 DOI: https://doi.org/10.12349/iser.v5i7.3118

The Dual-track Dynamic Balance Identification Method For the Boundary of Critical Information Infrastructure

Zijie Luo[1] Ruining Liu[1*] Wenzhuo Du[2] Peiyu Luo[2] Guancheng Chen[2]

1. Computer Science Department of Guizhou Police College, Guiyang, Guizhou, 550005, China
2. Xinwang College, People's Public Security University of China, Beijing, 100038, China

Abstract

Critical Information Infrastructure (CII) is pertinent to the stability and orderly development of a country. In the present era, countries worldwide are placing increasing emphasis on and stressing the protection of CII, and the safeguarding of CII is predicated on the precise identification of its boundaries. This paper draws inspiration from the DNA double-strand base complementary pairing model and biomimetically constructs the on-track synchronization model to alleviate the redundant identification tasks of CII operators and enhance the autonomy and sustainability of CII operators. In the CII boundary identification process, the business flow composed of the CII business requirement part and the hardware equipment part is utilized as the identification object and the carrier to circumvent the difficulty of a sole identification approach for the boundary identification of all types of CII. Moreover, in combination with the GAN algorithm strategy, the availability of this model is verified, with the aim of achieving effective identification of CII in significant industries and domains, enhancing the security of critical information infrastructure in multiple fields and industries, and promoting the development of information science.

Keywords

key information; infrastructure; boundary identification; GAN algorithm

基于 GAN 算法验证的关键信息基础设施边界识别模型研究

罗子杰[1] 刘瑞宁[1*] 杜文卓[2] 罗珮渝[2] 陈冠诚[2]

1. 贵州警察学院计算机科学系，中国 · 贵州 贵阳 550005
2. 中国人民公安大学信网学院，中国 · 北京 100038

摘　要

关键信息基础设施（critical information infrastructure，CII）关系到国家稳定与有序发展，当今世界各国越发重视与强调保护CII，而CII的保护基于对其边界的准确识别。论文借鉴DNA双链碱基互补配对模型，仿生构建对轨同步模型，用于减少CII运营者冗余识别任务，提高CII运营者自主性和持续性。在CII边界识别工作中以CII业务需求部分与硬件设备部分组成的业务流为载体作为识别对象，规避单一的识别方式对于全部种类CII边界识别的困难性。并结合GAN算法策略验证该模型的可用性，以期实现对重要行业和领域的CII有效认定，提升多领域、多行业关键信息基础设施安全，促进信息科学发展。

关键词

关键信息；基础设施；边界识别；GAN算法

1 引言

2017 年 6 月 1 日，《中华人民共和国网络安全法》正式实施。该法是中国站在全局性、前瞻性视角对网络社会进行治理的重要举措。该法提出网络安全战略和人才培养战略，明确了关键信息基础设施保护、国家安全审查、监测预警和信息通报等制度。关键信息基础设施相关保护机制建设提上日程。

【作者简介】罗子杰（2002–），男，中国贵州毕节人，在读本科生，从事网络安全、人工智能研究。

【通讯作者】刘瑞宁（1991–），女，中国河南安阳人，硕士，讲师，从事网络安全研究。

关键信息基础设施（critical information infrastructure，CII）是指公共通信和信息服务、能源、交通、水利、金融、公共服务、电子政务、国防科技工业等重要行业和领域的，以及其他一旦遭到破坏、丧失功能或者数据泄露，可能严重危害国家安全、国计民生、公共利益的重要网络设施、信息系统等[1]。如今中国各方面科技实力不断发展，CII 的保护愈发重要，而精准有效地实现对 CII 保护的前提是明确其范围边界。

CII 边界是指以关键业务为基础，由识别方法和关键信息基础设施元素构成，反映关键信息基础设施元素与关键业务之间的支撑、依赖关系以及关键信息基础设施元素的分布、部署情况，是开展保护、审查、应急处置等工作的重要依据[2]。

长期以来，CII 边界的定义包括“关键功能”“关键设备”“关键信息”“关键业务”“关键风险”等若干种界定方法。其中，任意一种单一的识别方式，其识别边界可能过宽或过窄，这是由于这项保护工作所要面对的攻击形式、类型、规模已经超越关键基础设施的传统防御范围。因此 CII 的界定方法或识别方式应当基于多重因子。

论文从 CII 边界业务流需求与硬件支撑变化适配管理出发，以关键业务为主，关键设备为辅，借鉴 DNA 双链碱基互补配对模型。仿生构建业务流构成模型用于识别 CII 边界时，规避单一的识别方式对于全部种类 CII 边界识别的困难性。仿生构建对轨同步模型用于减少 CII 运营者冗余识别任务，提高 CII 运营者自主性和持续性。确保在 CII 边界业务需求和硬件支撑在不断变化不断融合的过程中提高二者的适配度，以期实现对重要行业和领域的 CII 有效认定，提升多领域、多行业关键信息基础设施安全，促进信息科学发展。

2 CII 边界识别方法

2.1 业务流组成

论文对 CII 边界的识别从业务流角度出发，以 CII 业务需求部分与硬件设备部分组成的业务流为载体并作为识别对象，规避单一的识别方式对于全部种类 CII 边界识别的困难性。如图 1 所示构建业务流组成模型。业务流分为业务需求部分与硬件设备部分，业务需求部分由关键业务及其关键风险构成，硬件设备部分由关键设备及其关键功能（针对关键业务产生的关键风险缺口由关键设备更新其关键功能弥补）构成。

业务流是指在 CII 运作时，各设备共同作用形成的某项业务秩序。同一设备可参与不同业务流，但其在不同业务流中的重要程度也有所不同。在某业务流的运行过程中，剖析其各环节所用到的关键设备，筛选出该环节下最具关键性的关键设备。

关键风险是指针对某一业务流应用场景，关键设备的功能胜任力有限而产生的缺陷风险。例如，在高轨道卫星（GEO）移动通信业务流中，卫星环节的关键设备为高轨道卫星。针对高轨道卫星（GEO）移动通信业务流有电话信息传输时延、能源不足、太空垃圾碰撞卫星等关键风险。综合考虑以上关键风险的发生频率高低及其影响大小，在该业务流中最为代表的关键风险为电话信息传输时延。

2.2 构建对轨同步模型

对于 CII 的识别，大多基于各相关方多轮协商的最大共同认可确定，识别结果的客观性和有效性往往难以通过某单一验证机制确认。因此，需要采用持续性机制，在上次识别结果基础上不断迭代优化，根据信息基础设施的各关键风险组成、网络安全动态及组织管理调整等因素，进行定期或不定期的多次识别。但这种持续性机制对于 CII 运营者而言，往往缺少自主性和持续性。

由此论文借鉴 DNA 双链碱基互补配对模型，仿生构建如图 2 所示的对轨同步模型，该模型可快速反映关键业务与关键设备迭代更新是否同步，提升同步频率，减少 CII 运营者冗余识别任务，提高 CII 运营者自主性和持续性。对轨同步模型借鉴 DNA 双螺旋结构，由关键业务与其级联的关键风险作为 CII 主轨，联合关键设备检测方，结合关键设备与关键功能之间的支撑、依赖关系，获取最新关键设备情况作为副轨。关键风险与关键功能模仿 DNA 双螺旋结构碱基互补相互适配映射，并平衡业务需求与硬件设备约束（在各阶段工作结束或定性完成后，为保证成果的时效性、可用性、完整性、独立性，赋予该成果的规则限制）提升二者适配度，确保 CII 边界识别清晰准确、时效性长。

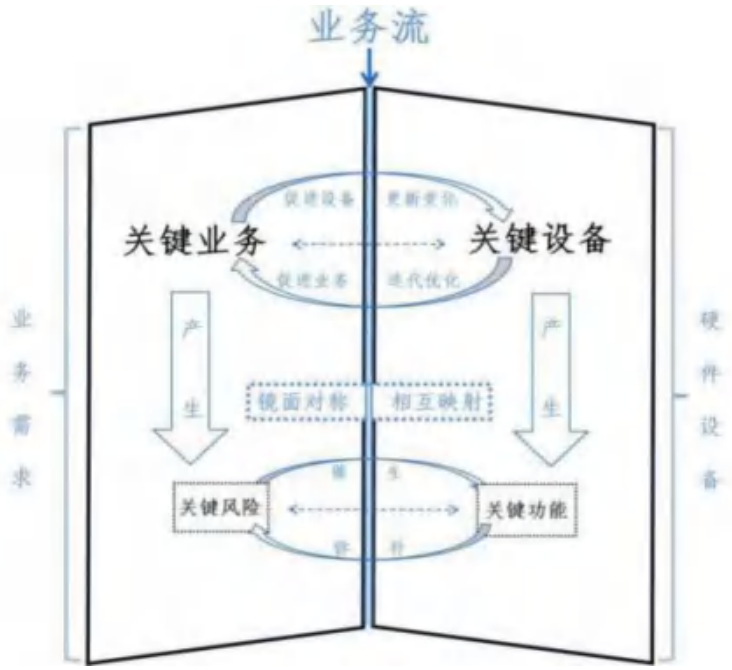

图 1 业务流构成模型

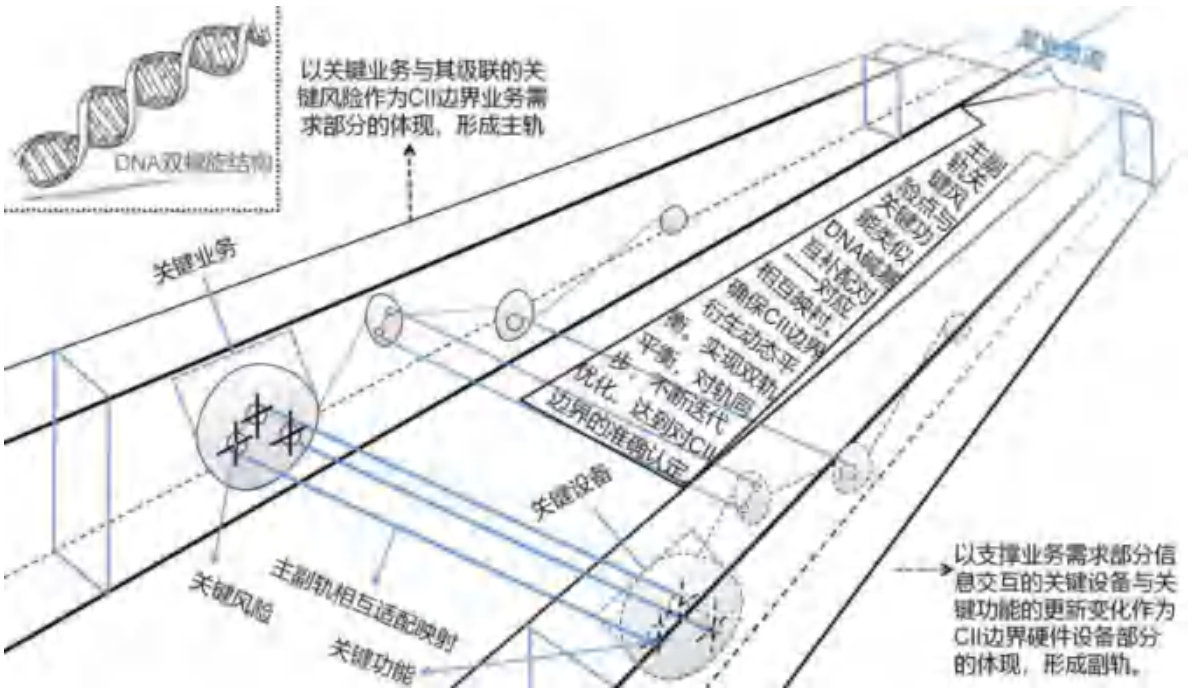

图 2 对轨同步模型

3 验证对轨同步模型的有效性

3.1 验证方法分析

CII 识别工作必须具备一定的专业和技术知识，需要 CII 运营者在国家有关部门的指导监督下完成。各重要行业、领域主管部门从宏观和整体角度对本行业、本领域 CII 运营者的重要性做出判断，但缺乏有关 CII 运营者业务运行具体情况的信息。CII 运营者因缺少对国家网络安全整体态势的深入认识，难以对自身运营业务的重要性进行客观判断。由于识别与被识别双方信息的不对称，造成 CII 的识别结果五花八门，甚至千差万别[3]。此外，以往边界识别存在识别时效短、边界识别模糊等问题。鉴于此，提出 CII 边界的对轨同步模型，以减少 CII 运营者冗余识别任务，提高 CII 运营者自主性和持续性，实现清晰识别边界，提高识别时效，从而实现对重要行业和领域的 CII 有效识别及通信安全。为通过具体数据可视化验证该模型运用于 CII 边界的识别是否有效，需选择一种能实现不断生成、不断验错、不断接近直到达到与真实数据相同的人工智能算法对该模型的有效性进行验证，我们可选择生成对抗网络（generative adversarial network，GAN）人工智能算法对轨同步模型的有效性进行验证。

3.2 具体验证方法

在对 GAN 进行介绍前，为了方便统一与理解，采用如表 1 所示的符号定义。

真实样本 x 服从数据分布 p_d（$x\sim p_d$），生成样本 x’ 服从生成数分布 P_g（$x'\sim p_g$），标签（条件）记为 $y\epsilon Y$，隐变量（也称噪声）记为 z，服从先验分布 P_z（$z\sim p_z$），样本的真实标签记为 y_p，否则记为 y_n[4]。

表 1 关键符号定义

符号	描述
G	生成器
D	鉴别器
x	真实样本
X	全体真实样本（真实样本空间）
x'	生成样本
$\overline{X}$	生成样本空间
y	真实标签（条件）
Y	标签数量
p_z	先验分布
z	隐变量（噪声）$z\sim p_z$
v	样本的特征向量（分类层输入端）$v=D_b(x)$
y_p	样本的正确标签
y_n	样本的错误标签

生成对抗网络（generative adversarial network，GAN），它是一种生成模型，由 1 个生成器（G）和 1 个判别器（D）组成，其模型架构如图 3 所示。生成器接收噪声 z 作为输入，其任务是生成尽可能接近真实数据 x 的数据样本。判别器的输入可以是真实数据 x 或是生成器的生成数据 x'，其输出是一个概率值，表示判别器识别输入是真实数据的概率。若判断输入为真实数据，则输出接近 1，否则输出接近 0。生成数据和判别结果相互对抗，不断提高生成器的性能。

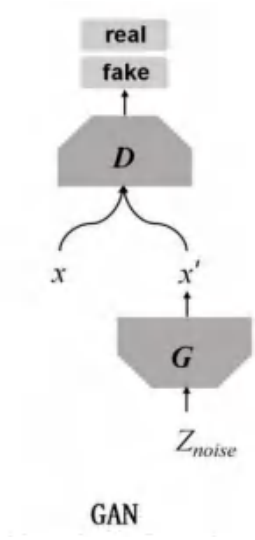

图 3 GAN 模型架构

GAN 的训练可以看作最大最小优化问题：$min_G max_D V(D,G)=E_{x\sim Pd}[log(D(x))+E_{z\sim pz}[log(1-D(G(z)))]$，其中，鉴别器 D 的训练目标为最大化 V(D,G)，若输入真实样本，希望 D（x）接近 1，若输入生成样本 G（z），D（G（z））接近 0，这意味着鉴别器能够正确辨别样本真伪。而对于生成器 G，训练目标正好相反，期望 G（z）接近 1，这意味着生成器完美欺骗鉴别器。

GAN 算法逻辑与对轨同步模型中的主轨、副轨识别工作逻辑相符。由此可将 GAN 算法逻辑与对轨同步模型结合，规避单一的识别方式对于全部种类 CII 边界识别的困难性。对轨同步模型中业务需求部分迭代优化生成新的关键业务，硬件设备部分针对新的关键业务所产生的新的关键风险提供多个不同的关键功能逐一弥补尝试，判别区分针对目前关键风险的最适关键功能，以期完善关键风险防御机制，不断达到边界识别清晰。

针对任一业务流建立对轨同步模型，分析出其关键业务、关键风险、关键设备、关键功能的特征值，根据预先设计好的模型对特征值进行计算，得出业务流中业务需求部分与硬件设备部分特征值。将硬件设备部分不同的关键功能分析得到的不同特征值作为生成样本特征值 G（z）逐一输入 GAN 算法中的判别器，直至 D（G（z））接近 0 时，意味着硬件设备部分可判别区分出关键设备针对目前的关键风险提供的最适关键功能。业务需求部分特征值输入 GAN 算法中的生成器，训练目标正好相反，期望 G（z）接近 1，意味着生成器完美欺骗判别器，关键设备提供的多个关键功能均无法弥补关键风险缺口，需再次更新关键功能，弥补关键风险缺口，实现双轨动态平衡。

4 结语

论文提出一种可清晰识别 CII 边界的对轨同步模型。该方法可针对任一业务流建立对轨同步模型，规避了单一识别方式对于全部种类 CII 边界识别的困难性，从宏观层面提供了全种类 CII 边界的统一识别方法，减少了 CII 运营者冗余识别任务，提高了 CII 运营者自主性和持续性。后续工作将持续跟进该方法的实际运用效能，不断收集实验数据，完善 CII 保护机制，以期实现对重要行业和领域的 CII 有效认定，提升多领域、多行业关键信息基础设施安全，促进信息科学发展。

参考文献

[1] 中国政府网.关键信息基础设施安全保护条例[Z].(2021.07.30)[2023.9.1].

[2] 河北网信网络安全宣传教育网安标准.信安标委发布关键信息基础设施边界确定方法[Z].(2022-05-16)[2023.9.1].

[3] 冯燕春,胡容铨,谭元翼,秦小伟.如何识别关键信息基础设施的边界[J].中国信息安全,2018(12):99-101.

[4] 熊海裕.条件生成对抗网络的生成质量与模式崩溃问题研究[D].长沙:中南大学,2022.