误预测. Goodfellow 等 [113] 首次提出了对抗样本的概念, 并基于梯度设计了一种无目标攻击方法, 即快速梯度符号方法 (fast gradient sign method, FGSM), 通过计算一次梯度来生成对抗样本. 实验表明, 该方法可以使模型以极高的置信度作出错误预测. 由于 FGSM 仅计算一次梯度, 通常会欠拟合网络. 为此, Dong 等 [114] 提出一种基于动量迭代的攻击算法 (momentum iterative FGSM, MI-FGSM) 来增强对抗攻击, 通过将动量项集成到攻击的迭代过程中, 使得算法在迭代期间稳定更新方向, 避免局部最优, 从而提高对抗样本的可迁移性并提升攻击的成功率. 对抗补丁攻击是对抗样本的一种特殊形式. 在对抗样本攻击中, 攻击者通常希望尽可能减少扰动程度, 以避免被发现, 但是在对抗补丁攻击中, 攻击者不再将自己限制在难以察觉的变化中. 该攻击方案会生成一个与图像无关的补丁, 然后可以将此补丁放置在图像中的任何位置以攻击分类器. Thys 等 [115] 提出了一种专门针对行人对象的对抗补丁生成方法, 通过将该补丁贴在行人对象前方, 可以成功地在人员检测器中隐藏人物.

投毒攻击主要发生在数据收集与预处理阶段, 攻击者通过有意识地投放不正确或有偏差的数据来降低数据可用性, 从而影响系统模型、扰乱分析结果. 由于优化中毒样本的梯度更新过程固有的复杂性, 现有的投毒攻击主要针对单个任务设计, 难以实现不同任务间的迁移. Muñoz-González 等 [116] 首次将投毒攻击的定义扩展到多类任务, 并在此基础上, 提出了一种类似 FGSM 优化过程的、基于反向梯度优化思想的新型攻击算法. 通过自动微分计算梯度, 同时反转学习过程来进行数据投毒, 从而大大降低了攻击复杂度, 提升了投毒攻击的跨任务迁移能力.

后门攻击是一种新兴的且危害性更大的投毒攻击方式, 攻击者在模型训练阶段通过某种方式对模

型植入一些后门. 在测试评估和推理应用阶段, 当后门未被激发时, 被攻击的模型具有和正常模型类似的表现; 而当模型中隐藏的后门被激活时, 模型的输出变为攻击者预先指定的标签以达到恶意攻击的目的. 后门攻击隐蔽性极强, 因此给很多安全相关的应用 (如生物认证或自动驾驶系统) 带来极大的安全隐患. Gu 等 [117] 通过将特殊的后门图像 (如炸弹、花朵) 贴在停止标志牌上并将其表示为限速标志以在路标识别模型中生成后门. 对于不包含后门触发器的输入, 该模型可以正确分类正常的路面标志, 但一旦攻击者将后门触发器放置在输入中, 该模型将会产生错误预测, 如将停车标志识别为限速标志等.

(2) 虚假信息防御: 虚假信息防御对确保深度神经网络模型的可靠性和安全性至关重要, 例如在自动驾驶中, 人们希望驾驶系统在遇到虚假路标或目标并无法做出安全决策时, 应选择保守策略并发出警报提示驾驶人员进行干预. 深度神经网络模型需要针对不同的虚假信息攻击方式采取相应的防御策略, 包括数据清洗、鲁棒模型设计、安全评估等. Saha 等 [118] 首次提出清洗自检督训练过程中的未标注数据, 通过设计基于知识蒸馏的防御方法, 抵消有毒数据的后门攻击. Jia 等 [119] 提出了一种新的对抗训练模型框架 LAS-AT, 使用可学习的策略生成网络生成攻击策略, 对抗训练深度学习模型, 提升模型的鲁棒性. 为了评估基于深度神经网络模型的虚假信息防御能力, Ren 等 [120] 提出了一种鲁棒性验证评估方法, 量化了深度神经网络模型对抗各种攻击方式的能力. Wang 等 [121] 首次提出防御补丁的概念, 通过将数据集中每个类别的表征信息作为防御补丁随机注入到训练图像中, 增强图像的语义判别性, 从而抵御不同类型的虚假信息攻击.

7.3 小结

关键基础设施与软硬件安全是虚实融合网络空间持久运行的基础. 关键基础设施的安全主要包括物理世界基础设施安全和数字化基础设施安全, 相关研究较为充分, 但对提供公共服务的物联网基础设施安全关注不足; 关键软硬件安全还在发展前期, 近年来人工智能模型的安全性得到了很多关注和研究, 对抗样本、数据投毒、后门攻击是虚假信息攻击的研究热点, 需要针对不同的攻击方式采取相应的防御策略.

8 应用安全与网络空间治理

应用安全与网络空间治理指虚实融合网络空间运行中在经济、社会活动及国家安全等方面出现的应用层安全风险, 以及虚拟社会空间在伦理道德、政策法律等方面出现的新问题. 由于虚实融合网络空间中将大量存在去中心化系统, 其治理是一个难点. 这种新型的网络空间虽然不存在实际的集中管理者, 避免了业务上的垄断, 但去中心化是需要异构系统实现互操作的, 技术体系、引擎和标准事实上将存在底层垄断, 内容生产端涉及资源库以及先进的技术工具壁垒, 也会产生生产力和资源垄断, 后来者为了接入系统将必须遵循先行者和主导者所建立的规则, 而主导者由于某些原因将某个地区或组织进行切割时也将更为容易, 例如停止核心基础引擎的授权将可能导致某些区域整体从数字世界中下线. 同时, 现有的政府治理体系对本国的产业、资金、文化、消费等具有各种成型的监管规则和手段, 并随着时代发展不断维护升级, 但数字世界 “新大陆” 中没有 IP 地址带来的物理边界识别, 用户可以很容易将数字资产在不同平台上迁徙来规避监管. 这方面的很多问题还有待于生态发展而开始明晰, 下面将分别从应用安全、数据分级和标准制定、伦理道德与法律风险等方面进行介绍, 如图 14 所示.

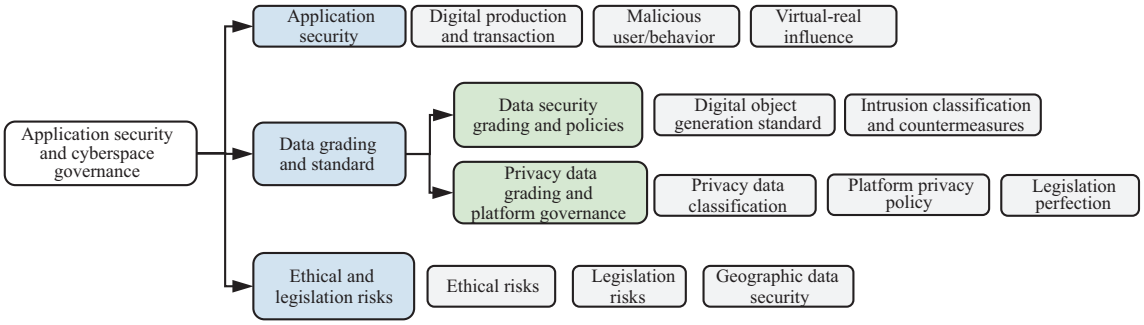

图 14 (网络版彩图) 虚实融合网络空间应用安全与网络空间治理的问题分类

Figure 14 (Color online) Classification of application security and cyberspace governance issues in virtual-real mixing

8.1 虚实融合网络空间应用安全

虚实融合网络空间利用虚拟现实/增强现实等技术, 提供了沉浸可视的场景和数字内容, 具有开放、共享和持久等特点, 支撑元宇宙等多种数字化生存和应用形态, 在社交、购物、游戏、文化、智慧城市等众多领域得到发展. 由于这些应用中用户的行为都在虚实融合空间中, 对各类数据和交互等的安全性也有高的要求, 包括机密性保护、完整性维护、可用性促进等方面.

虚实融合网络空间的应用往往涉及大量的数字内容生产、使用和交易等活动, 例如很多平台具有数字作品、数字内容和数字货币等, 具备数字市场, 并利用技术手段基于加密数字货币、NFT (non-fungible token, 非同质化通证) 等进行交易, 这些数字内容的生产制造交易等与现实世界具有大的差异 [122], 数字资产及其交易对于使用者、管理者有新的政策性约束, 对于存储、交换等也有高的技术要求, 如果缺乏有效手段, 在应用时会产生安全性问题 [4, 96]5). 与现实世界类似, 在虚实融合空间及其应用中存在恶意用户, 会利用应用平台/系统中的技术漏洞实施破坏和欺诈, 对应用安全带来隐患, 恶意用户还会利用目前政策的不完善, 通过对数字资产定价和交易实施洗钱、逃税等行为, 对安全造成影响; 恶意用户利用系统缺乏有效监管手段, 采用拒付款等获得合法数字资产, 或对合法数字资产实施非法复制等行为获取不当利益; 虚实融合网络空间应用中一些恶意行为也会削弱公平机制的作用 [123], 例如对市场垄断获取利益、对数字资产的不合理定价获得暴利、利用 AI 等技术窃取资源等.

虚实融合网络空间可在数字世界和物理世界建立桥梁, 这种互操作机制为虚实交互提供便利, 也对物理世界和人类世界带来安全性的隐患. 例如, 在医疗方面, Li 等 [124] 提出了一种沉浸式的个性化手术模拟系统, 可基于临床数据建立个性化人体器官可交互数字孪生模型, 通过对人体器官数字孪生模型进行防攻击、防篡改、防挟持保护, 可有效提升患者敏感生物信息安全, 并减少由于恶意攻击带来的医疗风险. 此外, 通过对数字世界中设施的攻击, 可对物理世界的基础设施 (如智能电网 [125]) 等造成侵害; 虚实融合网络空间的高沉浸感, 会使得用户脱离物理世界, 通过对用户接入虚实融合网络空间设备的攻击, 可窃取用户信息并在物理世界实施破坏 [85], 同时高沉浸感还容易造成用户的沉迷, 从而使得用户对物理世界封闭 [126] 等.

目前应用中已产生诸多安全性问题, 涉及大量资金, 例如 2021 年 11 月, “进化猿” NFT 负责人进行项目欺诈, 项目上线一周后消失, 其官方推特账号和网站均消失, 卷走价值约 270 万美元的以太币; 2022 年基于 binance smart chain (BSC) 的元宇宙项目 Paraluni 由于系统漏洞等损失 170 万美元; 2022 年 6 月 4 日, NFT 项目 BAYC 的聊天软件 Discord 服务器被攻击, 攻击者盗取了 BAYC 社区管理者

5) 邓亮. 元宇宙业态下新型经济犯罪风险及其治理. 《人民论坛》. 2022.9. http://www.71.cn/2022/0915/1179151.shtml.

的账号和价值 40 万美元的 NFT, 事后攻击者很快在交易平台卖掉盗窃的 NFT, 转换成以太币等主流数字货币, 并洗钱和切断资金链路逃避追踪, 这类安全性风险在国际上日益得到重视.

8.2 数据分级和标准制定

(1) 数据安全分级与安全策略: 海量数据需要可靠的安全分级进行甄别和分类, 并使用对应的安全策略进行加密和保护. 同时, 数据安全保护需要制定标准、规则和统一的协议, 以便不同设备和平台兼容安全策略. 数字孪生系统的标准规范主要包括孪生对象生成规范 [127] 以及对应不同入侵攻击的分类和策略规范 [128]. Alshammari 等 [127] 针对智慧城市领域数字孪生对象和物理实体间的双向映射安全问题, 提出了安全构建数字孪生对象的相关标准和可扩展规范. 他们对该标准的建立提出了 3 条具体的建议: 扩展已有的建筑信息模型 (building information model, BIM), 使其支持实时信息流; 改进 BIM 标准, 使其支持国际上现有的通用网络安全框架来保证信息的真实性、可识别性和可访问性; 确保数字孪生应用满足物联网的安全需求, 即开发一个新的 “安全层” 来支持和管理即将出现的智能应用架构. Alcaraz 等 [38] 分析了目前数字孪生的构建范式, 并将其分为 4 层. 其中物理层从物理空间捕获动态信息, 并为物理资产准备控制指令; 数据管理和数据同步层负责规范异构的多源数据形式; 数据建模和服务层运行数据处理模型并提供维护、监控、安全诊断等服务; 数据可视化和访问层: 允许最终用户和实体访问可视化或模拟结果. 针对不同攻击对数字孪生系统的影响, Hussaini 等 [128] 设计了一个系统分类法, 从四层体系结构的角度探讨了对基于数字孪生的信息物理系统的不同攻击以及它们是如何影响系统的. 他们提出了该类 CPS 的四层攻击空间 (即对象层、通信层、数字孪生层和应用层)、三层攻击对象 (即机密性、完整性和可用性), 以及结合强度和知识的攻击类型. 同时, 该团队还提出了一种称为安全数字孪生开发生命周期的防御机制, 并指出了综合利用其他技术 (入侵检测、区块链、建模、模拟和仿真) 来保护该类 CPS 的重要性.

(2) 隐私数据分级与平台治理: 用户隐私数据一般使用不同类型的设备采集, 其信息格式和隐私内容种类十分不同. 为了能够制定更加完备的隐私保护策略、安全管理不同类型的隐私信息, 对标目前互联网使用的成熟协议, 制定隐私数据的分类和分析标准是需要研究的问题. Maloney 等 [129] 指出, 大多数的虚拟现实社交平台 (如 AltspaceVR, RecRoom 和 VRchat) 并没有清楚地告知用户所采集信息的保密程度, 即隐私划分的标准. Alcaraz 等 [38] 分析了数字孪生范式的现状, 并对潜在威胁、功能层次、操作要求进行了有效分类. 同时, 他们还提供了一套安全建议和方法, 以保护数字孪生体的隐私信息. 同时, 虚拟世界的社会活动和用户之间的交互主要发生在由个人或国家所构建的虚拟平台中, 网络平台实质上掌握大量的个人隐私和社会隐私信息, 对平台隐私保护策略和治理方法的研究也是重要的一环. Nwaneri [130] 指出, 脸书有对用户进行试验的记录 (例如, 从订阅中删除负面或正面新闻会影响用户的情绪). 而 Oculus 公司的安全政策也允许脸书进行类似的实验. 考虑到泛虚拟现实系统收集的数据类型, 用户的隐私所面临的风险甚至会比在社交网络中发现的风险更高. 同时, 关于收集 “用户使用虚拟现实头盔时的身体运动和尺寸信息” 的政策声明有可能引发歧视和大规模监控. 因此, 出于对大规模用户隐私的保护, 需要这些泛虚拟现实平台明确自定义隐私保护策略、向用户科普隐私信息的敏感性、服从国家相关治理法规, 防止隐私数据大规模泄露. 然而, 对于虚实融合空间隐私相关法律制定是一个全新的问题, 理应完善相关法律约束范围和制定相对应的惩戒措施. 美国沃特金斯律师事务所的 Lake [131] 曾发文指出, 目前美国在线身份盗用相关的法律并没有充分保护虚拟环境中的用户隐私, 这些法律的结合使原告没有任何一方可以起诉: 匿名法律使得不可能起诉身份占用者, 而 ISP 豁免权也阻止了受害者起诉虚拟环境平台提供商. 在这些问题得到解决之前, 虚实融合环境中身份盗窃的受害者甚至无法向美国法院提出他们的索赔. 因此, 与虚实融合空间相关的法律制定是维护各平

台运行和人类隐私的重要保障工具, 是未来需要探讨的重要方向.

8.3 伦理道德与法律风险

虚实融合网络空间中伦理道德的风险类型与网络虚拟空间内容和服务提供商的理念设计, 空间参与群体的兴趣、风俗习惯与观念以及物理世界中社会发展状态等都有错综复杂的关系, 涉及的因素很多, 但目前比较容易形成共识的是虚拟形象需要进行行为伦理规范, 无论是虚拟化身或者是空间提供的智能角色都应该遵循. 例如在网络虚拟空间中进行不断注视或跟随、远处监视、化身或言语的骚扰、不正当的监视 [81], 在公共场合做出不雅行为或者违反医学伦理的虚拟实验等都应该禁止. 值得注意的是智能角色的存在会影响人类社会关系, 如果人工智能算法的服务商从经济利益、恶意心理或其他驱动出发, 提供一些引导用户价值导向、破坏社会文化传统、传播伪科学或邪教学说行为的智能模型, 这些模型并没有明确的关键词或显著行为特征, 不借助于技术辅助统计分析很难发现, 因而具有很强的隐蔽性, 将可能破坏物理世界的社会关系稳定性并造成损害.

虚实融合网络空间是跨国界、多种族、虚拟存在的拟真交互环境, 其立法、取证、限制与执行将存在较大的问题. Casey 等 [85] 提出黑客可以采集记录目标用户的生活规律以及实时位置, 方便在物理世界入室盗窃或伤害, 或在虚拟世界跟踪骚扰用户虚拟化身, 后者难以用法律实施制裁. 虚拟空间的行为可能导致物理世界的后果, 例如通过利用 VR 设备的缺陷来重置硬件的物理边界, 将用户 “推下” 楼梯或误导到马路上导致伤害, 是在虚拟空间还是物理世界的哪一方所在地发起诉讼则是一个问题. 比特币在 “暗网” 中已被广泛用作洗钱和进行非法交易, 这个问题在虚实融合网络空间中将同样存在 [132]. 由于人工智能技术的应用, 将带来虚拟角色和自动无人系统的违法犯罪问题, 其界定责任也是一个需要细化分类和专门规范的问题.

此外, 虚实融合网络空间将大量用到地理信息数据, 而这涉及国家安全, 各国法律都有对测绘数据的保密条例. 2022 年 8 月自然资源部发布了《促进智能网联汽车发展维护测绘地理信息安全的通知》, 将智能联网汽车归为测绘工具, 并将车企、地图提供商和智能驾驶软件提供商三类企业定义为测绘活动行为的主体, 车主和驾乘人员不需要为车辆从事的测绘活动承担法律责任. 实际上, 以人为主体的 AR 导航也可能被采集地理信息数据用于众包地图导航, 其设备和软件作为个人消费电子产品, 不会像无人驾驶车辆那样被严格注册和管理, 保密级测绘数据的采集和传递过程更加隐秘, 带来了地理信息空间数据的保密风险.

8.4 小结

虚实融合网络空间的应用安全与网络空间治理不是仅依靠技术突破和工程研发能解决的, 需要各国政府和国际组织来进行有组织的研究和协作解决. 在应用安全方面, 跨国合作追查案件将变得很必要和普遍, 这需要各国政府能认识到并形成共识, 突破现有的一些壁垒和限制. 数据安全分级和标准制定需要对虚实融合网络空间中的海量数据进行分类和甄选, 并使用不同等级的安全策略进行储存和保护. 特别地, 数据和隐私分级标准的制定有利于对大型网络平台进行监管和审查, 在保证隐私数据安全的同时, 也有助于对隐私侵犯事件进行调查和责任界定. 同时, 虚实融合网络空间的伦理道德与法律风险具有很强的隐蔽性, 其立法、取证、限制与执行需要技术社区与政府机构密切合作.

9 需解决的问题与未来方向

网络空间安全是有关国家安全、经济发展、社会稳定和人民利益的重大问题, 各国政府高度重视,

持续加强安全技术研发, 并制定各种管理办法确保自身的网络空间安全, 不断取得成效. 同时, 网络空间安全也是学科高度交叉的科学技术研究领域, 仍有许多尚未解决的技术问题, 而且随着网络空间边界的不断扩张, 新的安全问题不断涌现. 我们根究对虚实融合网络空间安全国际研究现状和未来发展趋势的深入分析, 结合我们的相关研究体会, 归纳出如下 10 个虚实融合网络空间需解决的问题或发展方向.

(1) 智能体用户身份认证的安全问题. 随着人工智能等技术的发展, 真实世界和数字世界的智能体将成为虚实融合空间的一类重要用户, 其对资源的访问和控制也需要身份认证和权限控制, 这类智能体的身份信息也可能被窃取, 或自身也可能被恶意用户操控以窃取空间资源. 如何有效对智能体用户进行身份认证, 以及制定相应的法规约束是未来需要考虑的重要内容.

(2) 虚实融合网络空间中违法行为的数字取证. 目前网络空间的数字取证可以通过过程记录、截屏录像、网页公证等方式进行. 但是, 在虚拟场景中, 虚拟化身之间的交互与交易可能不借助文字、语音、图标、图片等传统媒体文件, 而代以实时音视频和姿态动作等流式数据, 同时, 每个终端用户只能接收到整个虚拟世界中的可见视野信息. 多方音视频流的数据量巨大, 流式动作序列难以被准确识别和长期保存, 以及用户只接收到部分信息, 这些因素导致虚实融合网络空间中的交互和交易过程无法被完整保存到云端甚至本地, 数字取证是一个挑战性问题.

(3) 数字对象数据全生命周期的入侵智能检测. 数字对象的数据产生、传输、更新、存储、应用、销毁等全生命周期过程中都有被外界入侵的风险. 特别地, 对数字孪生体的攻击可能会对其物理实体产生影响. 因此, 如何快速准确识别网络入侵类型及攻击手段是进行数据保护、化解网络攻击、制定防御策略的基础. 同时, 如何对数字对象数据全生命周期内所受攻击的影响规模、入侵来源、攻击手段进行智能识别和诊断也亟待深入研究.

(4) 数字对象数据的安全分级测评与安全操控策略制定. 数字对象在构建、使用和维护过程中会涉及多时空数据, 这些数据不仅存在多模态、多粒度、多来源、多用途等特点, 而且也因重要性的不同存在不同级别的增、删、查、改安全防护要求. 因此, 如何对这些数据的潜在安全风险进行分级测评是需要解决的基础性问题. 同时, 对于不同类型和用途的数字对象, 由于所需采集的信息种类以及采集设备、储存格式、安全等级、应对攻击的防御策略均不尽相同, 如何制定统一的数字对象数据安全操控策略, 使其可以适配不同数据源和信息采集设备并能够融合现有网络底层协议, 是一个重要的研究方向.

(5) 去中心化架构下数字对象数据加密和安全存储. 在去中心化架构下, 数字对象的部署涉及物理实体和信息的分布式远程控制, 为了防止数字对象被挟持和篡改, 这对数字对象的身份管理、身份验证、授权及加密提出了越来越高的要求. 如何对数字对象数据进行去中心化加密和管理, 保证传输和访问的数据真实性以及防篡改和防非法访问能力, 都是待解决的重要问题.

(6) 数字对象身份的统一编码表征. 若没有统一、安全的标识符作为数字对象的唯一标识, 这将导致不同的系统之间无法相互认证, 难以形成安全可信的通信连接, 并限制数字对象的规模. 目前通常只根据单一因素生成数字对象的身份标识, 尚难以做到按区域、按系统、按内部职能等因素动态分配身份标识和进行统一整合管理. 因此, 随着虚实融合网络空间的扩展, 如何统筹考虑标识冲突问题、统一标识分配的资源消耗问题, 建立一种灵活、安全、能够避免标识冲突的数字对象身份统一编码表征, 并建立相应的安全保护机制, 是虚实融合网络空间安全的重要研究方向之一.

(7) 生物信息采集过程的标准规范建立. 虚实融合网络空间的人机交互一般需采集隐私相关的生物特征信息, 例如面部表情、眼睛/手部运动、语音和生物特征, 甚至脑电波模式. 虽然这些生物特征与交互设备可以建立准确的物理关联映射, 但通过旁路攻击、运动模式分析、视频分析等技术

手段也可能会造成个人隐私泄露问题. 因此, 建立生物信息采集过程的标准规范是保护个人用户隐私和生物信息的重要前提. 如何将不同类型的生物特征交互输入信息进行分类并据此进行软/硬件加密, 以及如何通过对抗攻击学习等方式对交互输入进行虚拟替换等, 都是具有重要实用价值的研究方向.

(8) 虚实融合网络空间交互过程中的物理环境探测和人身安全预警. 用户在虚实融合网络空间中的交互往往会隔绝自身对外部物理环境的感知, 以获得更加沉浸式的体验. 在用户与远程协作者进行交互过程中, 物理环境与虚拟环境的不一致性可能导致用户与物理空间的障碍物、边界以及其他用户发生意外碰撞, 甚至会导致严重的人身安全威胁. 如何通过计算机网络连接的多种传感器, 设计和部署自动感知真实环境的人工智能算法, 并利用视、听、触觉等多模态方式对用户进行安全预警, 是未来值得重点研究的安全问题之一.

(9) 传感交互设备防火墙系统的安全规范. 在虚实融合网络空间中, 用户各类交互数据的采集高度依赖传感设备. 为确保数据的安全和完整性, 传感交互设备需具有完善的防火墙系统, 通过监控网络流量和数据传输以防止未授权访问、数据泄漏和网络攻击等安全问题. 如何建立传感交互设备防火墙系统的安全规范, 保证传感设备间的兼容性和一致性, 提升传感交互的安全性和可靠性也是值得研究的问题.

(10) 虚实融合网络空间中的认知对抗. 网络空间中的媒体传播广泛并深刻地影响着人们的认知, XR 设备与交互、脑机接口、生成式 AI 等技术正推动产生元宇宙等新型社交环境, 这将进一步重构人类的数字化生存模式, 丰富网络环境的认知空间. 另一方面, 虚实融合结合人工智能为用户提供更多服务的同时, 也将带来新的问题, 如开放场景下虚假和异常信息的后门攻击和认知影响等. 如何检测识别虚实融合网络空间中的认知操纵意图, 防范对网络用户的认知渗透, 是一个十分具有挑战的问题.

10 结束语

在互联网、物联网以及新兴的虚拟现实、人工智能、数字孪生、区块链等技术的推动下, 未来将可以通过泛联网实现物理、人类和数字三界的贯通融合, 从而发展出虚实融合网络空间, 形成人、机、物三元混合空间, 极大地拓展网络空间的边界, 带来全新的大众体验、社交形态、生产模式和数字经济发展路径. 与此同时, 这一新型网络空间应用也必然地带来了新的安全与隐私保护问题. 本文全面分析综述了国际上对这些问题的研究现状和发展趋势, 并提出了需要解决的一些问题和未来方向. 当前, 对虚实融合网络空间相关的安全风险和隐私保护研究正处于起步阶段, 随着泛联网的逐步成熟和虚实融合网络空间应用的普及与发展壮大, 将会大量出现具有重要影响的虚实融合网络空间安全事件, 相关领域的同行们需要重视并提前开展相关研究, 为维护我国和人类共同的虚实融合网络空间安全作出贡献.

参考文献

1 Zhao Q P. A survey on virtual reality. Sci Sin Inform, 2009, 39: 2–46 [赵沁平. 虚拟现实综述. 中国科学: 信息科学, 2009, 39: 2–46]

2 Zhou Z, Zhou Y, Xiao J J. Survey on augmented virtual environment and augmented reality. Sci Sin Inform, 2015, 45: 157–180 [周忠, 周颐, 肖江剑. 虚拟现实增强技术综述. 中国科学: 信息科学, 2015, 45: 157–180]

3 Tao F, Zhang H, Liu A, et al. Digital twin in industry: state-of-the-art. IEEE Trans Ind Inf, 2018, 15: 2405–2415

4 Wang Y T, Su Z, Zhang N, et al. A survey on metaverse: fundamentals, security, and privacy. IEEE Commun Surv Tutorials, 2023, 25: 319–352

5 Lee L H, Braud T, Zhou P, et al. All one needs to know about metaverse: a complete survey on technological singularity, virtual ecosystem, and research agenda. 2021. ArXiv:2110.05352
6 Wang P Q, Luo H, Sun Y. Internet of Things oriented multi-identification mapping model. Sci Sin Inform, 2013, 43: 1244–1264 [王平泉, 罗红, 孙岩. 面向物联网的多元标识映射模型. 中国科学: 信息科学, 2013, 43: 1244–1264]
7 George C, Khamis M, von Zezschwitz E, et al. Seamless and secure VR: adapting and evaluating established authentication systems for virtual reality. In: Proceedings of the 24th Network and Distributed System Security Symposium (NDSS), 2017
8 Yu Z, Liang H N, Fleming C, et al. An exploration of usable authentication mechanisms for virtual reality systems. In: Proceedings of the IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), 2016. 458–460
9 Olade I, Liang H N, Fleming C, et al. Exploring the vulnerabilities and advantages of swipe or pattern authentication in virtual reality (VR). In: Proceedings of the 4th International Conference on Virtual and Augmented Reality Simulation (ICVARS), Sydney, 2020. 45–52
10 Mathis F, Williamson J H, Vaniea K, et al. RubikAuth: fast and secure authentication in virtual reality. In: Proceedings of the Extended Abstracts of CHI Conference on Human Factors in Computing System, Hawaii, 2020. 1–9
11 Mathis F, Williamson J H, Vaniea K, et al. Fast and secure authentication in virtual reality using coordinated 3D manipulation and pointing. ACM Trans Comput Hum Interact, 2021, 28: 1–44
12 Funk M, Marky K, Mizutani I, et al. LookUnlock: using spatial-targets for user authentication on HMDs. In: Proceedings of the Extended Abstracts of CHI Conference on Human Factors in Computing Systems, Scotland, 2019. 1–6
13 Schneegass S, Oualil Y, Bulling A. SkullConduct: biometric user identification on eyewear computers using bone conduction through the skull. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, San Jose, 2016. 1379–1384
14 Raja K B, Raghavendra R, Stokkenes M, et al. Multi-modal authentication system for smartphones using face, iris and periocular. In: Proceedings of the IEEE International Conference on Biometric (ICB), Phuket, 2015. 143–150
15 Venkatasubramanian K K, Banerjee A, Gupta S K S. PSKA: usable and secure key agreement scheme for body area networks. IEEE Trans Inform Technol Biomed, 2010, 14: 60–68
16 Mustafa T, Matovu R, Serwadda A, et al. Unsure how to authenticate on your VR headset? Come on, use your head! In: Proceedings of the ACM International Workshop on Security and Privacy Analytics, Tempe, 2018. 23–30
17 Sivasamy M, Sastry V N, Gopalan N P. VRCAuth: continuous authentication of users in virtual reality environment using head-movement. In: Proceedings of the 5th IEEE International Conference on Communication and Electronics Systems (ICCES), Coimbatore, 2020. 518–523
18 Pfeuffer K, Geiger M J, Prange S, et al. Behavioural biometrics in VR: identifying people from body motion and relations in virtual reality. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, 2019. 1–12
19 Kupin A, Moeller B, Jiang Y, et al. Task-driven biometric authentication of users in virtual reality (VR) environments. In: Proceedings of the 25th International Conference Multi-Media Modeling, Greece, 2019. 55–67
20 Ajit A, Banerjee N K, Banerjee S. Combining pairwise feature matches from device trajectories for biometric authentication in virtual reality environments. In: Proceedings of the 2nd IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), San Diego, 2019. 9–97
21 Liebers J, Abdelaziz M, Mecke L, et al. Understanding user identification in virtual reality through behavioral biometrics and the effect of body normalization. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, Yokohama, 2021. 1–11
22 Miller R, Banerjee N K, Banerjee S. Within-system and cross-system behavior-based biometric authentication in virtual reality. In: Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Atlanta, 2020. 311–316
23 Miller R, Banerjee N K, Banerjee S. Using siamese neural networks to perform cross-system behavioral authentication in virtual reality. In: Proceedings of the IEEE Virtual Reality and 3D User Interfaces (VR), Lisbon, 2021. 140–149
24 Miller R, Banerjee N K, Banerjee S. Combining real-world constraints on user behavior with deep neural networks for virtual reality (VR) biometrics. In: Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Christchurch, 2022. 409–418
25 Shen Y R, Wen H K, Luo C W, et al. GaitLock: protect virtual and augmented reality headsets using gait. IEEE Trans Dependable Secure Comput, 2018, 16: 484–497
26 Rogers C E, Witt A W, Solomon A D, et al. An approach for user identification for head-mounted displays.

In: Proceedings of the ACM International Symposium on Wearable Computers (ISWC), Osaka, 2015. 143–146
27 Holland C, Komogortsev O V. Biometric identification via eye movement scanpaths in reading. In: Proceedings of the International Joint Conference on Biometrics (IJCB), Washington, 2011. 1–8
28 Luo S Q, Nguyen A, Song C, et al. OcuLock: exploring human visual system for authentication in virtual reality head-mounted display. In: Proceedings of the 27th Network and Distributed System Security Symposium (NDSS), 2020
29 Lohr D, Berndt S H, Komogortsev O. An implementation of eye movement-driven biometrics in virtual reality. In: Proceedings of the ACM Symposium on Eye Tracking Research & Applications (ETRA), Warsaw, 2018. 1–3
30 Lohr D J, Aziz S, Komogortsev O. Eye movement biometrics using a new dataset collected in virtual reality. In: Proceedings of the ACM Symposium on Eye Tracking Research and Applications (ETRA), Stuttgart, 2020. 1–3
31 Olade I, Fleming C, Liang H N. BioMove: biometric user identification from human kinesiological movements for virtual reality systems. Sensors, 2020, 20: 2944
32 Liang Z, Tan F, Chi Z R. Video-based biometric identification using eye tracking technique. In: Proceedings of the IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC), Hong Kong, 2012. 728–733
33 Zhu H D, Jin W Q, Xiao M Y, et al. BlinKey: a two-factor user authentication method for virtual reality devices. Proc ACM Interact Mob Wearable Ubiquitous Technol, 2020, 4: 1–29
34 Mathis F, Fawaz H I, Khamis M. Knowledge-driven biometric authentication in virtual reality. In: Proceedings of the Extended Abstracts of CHI Conference on Human Factors in Computing Systems, Honolulu, 2020. 1–10
35 Wang J, Gao B Y. Analysis of multi-attribute user authentication to against man-in-the-room attack in virtual reality. In: Proceedings of the HCI International Conference, 2021. 455–461
36 von Willich J, Funk M, Müller F, et al. You invaded my tracking space! Using augmented virtuality for spotting passersby in room-scale virtual reality. In: Proceedings of the Designing Interactive Systems Conference (DIS), San Diego, 2019. 487–496
37 Liebers J, Schneegass S. Introducing functional biometrics: using body-reflections as a novel class of biometric authentication systems. In: Proceedings of the Extended Abstracts of CHI Conference on Human Factors in Computing Systems, Honolulu, 2020. 1–7
38 Alcaraz C, Lopez J. Digital twin: a comprehensive survey of security threats. IEEE Commun Surv Tutorials, 2022, 24: 1475–1503
39 Eckhart M, Ekelhart A. Digital twins for cyber-physical systems security: state of the art and outlook. In: Security and Quality in Cyber-Physical Systems Engineering. Cham: Springer, 2019. 383–412
40 Eckhart M, Ekelhart A. Towards security-aware virtual environments for digital twins. In: Proceedings of the ACM Workshop on Cyber-Physical System Security, Nagasaki, 2018. 61–72
41 Liu J, Li C L, Bai J P, et al. Security in IoT-enabled digital twins of maritime transportation systems. IEEE Trans Intell Transp Syst, 2021, 24: 2359–2367
42 Susila N, Sruthi A, Usha S. Impact of cloud security in digital twin. Adv Comput, 2020, 117: 247–263
43 Suhail S, Hussain R, Jurdak R, et al. Trustworthy digital twins in the industrial Internet of Things with blockchain. IEEE Internet Comput, 2021, 26: 58–67
44 Seth B, Dalal S, Jaglan V, et al. Integrating encryption techniques for secure data storage in the cloud. Trans Emerging Tel Tech, 2022, 33: e4108
45 Sajay K R, Babu S S, Vijayalakshmi Y. Enhancing the security of cloud data using hybrid encryption algorithm. J Ambient Intell Hum Comput, 2019. doi: 10.1007/s12652-019-01403-1
46 Aks S M Y, Karmila M, Givan B, et al. A review of blockchain for security data privacy with metaverse. In: Proceedings of the IEEE International Conference on ICT for Smart Society (ICISS), 2022. 1–5
47 Bécue A, Fourastier Y, Praça I, et al. CyberFactory# 1—securing the industry 4.0 with cyber-ranges and digital twins. In: Proceedings of the IEEE International Workshop on Factory Communication Systems (WFCS), Imperia, 2018. 1–4
48 Wu J Y, Guo J K, Lv Z H. Deep learning driven security in digital twins of drone network. In: Proceedings of the IEEE International Conference on Communications, Seoul, 2022. 1–6
49 de Benedictis A, Esposito C, Somma A. Toward the adoption of secure cyber digital twins to enhance cyber-physical systems security. In: Proceedings of the International Conference on the Quality of Information and Communications Technology, Talavera de la Reina, 2022. 307–321
50 Bitton R, Gluck T, Stan O, et al. Deriving a cost-effective digital twin of an ICS to facilitate security evaluation.

In: Proceedings of the 23rd European Symposium on Research in Computer Security, Barcelona, 2018. 3–7
51 Ramezanpour K, Jagannath J, Jagannath A. Security and privacy vulnerabilities of 5G/6G and WiFi 6: survey and research directions from a coexistence perspective. Comput Networks, 2023, 221: 109515
52 Sicari S, Rizzardi A, Coen-Porisini A. Security & privacy issues and challenges in NoSQL databases. Comput Networks, 2022, 206: 108828
53 Babun L, Denney K, Celik Z B, et al. A survey on IoT platforms: communication, security, and privacy perspectives. Comput Networks, 2021, 192: 108040
54 Martinovic I, Davies D, Frank M, et al. On the feasibility of side-channel attacks with brain-computer interfaces. In: Proceedings of the USENIX Security Symposium, Bellevue, 2012. 143–158
55 Bozkir E, Geisler D, Kasneci E. Person independent, privacy preserving, and real time assessment of cognitive load using eye tracking in a virtual reality setup. In: Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, 2019. 1834–1837
56 John B, Koppal S, Jain E. EyeVEIL: degrading iris authentication in eye tracking headsets. In: Proceedings of the ACM Symposium on Eye Tracking Research & Applications, Denver, 2019. 1–5
57 John B, Jörg S, Koppal S, et al. The security-utility trade-off for iris authentication and eye animation for social virtual avatars. IEEE Trans Visual Comput Graphics, 2020, 26: 1880–1890
58 Chen Y F, Du Y C, Xu C L, et al. ArmSpy: video-assisted PIN inference leveraging keystroke-induced arm posture changes. In: Proceedings of the IEEE Conference on Computer Communications, London, 2022. 1878–1887
59 Luo S Q, Hu X Y, Yan Z S. HoloLogger: keystroke inference on mixed reality head mounted displays. In: Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Christchurch, 2022. 445–454
60 Al Arafat A, Guo Z S, Awad A. VR-Spy: a side-channel attack on virtual key-logging in VR headsets. In: Proceedings of the IEEE Virtual Reality and 3D User Interfaces (VR), Lisbon, 2021. 564–572
61 Li S, He X X, Song W F, et al. Graph diffusion convolutional network for skeleton based semantic recognition of two-person actions. IEEE Trans Pattern Anal Mach Intell, 2023, 45: 8477–8493
62 Buck L E, Bodenheimer B. Privacy and personal space: addressing interactions and interaction data as a privacy concern. In: Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Lisbon, 2021. 399–400
63 Li S, Savaliya S, Marino L, et al. Brain signal authentication for human-computer interaction in virtual reality. In: Proceedings of the IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), New York, 2019. 115–120
64 Groshev M, Guimarães C, Martín-Pérez J, et al. Toward intelligent cyber-physical systems: digital twin meets artificial intelligence. IEEE Commun Mag, 2021, 59: 14–20
65 Ruth K, Kohno T, Roesner F. Secure multi-user content sharing for augmented reality applications. In: Proceedings of the USENIX Security Symposium, Santa Clara, 2019. 141–158
66 Ritzdorf H, Soriente C, Karame G O, et al. Toward shared ownership in the cloud. IEEE Trans Inform Forensic Secur, 2018, 13: 3019–3034
67 Lebeck K, Ruth K, Kohno T, et al. Securing augmented reality output. In: Proceedings of the IEEE Symposium on Security and Privacy (SP), San Jose, 2017. 320–337
68 Zhang S L, Li S, Hao A, et al. Point cloud semantic scene completion from RGB-D images. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2021. 35: 3385–3393
69 Zhang S L, Hao A, Qin H. Knowledge-inspired 3D scene graph prediction in point cloud. In: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), 2021. 34: 18620–18632
70 Sun J M, Shen Z H, Wang Y A, et al. LoFTR: detector-free local feature matching with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021. 8922–8931
71 Dusmanu M, Rocco I, Pajdla T, et al. D2-Net: a trainable CNN for joint description and detection of local features. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, 2019. 8092–8101
72 Pittaluga F, Koppal S J, Kang S B, et al. Revealing scenes by inverting structure from motion reconstructions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, 2019. 145–154
73 Vilk J, Molnar D, Livshits B, et al. SurroundWeb: mitigating privacy concerns in a 3D web browser. In: Proceedings of the IEEE Symposium on Security and Privacy, San Jose, 2015. 431–446
74 Lin P Y, You B, Lu X Y. Video exhibition with adjustable augmented reality system based on temporal psycho-visual modulation. J Image Video Proc, 2017, 2017: 1–11

75 Valluripally S, Gulhane A, Mitra R, et al. Attack trees for security and privacy in social virtual reality learning environments. In: Proceedings of the 17th IEEE Consumer Communications & Networking Conference (CCNC), Las Vegas, 2020. 1–9

76 Suhail S, Hussain R, Jurdak R, et al. Blockchain-based digital twins: research trends, issues, and future challenges. ACM Comput Surv, 2022, 54: 1–34

77 de Donno M, Giaretta A, Dragoni N, et al. Cyber-storms come from clouds: security of cloud computing in the IoT era. Future Internet, 2019, 11: 127

78 Speciale P, Schonberger J L, Sinha S N, et al. Privacy preserving image queries for camera localization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR), Long Beach, 2019. 1486–1496

79 Speciale P, Schonberger J L, Kang S B, et al. Privacy preserving image-based localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, 2019. 5493–5503

80 Do T, Miksik O, DeGol J, et al. Learning to detect scene landmarks for camera localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, 2022. 11132–11142

81 Falchuk B, Loeb S, Neff R. The social metaverse: battle for privacy. IEEE Technol Soc Mag, 2018, 37: 52–61

82 Ning H S, Wang H, Lin Y J, et al. A survey on metaverse: the state-of-the-art, technologies, applications, and challenges. 2021. ArXiv:2111.09673

83 Odeleye B, Loukas G, Heartfield R, et al. Detecting framerate-oriented cyber attacks on user experience in virtual reality. In: Proceedings of the International Workshop on Security for XR and XR for Security, 2021. 1–5

84 Gulhane A, Vyas A, Mitra R, et al. Security, privacy and safety risk assessment for virtual reality learning environment applications. In: Proceedings of the 16th IEEE Consumer Communications & Networking Conference (CCNC), Las Vegas, 2019. 1–9

85 Casey P, Baggili I, Yarramreddy A. Immersive virtual reality attacks and the human joystick. IEEE Trans Dependable Secure Comput, 2019, 18: 550–562

86 Chen Y. Olfactory display: development and application in virtual reality therapy. In: Proceedings of the IEEE International Conference on Artificial Reality and Telexistence–Workshops (ICAT), Hangzhou, 2006. 580–584

87 Maggioni E, Cobden R, Dmitrenko D, et al. Smell space: mapping out the olfactory design space for novel interactions. ACM Trans Comput Hum Interact, 2020, 27: 1–26

88 Hamed A, Khalek A A. Acoustic attacks in the era of IoT: a survey. In: Proceedings of the Amity IEEE International Conference on Artificial Intelligence (AICAI), Dubai, 2019. 855–858

89 Roesner F, Kohno T, Molnar D. Security and privacy for augmented reality systems. Commun ACM, 2014, 57: 88–96

90 Pfeiffer M, Duente T, Rohs M. Let your body move: a prototyping toolkit for wearable force feedback with electrical muscle stimulation. In: Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services, Florence, 2016. 418–427

91 Heartfield R, Loukas G. A taxonomy of attacks and a survey of defence mechanisms for semantic social engineering attacks. ACM Comput Surv, 2015, 48: 1–39

92 Guimarães M, Prada R, Santos P A, et al. The impact of virtual reality in the social presence of a virtual agent. In: Proceedings of the ACM International Conference on Intelligent Virtual Agents, Scotland, 2020. 1–8

93 Biancardi B, Wang C, Mancini M, et al. A computational model for managing impressions of an embodied conversational agent in real-time. In: Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII), Cambridge, 2019. 1–7

94 Adams D, Bah A, Barwulor C, et al. Ethics emerging: the story of privacy and security perceptions in virtual reality. In: Proceedings of the USENIX Security Symposium, Baltimore, 2018. 427–442

95 Krämer N, Sobieraj S, Feng D, et al. Being bullied in virtual environments: experiences and reactions of male and female students to a male or female oppressor. Front Psychol, 2018, 9: 253

96 Sun J Y, Gan W S, Chao H C, et al. Metaverse: survey, applications, security, and opportunities. 2022. ArXiv:2210.07990

97 Narendra M, Valarmathi M L, Anbarasi L J. Watermarking techniques for three-dimensional (3D) mesh models: a survey. Multimedia Syst, 2022, 28: 1–19

98 Ueda J, Okajima K. AR food changer using deep learning and cross-modal effects. In: Proceedings of the 2nd IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), San Diego, 2019. 110–117

99 Nakano K, Horita D, Isoyama N, et al. Ukemochi: a video see-through food overlay system for eating experience in the metaverse. In: Proceedings of the CHI Conference on Human Factors in Computing Systems Extended Abstracts, New Orleans, 2022. 1–8

100 Ahn S, Gorlatova M, Naghizadeh P, et al. Adaptive fog-based output security for augmented reality. In: Proceedings of the Morning Workshop on Virtual Reality and Augmented Reality Network, Budapest, 2018. 1–6

101 Valluripally S, Gulhane A, Hoque K A, et al. Modeling and defense of social virtual reality attacks inducing cybersickness. IEEE Trans Dependable Secure Comput, 2021, 19: 4127–4144

102 Rafique M U, Cheung S S. Tracking attacks on virtual reality systems. IEEE Consumer Electron Mag, 2020, 9: 41–46

103 Saber O, Mazri T. Smart city security issues: the main attacks and countermeasures. Int Arch Photogramm Remote Sens Spatial Inf Sci, 2021, 46: 465–472

104 Ferraz F S, Ferraz C A G. Smart city security issues: depicting information security issues in the role of an urban environment. In: Proceedings of the IEEE/ACM International Conference on Utility and Cloud Computing, Dresden, 2014. 842–847

105 Adams D, Bah A, Barwulor C, et al. Perceptions of the privacy and security of virtual reality. In: Proceedings of the iConference, Sheffield, 2018

106 Ling Z, Li Z P, Chen C, et al. I know what you enter on gear VR. In: Proceedings of the IEEE Conference on Communications and Network Security (CNS), Washington, 2019. 241–249

107 Tseng W J, Bonnail E, Mcgill M, et al. The dark side of perceptual manipulations in virtual reality. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, New Orleans, 2022. 1–15

108 Kudo Y, Tang A, Fujita K, et al. Towards balancing VR immersion and bystander awareness. Proc ACM Hum Comput Interact, 2021, 5: 1–22

109 Sinnett S, Spence C, Soto-Faraco S. Visual dominance and attention: the Colavita effect revisited. Percept Psychophys, 2007, 69: 673–686

110 Kohli L. Redirected touching: warping space to remap passive haptics. In: Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI), Waltham, 2010. 129–130

111 Romanoff T, Neschke S, Draper D, et al. Top Risks in Cybersecurity 2023. Bipartisan Policy Center, 2023. https://bipartisanpolicy.org/report/top-risks-cybersecurity-2023/

112 Gao Y S, Doan B G, Zhang Z, et al. Backdoor attacks and countermeasures on deep learning: a comprehensive review. 2020. ArXiv:2007.10760

113 Goodfellow I J, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. 2014. ArXiv:1412.6572

114 Dong Y P, Liao F Z, Pang T Y, et al. Boosting adversarial attacks with momentum. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, 2018. 9185–9193

115 Thys S, van Ranst W, Goedemé T. Fooling automated surveillance cameras: adversarial patches to attack person detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops), Long Beach, 2019

116 Muñoz-González L, Biggio B, Demontis A, et al. Towards poisoning of deep learning algorithms with back-gradient optimization. In: Proceedings of the ACM Workshop on Artificial Intelligence and Security, Dallas, 2017. 27–38

117 Gu T, Dolan-Gavitt B, Garg S. BadNets: identifying vulnerabilities in the machine learning model supply chain. 2017. ArXiv:1708.06733

118 Saha A, Tejankar A, Koohpayegani S, et al. Backdoor attacks on self-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, 2022. 13327–13336

119 Jia X J, Zhang Y, Wu B Y, et al. LAS-AT: adversarial training with learnable attack strategy. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, 2022. 13388–13398

120 Ren C, Xu Y. Robustness verification for machine-learning-based power system dynamic security assessment models under adversarial examples. IEEE Trans Control Netw Syst, 2022, 9: 1645–1654

121 Wang J K, Yin Z X, Hu P F, et al. Defensive patches for robust recognition in the physical world. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, 2022. 2456–2465

122 Yang Q L, Zhao Y T, Huang H W, et al. Fusing blockchain and AI with metaverse: a survey. IEEE Open J Comput Soc, 2022, 3: 122–136

123 Zhang M Y, Yang L, He S B, et al. Privacy-preserving data aggregation for mobile crowdsensing with externality: an auction approach. IEEE ACM Trans Networking, 2021, 29: 1046–1059

124 Li S, Cui J H, Hao A M, et al. Design and evaluation of personalized percutaneous coronary intervention surgery simulation system. IEEE Trans Visual Comput Graphics, 2021, 27: 4150–4160

125 Ahmed S M, Guo C, Zhao Y D. A survey on the detection algorithms for false data injection attacks in smart grids. IEEE Trans Smart Grid, 2019, 11: 2218–2234

126 Spiegel J S. The ethics of virtual reality technology: social hazards and public policy recommendations. Sci Eng Ethics, 2018, 24: 1537–1550

127 Alshammari K, Beach T, Rezgui Y. Cybersecurity for digital twins in the built environment: current research and future directions. ITcon, 2021, 26: 159–173
128 Hussaini A, Qian C, Liao W X, et al. A taxonomy of security and defense mechanisms in digital twins-based cyber-physical systems. In: Proceedings of the IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics), Espoo, 2022. 597–604
129 Maloney D, Zamanifard S, Freeman G. Anonymity vs. familiarity: self-disclosure and privacy in social virtual reality. In: Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology, 2020. 1–9
130 Nwaneri C. Ready lawyer one: legal issues in the innovation of virtual reality. Harvard J Law & Technol, 2016, 30: 601
131 Lake J. Hey, you stole my avatar!: virtual reality and its risks to identity protection. Emory Law J, 2019, 69: 833
132 Li Y N, Susilo W, Yang G M, et al. Toward privacy and regulation in blockchain-based cryptocurrencies. IEEE Network, 2019, 33: 111–117

Security in virtual-real mixing cyberspaces: a survey

Qinping ZHAO, Zhong ZHOU*, Xiaohui LIANG, Shuai LI, Miao WANG & Yan WANG

Zhongguancun Laboratory, Beijing 100094*, China*
* Corresponding author. E-mail: zz@buaa.edu.cn

Abstract Driven by the continuous development of computer and network infrastructure, more and more human activities are migrating from the physical world to the digital world, bringing the motivation and thought of building a new type of virtual-real mixing cyberspace. Virtual reality, augmented reality, digital twins, metaverse, etc. become world-wide hot spots. Based on the Internet and the Internet of Things, the virtual-real mixing network further interconnects computers with independent identities, various physical objects, and their digital twins, as well as computer-generated digital native objects. In this way, the physical world, the digital world, and the human world are mixed and merged together. We name this kind of emerging network as "Pervasive Internet", which forms a virtual-real mixing cyberspace where humans, machines, and physical/virtual things are interlinked and interconnected, bringing totally new public experiences, social forms, production patterns, and digital economic development paths. This new type of cyberspace has greatly expanded the space boundaries and application fields of the Internet and the Internet of Things, whereas it brings new security and privacy protection issues. This article firstly introduces the concept and architecture of Pervasive Internet and virtual-real mixing cyberspace, analyzes their security and privacy risks, and then reviews the latest international status and trends in user authentication and authorization control, data security, privacy protection, perception and interaction security, critical infrastructure and hardware/software security, and application security and cyberspace governance. We proposed ten problems that need to be solved at the end.

Keywords virtual-real mixing cyberspace, Pervasive Internet, digital twins, security, privacy