



卫星互联网安全威胁分析与体系化防御策略

高航^{1,3}, 赵军锁^{2,3*}, 匡文渊⁴

1. 中国科学院软件研究所基础软件与系统重点实验室, 北京 100190

2. 中国科学院软件研究所, 北京 100190

3. 中国科学院大学, 北京 101408

4. 三六零纵横信息技术有限公司, 北京 100102

* 通信作者. E-mail: junsuo@iscas.ac.cn

收稿日期: 2025-11-17; 修回日期: 2025-12-28; 接受日期: 2026-02-04; 网络出版日期: 2026-03-03

北京市自然科学基金 (批准号: 4264133) 资助项目

摘要 本文针对全球卫星互联网爆发式增长背景下, 低轨巨型星座呈现出的规模化、网络化、智能化、服务化及生态化五大关键演进趋势, 深入分析了在“十万卫星”级超大规模组网环境下, 空天核心节点、跨域链路及星地融合边界所面临的供应链攻击、侧信道窃取及跨域协同打击等非对称威胁挑战. 针对传统防御模式在应对异构网络互操作、协同指挥与系统可信性方面的局限性, 本文提出了一套以“卫星互联网安全大脑”为决策核心、“星云站端”四级智能体为执行单元, 并融合 NodeMatrix, DataMatrix, AppMatrix 与 CodeMatrix 四大治理框架的体系化防御策略. 该体系有效解决了异构环境下的资源统一抽象、多源数据互操作、敏捷战术响应以及软件全生命周期可信演进等关键难题, 为保障复杂巨型星座系统的长期可靠运行提供了系统化解决方案.

关键词 卫星互联网, 复杂异构巨型星座, 网络安全, 治理框架, 体系化防御

1 引言

以复杂异构巨型星座为基础的行星级网络体系——卫星互联网正在迎来前所未有的爆发式增长, 一个以“十万卫星、百万应用、千亿链接、万亿市场”为预期的全新信息网络时代正在加速到来. 以 Starlink, OneWeb 等为代表的低轨 (LEO) 巨型星座正在快速部署, 我国的“GW”星座与“千帆”星座计划也在稳步推进. 这一发展浪潮呈现出鲜明的规模化、网络化、智能化、服务化和生态化的“五化”趋势, 卫星的功能也不再局限于传统的通、导、遥, 而是日益成为全球信息网络的天基核心节点^[1]. 技术层面, 这一变革得益于“软件定义卫星” (software-defined satellite, SDS) 技术的日益成熟, 以及软件定义网络 (SDN)、网络功能虚拟化 (NFV) 和人工智能等先进信息技术的深度融合^[2,3]. 这些技术使得卫星星座的功能可以通过软件重构和在轨升级, 实现了前所未有的灵活性和可扩展性^[4]. 在这种趋势下, 卫星互联网的边界日益模糊, 特别是以美国为首的“军民商盟”一体化融合正在加速, 商业星座 (如星链、星盾) 被深度整合进其“联合作战全域指挥与控制” (CJADC2) 体系, 与

引用格式: 高航, 赵军锁, 匡文渊. 卫星互联网安全威胁分析与体系化防御策略. 中国科学: 信息科学, 2026, 56: 620–642, doi: 10.1360/SSI-2025-0491

Gao H, Zhao J S, Kuang W Y. Analysis of security threats and systematic defense strategies for satellite Internet. Sci Sin Inform, 2026, 56: 620–642, doi: 10.1360/SSI-2025-0491

地面互联网基础设施(如谷歌、微软)和军事网络无缝对接,形成了一个广义上的、跨越陆海空天网的超级信息系统^[5,6].

这种高度融合与软件定义的特性,彻底改变了卫星互联网的系统本质.它已不再是传统航天领域中功能固化、边界清晰、以“黑盒为主”的封闭系统,而是演变为一个网络节点、组合关系、数据流向持续动态演化的“超大规模分布式软件系统”(Ultra Large Scale Distributed Software Systems)^[7].该复杂异构巨系统将天基信息网、地面互联网、移动通信网乃至物联网融为一体,形成了一个规模空前、结构极度复杂、攻防对抗激烈的全新网络空间疆域.然而,这种开放性和复杂性也带来了严峻的体系性安全风险^[8~10].当前,卫星互联网发展普遍面临“建不起来、连不起来、管不起来、用不起来、迭代不起来”的五大瓶颈,其中“管不起来”的安全问题尤为突出^[11,12].其根本矛盾在于“体系化构建需求与碎片化产品之间的矛盾”.传统的安全防护思想,如针对单颗卫星的 TT&C 链路加密或地面站的边界防护,在这种高度动态和开放的复杂异构巨系统中已捉襟见肘.

现有的网络安全理论、方法和技术体系,大多源于地面互联网,已难以完全适应卫星互联网在物理层、网络层和应用层上的全新安全挑战.一方面,卫星“天生脆弱”,在轨资产面临激光致盲、动能碰撞、电子干扰等多种物理威胁^[13,14];另一方面,其“星、云、站、端”的全链路均暴露在体系化的网络攻击之下,从终端劫持、地面站入侵到云服务泄露,乃至软件供应链攻击,形成了立体的、跨域的攻击向量^[15,16].近年来,虽然针对特定攻击(如 GPS 欺骗、星间链路窃听)的研究已取得进展,但大多停留在“碎片化”的单点防御,缺乏一个能够覆盖全域、动态演进、智能响应的体系化防御体系^[17,18].本文旨在分析卫星互联网所面临的安全威胁进行全景分析,并重点探讨一种基于“卫星互联网安全大脑”、四大治理框架、四级“智能体”的新型防御体系.其中,安全大脑指的是基于认知智能的集中式安全编排与管控中心.该体系借鉴了地面网络中 AI 驱动的“安全大脑”和自主安全运维的理念,以节点、数据、应用、代码四类治理框架^[19~22]为基础,通过构建“云安”(指挥中心)、“站安”(指挥官)、“星安”(哨兵)和“端安”(情报源)四级协同的智能代理体系,试图解决复杂异构巨星座在开放、动态环境下的自主安全管理难题.本文的后续章节将详细阐述这一体系.第2节深入分析卫星互联网的脆弱性与安全挑战;第3节重点设计并论述“安全大脑”与四级智能体的体系化防御策略;第4节探讨实现该策略所需的 LVC 试验床、安全软件工厂等关键支撑基础设施.

2 卫星互联网面临的体系化攻击

2.1 卫星互联网具有的关键作用分析

在全球信息基础设施加速向空天一体化迈进的历史进程中,卫星互联网作为一项具有划时代意义的战略性新兴产业,其重要性已上升到关乎国家发展和全球治理的关键层面.这一技术革命的核心在于,它打破了传统地面通信的地理限制,正在以前所未有的速度和规模,构建一个覆盖全球、连接万物的空天核心信息网络,为经济社会发展和国家安全提供了坚实的基础支撑.卫星互联网是推动全球信息普惠和经济增长的强大引擎.随着以低轨(LEO)巨型星座为代表的系统投入大规模部署,其高通量、低时延的特性有效弥补了传统地面网络在偏远山区、广袤海洋、空中航线以及沙漠地带的覆盖盲区,极大地提升了全球信息的互联互通水平.这不仅能够加速数字鸿沟的弥合,让欠发达地区的民众共享信息时代的红利,更开辟了一个估值万亿级的太空信息新市场,支撑着“十万卫星”的制造、发射与运营,驱动着“百万应用”的创新与落地,形成了从上游制造到下游服务的完整产业链生态,是未来国家经济新的增长极.同时,它确立了空天信息网络的战略核心地位,是构建未来战争和治理体系的关键支撑.卫星的功能已不再局限于单一的通信中继或遥感观测,而是演变为全球信息网络的神经末梢和核心节点.这种转变加速了以美国为首的“军民商盟”一体化融合,将商业化星座深度整合进联合作战体系,与地面互联网基础设施、移动通信网络乃至物联网实现无缝对接,形成了一个跨越陆、海、空、天、网的广义超级信息系统.卫星互联网不仅是实现全域感知、精确指挥、快速响应的关键信息传输载体,更是提升国家在复杂电磁环境和极端自然灾害下应急通信保障能力、维护国家主权和信息安全的战略资源.此外,它代表了信息技术融合与创新的前沿方向,其发展趋势鲜明地体现为规模化、网络化、智能化、服务化和生态化的“五化”特征.

通过深度融合“软件定义卫星”(SDS)、软件定义网络(SDN)、网络功能虚拟化(NFV)以及人工智能等先进技术,卫星互联网实现了网络架构的灵活重构、资源的动态分配和服务的智能管理,极大地提高了系统的灵活性和可扩展性。这种技术上的变革不仅为实现天地一体化信息网络提供了可行路径,更标志着人类社会进入了一个全新的网络空间疆域。因此,牢牢把握卫星互联网的发展主动权,是抢占信息技术制高点、维护国家在未来网络空间中竞争优势和主导权的核心战略任务。

2.2 卫星互联网遭受的体系化攻击现状

然而,在全球信息基础设施加速向空天一体化的同时,卫星互联网的固有脆弱性正使其成为体系化攻击的重点目标。其固有安全脆弱性根植于体系架构的全环节,这构成了一个无可回避的“全环节安全不足”困境^[13,23],为高级持续性威胁(APT)组织发动复杂、多阶段的体系化攻击提供了可乘之机。这种攻击模式通过供应链攻击、侧信道攻击以及跨域协同攻击等手段,全面渗透星、云、站、端,形成难以防御的攻击链。首先,卫星天生脆弱,它们运行在严苛的太空环境中,面临辐射损伤、高能粒子干扰等独特的物理威胁,同时受到体积、重量和功耗(SWaP)的严格限制,这导致星上计算机(OBC)的算力极其有限,难以部署复杂的入侵检测系统、先进加密算法或现代安全监控代理^[24,25]。传统的航天设计观念倾向于静态、封闭的系统,但在软件定义卫星(SDS)技术的驱动下,卫星必须通过在轨软件重构和远程更新来保持灵活性,这却直接将地面互联网的软件供应链风险和零日漏洞引入了在轨资产,形成了体系化攻击的首要渗透向量,使得一旦恶意代码被注入,修复成本极高甚至不可能,极大加剧了卫星的数字脆弱性^[26,27]。其次,地面站的安全性也严重不足。随着卫星互联网向广域、低延迟服务发展,测控站、数传站、信关站正从专用硬件转向基于商用现货(COTS)硬件和通用操作系统的软件化架构,并越来越多地依赖于第三方云计算平台进行数据处理和任务调度。这种“云化”和“软件定义”的趋势,虽然提高了效率和灵活性,却将天基系统无缝连接到地面互联网的 attack 面。地面站网一旦被攻破,攻击者便能通过遥控遥测命令或固件更新来危害整个星座,实现跨域指挥与控制(C2)。2022年Viasat KA-SAT遭受的网络攻击事件便是地面站网脆弱性被利用的典型例证,它表明单一地面节点的失陷可以导致大规模用户终端的失效和服务中断^[14,28~30]。最后,用户终端作为卫星互联网的第一道数字防线,但也面临巨大的威胁。卫星用户终端作为大规模部署的消费级产品,其物理可接触性为攻击者提供了便利的逆向工程和固件提取机会。国内外研究已证实,通过低成本的物理攻击手段,可以成功提取终端固件,分析其通信协议,进而实现对终端的劫持甚至反向渗透到星座网络^[31,32]。由于终端数量庞大且分散,其很容易成为攻击者构建大规模僵尸网络、发起分布式拒绝服务(DDoS)攻击或窃取用户数据的跳板,构成体系化攻击的末端火力网。

2.3 卫星互联网安全面临的当前研究困境

在对卫星互联网的固有脆弱性及其遭受体系化攻击的现状进行深入剖析后,必须正视当前在体系化安全研究和实战化验证方面存在的显著困境。这些困境主要体现在理论体系的滞后性、数据的稀缺性和验证环境的缺失性,严重制约了防御能力的提升^[33~35]。一方面,针对太空系统的体系化安全研究尚处于起步阶段,远未形成成熟的理论体系。尽管国际上如美国国防高级研究计划局(DARPA)的SPARTA(太空攻击研究与战术分析)、欧洲空间局(ESA)的SPACE-SHIELD等项目已开始尝试构建针对天基系统的安全知识库和威胁建模框架,旨在建立类似地面网络MITRE ATT&CK的太空网络安全知识体系^[36],但这些努力大多停留在理论概念和初期验证层面^[37]。太空网络的高动态性、强约束性以及跨域耦合性,使得地面网络安全理论(如零信任、纵深防御)难以直接简单移植,例如,如何在高延迟、低带宽的星地链路环境下实现零信任的“永不信任,始终验证”原则,如何在资源受限的卫星OBC上运行复杂的机器学习检测模型,都是尚未完全解决的根本性难题。这种理论和框架的滞后,导致在面对体系化、智能化、持续演进的攻击时,防御方缺乏统一的“战法”和标准化的应对策略。另一方面,公开的可在数字环境中完整复现的攻击链案例极少^[38]。卫星互联网作为一个高度敏感的关键基础设施,无论是商业运营商还是国家航天机构,都将安全事件报告视为机密。与地面网络中黑客攻击、漏洞利用和应急响应报告的相对公开性不同,卫星系统一旦遭受攻击或出现异常,信息往往被严格保密或归咎于“技术故障”。这种信息的高度不透明,导致安全研究人员严重缺乏真实世界的攻击数据、完整的攻击路径日志、被利用的漏洞细节

以及受损系统的运行快照^[38]. 这种数据和案例的稀缺性产生了两个直接的恶果: 一是阻碍了基于大数据和深度学习技术的威胁检测模型的发展, 缺乏真实、高质量的标注数据, 使得人工智能驱动的“安全大脑”难以有效训练和部署; 二是无法建立有效的 LVC (live, virtual, constructive) 试验床或网络靶场, 没有可信、完整的攻击链案例, 防御策略的实战化、端到端验证就无法进行, 安全技术只能停留在实验室的理论推演和局部功能测试, 无法证明其在复杂、高动态的巨系统环境中的实际有效性. 研究困境还体现在缺乏一个统一、标准化的安全开发与评估环境. 卫星互联网的快速迭代要求安全必须融入 DevSecOps 流水线中, 但目前缺乏统一的“节点治理框架 (NodeMatrix)”、“数据治理框架 (DataMatrix)”、“应用治理框架” (AppMatrix) 和“代码治理框架” (CodeMatrix), 导致安全功能的开发和部署仍然是碎片化的. 同时, 缺乏一套能够模拟大规模星座高动态拓扑、带宽时延约束、软硬件异构特性以及真实物理环境影响 (如轨道、太阳活动) 的数字孪生或虚拟化验证平台. 这种评估环境的缺失, 使得任何提出的体系化防御策略 (如本文提出的四级智能体协同) 都难以进行科学、等效、全面的性能评估和验证, 更无法进行具有说服力的“体系化攻击 VS 体系化防御”的攻防对抗演习^[39]. 综上, 突破当前研究困境, 需要国际合作推动信息共享, 并投入大量资源建设兼具数字模拟和硬件在环 (hardware-in-the-loop) 能力的综合安全试验验证平台.

3 体系化防御策略

3.1 指导思想

面对卫星互联网体系化、全链路的安全挑战, 解决关键在于从依赖单点防护的“碎片化堆叠”模式, 向主动、持续演进的“体系化对抗”根本转变. 该战略的核心在于构建具备认知智能的防御体系, 以对等的“体系化防御”反制跨越星、云、站、端的“体系化攻击”, 实现对威胁意图的主动预测与动态适应. 为此, 必须建立融合零信任 (ZTA) 与安全编排自动化 (SOAR) 的标准化框架: 一方面遵循“永不信任, 始终验证”原则, 强化异构信任域间的实时授权; 另一方面通过智能化编排联动“安全大脑”与多级智能体, 将跨域威胁响应时间从数小时缩短至数秒, 实现高动态环境下的高效闭环防御.

3.2 总体架构

在体系化对抗的指导思想下, 卫星互联网的安全防御必须超越传统的网络管理范畴, 构建一个具备中枢神经系统功能的“卫星互联网安全大脑”, 将其视为应对全域威胁的新型战略基础设施. 这一架构的提出, 旨在打破传统天基安全中信息孤岛、能力分散的困境, 实现防御能力的逻辑集中与物理分布的协同统一. 图 1 给出了总体架构图.

根据该架构图, 卫星互联网安全防御体系是一个以“安全大脑”为中枢控制的智能化闭环系统, 它首先作为“逻辑集中的智能安全中枢”, 承载于地面段的高性能计算中心之上; 然后通过“认知智能与掌控”四大职能 (包括情报汇聚、分析融合、智能决策、策略下发) 实现体系化的感知、分析与响应闭环; 最终, 这些决策和策略被下发到“全层级全域防御架构”中, 由部署在空间段 (卫星)、地面段 (测控站/信关站) 和用户段 (终端) 的治理框架 NodeMatrix, DataMatrix, AppMatrix, CodeMatrix, 结合分布式智能体 (节点控制) 来执行具体的防御动作, 如隔离、重路由或审计, 从而实现对新型战略基础设施的体系化防御. 表 1 则给出了卫星互联网安全防御体系中的理论模型与具体工程实体的映射关系.

3.3 安全大脑

“安全大脑”在概念上是一个逻辑集中的智能安全中枢, 是整个体系化防御框架的认知核心. 其物理实现的核心载体是“地面段智能管控计算中心”. 选择地面段作为核心, 是基于天基系统的固有约束; 尽管部分计算能力已向在轨卫星前置 (如“星安”智能体), 但卫星 (V_{star}) 在功耗、存储和算力 (SWaP) 上的严格限制, 无法支撑高性能的深度学习模型、海量历史数据的存储和复杂的全局图计算. 因此, “安全大脑”的运算中枢必须部署在

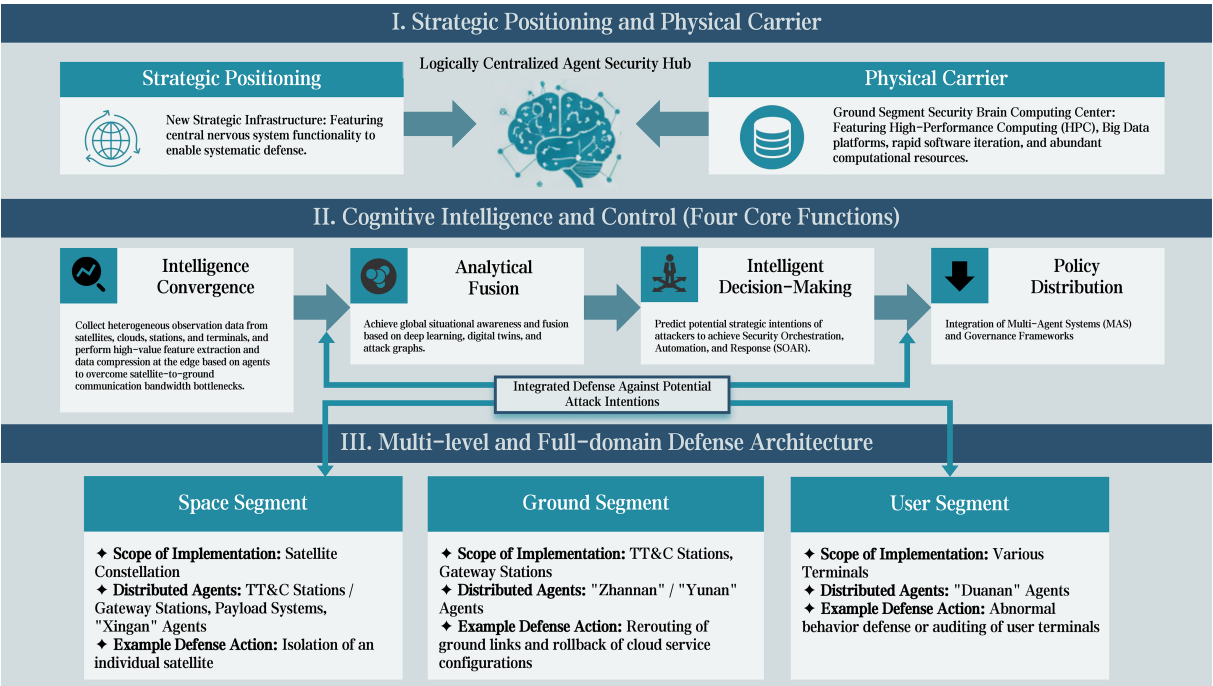


图 1 (网络版彩图) “卫星互联网安全防御体系” 总体架构图。

Figure 1 (Color online) “Satellite Internet security architecture” overall architecture diagram.

表 1 卫星互联网安全防御体系理论模型与工程实体映射表。

Table 1 Mapping of satellite Internet security defense system: theoretical model to engineering entities.

Symbol	Set definition	Engineering entity example	Key attributes & constraints	Corresponding agent
$v \in \mathcal{V}_{\text{star}}$	Space segment node	LEO broadband satellite (e.g., AT8000 platform) Payload: on-board router/switch	SWaP-constrained: • CPU: PowerPC/RISC-V (1~2 cores); • Memory: <4 GB; • Power budget: <15 W (security dedicated).	“Star-An” (lightweight probe)
$v \in \mathcal{V}_{\text{gnd}}$	Ground segment node	Gateway/TT&C station Payload: Baseband processing server	High computing power: • Server cluster (x86 arch); • Virtualized container environment.	“Station-An” (regional commander)
$v \in \mathcal{V}_{\text{user}}$	User segment node	User terminal (VSAT/handheld) Payload: baseband chip/SoC	High exposure surface: • Physical environment is uncontrollable; • Firmware is vulnerable to extraction and analysis.	“Terminal-An” (intelligence source)
$v \in \mathcal{V}_{\text{cloud}}$	Cloud node	Ground operation center/cloud platform	Elastic computing: • HPC clusters; • Threat intelligence database.	“Cloud-An” (global brain)
$e \in \mathcal{E}_{\text{isl}}$	Inter-satellite link (ISL)	Laser/microwave ISL	High dynamics: • Topology reconfigures every 10~90 min; • Anti-interception capabilities are required.	Link encryption monitoring
$e \in \mathcal{E}_{\text{sgl}}$	Satellite-ground link (SGL)	Ku/Ka/Q/V band feeder link	High latency and asymmetry: • Propagation delay: 20~50 ms; • Bandwidth B_{max} is the primary bottleneck.	Traffic compression audit

拥有充沛计算资源、可进行快速软件迭代和维护的地面数据中心或云平台. 这使得它能够充分利用高性能计算(HPC)、大数据平台以及最新的威胁情报云服务, 为整个系统提供强大的智能支撑.

作为新型战略基础设施,“安全大脑”通过高速、可信的链路连接空间段(卫星星座)、地面段(测控站、信关站)和用户段(各类终端),形成了覆盖全域的感知和指挥网络. 其核心功能可以概括为情报的汇聚、分析、决策和下发四大战略职能,从而将传统的、分散的网管功能(NMS)和安全信息与事件管理(SIEM)系统,升级为具备认知智能和自主响应能力的中央指挥平台.

在威胁情报的汇聚方面,“安全大脑”是全网数据流的汇集点. 它实时接收来自星、云、站、端四级智能体采集的异构观测数据 O_i . 这些数据不仅包括传统的网络日志和流量信息,还包括卫星遥测遥控(TT&C)数据、星上OBC负载、链路质量异常以及用户终端的行为画像等. 根据信息论原理,智能体在边缘对原始数据进行了高价值特征提取和压缩,以克服星地通信带宽($B_{\max}$)的瓶颈,确保上传的信息流在满足速率约束的前提下最大化保留与真实风险状态 $R(t)$ 之间的互信息 $I(R(t); O)$. 这种机制确保了“安全大脑”能够获得一张低延迟、高保真度的全域安全态势图.

在分析与融合方面,“安全大脑”负责对汇聚的威胁情报进行深度学习驱动的认知分析. 它超越了传统的基于规则或签名的检测模式,专注于全局态势感知与融合、高级威胁(APT)的检测与预测. 通过构建系统的数字孪生和攻击图模型,它可以实现跨域攻击路径的实时重建和脆弱性预测,例如,分析一个从用户终端发起、经地面站中转、最终针对星上载荷系统的复合攻击链. 这种能力体现了当前安全领域“认知网络”和“意图驱动安全”的研究前沿,即预测攻击者的潜在战略意图和下一步行动,而非仅仅对已发生的事件进行被动响应.

在智能决策方面,“安全大脑”根据态势分析结果,生成最优的全局防御策略. 这体现了安全编排、自动化与响应(SOAR)的理念. 它利用复杂的优化算法(如基于图论的最小风险路径覆盖算法,如前述定理所示),快速确定最有效的防御动作组合,例如是采取对单个卫星的隔离(由“星安”执行)、对地面链路的重路由(由“站安”执行),还是对云服务进行配置回滚(由“云安”执行). 这种决策能力保证了防御动作的有效性和协同性,能够将响应时间从传统的人工决策的数小时级别压缩至分钟甚至秒级.

在策略下发与协同方面,“安全大脑”是多智能体系统(MAS)的战略指挥官^[40]. 它将生成的战略决策转化为战术指令,并通过统一的治理框架(NodeMatrix, AppMatrix)实时、精确地分发给分散在各安全域的星、站、端智能体. 这种分级指挥体系,确保了防御指令能够在高动态、高延迟的环境下被可靠执行,实现星、云、站、端的无缝协同响应,最终达成体系化防御的目标. 因此,“卫星互联网安全大脑”不仅是一个计算中心,更是实现卫星网络弹性、自主安全运营和战略级防御能力的基石.

针对“安全大脑”可能面临的单点失效或被攻陷风险,体系在物理载体上构建了高韧性的分布式共识架构. 我们假设攻击者具备对地面基础设施的渗透能力,能够完全控制不超过 f 个计算节点. 基于PBFT(实用拜占庭容错)算法,系统部署 $N = 3f + 1$ 个异地计算节点. 在本方案中,设定 $N = 4, f = 1$,即允许任意一个地面计算中心发生宕机或被恶意劫持. 在策略下发环节,采用 (t, n) 门限签名方案(threshold signature),设定签名阈值 $k = 3$,即任何全局性防御指令(如全网路由重构)必须获得至少3个异地节点的私钥签名片段才能合成有效指令,从而杜绝了单一节点私钥泄露导致的灾难性后果. 为防止长期密钥泄露,系统引入Epoch机制. 每个Epoch(设为24小时)结束后,通过分布式密钥生成(DKG)协议自动轮换各节点的私钥分片,旧签名的公钥在Epoch切换后自动失效.

为落实“安全大脑”与四级智能体之间的高效协同,本体系设计了基于UDP/QUIC改造的轻量级安全协同协议(Lightweight Security Coordination Protocol, LSCP). 该协议消息帧采用紧凑的TLV(type-length-value)格式,以最小化协议开销. 关键消息结构定义如下:

$$\text{Msg} = \{\text{Ver} \parallel \text{Type} \parallel \text{SessionID} \parallel \text{Timestamp} \parallel \text{Priority} \parallel \text{Payload} \parallel \sigma_{\text{Key}}\}, \quad (1)$$

其中,Type区分消息类型(如0x01:威胁心跳,0x02:策略下发,0x03:固件更新);Priority用于QoS调度,确保战时高等级威胁情报抢占带宽; σ_{Key} 为基于门限签名的完整性校验字段. 针对星地链路高动态导致的频繁中断问题,体系引入“分级自治决策算法”. 当且仅当星地链路连通性 $L_{\text{status}} = 1$ 时,执行全局最优策略;当链路中

断 ($L_{status} = 0$) 持续超过阈值 $T_{timeout}$ 时, 系统自动切换至“降级自治模式”. 此时, 地面站的“站安”智能体 (Station-An) 权限升级, 作为区域临时根节点, 依据本地缓存的规则库接管其覆盖范围内“星安”智能体的指挥权. 自治决策逻辑的伪代码描述如算法 1.

算法 1 智能体分级自治决策逻辑.

输入: 本地观测数据 O_i , 链路连接状态 L_{status} ;

主迭代: 最终防御动作 A ;

```

1:  $O_{compressed} \leftarrow \text{FeatureExtract}(O_i)$ ; {边缘特征提取}
2: if  $L_{status} == \text{CONNECTED}$  then
3:   Send(Cloud_An,  $O_{compressed}$ ); {上报安全大脑}
4:    $A \leftarrow \text{WaitPolicy}(\text{Cloud\_An})$ ; {等待全局策略}
5:   if Role == Star_An then
6:     Send(Station_An,  $O_{compressed}$ ); {向过境信关站汇报}
7:      $A \leftarrow \text{LocalRuleMatch}(O_i) \vee \text{Wait}(\text{Station\_An})$ ;
8:   else if Role == Station_An then
9:     ContextLocal  $\leftarrow \text{Aggregate}(\text{Star\_An\_Group})$ ; {区域融合}
10:     $A \leftarrow \text{GenerateTacticalPolicy}(\text{Context}_{Local})$ ;
11:   end if
12: end if
13: Execute( $A$ ).

```

引入共识与多重签名机制不可避免地带来了额外的计算与通信开销. 我们对此进行了严格的代价评估. 一次全局策略生成的总时延定义为 $T_{total} = T_{consensus} + T_{sign} + T_{verify}$. 在千兆光纤互联的地面节点间, 4 节点 PBFT 共识过程 (三阶段交互) 平均耗时约 120 ms; $k = 3$ 的门限签名合成耗时约 40 ms; 星上边缘节点的验签 (ECDSA) 耗时约 20 ms. 因此, 总额外时延约为 180 ms. 虽然 180 ms 的延迟对于毫秒级的实时包交换是不可接受的, 但对于“安全大脑”负责的战略级防御 (如隔离卫星、全网补丁分发) 而言, 其时间粒度通常在秒级以上. 因此, 用 180 ms 的时延换取对拜占庭错误的免疫力是符合系统设计目标的. 对于需毫秒级响应的战术动作 (如丢弃恶意帧), 则由“星安”智能体依据本地预置规则 (local rules) 直接执行, 无需经过全局共识, 从而实现了安全韧性与实时性的解耦.

3.4 治理框架: NodeMatrix, DataMatrix, AppMatrix, CodeMatrix

为了使“卫星互联网安全大脑”这一新型战略基础设施在复杂场景上高效、可靠地运行, 并实现理论模型中所追求的全局最优观测和跨域检测完备性, 必须建立一套统一、标准化的治理框架 (governance framework). 这套框架相当于安全大脑的“操作系统”和“宪法”, 解决了执行单元在异构环境下的互操作、协同指挥和可信性难题. 图 2 给出了治理框架的总体原理. 该体系的运行依赖于四类核心的治理矩阵.

为深入分析此架构的有效性, 首先建立卫星互联网的理论模型. 卫星互联网系统可被建模为一个时变异构图 $G^{(t)} = \{\mathcal{V}, \mathcal{E}^{(t)}\}$. 其中, 顶点集 $\mathcal{V}$ 代表系统中的所有实体节点, 并可划分为 4 个不相交的子集: $\mathcal{V} = \mathcal{V}_{\text{star}} \cup \mathcal{V}_{\text{gnd}} \cup \mathcal{V}_{\text{user}} \cup \mathcal{V}_{\text{cloud}}$, 分别代表空间段卫星节点、地面段节点 (如信关站、运控中心)、用户终端节点和云服务节点. 边集 $\mathcal{E}^{(t)}$ 代表节点间的通信链路, 其拓扑结构随卫星运行和网络任务 (如星间链路 $\mathcal{E}_{\text{isl}}$ 、星地链路 $\mathcal{E}_{\text{sgl}}$) 的变化而动态变化. 实际上, $\mathcal{V}$ 也会随时间发生一定变化, 然而这种变化相对于 $\mathcal{E}^{(t)}$ 来讲较为偶发, 为简化问题在这里不再考虑 $\mathcal{V}$ 中的变化.

NodeMatrix 进行硬件整合, 能够解决顶点集 $\mathcal{V} = \mathcal{V}_{\text{star}} \cup \mathcal{V}_{\text{gnd}} \cup \mathcal{V}_{\text{user}} \cup \mathcal{V}_{\text{cloud}}$ 的高度异构问题. 它通过定义统一的资源抽象层 (HAL) 和虚拟化接口, 将资源受限的“星安”节点、计算能力强大的“站安”节点以及商业化的云服务节点进行抽象, 实现“计算即服务”的安全架构. 这种抽象与地面网络中的软件定义网络 (SDN) 理念一致, 将卫星互联网的基础设施软件化, 使“云安”智能体能够像调度地面资源一样, 对天基和地基的安全计算资源进行按需配给和统一调度. 例如, 当检测到潜在的高级威胁时, NodeMatrix 能够迅速为地面站 $v \in \mathcal{V}_{\text{gnd}}$ 分

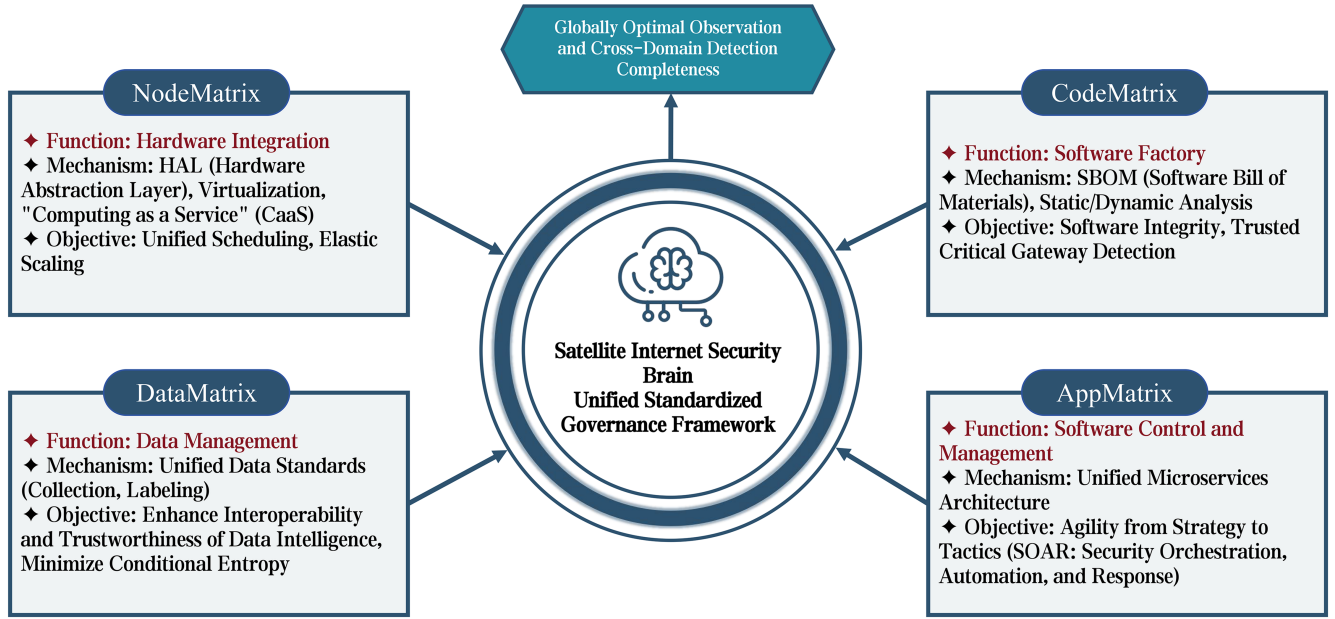


图2 (网络版彩图) 治理框架原理图.

Figure 2 (Color online) Governance framework schematic.

配额外的容器或虚拟机资源, 以运行高算力分析任务, 从而提升该节点在关键网关上的检测概率 $\rho_v^{(t)}$, 严格提高跨域检测完备性定理中的下界, 实现防御能力的弹性扩展. 国际上对于天基系统资源虚拟化和容器化的研究, 正是 NodeMatrix 的理论基础.

DataMatrix 进行数据管理, 是信息论框架下实现最优观测的前提. 根据安全大脑最优观测定理 (21) 和 (22), 全网观测数据 O_i 经过压缩编码 U_i 必须满足带宽约束 $B_{\max}$, 同时最大化与真实风险状态 R 之间的互信息. DataMatrix 通过定义统一的数据采集、处理、标注与接口标准, 确保了所有智能体, 尤其是资源受限的“星安”和“端安”智能体, 在本地进行数据预处理和压缩后, 汇聚到“云安”的威胁情报 $U = \{U_i\}$ 是可互操作、可信赖的. 它解决了异构日志格式和遥测数据的语义鸿沟, 为“云安”智能体应用图神经网络 (GNN) 和深度学习模型进行全域态势融合提供了高质量、统一格式的输入. 这有效避免了因数据质量或格式不一致导致的互信息损失, 确保安全大脑对全局风险状态 R 的条件熵 $H(R|Y)$ 达到理论上的最小化.

此外, 针对卫星互联网天然具备的跨国部署与军民融合特征, DataMatrix 在数据治理中强制内嵌了数据主权与合规性策略. 框架采用基于地理围栏 (geo-fencing) 的标签化技术, 自动识别数据产生地与过境节点的法律管辖权 (如欧盟 GDPR 或各国数据安全法), 据此执行本地化存储或脱敏传输决策. 对于必须跨境共享的威胁情报, DataMatrix 规定需采用联邦学习或多方安全计算 (MPC) 模式, 确保在不交换原始敏感数据的前提下实现全球防御协同, 并提供不可篡改的审计日志以满足国际合作中的法律追溯要求.

AppMatrix 进行软件管控, 负责将“云安”智能体的战略决策转化为可执行的战术行动, 是实现跨域协同响应的关键. 它定义了安全功能微服务化和应用程序接口 (API) 的规范, 确保了防御策略 (如隔离、重配置、固件热补丁) 的抽象与分发. 当“云安”生成一个全局策略后, AppMatrix 能够保证该策略指令能够在不同计算环境和操作系统上的“星安”、“站安”和“端安”智能体中被正确、高效地解释和执行. 这种标准化接口和微服务架构极大地提升了防御体系的敏捷性和自动化程度, 与安全编排、自动化与响应 (SOAR) 理念高度一致. 例如, 面对一次复杂的跨域攻击, AppMatrix 可迅速编排“站安”进行流量重路由、“星安”进行链路丢包率调整, 共同将攻击路径 P 上的检测概率 $\rho_v^{(t)}$ 提升至 $1 - \epsilon$, 从而满足跨域检测完备性要求.

CodeMatrix 实际上是一个软件工厂, 是保障整个体系可信根的基石, 主要应对最隐蔽的供应链攻击向量. 它

通过规范化代码开发、自动化构建、代码静态/动态检查以及软件物料清单 (SBOM) 管理等流程, 将安全保障前置到 DevSecOps 流水线的每一个环节. 这对于保证部署在 $\mathcal{V}_{\text{star}}$ 和 $\mathcal{V}_{\text{user}}$ 等资源受限节点上的智能体软件 (如“星安”固件、“端安”应用) 的完整性和可信性至关重要. CodeMatrix 确保了智能体软件不受恶意代码或硬件木马的污染, 保障了检测概率 $\rho_v^{(t)}$ 的可信度. 若智能体软件本身被攻陷, 则其报告的检测结果 $\rho_v^{(t)}$ 的真实性将趋于零, 体系化防御将彻底失效. 因此, CodeMatrix 是维持整个多智能体系统 (MAS) 安全和长期演进能力的根本防线, 确保了所有的决策和执行都建立在一个可信的软件基础之上. 四类治理框架共同构成了“卫星互联网安全大脑”这一战略基础设施的“宪法”, 它们将理论模型和优化目标转化为可操作、可扩展的工程实践. 然而, 鉴于卫星互联网系统固有的高动态性、跨域耦合性, 以及星地链路固有的长时延与带宽限制, 尽管拥有了统一的治理标准, 安全大脑的算力中枢依然受限于星地通信的时延和带宽约束, 无法对 $\mathcal{V}_{\text{star}}$ 和 $\mathcal{V}_{\text{user}}$ 中的节点进行低延迟、集中式的决策和“微操”. 因此, 为将治理框架的战略意图和防御逻辑转化为全网范围内的即时、分布式执行, 安全大脑必须将其智能“前置”部署, 依赖于物理分散、逻辑协同的执行单元.

为进一步保障治理框架在强对抗环境下的底层可信与容错能力, NodeMatrix 与 CodeMatrix 必须建立在硬件可信根 (hardware root of trust, HRoT) 之上. 体系在物理层集成 TPM/TEE 安全模块, 实现密钥的全生命周期管理与基于硬件的远程证明 (remote attestation), 严禁私钥明文导出, 从而构建不可篡改的系统启动信任链. 同时, 针对个别节点被攻陷可能导致的检测概率 $\rho_v^{(t)}$ 报告失真问题, 防御体系引入实时信任系数 $\alpha_v^{(t)} \in [0, 1]$. 该系数综合远程证明结果与邻居节点的一致性校验, 对各智能体上报的检测指标进行加权修正 ($\hat{\rho}_v^{(t)} = \rho_v^{(t)} \times \alpha_v^{(t)}$), 确保在部分节点密钥泄露或遭受拜占庭攻击的情况下, 安全大脑仍能维持全局态势判定的鲁棒性与可审计性.

为确保治理框架的工程落地与跨厂商互操作性, 并解决异构设备间的语义鸿沟, 本体系建立了标准化的接口契约与数据语义规范, 包括“软件定义卫星节点治理框架 (NodeMatrix) 和协议规范”^[19]、“软件定义卫星数据治理框架 (DataMatrix) 和硬件抽象层 (HAL) 规范”^[20]、“软件定义卫星应用治理框架 (AppMatrix) 和应用编程接口 (API) 规范”^[21]、“软件定义卫星代码治理框架 (CodeMatrix) 和编码规范”^[22] 四份技术规范. 在此基础上, 于 DataMatrix 中, 后续将采用轻量级 CBOR 格式承载压缩后的遥测数据, 并基于 STIX 2.1 标准构建跨域威胁情报模型, 从而将理论模型中的编码映射 f_i 工程化为边缘侧的高效特征提取算子, 确保在满足 B_{max} 带宽约束的同时保留关键语义. 在 AppMatrix 与 NodeMatrix 层面, 将通过定义统一北向接口, 屏蔽异构卫星 (如不同厂商的 OBC) 与地面设备的私有协议差异; 同时引入声明式策略模板与语义版本控制, 确保融合映射 g 所生成的防御策略在云端的动态编排、灰度发布与无损回滚, 从而实现从“逻辑集中管控”到“异构物理执行”的确定性映射.

下面, 给出一个具体的工程实例实现. NodeMatrix 采用容器化技术屏蔽底层硬件差异. 依据规范中对组件资源隔离的要求, 工程上通过资源配额模板严格限制安全组件的算力与内存占用, 防止其在异常状态下抢占卫星关键业务资源. 下面给出其服务声明文件.

```
services:
  star-agent-core:
    image: star-sec/agent:v2.1
    deploy:
      resources:
        limits:
          # 限制占用不超过单核20%
          cpus: '0.2'
          # 内存硬顶,防止OOM
          memory: 128M
    security_opt:
      - no-new-privileges:true
```

DataMatrix 针对星地链路的带宽瓶颈,建立了跨域统一的数据访问标准,定义了基于 Protocol Buffers (proto3) 的紧凑型威胁情报序列化协议. 该规范将异构传感器的原始物理信号规约为标准语义数据,确保在低带宽环境下实现高保真的实时报警上报. 下面给出其消息传递格式.

```
message ThreatAlert {
  // HAL节点标识符
  uint32 satellite_id = 1;
  // 统一时统格式
  uint64 timestamp = 2;

  enum ThreatType {
    UNKNOWN = 0;
    DDOS_FLOOD = 1;
    ILLEGAL_CMD = 2;
  }
  ThreatType type = 3;

  // 原始日志摘要用于审计
  bytes evidence_hash = 4;
  // 本地判定置信度
  float confidence_score = 5;
}
```

AppMatrix 定义了统一的南向防御接口 (SPI), 实现了从战略决策到战术执行的自动化映射. 通过标准化的服务调用契约, 该框架屏蔽了底层执行设备 (如 Linux 网关或 FPGA 载荷) 的实现差异, 使得阻断等防御策略能够跨异构节点一致执行. 下面给出部分执行定义.

```
// 统一阻断接口:屏蔽异构执行单元实现差异
Result BlockTarget(
    TargetID t_id,
    Duration timeout
);
```

CodeMatrix 确立了面向航天环境的安全编码规范, 强制推行静态代码分析与供应链审计机制. 该框架在 CI/CD 流水线中集成规范化的检查工具, 严禁动态内存分配与递归调用等高风险操作, 从源头保障智能体软件在复杂环境下的可信度与健壮性. 以下为部分配置示例.

3.5 核心组成: 星、云、站、端四级智能体

尽管 NodeMatrix, DataMatrix, AppMatrix 和 CodeMatrix 四大核心治理框架为“卫星互联网安全大脑”提供了统一的资源抽象、数据互操作、策略执行和可信根基石, 将防御战略转化为标准化的工程规范, 但卫星互联网系统固有的高动态拓扑、跨域耦合特性, 以及星地链路在时延和带宽方面的严苛限制, 从根本上阻碍了中心化的实时决策. 具体而言, 安全大脑的算力中枢——云端和地面计算中心, 由于受到长距离通信延迟和带宽预算 $B_{\max}$ 的约束, 无法对数量庞大且资源受限的在轨卫星节点 ($\mathcal{V}_{\text{star}}$) 和分散的用户终端 ($\mathcal{V}_{\text{user}}$) 进行低延迟的“微操”和集中式指挥. 因此, 为确保防御指令能够即时响应高动态的威胁演变, 并将安全大脑的战略意图转化为全网范围内的即时、分布式执行, 必须突破集中式管理的物理瓶颈, 将其智能能力“前置”部署.

```
# .clang-tidy 配置示例 (CodeMatrix规范实现)
Checks: >
-*,
cppcoreguidelines-no-malloc,      # 禁止 malloc/free
cppcoreguidelines-pro-bounds-*,   # 强制边界检查
misc-no-recursion                  # 禁止递归调用

# 警告视为错误, 阻断构建
WarningsAsErrors: '*'
```

这便催生了星、云、站、端四级智能体这一分布式执行单元, 它们作为安全大脑在各个异构节点上的代理和执行者, 依据治理框架所确立的协同规范, 在各自的区域或领域内完成态势感知、战术决策与响应, 实现低延迟、高效率的区域协同决策, 从而在信息论意义上以最小的带宽代价 (最小化 $\sum_i \mathbb{E}[L(U_i)]$) 最大化对全局风险状态 R 的观测互信息 $I(R; Y)$, 确保整个防御体系能够在严苛的物理约束下, 依然满足跨域检测完备性的战略要求. 这四种智能体根据其部署位置、计算能力、通信延迟和安全职能的不同, 被赋予了明确的定位和任务, 共同构成一个自感知、自决策、自适应的多智能体安全系统 (MAS), 在全网范围内贯彻安全大脑的协同防御意图.

星、云、站、端四级智能体是“安全大脑”在 $G^{(t)}$ 中各个顶点上的“代理”和“执行者”, 各司其职, 协同作战. 第一级, “云安” (Cloud-An) 智能体, 部署于云端或地面计算中心, 是“指挥中心”. 它承担全局态势感知与信息融合、基于海量数据和复杂模型的战略决策与策略生成、高级威胁 (如 APT) 的检测与预测, 以及跨域协同调度. 第二级, “站安” (Station-An) 智能体, 部署于地面站 $v \in \mathcal{V}_{\text{gnd}}$, 是“指挥官”. 它负责执行“云安”的战略意图, 承担其管辖区域 (如一个信关站覆盖区) 的战术指挥, 负责地面网络边界防护、区域协同防御, 并利用其较强的算力资源, 为 $\mathcal{V}_{\text{star}}$ 和 $\mathcal{V}_{\text{user}}$ 节点卸载高算力的安全任务 (如复杂加解密、恶意代码沙箱分析). 第三级, “星安” (Star-An) 智能体, 轻量化部署于卫星节点 $v \in \mathcal{V}_{\text{star}}$, 是“哨兵”. 受限于星上 (on-board) 的功耗、存储和算力 (SWaP) 约束, 其功能高度优化, 专注于本地态势感知 (如链路信号质量、OBC 负载)、数据采集与预处理 (如日志过滤与压缩)、轻量级的实时响应 (如丢弃恶意遥控帧), 以及保障卫星的生存性与机动规避 (如响应地面指令执行规避动作). 第四级, “端安” (Terminal-An) 智能体, 部署于用户终端 $v \in \mathcal{V}_{\text{user}}$, 是“第一道防线”和“情报源”. 它负责感知用户侧的异常行为、执行终端安全防护 (如固件完整性校验)、处理链路加密与认证, 并作为“安全大脑”感知网络末梢威胁的最重要情报来源.

该四级智能体体系 (agent-based system, $\mathcal{A}$) 构成了一个分布式的信息处理网络. 设 S 为系统 $G^{(t)}$ 在某一时刻的真实安全状态 (例如, 一个高维向量, 表示每个节点 $v \in \mathcal{V}$ 是否被入侵). “安全大脑”的目标是最大化对 S 的认知, 即最小化关于 S 的不确定性, 这里表示为熵 $H(S)$. 各级智能体 $A_i \in \mathcal{A}$ 持续采集其本地观测数据 O_i . 在没有智能体时, “安全大脑”对 S 的熵为 $H(S)$. 在部署智能体后, “安全大脑”的后验熵为 $H(S|O_{\text{cloud}}, O_{\text{gnd}}, O_{\text{star}}, O_{\text{user}})$. 该体系的价值在于, 它通过“星安”和“端安”等轻量级“哨兵”进行本地信息处理与压缩 (如特征提取), 在保证关键安全信息 (即最大化互信息 $I(S; O_i)$) 不丢失的前提下, 极大地减少了需要通过带宽受限的星地链路上传的数据量 $R(O_i)$. 这实质上是在求解一个分布式的率失真 (rate-distortion) 问题:

$$\min_{\mathcal{A}} \sum_i R(O_i) \quad \text{s.t.} \quad I(S; \{O_i\}_{i \in \mathcal{A}}) \geq I_{\text{threshold}}. \quad (2)$$

通过将原始观测数据在 $\mathcal{V}_{\text{star}}$ 和 $\mathcal{V}_{\text{user}}$ 上进行预处理, 该架构显著降低了对通信带宽的依赖, 实现了低成本、高效率的全局态势感知.

下面进一步验证该体系化策略在应对跨域攻击时的完备性. 跨域攻击是卫星互联网面临的最严峻挑战之一, 攻击者利用不同域 (如地面互联网、空间段) 的信任关系和协议漏洞, 构建复杂的攻击链. 我们提出以下定理来证明该防御体系对跨域攻击的检测能力.

定理1 (跨域风险检测完备性) 设离散时间 $\mathbb{T}$, 系统于时刻 t 由 $G^{(t)} = (\mathcal{V}, \mathcal{E}^{(t)})$ 描述, $\mathcal{V} = \bigsqcup_{i=1}^n \mathcal{V}_i$. 记跨域边集 $\mathcal{E}_c^{(t)} = \{(u, v) \in \mathcal{E}^{(t)} : \exists \mathcal{V}_a \neq \mathcal{V}_b, u \in \mathcal{V}_a, v \in \mathcal{V}_b\}$, 网关 $\mathcal{V}_{gw}^{(t)}$ 为 $\mathcal{E}_c^{(t)}$ 的端点集, $\mathcal{V}_{gw}^{[*]} = \bigcup_t \mathcal{V}_{gw}^{(t)}$. 对任意跨域时间关联路径 $P = ((v_1, t_1), \dots, (v_k, t_k))$ 定义 $\mathbb{I}(P) = \{i; (v_i, v_{i+1}) \in \mathcal{E}_c^{(t_i)}\}$, $S_i = \{v_i, v_{i+1}\} \cap (\mathcal{V}_{A_{\text{gnd}}}^{(t_i)} \cup \mathcal{V}_{A_{\text{cloud}}}^{(t_i)})$. 若存在路径命中集 $\mathcal{H} \subseteq \mathcal{V}_{gw}^{[*]}$ 使得

$$\forall P \exists i \in \mathbb{I}(P), \exists v \in S_i \cap \mathcal{H} : \rho_v^{(t_i)} \geq 1 - \epsilon, \quad \epsilon \rightarrow 0, \quad (3)$$

则检测完备性 $\mathcal{C} = 1$, 并且对任意跨域路径 P 有检测成功概率下界

$$P_{\text{det}}(P) \geq \max_{i \in \mathbb{I}(P)} \max_{v \in S_i} \rho_v^{(t_i)}, \quad (4)$$

且若同一步内各被监控端点独立、跨步独立, 则

$$P_{\text{det}}(P) \geq 1 - \prod_{i \in \mathbb{I}(P)} \prod_{v \in S_i} (1 - \rho_v^{(t_i)}). \quad (5)$$

“安全大脑”通过增大 $|S_i|$ 与提升 $\rho_v^{(t)}$ 可严格提高以上下界.

证明 为形式化推导, 先以等式给出所需的事件与概率对象, 对任一被监控网关节点与时刻, 定义“在该时刻于该节点检测成功”的事件为

$$\mathbf{D}_{v,t} \equiv \{\text{在时刻 } t \text{ 于节点 } v \text{ 发生检测成功}\}, \quad (6)$$

并将其成功概率记为

$$\mathbb{P}(\mathbf{D}_{v,t}) = \rho_v^{(t)} \in [0, 1]. \quad (7)$$

把路径上第 i 个跨域跳步的“步检测事件”定义为

$$\mathbf{D}_i \equiv \bigcup_{v \in S_i} \mathbf{D}_{v,t_i}, \quad (8)$$

而“沿整条路径至少一次被检测”的总事件为

$$\mathbf{D} \equiv \bigcup_{i \in \mathbb{I}(P)} \mathbf{D}_i, \quad (9)$$

据此路径的检测成功概率写作

$$P_{\text{det}}(P) = \mathbb{P}(\mathbf{D}). \quad (10)$$

利用并集概率的基本不等式

$$\mathbb{P}\left(\bigcup_j A_j\right) \geq \max_j \mathbb{P}(A_j), \quad (11)$$

将式 (9) 与 (8) 代入可得

$$P_{\text{det}}(P) = \mathbb{P}\left(\bigcup_{i \in \mathbb{I}(P)} \mathbf{D}_i\right) \geq \max_{i \in \mathbb{I}(P)} \mathbb{P}(\mathbf{D}_i) \geq \max_{i \in \mathbb{I}(P)} \max_{v \in S_i} \rho_v^{(t_i)}, \quad (12)$$

这就给出了式 (4) 的下界. 若同一步内各端点的检测相互独立, 则

$$\mathbb{P}(\mathbf{D}_i) = 1 - \prod_{v \in S_i} (1 - \rho_v^{(t_i)}), \quad (13)$$

进一步若跨步之间也独立, 由式 (9) 与 (13) 得到

$$P_{\text{det}}(P) = 1 - \prod_{i \in \mathbb{I}(P)} \left(1 - \mathbb{P}(\mathcal{D}_i)\right) = 1 - \prod_{i \in \mathbb{I}(P)} \prod_{v \in S_i} (1 - \rho_v^{(t_i)}), \quad (14)$$

从而给出式 (5) 的下界. 为证明完备性, 设存在一个被监控的路径命中子集, 使得每一条跨域路径在其某一步都命中其中某个网关节点, 该节点在对应时刻的检测概率可以任意接近 1, 形式化为存在某个命中节点与跳步满足

$$\rho_{v^*}^{(t_i)} \geq 1 - \epsilon \quad (\epsilon > 0 \text{ 可任意小}). \quad (15)$$

把式 (15) 代入式 (12) 立得

$$P_{\text{det}}(P) \geq 1 - \epsilon. \quad (16)$$

对所有跨域路径取下确界并令 $\epsilon \rightarrow 0$, 便得到

$$\lim_{\epsilon \rightarrow 0} \inf_P P_{\text{det}}(P) = 1. \quad (17)$$

即检测完备性成立. 最后刻画“安全大脑”对下界的提升: 由式 (13) 可直接计算步检测概率对任一单点检测概率的偏导

$$\frac{\partial}{\partial \rho_w^{(t_i)}} \left[1 - \prod_{v \in S_i} (1 - \rho_v^{(t_i)}) \right] = \prod_{u \in S_i \setminus \{w\}} (1 - \rho_u^{(t_i)}) \geq 0, \quad (18)$$

从而步检测概率关于各单点检测概率单调不减, 若通过扩展覆盖把某一步的被监控端点集合由 S_i 扩大为 $S_i \cup \{w\}$ 并赋予新节点正的检测概率, 则有

$$1 - \prod_{v \in S_i \cup \{w\}} (1 - \rho_v^{(t_i)}) = 1 - (1 - \rho_w^{(t_i)}) \prod_{v \in S_i} (1 - \rho_v^{(t_i)}) \geq 1 - \prod_{v \in S_i} (1 - \rho_v^{(t_i)}), \quad (19)$$

进而将式 (19) 代入路径级表达式 (14) 可得

$$1 - \prod_{i \in \mathbb{I}(P)} \prod_{v \in S'_i} (1 - \rho_v^{(t_i)}) \geq 1 - \prod_{i \in \mathbb{I}(P)} \prod_{v \in S_i} (1 - \rho_v^{(t_i)}) \quad (S'_i \supseteq S_i), \quad (20)$$

这表明扩大覆盖或提升单点准确率都会严格抬升定理给出的两类下界, 从而体现“安全大脑”的增益.

同样地, 我们提出以下定理来证明“安全大脑”对于全局风险的分析观测能力.

定理2 (安全大脑最优观测定理) 在总上传带宽预算约束下, 存在编码映射与云侧融合映射使得“安全大脑”方案实现对全局风险状态 R 与“安全大脑”针对于 R 的融合分析输出 Y (即安全看到的“风险影像”) 之间互信息的全局最优化:

$$\max_{\{f_i\}, g} I(R; Y), \quad (21)$$

并满足带宽约束

$$\sum_{i \in \{\text{gnd}, \text{star}, \text{user}\}} \mathbb{E}[L(U_i)] \leq \mathcal{B}_{\text{max}}, \quad (22)$$

其中 $L(U_i)$ 表示 U_i 的码长, $\mathbb{E}[L(U_i)]$ 为其期望码长 (等价于平均上传速率), $\mathcal{B}_{\text{max}}$ 为总上传带宽预算.

证明 设各层观测 O_i 经编码 $U_i = f_i(O_i)$ 形成 $U = (U_{\text{gnd}}, U_{\text{star}}, U_{\text{user}})$, 云侧输出为 $Y = g(U)$. 上传受预算 $\sum_i \mathbb{E}[L(U_i)] \leq \mathcal{B}_{\text{max}}$ 约束. 证明的目标是

$$\max_{\{f_i\}, g} I(R; Y) \quad \text{s.t.} \quad \sum_i \mathbb{E}[L(U_i)] \leq \mathcal{B}_{\text{max}}. \quad (23)$$

由马尔可夫 (Markov) 链 $R \rightarrow U \rightarrow Y$ 的数据处理不等式

$$I(R; Y) \leq I(R; U). \quad (24)$$

取 g 为恒等映射 $Y = U$ 或充分统计映射 $Y = T(U) \approx p(R | U)$ 可使等号成立, 即对任意固定的编码族 $\{f_i\}$, 有

$$\max_g I(R; Y) = I(R; U). \quad (25)$$

因此式 (23) 等价缩约为仅对编码族求解

$$\max_{\{f_i\}} I(R; U) \quad \text{s.t.} \quad \sum_i \mathbb{E}[L(U_i)] \leq \mathcal{B}_{\max}. \quad (26)$$

在常见建模下可令 U_i 取有限字母表并允许随机化编码, 平均码长约束诱导的可行集在适当拓扑下是紧的, 而互信息 $I(R; U)$ 关于联合分布连续, 故由 Weierstrass 极值定理, 存在 $\{f_i^*\}$ 达到式 (26) 的最优值. 配合使式 (25) 取等的 g^* (如恒等映射), 得到一组 $(\{f_i^*\}, g^*)$ 达到式 (23) 的最优值:

$$\max_{\{f_i\}, g} I(R; Y) = I(R; U^*) \quad \text{with} \quad U^* = (f_i^*(O_i))_i. \quad (27)$$

记“安全大脑”策略集为 Π_{ST} , 去中心化无协同策略集为 Π_{dec} . 显然

$$\Pi_{\text{dec}} \subseteq \Pi_{\text{ST}}, \quad (28)$$

从而

$$\max_{\pi \in \Pi_{\text{ST}}} I(R; Y_\pi) \geq \max_{\pi \in \Pi_{\text{dec}}} I(R; Y_\pi). \quad (29)$$

这意味着在同一带宽预算下, 该安全防御体系的联合优化不劣于任何去中心化方案, 且通常更优.

之后, 由互信息的单调性给出下界

$$I(R; U) = I(R; U_{\text{gnd}}, U_{\text{star}}, U_{\text{user}}) \geq \max\{I(R; U_{\text{gnd}}), I(R; U_{\text{star}}), I(R; U_{\text{user}})\}, \quad (30)$$

数据处理与码长预算给出上界

$$I(R; U) \leq I(R; O_{\text{gnd}}, O_{\text{star}}, O_{\text{user}}), \quad (31)$$

$$I(R; U) \leq H(U) \leq \mathbb{E}[L(U)] = \sum_i \mathbb{E}[L(U_i)] \leq \mathcal{B}_{\max}. \quad (32)$$

由式 (24)~(32) 可知, 该安全防御体系通过在式 (26) 上联合优化编码与在式 (25) 中选择最优融合, 使 $I(R; Y)$ 在观测天花板与带宽天花板之间达到可实现的最大值: 等价地最小化条件熵

$$H(R | Y) = H(R) - I(R; Y), \quad (33)$$

从而在信息论意义上最大化对全局风险的覆盖与检测能力.

上述理论最优解在实际工程运维中面临对抗性攻击与资源竞争的双重挑战. 为此, 体系在模型层面引入对抗训练与基于统计显著性检验的概念漂移监控机制, 一旦检测到数据分布偏移或投毒攻击迹象, 即自动触发模型回滚与在线增量学习, 确保“安全大脑”在长期运行中的鲁棒性. 同时, 为了量化评估运维效能, 系统建立包含误报率 (P_{fp})、检测延迟 (T_{lat}) 与计算能耗 (E_{cost}) 的多维指标体系, 并引入业务质量 (QoS) 动态权衡因子 λ . 在带宽预算 $\mathcal{B}_{\max}$ 的分配上, 根据威胁等级 (如 DEFCON) 动态调整 λ , 在常态下优先保障用户通信业务吞吐量, 在战时则允许安全情报数据抢占带宽, 以实现“安全效能”与“业务质量”的弹性平衡.

在工程实践中, 单点检测概率 $\rho_v^{(t)}$ 并非抽象的静态数值, 而被实例化为一个由静态特征覆盖率与动态行为异常指数构成的加权函数. 具体而言, 该指标的计算依赖于 CodeMatrix 治理框架与星上 FDIR (故障检测、隔

离与恢复) 机制的深度协同. 其中, 静态分量由软件构建阶段的自动化故障注入测试确定, 表征智能体对已知攻击特征 (如特定恶意 C&C 指令序列) 的基线拦截能力, 动态分量则由部署于星务计算机 (OBC) 的“星安”智能体实时计算, 通过监控 CPU 任务调度的异常负载以及 1553B/SpaceWire 总线上的违规时序指令来量化系统的实时异常程度. 最终, 地面“安全大脑”利用贝叶斯 (Bayes) 推断对历史误报数据进行后验分析, 并通过遥控指令周期性更新各分量的权重参数, 从而在资源受限的星上计算环境中实现对 $\rho_v^{(t)}$ 的高置信度动态估算.

为直观阐述体系化防御在应对复杂威胁时的关键作用, 本文构建了“基于终端劫持的卫星指令注入攻击链”典型场景: 攻击者首先物理入侵用户终端 (v_{user}), 通过 JTAG 接口篡改固件以植入恶意代码 (T0-T1); 随即利用受控终端伪装合法用户连接信关站 (v_{gnd}), 借助协议模糊测试 (Fuzzing) 触发解析漏洞以绕过边界防火墙 (T2-T3); 进而向过境卫星 (v_{star}) 的星务计算机 (OBC) 注入伪造的特权级遥控指令 (T4-T5), 最终引发载荷关机或姿态翻转等灾难性后果 (T6). 针对此类传统防御因流量合法性掩护而难以感知的跨域攻击, 本文提出的体系化防御架构依据跨域检测完备性 (定理 1) 实施多维协同阻断: 由“端安”智能体在边缘侧实时校验固件哈希完整性, 由“站安”智能体在链路层识别异常协议封装行为, 并由“星安”智能体在终端层验证指令执行的上下文逻辑, 所有异常情报经由“安全大脑”进行全局关联分析, 从而在攻击链尚处信道跳板阶段 (阶段二) 即可实现精准切断.

4 关键支撑基础设施

4.1 LVC 引擎与试验床

在卫星互联网安全防御体系中, 无论是安全大脑和四级智能体的战略决策、四大治理框架的标准化推行, 还是多智能体系统 (MAS) 的跨域协同能力, 其有效性都必须在一个高度逼真、能够捕捉到时变异构图 $\mathcal{G}^{(t)}$ 动态特性的环境中进行开发、验证和持续演进. 因此, LVC (live-virtual-constructive, 实兵 - 虚拟 - 构造) 引擎与试验床被确立为保障整个体系化防御策略能够落地和成熟的不可或缺的新型基础设施.

LVC 范式为太空网络安全提供了独特的、分层级的验证环境, 它将昂贵的、高风险的实战环境 (live) 与灵活、可重复的虚拟仿真环境 (virtual/constructive) 相结合. 在实际部署前, 体系化防御策略的效能主要通过 LVC 引擎进行预测和优化. 首先, 试验床必须构建高度逼真的虚拟与构造环境. 这要求建立卫星互联网系统的数字孪生 (digital twin), 能够精确模拟复杂异构巨星座的轨道动力学、动态链路拓扑 $\mathcal{E}^{(t)}$ 、星上 OBC 的算力与功耗约束 (SWaP), 以及星地、星间链路的高时延和带宽变化. 在这个数字孪生环境内部署四级智能体的虚拟副本, 并利用数据虚实混成技术, 将真实的遥测数据 O_i 与构造的攻击流量和故障注入数据进行混合, 用于训练和测试“云安”智能体的高级威胁检测与预测模型. 这种环境尤其擅长对跨域检测完备性定理进行验证, 通过注入复杂的、跨越 $v_{\text{star}}, v_{\text{gnd}}, v_{\text{user}}$ 的攻击路径 P , 反复测试在不同带宽约束和拓扑下, 系统通过提升 $\rho_v^{(t)}$ 从而实现检测成功概率 $P_{\text{det}}(P)$ 的能力. 此外, 硬件在环的构造性验证, 是将真实的“星安”和“站安”智能体硬件平台与虚拟的星座环境连接起来, 对 CodeMatrix 和 AppMatrix 所定义的安全应用进行实际性能和可靠性测试, 确保防御指令在硬件层面能够低延迟、无差错地执行.

其次, LVC 引擎的关键价值在于提供在轨新技术试验 (Live) 环境, 以验证和迭代那些无法在地面完全模拟的前沿安全技术. 例如, 针对卫星通信面临的窃听和激光致盲威胁, 我国及国际上都在积极利用新技术试验卫星 (如“天智一号”等), 开展抗量子密钥交换 (quantum key distribution, QKD) 等技术的在轨验证. 这种技术将“星安”和“站安”智能体硬件平台与虚拟的复杂异构巨星座环境连接起来, 对四大治理框架所定义的安全对象进行实际性能和可靠性测试, 确保了其能够成为防御体系中各级智能体之间的核心加密支撑技术.

此外, 为了实现“体系化攻击 VS 体系化防御”的最高级别对抗, LVC 试验床需要整合在轨网络安全测试平台. 美国的 DARPA/空军研究实验室 (AFRL) 近年来通过“Hack-A-Sat”等项目, 并计划部署 Moonlighter 等卫星, 以构建了一个“太空黑客沙箱”. 这种专用的、可编程的在轨平台允许红队 (Red Team) 在受控的、真实的太空环境中对目标卫星进行模拟网络攻击. 这种实战攻防演练对于验证“星安”智能体的轻量级实时响应和机动

规避能力具有不可替代的价值,能够揭示在极端条件下,如星上资源竞争、延迟突增时,防御策略的实际性能边界.这些来自真实攻防的数据和经验教训,反过来能够回馈到“云安”智能体的学习模型和治理框架的优化迭代中,形成一个闭环的 DevSecOps 流程,最终使整个安全防御体系具备持续优化与性能调优的能力.

LVC 引擎与试验床不仅是一个验证工具,更是推动卫星互联网安全从理论研究走向工程实践的桥梁,它通过虚实结合的方式,为复杂、动态、高风险的太空网络安全问题提供了一个安全、可控、可持续的演进环境.

4.2 安全软件工厂

卫星互联网作为新型战略基础设施,其核心的挑战在于快速技术迭代与海量软件缺陷之间的矛盾.传统的航天软件开发模式采用周期长、僵化的“瀑布式”流程,与当前卫星互联网追求的“数周”或“数月”的极速研制周期完全不兼容.此外,体系化防御的核心——星、云、站、端四级智能体,本质上是大量分布式、高复杂度的安全软件集合.任何一个智能体软件中的漏洞,都可能成为攻击者利用的供应链攻击向量,从而破坏整个防御体系的可信根.因此,为应对这一挑战,必须建立面向卫星互联网安全的安全软件工厂.

安全软件工厂是实现软件定义卫星 (SDS) 和体系化安全防御的关键使能器.它的核心变革在于彻底实现从传统的“瀑布式”开发模式向 DevOps (开发、运维一体化) 的转变,并进一步融入 DevSecOps 理念,将安全活动内嵌到软件生命周期的每一个环节.这种转变不仅仅是工具链的简单升级,而是组织架构、文化和流程的深层革命.

该架构可以推动开发流程向 DevOps 演进,支持高度自动化的 CI/CD 流水线 (持续集成/持续部署).在卫星互联网环境中,这意味着安全软件工厂必须能够为 $\mathcal{V}_{\text{star}}, \mathcal{V}_{\text{gnd}}, \mathcal{V}_{\text{user}}$ 各域的智能体提供统一的、高频次的软件构建和部署能力. CI/CD 流水线自动化实现了从需求获取、代码提交、自动化构建、集成测试、代码静态/动态安全检查,到最终部署包签名的全过程.对于空间段的“星安”智能体而言,自动化部署流水线能够大幅缩短软件更新周期,使卫星具备在轨快速打补丁的能力,从而有效应对零日漏洞或新的高级威胁.这种自动化和快速迭代是应对复杂攻击者所需敏捷防御能力的关键.

安全软件工厂驱动应用架构将从传统的“单体”结构向“微服务化”演进.传统的航天软件常采用耦合度高、维护困难的单体架构.而微服务化架构将复杂系统分解为一组独立、松耦合的服务,每个服务都可以由 AppMatrix 定义的标准化接口进行管理和部署.对于安全大脑而言,这意味着“云安”的每一个核心功能 (如威胁情报融合、攻击图建模、策略生成) 都可以作为一个独立的微服务存在,它们可以独立开发、测试、部署和扩展.这种架构极大地提升了安全功能的快速迭代和部署能力.例如,当发现一种新的 GNSS 欺骗攻击时,安全软件工厂可以仅针对“端安”智能体中负责 PNT 数据验证的微服务进行快速重构和热更新,而无需重新编译和部署整个系统,这对于提升整个体系的敏捷性和故障隔离能力至关重要.同时,微服务和容器化 (如 Docker/Kubernetes) 的结合,也为 NodeMatrix 提供了统一的运行环境抽象,便于将安全功能从地面站 $v \in \mathcal{V}_{\text{gnd}}$ 迁移或卸载到云端 $v \in \mathcal{V}_{\text{cloud}}$.

对于软件工厂的建设,安全必须内嵌到整个工厂流程中,实现 DevSecOps.这与 CodeMatrix 治理框架的目标紧密相连.安全软件工厂必须在 CI/CD 流水线中集成静态应用安全测试 (SAST)、动态应用安全测试 (DAST) 以及软件成分分析 (SCA),以确保代码的安全性.这有助于识别和消除代码中的已知漏洞,并对所有依赖的开源组件 (如 COTS 软件库) 进行追踪和管理,从源头切断供应链攻击的路径.特别是在航天领域,软件工厂需要集成形式化验证技术,对涉及遥测遥控、姿态控制等关键任务代码的逻辑正确性和安全性进行数学证明,确保部署到“星安”智能体上的代码满足最高的可靠性和安全需求.通过这种方式,安全被视为质量而非事后弥补的环节,有效保障了四级智能体代码的可信度 $\rho_v^{(t)}$,从根本上支撑了体系化对抗的可靠性.

安全软件工厂通过高度的自动化、微服务化的架构以及内嵌的 DevSecOps 流程,将卫星互联网的安全能力研制周期从传统的数年或数月大幅压缩至数周,实现了持续优化与性能调优的目标,确保了整个安全防御体系能够以超前的速度和敏捷性应对不断演进的太空网络威胁.

4.3 数字资产资源池

构建卫星互联网安全防御基础设施,实现从“碎片化产品”到“体系化对抗”的战略转变,其核心驱动力在于智能,而智能的基石则在于对海量、高质量数字资产的积累和有效利用.因此,建立数字资产资源池(digital asset resource pool)是构建卫星互联网安全防御基础设施的首要任务.该资源池是为星、云、站、端四级智能体(A)提供数据支撑和知识体系的“燃料库”和“知识图谱”.

数字资产资源池的核心在于建立安全大数据平台,以汇聚各域($\mathcal{V}_{\text{star}}, \mathcal{V}_{\text{gnd}}, \mathcal{V}_{\text{user}}, \mathcal{V}_{\text{cloud}}$)智能体采集的异构观测数据 O_i .这些数据包括:卫星遥测遥控(TT&C)日志、星间链路(ISL)和星地链路(SGL)的实时流量特征、地面运控网络的访问记录、云服务的API调用和配置日志,以及“端安”智能体上报的用户终端异常行为数据.由于这些数据具有高维度、高动态、强时序关联的特点,DataMatrix需要提供统一的规范,确保这些底层数据在汇聚到安全大脑后能够进行有效的时空对齐和语义融合.这种大数据平台是“云安”智能体进行全局态势感知与信息融合的基础,特别是利用图神经网络(GNNs)等先进技术对 $\mathcal{G}^{(t)}$ 上的动态关联关系和潜在攻击路径进行建模和预测时,高质量的数据输入是决定模型性能上限的关键.具体而言,模型将资源池中汇聚并标准化的异构数据 O_i 映射为时变图 $\mathcal{G}^{(t)}$ 的节点特征向量,利用GNN的消息传递(message passing)机制聚合邻域节点的时空上下文信息,从而在动态拓扑中精准捕获跨域攻击的传播特征与级联效应.

资源池必须构建一个动态、实时的漏洞库和威胁情报库.这包括但不限于:通用和专业CVE漏洞,即针对卫星互联网广泛采用的商业现货(COTS)硬件、开源软件(OSS)、特定的航天操作系统和卫星网络协议(如SpaceWire, CCSDS)的已知漏洞.高级持续性威胁(APT)组织情报,即收集和分析针对关键基础设施和航天目标的APT组织的攻击工具、战术、程序(TTPs)和历史活动日志.这些情报为“云安”智能体提供了高级威胁检测与预测所需的知识框架,使其能够识别出潜伏期长、跨域渗透性强的复杂攻击.还有恶意样本库,专门针对星上计算机(OBC)架构(如PowerPC, RISC-V)、地面站控制器以及用户终端固件的病毒、木马和后门样本.这些独特的恶意代码样本是训练“站安”智能体进行恶意流量深度检测和“星安”智能体进行轻量级异常行为检测模型的关键.

这一威胁情报和样本库的建设,直接支撑了以安全大脑为核心实现“体系化对抗”的能力.国内外先进研究,如网络靶场中采用的威胁情报平台和沙箱技术,需要将这些数字资产集成到LVC引擎中,用于模拟真实的攻防场景,持续验证防御策略的有效性.通过在LVC试验床上注入真实的恶意样本和APT攻击链,可以有效评估四级智能体在复杂攻击下的检测成功概率 $P_{\text{det}}(P)$.

数字资产资源池还需包含模型资产与知识图谱.除了原始数据和情报,已训练好的、性能得到验证的机器学习模型(如用于“端安”的用户行为基线模型、用于“星安”的在轨资源异常检测模型)也应作为可复用资产纳入管理.同时,将APT情报、漏洞信息和系统拓扑信息构建成安全知识图谱,能够提升“云安”智能体进行因果推理和决策解释的能力,将“黑箱”模型转化为可信赖的辅助决策系统.

通过CodeMatrix治理框架,该资源池还将存储所有经过安全验证和形式化证明的安全代码组件和配置模板,包括“星安”的轻量级加密模块、“端安”的固件完整性校验代码等.这不仅是安全软件工厂持续集成/持续部署(CI/CD)流水线所需的输入,也是确保整个系统软件供应链可信的关键保障.资源池的最终目标是提供一个高度整合、持续更新、可供AI模型训练和安全专家分析的统一数据环境,为整个卫星互联网安全防御体系提供自主安全运营和持续优化能力.

在卫星互联网的长周期运行中,攻击者的战术手段(TTPs)会不断演变,导致静态模型的检测能力随时间衰减(即概念漂移).为此,本体系设计了“影子模式(shadow mode)”的在线演进机制.新训练的AI模型首先以“影子模式”推送到“星安”智能体.在此模式下,模型对实时流量进行推理并生成日志,但不执行任何阻断操作.地面“安全大脑”对比影子模型与当前主模型的输出差异.只有当影子模型在保持误报率(FPR)不增加的前提下,显著提升了对新样本的检出率,才通过SwitchOver指令将其提升为主模型.若新模型上线后导致CPU负载异常(如 $>30\%$)或触发看门狗复位,NodeMatrix的健康检查机制将在1s内自动回滚至上一版本的镜像,确保防御系统的升级过程不会危及卫星平台的生存安全.这一机制构成了DevSecOps的闭环,解决了传统航天软件

表 2 LVC 数字孪生仿真平台关键参数设置.

Table 2 Key parameter settings for LVC digital twin simulation platform.

Parameter category	Value/configuration
<i>Constellation configuration</i>	
Configuration	Walker Delta (144/12/1), altitude 1200 km, inclination 87°
Inter-satellite link (ISL)	4 × laser links, bandwidth 5 Gbps, BER 10 ⁻⁹
Satellite-ground link (SGL)	Ka/V band, uplink 500 Mbps, downlink 1 Gbps
<i>Node performance</i>	
Satellite capacity (V_{star})	PowerPC arch, 4 GB RAM, 15% reserved for security container
Number of terminals (V_{user})	10000 (dynamically distributed across land and maritime routes)
<i>Traffic model</i>	
Background traffic	HTTP (40%), video stream (40%), VoIP (20%)
Attack traffic	DDoS (UDP flood), command injection

“敢发不敢改”的运维难题.

5 仿真实验与分析

5.1 实验环境与参数设置

为了验证本文提出的体系化防御策略在动态、异构环境下的有效性,构建基于 LVC 架构的高保真数字孪生仿真平台. 仿真实验的物理时间跨度设定为 24 h, 仿真步长精度为 1 ms, 以捕捉毫秒级的网络突发流量特征. 平台采用 STK 进行高精度的轨道动力学计算, 模拟了一个包含 144 颗卫星 (Walker Delta 构型, 高度 1200 km) 和 10 个地面信关站的动态网络; 同时通过 OPNET Modeler 进行全协议栈通信仿真. 关键仿真参数详见表 2. 背景流量基于 HTTP/Video/VoIP 混合模型生成 (总负载 60%), 并混入了 5% 的恶意攻击流量, 涵盖 DDoS 洪水攻击与跨域指令注入等典型威胁.

为了确保仿真实验的可复现性, 本研究对四级智能体的检测判定与响应逻辑进行了严格定义. 在 LVC 仿真引擎中, 防御体系采用“分级置信度判定”机制, 具体规则如下. (1) 端侧特征提取 (V_{user}): 端安智能体实时监控固件 Hash 完整性与 JTAG 接口电平. 定义异常特征向量 F_{term} , 当 Hash 校验失败或接口出现非授权高电平时, 标记本地威胁置信度 $C_{\text{term}} = 1.0$, 并触发本地“熔断”机制 (即物理切断数据发送电路). (2) 星侧异常检测 (V_{star}): 星安智能体监控 OBC 指令序列与总线时序. 设定滑动窗口 $W = 50$ ms, 计算指令序列的信息熵 $H(\text{seq})$. 当 $H(\text{seq})$ 偏离基线超过 3σ 时, 判定 $C_{\text{star}} = 0.8$, 并执行“指令丢弃”动作. (3) 跨域融合判定 (Cloud/Station): 安全大脑基于贝叶斯网络融合多源情报. 当接收到“端安”与“星安”的时空关联报警时, 计算全局风险值 $R_{\text{global}} = \alpha C_{\text{term}} + \beta C_{\text{star}} + \gamma C_{\text{net}}$. 设定双重阈值: 若 $R_{\text{global}} > 0.9$, 触发全网自动阻断 (下发黑名单至所有信关站); 若 $0.6 < R_{\text{global}} \leq 0.9$, 触发灰度审计模式, 对相关链路流量进行全量镜像存储.

5.2 对比方案与评估

指标实验设置了两组对比方案以评估防御效能. 由于研究对象涉及庞大航天系统, 并且卫星互联网为一个新领域, 因而融合多个相关系统工作^[41~43], 构建传统边界防御 (基准组), 以代表传统边界防御模式, 其仅在地面网关 (V_{gnd}) 和云端 (V_{cloud}) 部署防火墙与 IDS, 卫星与终端不具备深度检测能力; 体系化防御 (实验组) 则启用了本文提出的星、云、站、端四级智能体协同体系. 在评估指标方面, 重点考察跨域威胁检测率, 其定义为针对跨越 3 个以上逻辑域 (如终端 → 地面 → 卫星) 的复杂攻击链的成功阻断比例. 此外, 还对比了系统的平均响应时间和资源开销.

同时, 为了对于“星、云、站、端四级智能体协同防御体系”引入的额外资源开销进行评估, 并通过与航天

表 3 传统边界防御与四级智能体体系化防御性能对比实验结果. 加粗表示最优表现.

Table 3 Experimental results of performance comparison between traditional boundary defense and four-level agent-based systematic defense. The best performance is in bold.

Evaluation metric	Traditional boundary defense	Systematic defense	Performance diff. ratio
<i>Scenario I: cross-domain command injection</i>			
Single-domain detection rate	34.2%	92.5%	+58.3%
False positive rate (FPR)	12.4%	4.3%	-65.3%
<i>Scenario II: distributed denial of service (DDoS)</i>			
Average response time	2420 ms	45 ms	-98.1%
SGL bandwidth occupancy	95.6%	62.8%	-32.8%
On-board CPU load	15%	28%	+13%
<i>Scenario III: zero-day worm propagation</i>			
Network-wide infection rate	68.4%	4.2%	-64.2%
Containment time	125 min	12 s	-99.8%

工程标准阈值进行对比, 验证其在实际部署中的可行性, 额外设计了评估实验, 采用多维度量化分析方法, 涵盖计算资源、存储资源和通信带宽三个关键方面.

5.3 实验结果分析

实验分为多组实际场景与相应数据分析, 具体内容如下.

针对场景 I 的跨域指令注入攻击, 实验结果 (见表 3) 显示, 传统边界防御的检测率仅为 34.2%. 这是因为传统防御仅能看到地面站视角, 无法感知终端侧的固件篡改和星上侧的指令上下文异常. 相比之下, 体系化防御凭借“端安”在源头的固件校验和“星安”在末端的指令审计, 结合定理 1 的全域关联机制, 将跨域检测率显著提升至 92.5%, 同时将误报率 (FPR) 从 12.4% 降低至 4.3%.

针对场景 II 的 DDoS 攻击, 数据表明体系化防御具有显著的边缘阻断优势. 传统模式 (传统边界防御) 必须将恶意流量回传至地面清洗, 受限于星地链路的长往返时延, 平均响应时间高达 2420 ms, 且导致星地链路带宽占用率飙升至 95.6% (严重拥塞). 而体系化防御利用“星安”智能体在卫星节点本地执行丢包策略, 将平均响应时间缩短至 45 ms, 并成功节省了 32.8% 的下行链路带宽. 尽管这使星上 CPU 负载从 15% 上升至 28%, 但该计算代价对于保障战略级链路的安全畅通是完全可接受的.

针对场景 III 中模拟的利用卫星操作系统漏洞而通过星间链路横向传播的蠕虫病毒, 传统边界防御依赖地面中心更新病毒特征库, 受限于遥控上注的窗口期 (Contact Window), 全网补丁分发耗时 125 min, 导致星座感染率高达 68%. 体系化防御的“星安”智能体采用基于行为的邻居监测机制, 一旦发现相邻卫星表现出异常的扫描行为, 立即触发区域隔离策略. 实验显示, 体系化防御成功将病毒传播限制在初始感染节点周围的 3 跳范围内, 全网感染率被控制在 4.2% 以下, 证明了分布式智能体在切断“横向移动”方面的决定性优势.

系统开销与工程可行性分析实验则详细评估了“星安”智能体的资源占用情况 (见表 4). 在 PowerPC 架构的星务计算机上, 完整的“星安”容器镜像 (含推理引擎与策略库) 占用内存 128 MB, 仅占 OBC 总内存 (4 GB) 的 3.1%, 远低于航天软件通常设定的 20% 安全余量. 当“安全大脑”下发全网紧急封锁指令时, 利用星间链路的 Gossip 广播协议, 体系化防御实现全星座 (144 颗星) 策略同步仅需 3.2 s, 而传统地面逐星上注模式需等待过境, 完全同步需耗时 90 min 以上.

结合前述结果, 针对星上计算资源 (SWaP) 受限的假设, 实验表明即便在 DDoS 防御导致 CPU 负载升至 28% 的峰值压力下, 内存占用仍稳定在 3.1% (128 MB), 远低于 20% 的工程安全余量, 证实了模型对资源消耗不敏感, 满足工程可行性. 其次, 针对高动态拓扑导致的通信时延假设, 框架在 3.2 s 内完成全网同步的数据, 证明了分布式架构能有效克服对长链路通信的依赖. 此外, 在 5% 恶意流量的混合压力下, 系统仍维持 92.5% 的高检测率, 验证了防御体系在网络流量波动和复杂攻击环境下的鲁棒性, 支持了定理 1 关于跨域检测完备性的理论

表 4 星上智能体资源开销与系统扩展性测试结果.

Table 4 Testing results of on-board agent resource overhead and system scalability.

Overhead metric	Measurement value	Constraint threshold	Engineering feasibility
<i>Single-satellite resource consumption</i>			
Memory footprint	128 MB	4096 MB (total)	✓ (3.1% ratio)
Peak CPU load	28%	80% (safe margin)	✓ (secure)
Image storage space	256 MB	32 GB (flash)	✓ (negligible)
<i>Network-wide scalability</i>			
Policy sync time	3.2 s	< 10 s (req.)	✓ (compliant)
Threat intel reporting BW	12 kbps/sat	500 Mbps (uplink)	✓ (extremely low)

表 5 防御组件效能消融实验结果. 完整体系数据基准引用自表 3 “体系化防御” 结果: 去星安化/去端安化数据源自消融实验实测. 加粗表示最优表现.

Table 5 Ablation study results of defense component effectiveness. Complete system baseline refers to the “Systematic defense” results in Table 3; “w/o Star-An” and “w/o Terminal-An” data are derived from empirical ablation experiments. The best performance is in bold.

Defense architecture	Ablated component	Detection rate (%)	Avg. blocking latency (ms)	False positive rate (FPR) (%)
Full system	–	92.5	45	4.3
w/o Star-An	Star-An agent	68.0	320	6.5
w/o Terminal-An	Terminal-An agent	42.5	48	18.2

推导.

值得注意的是, 本实验虽未单独设置针对“安全大脑”AI 内核 (如攻击图构建与意图预测) 的独立测试集, 但场景 I (跨域指令注入) 的高检测率 (92.5%) 实际上是认知智能效能的直接体现. 在场景 I 中, 攻击者利用了合法的通信协议和被劫持的合法终端, 单一维度的检测 (如仅靠“站安”查流量或“星安”查日志) 均无法识别此类隐蔽的高级威胁 (检测率仅 34.2%). 体系化防御的高成功率, 归功于“安全大脑”在云端成功将“端安”的固件哈希微小变动与“星安”的异常指令序列在时空图谱上进行了关联重构, 这间接验证了系统在跨域威胁情报融合与攻击意图推演方面的有效性.

针对更为隐蔽的供应链攻击 (如预置后门) 与侧信道威胁, 目前的实验主要侧重于网络层面的对抗. 未来的工作将基于 CodeMatrix 治理框架, 在 LVC 试验床中引入“软件故障注入”模块, 通过在智能体编译构建阶段随机植入逻辑炸弹, 进一步量化评估防御体系在面对内生代码缺陷时的自检与清洗能力.

5.4 消融实验

为了验证四级智能体协同架构的必要性, 我们设计了消融实验 (ablation study), 在场景 I (跨域指令注入) 下分别测试了 3 种防御配置的性能. 完整体系: 启用星、云、站、端全链路智能体. 去星安化: 关闭卫星端“星安”智能体, 仅依赖地面站与终端防御, 模拟传统“天地分离”防御模式. 去端安化: 关闭终端侧“端安”智能体, 假设终端为不可信黑盒.

实验结果 (见表 5) 显示, 在移除“星安”智能体后, 虽然检测率仍维持在 68% 左右, 但平均阻断时延从 45 ms 激增至 320 ms. 这是因为缺乏星上处理能力, 所有异常流量必须回传至地面站进行清洗, 导致了巨大的往返时延代价. 而在移除“端安”智能体后, 针对跨域指令注入的检测率骤降至 42.5%, 且误报率上升至 18.2%. 这表明, 缺失了源头侧的物理层监控, 仅靠网络层流量分析难以精准区分伪造的合法指令. 消融实验有力证明了: 该体系的高检测率来源于“端安”的源头感知, 而低时延响应则高度依赖于“星安”的边缘计算能力, 二者缺一不可.

6 结论

卫星互联网作为软件定义和云原生技术驱动下的产物,其本质已演变为一个网络节点、组合关系和数据流向持续演进的开放复杂巨系统。本文的分析表明,其在星、云、站、端全链路均暴露出系统性的安全脆弱性,面临着物理、电磁和网络空间的多重威胁向量。这种“体系化攻击”的态势,决定了传统的、依赖“碎片化产品”堆叠的防御思想已然失效,其安全防护必须采用体系化的思路,转向“体系化攻击 VS 体系化防御”的持续对抗。

为应对这一挑战,本文提出构建“卫星互联网安全防护体系”这一新型战略基础设施。该体系以安全大脑为核心,以节点、数据、应用、代码四大治理框架为基础,通过部署“云安”(指挥中心)、“站安”(指挥官)、“星安”(哨兵)和“端安”(第一道防线)四级智能体,构建了一个逻辑集中、物理分布、全域协同的可持续演化主动防御系统。这一架构旨在解决复杂异构巨星座在高动态、强约束、广覆盖条件下的自主安全运营难题,是实现智能态势感知、跨域协同响应和体系化对抗的一种有益探索。

参考文献

- Huang T, Fang Z, Tang Q, et al. Integrated computing and networking for LEO satellite mega-constellations: architecture, challenges and open issues. *IEEE Wireless Commun*, 2024, 31: 92–100
- Jiang W. Software defined satellite networks: a survey. *Digital Commun Netws*, 2023, 9: 1243–1264
- Cao Y C, Sun Y F, Zhu Y G, et al. Game-theoretic clustering and scalable beamforming for multi-RIS-assisted cohesive satellite anti-jamming systems. *Sci China Inf Sci*, 2025, 68: 190304
- Qin X, Ma T, Tang Z, et al. Service-aware resource orchestration in ultra-dense LEO satellite-terrestrial integrated 6G: a service function chain approach. *IEEE Trans Wireless Commun*, 2023, 22: 6003–6017
- Soto D M. What Marine Corps stand-in forces should learn from Ukrainian applications of commercial satellite communication. Dissertation for Master's Degree. Monterrey: Naval Postgraduate School, 2023
- Taylor D, Stewart T, Kheifets E. Harnessing advanced technologies for swarm operations within CJADC2. In: *Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium*. 2023. 13–15
- Li T, Hu J P, Xu H Z. Analysis of software defined network applications in space-based information networks. *Telecommun Eng*, 2016, 56: 259–266 [李婷, 胡建平, 徐会忠. 天基信息网络的软件定义网络应用探析. *电信工程*, 2016, 56: 259–266]
- Feiler P, Gabriel R P, Goodenough J, et al. Ultra-large-scale systems: the software challenge of the future. *Software Engineering Institute*, 2006. <https://www.sei.cmu.edu/library/ultra-large-scale-systems-the-software-challenge-of-the-future/>
- Falco G. Cybersecurity principles for space systems. *J Aerospace Inf Syst*, 2019, 16: 61–70
- Kareem K M. Cyber threat landscape analysis for starlink assessing risks and mitigation strategies in the global satellite internet infrastructure. 2024. [ArXiv:2406.07562](https://arxiv.org/abs/2406.07562)
- Zhao J S, Wu F G, Liu G M, et al. Overall concept and technical practice of developing software defined satellite. *Satellite Network*, 2018, 4: 44–49 [赵军锁, 吴凤鸽, 刘光明, 等. 发展软件定义卫星的总体思路与技术实践. *卫星与网络*, 2018, 4: 44–49]
- Du X K, Shu N N, Liu C S, et al. Security challenges and defense technologies in low Earth orbit satellite networks: a review. *J Electron Inform Technol*, 2025, 47: 1–14 [杜星葵, 束妮娜, 刘春生, 等. 低轨卫星网络安全问题及防御技术综述. *电子与信息学报*, 2025, 47: 1–14]
- Bi Y, Han G, Xu S, et al. Software defined space-terrestrial integrated networks: architecture, challenges, and solutions. *IEEE Netw*, 2019, 33: 22–28
- Casari F, Galletta L. Space cybersecurity governance: assessing policies and frameworks in view of the future European space legislation. *J Cybersecurity*, 2025, 11: tyaf013
- Zhu X, Jiang C. Integrated satellite-terrestrial networks toward 6G: architectures, applications, and challenges. *IEEE Internet Things J*, 2021, 9: 437–461
- Yuan S, Peng M, Sun Y, et al. Software defined intelligent satellite-terrestrial integrated networks: Insights and challenges. *Digital Commun Netws*, 2023, 9: 1331–1339
- Gopal R, Ravishankar C. Software defined satellite networks. In: *Proceedings of the 32nd AIAA International Communications Satellite Systems Conference*, 2014. 4480
- Xu S, Wang X W, Huang M. Software-defined next-generation satellite networks: architecture, challenges, and solutions. *IEEE Access*, 2018, 6: 4027–4041
- Software Defined Satellite Technology Alliance. Software defined satellite node governance framework (NodeMatrix) and protocol specification. 2024 [软件定义卫星技术联盟. 软件定义卫星节点治理框架 (NodeMatrix) 与协议规范. 2024]

- 20 Software Defined Satellite Technology Alliance. Software defined satellite data governance framework (DataMatrix) and hardware abstraction layer (HAL) specification. 2024 [软件定义卫星技术联盟. 软件定义卫星数据治理框架 (DataMatrix) 与硬件抽象层 (HAL) 规范. 2024]
- 21 Software Defined Satellite Technology Alliance. Software defined satellite application governance framework (AppMatrix) and application programming interface (API) specification. 2024 [软件定义卫星技术联盟. 软件定义卫星应用治理框架 (AppMatrix) 与应用编程接口 (API) 规范. 2024]
- 22 Software Defined Satellite Technology Alliance. Software defined satellite code governance framework (CodeMatrix) and coding specification. 2024 [软件定义卫星技术联盟. 软件定义卫星代码治理框架 (CodeMatrix) 与编码规范. 2024]
- 23 Ahmad I, Suomalainen J, Porambage P, et al. Security of satellite-terrestrial communications: challenges and potential solutions. *IEEE Access*, 2022, 10: 96038–96052
- 24 Mohammed A. Protecting space assets: cybersecurity challenges and solutions for the final frontier. *Baltic J Eng Technol*, 2023, 2: 55–61
- 25 Huang Y Q, Mei L, Gao P Y, et al. Waveform cooperative communication for cohesive clustered satellite systems. *Sci China Inf Sci*, 2025, 68: 190307
- 26 Willbold J, Schloegel M, Vögele M, et al. Space odyssey: an experimental software security analysis of satellites. In: *Proceedings of IEEE Symposium on Security and Privacy (SP)*, 2023. 1–19
- 27 Qiao L, Wu W, Yan H, et al. Onboard reconfiguration method for routing FPGA programs in Leo satellite constellations. In: *Proceedings of the 15th International Conference on Advanced Infocomm Technology (ICAIT)*, 2023. 102–108
- 28 Sawik B. Space mission risk, sustainability and supply chain: review, multi-objective optimization model and practical approach. *Sustainability*, 2023, 15: 11002
- 29 Bailey B. Establishing space cybersecurity policy, standards, and risk management practices. El Segundo: The Aerospace Corporation, 2020. https://aerospace.org/sites/default/files/2020-10/Bailey%20SPD5_20201010%20V2_formatted.pdf
- 30 Frøseth E K, Kavallieratos G, Katsikas S. Threat modeling in satellite communications for maritime operations. In: *Proceedings of International Workshops on Computer Security*, 2024
- 31 Kang M, Park S, Lee Y. A survey on satellite communication system security. *Sensors*, 2024, 24: 2897
- 32 Zhan Y, Zeng G, Pan X. Networked TT&C for mega satellite constellations: a security perspective. *China Commun*, 2022, 19: 58–76
- 33 Harrison T, Johnson K, Roberts T G. *Space Threat Assessment 2019*. Washington, DC: Center for Strategic & International Studies, 2019
- 34 Liu Z, Lin C, Chen G. Space attack technology overview. *J Phys-Conf Ser*, 2020, 1544: 012178
- 35 Wang G, Wan G, Su Z, et al. High-performance on-orbit intelligent computing and real-time services for remote sensing satellites based on large-scale computing power in space. *IEEE Access*, 2025, 13: 92114–92133
- 36 Al-Sada B, Sadighian A, Oligeri G. MITRE ATT&CK: state of the art and way forward. *ACM Comput Surv*, 2024, 57: 1–37
- 37 Khan S K, Shiwakoti N, Diro A, et al. Space cybersecurity challenges, mitigation techniques, anticipated readiness, and future directions. *Int J Crit Infrastructure Protection*, 2024, 47: 100724
- 38 Carlo A, Obergfaell K. Cyber attacks on critical infrastructures and satellite communications. *Int J Crit Infrastructure Protection*, 2024, 46: 100701
- 39 Liu W, Wu M, Wan G, et al. Digital twin of space environment: development, challenges, applications, and future outlook. *Remote Sens*, 2024, 16: 3023
- 40 Kuang W Y, Zhang S S, Zhao J S. Technical challenges and conceptual design of the security brain for mega-constellation networks. *J Aerospace Tech Eng*, 2025, 2: 16–24 [匡文渊, 张珊珊, 赵军锁. 巨星座网络安全大脑的技术挑战与概念设计. *航天技术与工程学报*, 2025, 2: 16–24]
- 41 Zhu L D, Zhang Y, Jia G Y. Status and prospects of routing technology for satellite internet. *J Commun*, 2024, 42: 33–42 [朱立东, 张勇, 贾高一. 卫星互联网路由技术现状及展望. *通信学报*, 2024, 42: 33–42]
- 42 Li J Q, Chen Q, Yang L. Research status and prospects of routing technology for low Earth orbit (LEO) mega-constellations. *Space Electron Technol*, 2025, 22: 1–12 [李佳奇, 陈全, 杨磊. 低轨巨型星座路由技术研究现状及展望. *空间电子技术*, 2025, 22: 1–12]
- 43 Zuo Y, Yue M, Yang H, et al. Integrating communication, sensing and computing in satellite Internet of Things: challenges and opportunities. *IEEE Wireless Commun*, 2024, 31: 332–338

Analysis of security threats and systematic defense strategies for satellite Internet

Hang GAO^{1,3}, Junsuo ZHAO^{2,3*} & Wenyuan KUANG⁴

1. *Key Laboratory of Basic Software and Systems, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China*

2. *Institute of Software, Chinese Academy of Sciences, Beijing 100190, China*

3. *University of Chinese Academy of Sciences, Beijing 101408, China*

4. *360 Information Technology Co., Ltd., Beijing 100102, China*

* Corresponding author. E-mail: junsuo@iscas.ac.cn

Abstract In response to the explosive growth of the global satellite internet and the five key evolutionary trends of LEO mega-constellations-scalability, networking, intelligence, servitization, and ecosystem expansion-this paper conducts an in-depth analysis of the asymmetric threats facing “100000-satellite” scale networks. Specifically, it examines challenges such as supply chain attacks, side-channel exploitation, and cross-domain coordinated strikes that target core aerospace nodes, cross-domain links, and space-ground integrated boundaries. To overcome the limitations of traditional defense models regarding heterogeneous interoperability, collaborative command, and system trustworthiness, this paper proposes a systematic defense strategy. This architecture establishes the “Satellite Internet Security Brain” as the decision-making core and the “Star-Cloud-Station-Terminal” four-tier agents as execution units, integrated with the NodeMatrix, DataMatrix, AppMatrix, and CodeMatrix governance frameworks. This approach effectively resolves critical challenges involving unified resource abstraction in heterogeneous environments, multi-source data interoperability, agile tactical response, and trusted full-lifecycle software evolution, thereby providing a robust solution for the long-term reliability of complex mega-constellation systems.

Keywords satellite Internet, complex and heterogeneous mega-constellations, cybersecurity, governance framework, systematic defense