arXiv:2603.03526v1 [cs.MA] 3 Mar 2026

多智能体影响图到混合威胁建模

Maarten Vonk[1,2],Anna V. Kononova[2],Thomas Thomas B¨ack[2] andTim Sweijs[1]

[1]海牙战略研究中心，Lange Voorhout 1，海牙，2514EA，荷兰。[2]LIACS，莱顿大学，Einsteinweg 55，莱顿，2333CC，荷兰。

摘要

西方政府已采取各种反混合威胁措施，以防御低于常规军事门槛的敌对行动。由于混合威胁本身含糊不明、具有跨领域性质，而且反制措施如何影响对手行为也存在不确定性，这些措施的影响尚不明确。本文提出一种新方法，通过（多主体）影响图框架，将以往彼此分离的混合威胁建模方法整合起来，从而阐明这些措施的影响。该模型权衡反制措施的成本、其劝阻对手实施混合威胁的能力，以及其减轻混合威胁影响的潜力。我们针对一个受现实启发的情境运行了 1000 个半合成变体，模拟攻击方智能体 A 与防御方 B 围绕针对关键基础设施的网络攻击所展开的战略互动，以探究五种不同反混合威胁措施的有效性。反混合措施涵盖从增强韧性、拒止对手实施混合威胁的能力，到以惩罚威胁进行劝阻等手段。我们的分析主要评估反混合威胁措施的总体特征。此方法使我们能够概括这些措施的有效性，并考察参数影响敏感性。此外，我们还讨论政策相关性，并概述未来的研究方向。

关键词： 混合威胁、对手行为分析、战略互动建模、威慑、影响图、因果推断、博弈论

1 引言

混合威胁是指协调运用暴力与非暴力手段，在武装冲突门槛以下利用漏洞并影响对手。随着全球互联互通日益增强，这类威胁正成为日益严峻的挑战。为此，各国已实施范围广泛的潜在反混合措施，包括经济制裁、网络防御策略、信息宣传活动和外交倡议。然而，由于混合威胁性质复杂且不透明，往往跨越多个领域并难以归因，这些措施的有效性仍不确定。因此，研究人员转而采用建模方法，运用博弈论方法（Balaban 和 Mielniczek，2018）或概率方法（Balcaen 等人，2022年），因为这些方法能够通过系统评估不同措施在各种条件下的影响来应对这种不确定性，同时为理解威胁与反制措施之间的相互作用提供可行见解。

在这些方法论各异的研究路径基础上，本文提出一种整合概率论与博弈论的模型，以全面评估反混合威胁措施的有效性。我们以概率方式对两个类国家主体之间的互动进行建模，以考虑认知和心理威慑因素（Berejikian，2002），并以博弈论方式进行建模，以刻画战略决策。防御方的收益取决于反制措施成本与混合攻击潜在损害之间的权衡，同时考虑此类攻击可能已成功威慑，也可能未能成功威慑。通过最大化期望收益，推导出最优反制措施；同时，结合对手的战略应对，归纳出博弈均衡。

为检验这一建模方法，我们借鉴真实事件，构建了一个针对关键基础设施的网络威胁情景。我们咨询了政策专家并查阅现有文献，以确定应对此网络威胁的相关反制措施，并汇总各项反混合措施的成本、损害缓解能力和威慑能力估算值。这些估算为分析反混合措施提供了依据，使我们能够评估其在不同情景下的有效性，包括对手参与战略竞争的情景。为验证所提出的方法，我们将研究结果置于现有研究框架中加以考察，并开展敏感性分析，以识别并量化影响模型结果的关键变量。由此，我们得以探讨以下研究问题：

•在不确定以下因素的情况下，反混合威胁措施具备哪些特征，最能帮助防御方应对攻击方对关键基础设施构成的网络威胁：1）这些措施能否阻止对手发动攻击；2）这些措施能否减轻攻击造成的影响；3）这些措施的成本。

• 在针对关键基础设施的网络威胁情境下，反混合威胁措施具备哪些特征，能够促使参与者的策略形成均衡？需要考虑以下因素：1）反混合措施和混合行动的成本；2）这些措施减轻攻击影响的能力；3）采取战略行动的对手。

为回答这些研究问题，本文结构安排如下：下一节介绍威慑与混合威胁的背景信息。随后，我们概述建模方法论，并讨论如何通过概率分布将文献综述和专家意见中的见解转化为模型输入。接下来一节介绍混合威胁情景的内容及相关的跨域反混合威胁措施。倒数第二节将呈现并分析结果，并将其置于类似研究的更广泛框架中加以阐释。最后一节探讨研究发现的政策相关性，并指出未来研究方向。

2 背景

技术的快速发展、经济一体化的深化以及社会的数字化互联，重塑了当代国家间竞争，为修正主义国家提供了创新工具，使其得以在大规模武装冲突门槛以下追求战略目标。在欧洲，观察人士使用混合威胁这一术语，广义上指“协调一致”的运用暴力与非暴力手段的行为；这些行为往往难以检测和归因，旨在将民主进程武器化，并对对手施加影响。[1] 尽管“混合冲突”一词常被交替使用，但在概念上，它不同于美国学者和战略家所说的“灰色地带”策略（Paul，2016）。灰色地带战略指的是一种独特的、近乎持续的国家间冲突状态：通过法律正当性依据加以掩饰的侵略性行动，被用来胁迫对手并追求有限目标，同时利用对手的漏洞，在检测与归因门槛以下采取行动（Morriset al，2019）。然而，这两个术语都指代冲突门槛以下的侵略性行为，包括信息与心理行动、政治、外交和经济胁迫、进攻性网络行动，以及利用代理人破坏对手稳定。

各国目前正寻求制定一系列反混合威胁政策，旨在增强韧性并提高敌对国家的成本，以威慑此类行为。然而，与常规与核领域中公开的侵略或进攻行为相比，混合行动因其不透明且跨领域的性质而带来独特挑战，难以加以威慑和防御（Mallory，2018）。

[1] Hybrid CoE，《混合威胁：一种概念》，访问于2023年6月7日，https://www.hybridcoe.fi/hybrid-threats-as-a-phenomenon/

传统上，威慑旨在“阻止或约束一个国家采取不受欢迎的行动”（Mazarr，2021），其方式包括剥夺对手实现目标的能力（拒止威慑）（Wilner 等，2021；Lupovici，2023；Arce，2023），或威胁在对方采取行动后施加成本（惩罚性威慑）（Freedman，2004；George 和 Smoke，1974）。关于常规与核威慑的研究普遍承认，认知是威慑行动成功的关键（Mazarr，2021），也就是说，对手必须认识到“其主动行动可能招致的成本将超过潜在收益”（Nyemann 和 Sørensen，2018）。为实现目标，威慑行动应当清晰、适度且可信。清晰性意味着能够明确传达防御方可能采取哪些措施予以回应；比例性指用于威慑的手段与防御方所追求目标之间的相称关系；最后，可信性是清晰性与比例性的函数，根植于威慑方应对外部侵略的能力和意愿（Mazarr，2021）。因此，威慑的传统原则建立在可信的惩罚威胁或可信的收益拒止之上。

这些经典威慑有利条件的应用在混合情境中受到严重阻碍。首先，灰色地带中的攻击性行为并不仅仅表现为军事对抗，而是军事与非军事、公开与隐蔽行动的复杂组合，涉及经济胁迫、虚假信息宣传活动、进攻性网络行动，甚至武装团体的部署。[2] 其次，混合威胁带来了额外且独特的问题，显著模糊了对抗发生的方式。其中，应对混合威胁时最紧迫的两个问题源于灰色地带活动持续发生，且往往难以归因于特定对手。这使得威慑更加复杂（Sweijs and Zilincik，2021）。

传统的域内惩罚和拒止战略虽然仍有一定作用，但已不足以应对灰色地带中发生的复杂互动（Gartzke 和 Lindsay，2019）。传统威慑战略应发展为复杂的跨域战略；除了威胁施加成本和拒止潜在收益外，还应包括向对手提供保证和激励（保证）、促进国际合作与规范制定（规范），以及利用经济和系统性相互依赖关系（纠缠）来影响对手行为。此外，鉴于灰色地带始终处于紧张状态，威慑行动应具有“累积性”：防御方应将反混合威胁战略视为“一个长期过程，单次越轨并不意味着失败”，在这一过程中，对手行为“由威慑方通过协调一致的努力加以塑造”（Sweijs 和 Zilincik，2019；Tor，2017）。在

[2]北约对混合威胁的应对，访问于2023年4月19日，https://www.nato.int/cps/en/ natohq/topics 156338.htm -

这一意义上，作者认为，更广泛的劝阻战略更适合应对混合威胁。劝阻在此被理解为一种总体战略，涵盖惩罚和拒止响应，并辅以先进的反制措施，以利用同侪之间的政治、外交和经济关系（NyeJr，2016）。因此，要劝阻灰色地带中的对手，只有战略性地运用外交—信息—军事—经济—金融—情报—执法（DIMEFIL）领域中的所有国家权力工具，同时管理升级动态以及对手可能采取的报复性响应，反混合威胁战略才能奏效。[3]

然而在实践中，对于如何制定成功的跨域响应，并没有公认的原则、指标或指导方针，而且行为失衡且模棱两可的风险始终存在（Sweijs 和 Zilincik，2021）。由于威慑和劝阻最终根植于认知，反混合威胁战略的有效性取决于对手的现实动机及其感知到的核心利益。这包括其采取进攻行动的倾向，以及它试图保护、以免遭可能报复的漏洞。[4] 然而，在大多数情况下，政策制定者缺乏充分的信息，无法了解对手在灰色地带追求的更广泛战略目标、驱动战争门槛以下行动的决策过程和收益计算，以及反混合威胁政策在多大程度上会影响混合威胁行为；这在很大程度上是因为大量混合威胁发生在检测与归因门槛以下（Kilcullen，2019）。因此，要衡量反混合威胁政策在现实世界中的有效性极其困难。

在尝试对混合威胁动态建模时，一些作者求助于博弈论，以研究敌对国家之间的战略互动，试图克服可用信息匮乏的问题（Balcaen 等，2022年；Attiah 等，2018）。其他人则将稀缺的数据来源纳入贝叶斯建模技术（Davis 等，2021；Balaban 和 Mielniczek，2018），旨在利用现有数据完善领域知识。尽管当前的博弈论方法难以捕捉混合威胁动态中固有的复杂性与不确定性，贝叶斯建模技术虽能有效处理不确定性，却无法充分呈现战略互动。

本文提出了一项创新性贡献：采用影响图方法，将反混合威胁政策（例如威胁检测、归因和缓解效果）所涉及的深层不确定性建模为概率关系（Howard 和 Matheson，2005）。通过将该方法扩展至多智能体框架（Koller 和 Milch，2003），本文纳入了博弈论考量，提出了一种统一的方法论，不仅弥合了这两种方法之间的差距

[3]《战略传播：混合威胁工具包》，Monika Gill、BenHeap 和 Pia Hansen（北约战略传播卓越中心），第 30–36 页，访问于 2023年4月26日，https://stratcomcoe.org/publications/strategic-communications-hybrid-threats-toolkit/213。[4]《反制混合威胁的九诫》，《国际政治季刊》，Michael R¨uhle，访问于 2023年5月20日，https://ip-quarterly.com/en/nine-commandments-countering-hybrid-threats

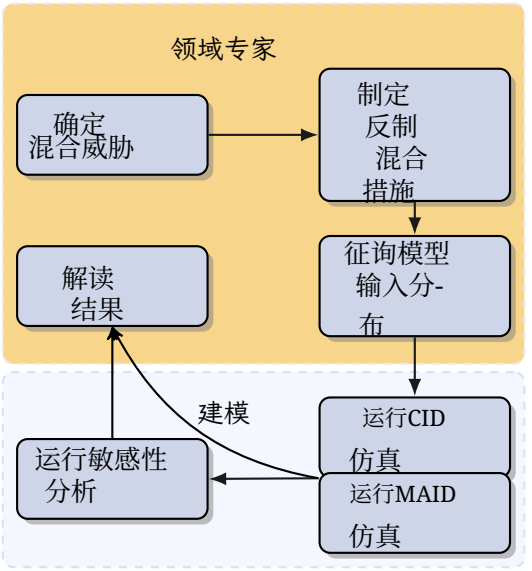

图 1：该图展示了本研究提出的反混合威胁分析所涉及的流程。首先，领域专家识别混合威胁，并制定相应的反混合措施。他们还提供关键输入参数，用于构建模型的概率输入分布。从这些分布中抽取样本，用于运行因果影响图（CID）和多智能体影响图（MAID）模型的仿真。最后，进行敏感性分析，并对模型结果进行解读，再与现有研究进行比较。

，还增强了它们在复杂、多行为体混合威胁情景中的适用性。

3 方法论

本节概述仿真模型的底层机制，以及如何获取运行该模型仿真所需输入的流程。拟议方法的完整范围（包括专家参与程度）见图 1。

该模型考虑了两个具备主权国家特征的主体的行为，并采用 Balcaen 等人的双主体方法。Balcaen 等人 (2022年) 和 Attiah 等人 Attiah 等人 (2018)。一方的代理人 A 旨在通过混合攻击实现其战略目标。另一方的代理人 B 则希望保护其国家利益，并威慑和防御混合攻击。代理人 B （防御方）选择一种反混合态势，以威慑或劝阻代理人 A 发动混合行动。因此，这种策略被称为反混合措施。为此，代理人 B 探索可用的反混合措施，通过改变成本效益计算来劝阻对手

发动混合攻击（Brantly，2018）。防御方还可以采取措施，例如增强检测和/或归因能力（Schneider，2019），以提升韧性并减轻混合行为可能造成的影响（Gilad 和 Tishler，2023）。

反混合措施和混合攻击都会产生直接成本。这些成本不仅包括资源成本，还包括诸如动员国内和国际支持所需的政治资本，以及可能与升级相关的成本。主体 A 与主体 B 之间的互动具有零和性质。使用概率来反映关键变量值存在相当程度的不确定性，而这些变量会导致不同结果。例如，反混合措施对主体 A 策略权衡的影响存在不确定性，检测、归因以及反混合措施的缓解效果也存在不确定性。

首先，本文介绍一种因果影响图方法，当对手的响应性以概率方式估计时，该方法可用于优化反混合威慑策略。随后，通过在博弈论框架中将两个主体都建模为参与者，将这种方法扩展为多智能体影响图，从而能够分析博弈均衡。

3.1 优化反混合策略

贝叶斯网络建模技术常用于处理概率关系与确定性关系的结合（Balaban和Mielniczek，2018；Zhu等，2020；Yan等，2012年；Johansson和Falkman，2008；Wang等，2012）。这使得联合分布可以根据图的结构，分解为条件概率的乘积（Pearl，2009年）。贝叶斯网络可以扩展为因果影响图，进一步将图中的节点细分为随机变量、效用节点和决策节点（Howard和Matheson，2005），并支持因果干预。在第一种提出的建模方法中，防御方预先承诺采取选定的反混合态势，以威慑或劝阻主体 A 实施混合行动，该态势由一个决策节点表示。对手的响应取决于成功劝阻的估计概率。

更正式地，令 D 为描述反混合措施的决策变量，其后是表示混合行为的随机变量 C。 D和 C可采用的策略及状态空间分别为 $\Omega_D = \{d_1, \ldots, d_n\}$和 $\Omega_C = \{c_1, \ldots, c_m\}$。反混合措施变量与混合行动变量的每种组合都会产生一个离散收益随机变量Θ，其状态空间为 $\Omega_\theta = \{\theta_1, \ldots, \theta_k\}$，表示不包括直接成本的交互收益结构。由于我们将混合行为与反混合行为之间的收益结构，与混合行动和反混合措施的直接成本分开，因此假设该收益结构包含一个零和博弈成分，即正值 $\theta \in \Theta$ 对应于代理人 B 的收益和代理人 A 的损失，而负值对应于

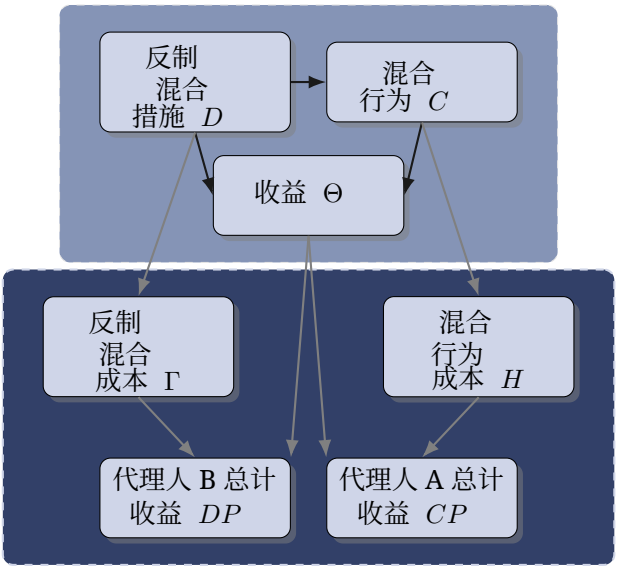

图 2：（多主体）因果影响图，用于表示混合威胁建模。底部背景层将确定性变量归为一组，而顶层则表示概率变量。概率关系以黑色箭头表示，确定性关系以灰色箭头表示。

$\theta \in \Theta$ 表示代理人 A 的收益和代理人 B 的损失。反混合措施和混合攻击的直接成本取自成本概率函数，分别记为 Γ和 H，其中反混合措施的成本状态空间为 $\Omega_\Gamma = \{\gamma_1, \ldots, \gamma_n\}$，且$\gamma_i \geq 0$适用于 $i = 1$、.....、 n；混合攻击的成本状态空间为 $\Omega_H = \{\eta_1, \ldots, \eta_m\}$，且 $\eta_i \geq 0$适用于 $i = 1$、.....、 m。最终可计算出总收益：代理人 B 的总收益为 $DP = \Theta - \Gamma$，代理人 A 的总收益为 $CP = -\Theta - H$。与这些关系相对应的因果影响图如图 2所示。[5] 请注意，若在初始收益节点中将直接成本分解出来，影响图本可成为纯概率图。不过，为了便于说明，我们将直接成本与交互成本分开。在影响图中，所有概率分布通常都假定为分类分布，因此条件概率可通过条件概率表（条件概率表）表示。根据影响图的结构，联合概率可分解为：

$$P(d, c, \theta, \gamma, \eta, dp, cp) = P(d)P(c \mid d)P(\theta \mid d, c)P(\gamma \mid d) \\ P(\eta \mid c)P(dp \mid \gamma, \theta)P(cp \mid \theta, \eta).$$

[5]虽然成本变量最初是从概率分布中抽取的，但在影响图中它们变为确定性变量，因为每次实验中，每种（反制）混合措施都对应一个成本值。

为了计算代理人 B 的总收益概率，我们必须对所有其他变量进行边缘化（可以忽略网络其他部分中的确定性变量（Koller 和 Friedman，2009年））：

$$
\begin{aligned}
P(dp) &= \sum_{d}\sum_{c}\sum_{\theta}\sum_{\gamma} P(d,c,\theta,\gamma,dp) \\
&= \sum_{d}\sum_{c}\sum_{\theta}\sum_{\gamma} P(d)P(c \mid d)P(\theta \mid d,c) \\
&P(\gamma \mid d)P(dp \mid \gamma,\theta).
\end{aligned}
$$

假设代理人 B 可以获得反混合措施的潜在成本 $\Omega_\Gamma = \{\gamma_1,\ldots,\gamma_n\}$、其威慑能力 $P(c \mid d$以及其减轻潜在损害的能力 $P(\theta \mid d,\ c)$。代理人 B 的目标是在假定的概率分布下，计算使其总收益最大化的反混合措施（Andrew 等人，2022年）。具体而言，我们要找到干预 $do(D = d)$，使代理人 B 的预期总收益最大化。由于反混合措施节点没有入箭头（见图 2），对变量进行干预等同于以该变量为条件（Pearl，2009年）：

$$
\begin{aligned}
&\max_{d_1,\ldots,d_n} \mathbb{E}[dp \mid do(D = d)] = \\
&\max_{d_1,\ldots,d_n} \sum_{c}\sum_{\theta}\sum_{\gamma} dpP(c,\theta,\gamma,dp,\mid do(D = d)) = \\
&\max_{d_1,\ldots,d_n} \sum_{c}\sum_{\theta}\sum_{\gamma} dpP(c \mid d)P(\theta \mid d,c) \\
&P(\gamma \mid d)P(dp \mid \gamma,\theta).
\end{aligned}
$$

这可以表述为一个整数线性规划（ILP）：

$$
\begin{aligned}
&\max \sum_{h=1}^{k}\sum_{j=1}^{m}\sum_{i=1}^{n} \theta_h w_{ijh} q_{ij} p_i + \sum_{i=1}^{n} \gamma_i p_i \\
\text{主体 t} \quad \text{o} \quad &\sum_{i=1}^{n} p_i = 1 \\
&p_i \in \{0,1\} \quad i = 1,\ldots,n
\end{aligned}
$$

3.2 子博弈完美均衡

在优化反混合策略时，我们使用概率估计每项措施实施后成功威慑对手的可能性。这些概率是事先确定的，即在选择反混合措施之前得出。因此，这些概率并未考虑任何短期

表1：从概率分布中抽取的值。反混合措施的成本和混合攻击造成的破坏性影响以百万美元计。反混合措施的成本和混合攻击造成的破坏性影响取自正态分布的变体，而威慑能力和减轻破坏性影响的能力对应的概率值，则分别取自其对应的共轭先验（贝塔分布和狄利克雷分布）。

值	含义	概率分布
θ_h	潜在 损害 of a 混合攻击	绘制 from 不同的截断正态分布 each 类别 of 造成损害的影响（Norgaard和Killeen，1980）。
γ_i	Cost for 开展 反混合措施 d_i	从不同的截断正态分布中分别抽取 反混合措施 d_i （Norgaard 和 Killeen，1980）
q_{ij}	对手的概率 实施 混合 行动 tion c_j 反混合 措施 d_i	取自不同的贝塔分布 分布 for each 反混合措施。
w_{ijh}	概率 of 潜在 损害 θ_h 之后 对手实施混合 行为 c_j 和防御方 反混合措施 d_i	从不同的狄利克雷分布中抽取 用于每项反混合措施和混合 措施与混合操作的组合 组合（Spiegelhalter和Lauritzen，1990）。

互动过程中出现的收益调整，而这些调整最终会决定均衡结果。为此，我们将因果影响图方法扩展为多智能体影响图（MAIDS）（Koller 和 Milch，2003；Hammond 等人，2021），以纳入这些战略考量并探讨均衡的概念。

形式上，混合行为节点 C （图2 ）不再以概率方式建模，而是被表示为决策节点。此外，两个智能体 A 和 B之间的区分通过为各自的参与者分配决策节点和效用节点而得以明确。具体而言，决策节点D 和效用节点 DP 被分配给主体 B，而决策节点 C 和效用节点 CP 与主体 A相关联。

我们希望找到一种解概念，能够在智能体理性行动时，确定一部分可能的结果。纳什均衡（NE）是非合作博弈中广泛使用的一种解概念，指没有任何主体能够通过单方面改变策略来获得更高收益（Nash Jr，1950），但它可能导致不可信威胁。这类威胁是指主体作出的决定：如果相关情形真的出现，执行该决定并不符合其自身利益（Selten and Bielefeld，1988）。

在混合威胁博弈中，当攻击者威胁要采取混合行动，尽管从收益角度看这并不符合其自身利益时，就会出现不可信威胁均衡。

为此，我们采用子博弈完美均衡（SPE）（Hammond 等人，2021）。子博弈完美均衡被定义为对纳什均衡的一种精炼，即纳什均衡条件不仅适用于整个博弈，也适用于其中的每个子博弈。因此，SPE 消除了不可信威胁的存在，并确保有关反混合措施或混合行动的决策在互动的每个阶段都保持理性且最优。通过对所有已识别的子博弈应用逆向归纳法，可以确定子博弈完美均衡。

3.3 概率分布与征询

人们普遍认为，使用准确的条件概率填充影响图是一项艰巨的任务（Johansson和Falkman，2008），因此应制定严谨的征询过程，以确保输入数据尽可能准确。为使估算保持现实，我们基于深入的文献综述，并辅以对七名具有战略研究背景的（初级）分析师开展的迷你德尔菲法，尝试估算每项反混合措施的成本、潜在威慑能力（即拒止或惩罚能力）以及韧性增强能力。由此得到概率分布，并从中抽取样本开展实验。起初，这些分布的参数参考了文献综述。随后，分析师根据所得分布的可视化结果，对参数进行了一次性调整。表1 汇总了各变量对应的概率分布详情，而具体参数见附录。从这些概率分布中较有可能抽取到的数值，表明其与文献共识及迷你德尔菲调查的结果高度一致；而不太可能抽取到的数值，则与之契合度较低。通过反复独立地从这些分布中抽取输入变量，我们共生成了1000个实验情景。由于缺乏严谨且标准化的方法来构建这些估算的先验分布（Mikkola等人，2023），这些实验可视为半合成实验，因此我们不得不依赖所构建的概率表示。

尽管实验具有半合成性质，本文所考虑的所有反混合措施均源自现实案例，其影响也由专家评分，从而确保反映现实世界的多样性和现有实证证据。由于该建模方法能够探索无法在现实世界中通过实证检验的动态过程，因此，数据的半合成性质是开展此项分析的必要手段。此外，所提出框架的灵活性确保其适用于其他领域和混合威胁类型，因为其底层原则和相互作用具有普适性，并不局限于特定的

表2：五项反混合措施

领域	标题	能力	依据
网络	提升网络韧性 at the 更广泛的社会层面	出台立法或开展合作，以要求个体和公司采取充分的级别 of 网络韧性，基于具体风险 of the 暴露主体。	提升网络韧性 at the 更广泛的社会层面是其中之一 核心整体政府方针的原则。社会方法 to 混合/网络威胁。
网络	运用进攻性网络能力	运用进攻性网络行动，以便 to 削弱 the 目标。	通常情况下，网络攻击并不能直接归因。在短期内，it 提供了一种隐蔽的方式以影响目-标。一次所产生的效果进攻性网络攻击是可扩展的。
法律	市场限制	出台立法以限制对手访问您在某个市场中的特定行业（例如信息与通信技术）。	此类措施减少了 the 可能性 for an 对手 to 利用漏洞-性，但它也可能与其他法律相冲突承诺（see 国际贸易反对.....的承诺 市场限制基于国籍）。
外交	Open 威慑信息传递通过战略沟通-沟通	某人的战略态势以便 to 说服目标服从与其战略目标。	保持透明与恶意行为者关于 one's own 战略优势 and 可能的行动。This 提高 the 做出更明智决策的可能性明智的决策 by 恶意行为者。

测试场景。这种适应性增强了其应对广泛混合威胁挑战的实用性。

4 实验设计

我们考虑这样一种情景：防御方 B 担心修正主义方代理 A 试图通过混合攻击破坏并损害代理人 B。具体而言，防御方了解代理人 A 在网络和信息领域的进攻能力，并担心后者会对其关键基础设施发动大规模网络攻击，例如发电厂、电网、水务管理设施、港口、医疗系统和／或其他基本服务。进攻性网络行动是低于大规模武装冲突门槛的一种明显的混合威胁。事实上，近年来，网络行动日益普遍，这归因于网络空间的技术、物理和逻辑层，以及网络和技术在日常生活中的广泛应用（Arce，2023）。[6] 此外，进攻性网络行动很可能造成实质性后果，导致严重的物理损害，例如 2009年 伊朗遭遇的 Stuxnet 攻击、2012年 沙特阿拉伯遭遇的 Shamoon 攻击，或 2017年 波及全球六十多个国家的 NotPetya 攻击。

基于 Valeriano 和 Maness Valeriano 和 Maness (2022年)、Stirparo 等人 Roth 等人(2019年) 编制的更新数据集中的一系列真实恶意网络行动，构建了一份经过匿名化、合理可信的混合行动清单；[6] 此外，还参考了相关文献。鉴于新技术呈指数级发展，且当前冲突的动态局势不断演变，我们还结合专家的想象力，力求预判潜在的行动（及应对）路径，并提炼出网络领域的关键变量。由此形成了一个真实可信的网络威胁情景，详见附录。类似地，我们也拟定了合理可信的反混合应对措施，专家选出了五项应对恶意网络攻击的首要跨领域措施，汇总于表 2。

通常很难确定针对关键基础设施的网络攻击所造成的确切成本和损失（Greenberg，2019年），但专家认为，防御方及时发现攻击并从中恢复的能力，会显著影响恶意行为造成的总体影响（Morato 等人，2018；Sullivan 和 Kamensky，2017年）。因此，我们根据网络攻击对关键基础设施可能造成的影响，划分了三类影响：

- θ_1 网络攻击关键基础设施，因缺乏有效的检测能力和足够的恢复能力，导致基本服务遭受大规模中断所造成的影响。
- θ_2 网络攻击关键基础设施，造成重大损失；这是由于防御方能够部分地，或至少最终能够检测到攻击或缓解其影响[g]及其后果。

[6] 外交关系协会，《网络行动追踪器》，查阅于2022年12月1日，https: //www.cfr.org/cyber-operations/

- θ_3 由于防御方能够预先检测攻击和/或从攻击中恢复，网络攻击对关键基础设施造成的影响有限或微不足道。

由于遭受恶意网络攻击的实体不愿分享攻击造成的确切影响信息，我们根据 Lis andMendel (2019) 的研究构建了重尾半正态分布，用于抽样潜在影响，θ_1，θ_2，θ_3 分别对应三类恶意网络攻击。采用附录中指定的高方差分布进行抽样，有助于考虑网络攻击潜在影响的可变性以及对此缺乏共识的情况。

4.1 反网络攻击措施

在分析现有的网络攻击应对措施时，重点关注域内应对措施（网络空间中的措施）以及域外应对措施（其他领域中的措施），例如执法、规范制定、公共外交和经济制裁。从累积威慑的角度来看，其中一些反混合措施旨在减轻敌对网络攻击可能造成的损害，另一些则旨在通过提高对手的成本效益计算，劝阻其采取攻击性行为。本文考察的五种反网络攻击措施分别是主动情报共享d_1、在更广泛的社会层面提升网络韧性 d_2、运用进攻性网络能力 d_3、实施市场限制 d_4 以及通过战略传播开展公开威慑宣传 d_5。这些措施列于表2。防御方也可以选择不采取反混合措施 d_6。附录列出了概率分布的具体设定，用于推导各类反网络攻击措施的成本、成功威慑的概率以及减轻损害影响的概率。

4.1.1 主动情报共享

主动情报共享包括在盟友之间共享情报，以帮助检测和归因攻击。情报共享可带来多项益处，从促进和改善态势感知（Wagner 等人，2019）到制定更完善的网络安全战略（Feledi 等人，2013）。这些益处有助于减轻网络攻击造成的损害。然而，情报共享也伴随着相关成本，例如意外向恶意利益相关方泄露机密信息的成本（Wagner 等人，2019）。当对手没有意识到情报共享时，这不会影响其实施混合行动的决策。

4.1.2 提升全社会层面的网络韧性

由于关键基础设施可能高度依赖私营部门利益相关方，因此提升更广泛社会层面的网络韧性可以降低潜在影响

网络攻击对关键基础设施造成的影响。可以通过立法或公私合作伙伴关系来实现这一目标。相比之下，公共部门与私营部门之间的合作更受青睐，因为单方面立法可能会导致过度关注徒劳的合规工作（Tiirmaa-Klaar，2016）。该措施增强网络韧性，从而降低出现损害影响的概率。因此，该措施也有助于实现拒止性威慑。然而，建立此类合作伙伴关系可能成本高昂，因为这会大量占用稀缺的网络安全专家资源。

4.1.3 运用进攻性网络能力

进攻性网络攻击旨在扰乱、削弱或拒止对手的进攻能力。进攻性网络行动可用于渗透网络，使对手暂时下线，阻止其利用此类网络开展恶意网络活动。或者，这一措施也可作为采取进一步进攻措施的威胁。执行网络行动需要承担组建网络部队和实施攻击的成本。这涉及拒止威慑和惩罚性威慑。进攻性网络行动作为一种威慑措施的有效性仍有争议，因为一些学者认为，要使进攻行动具有可信度，就必须利用对手网络中的漏洞；而这些漏洞一旦修补，进攻性网络行动的效用就会降低（McKenzie，2017年）。

4.1.4 市场限制

对于关键基础设施，国家可能会决定禁止使用据称由敌对行为体生产或提供的软件、硬件或其他信息通信技术产品和服务。作为较温和的做法，国家可能会对信息通信技术服务和产品的采购施加严格的尽职调查和风险评估义务。市场限制不可避免地会产生高昂成本（Krolikowski and Hall，2023）。然而，只允许可信方运营关键基础设施，可以限制恶意行为者控制关键基础设施的能力，从而增强韧性。同时，由于对手越来越难以在恶意行为者不熟悉的系统上开展混合行动，这一措施也能显著助力拒止威慑。如此一来，便可拒止对手实施攻击的能力。

4.1.5 通过战略传播开展公开威慑信息传递

防御方国家可以发布国家安全理论文件，说明针对特定威胁所采取的应对措施。这些措施可以通过战略传播渠道主动传达。从惩罚性威慑的角度来看，这一措施的有效性显然在很大程度上取决于

细节的详尽程度和所使用的措辞（即文件中提及的报复性威胁程度）。

5 结果

本节将讨论基于上一节实验设置得出的结果。首先，我们展示使用估算的威慑概率来优化反混合措施的结果。随后，我们分析子博弈完美均衡，其中将是否实施混合攻击建模为主体的策略选择。[7]

对于 1000 次实验中的每一次，我们按防御方 B 的总收益从最低到最高，对反混合措施的有效性进行排序。图 3 展示了所有实验中反混合措施排名的计数图。

总之，尽管施加市场限制的成本高昂（d_4），但由于它们有可能减轻损害并阻止对手发动攻击，因此往往被视为最优的反混合措施。情报共享（d_1）因其成本效益而受到重视，它能够促成及时的防御行动并推动对集体应对措施的政治支持，在减轻攻击造成的损害方面发挥着关键作用。进攻性网络行动（d_3）虽然可能扰乱敌方能力，但升级风险很高，成功率也不稳定。提升网络韧性（d_2）虽成本高昂，却始终被评为在减轻损害和威慑方面都有效。公开威慑沟通（d_5）取决于对手对威胁的响应，需要具备检测、归因和通信能力才能奏效。最后，在极少数情况下，不采取反混合措施（d_6）会成为最有效的策略。

为了得出更有意义的洞见，我们现在将关注点从各项措施的具体结果，转向可以从中得出的更广泛结果

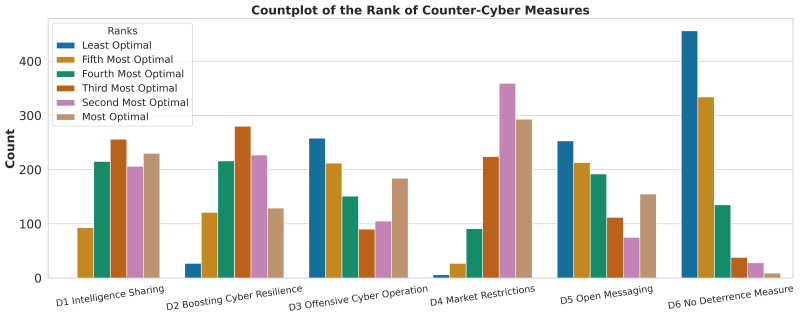

图 3：计算每项反混合措施旨在达到的特定等级的数量。

[7]建模工作已公开发布于 https://github.com/HCSS-Data-Lab/ Hybrid-Threat-Implementation。

这些措施的总体特征，尤其是考虑到，设计良好的引导方案会显著提升各项结果的吸引力与可靠性，而这些总体特征仍将保持不变。

总体而言，这些措施在几个方面有所不同：有些措施依靠通过惩罚来劝阻对手（公开传讯、进攻性网络行动），有些则着眼于在遭受攻击时减轻潜在损害（情报共享），还有一些措施兼具拒止威慑和增强韧性的作用（提升网络韧性和实施市场限制）。尽管这些特征在各项最优措施中分布均衡，但在大多数情况下被认定为最优的措施——即实施市场限制——在劝阻和增强韧性方面也最为灵活。此外，由于有利条件会影响各项措施的效果，因此它们在成本、减轻损害的能力和威慑能力方面的差异各不相同。例如，尽管以惩罚为手段的威慑措施（公开传讯和进攻性网络行动）可以是非常有效的反混合措施，但如图3所示，在某些实验中，它们也是效果最差的措施之一。这是因为这些措施的效果很大程度上取决于其对对手战略权衡的影响。如果劝阻未能奏效，它们就无法减轻混合行为造成的破坏性影响，使防御方暴露于风险之中。

为了检验输入参数的变化如何影响模型的输出，我们使用 van Stein 等人开发的最先进工具，对每项反混合措施进行了敏感性分析。Stein 等人（2022年），针对每项反混合措施开展了分析。

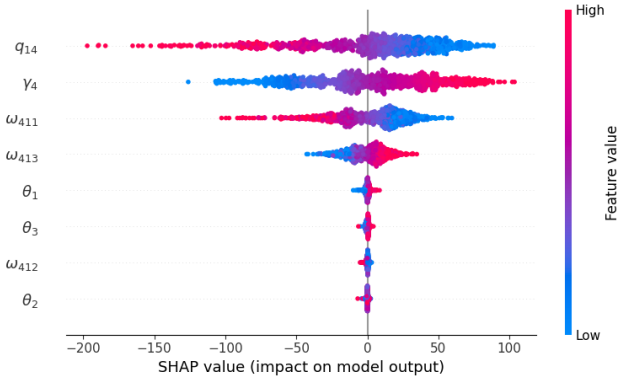

图4：实施市场限制的反混合措施的 SHAP 汇总图：纵轴表示按其对模型输出重要性排序的特征。横轴显示 SHAP 值，表明每个特征对模型输出影响的大小和方向。颜色渐变反映特征值。

虽然每项反混合措施的完整敏感性报告均可获取[8]，且表中汇总了每项反混合措施贡献最大的前三项特征3，图4 展示了实施市场限制的 SHAP 汇总图，作为所开展的更广泛敏感性分析的代表性示例。从图中可以看出，成功遏制对手的概率 q_{14}以及该措施的成本 γ_4 是影响反混合措施有效性的最重要因素。尽管该措施减轻负面影响的能力总体而言影响较小，但它在提高影响可忽略的概率（ω_{413}）或降低影响严重的可能性（ω_{411}）方面的作用，仍然是评估其有效性的重要因素。由于该措施的有效性受到其威慑对手能力的显著影响，因此不仅需要从概率角度考量这一因素，还应将其纳入博弈论框架。为此，我们计算了上一节详述的子博弈完美均衡，从而在战略情境下更全面地评估这些措施。

这1000次相同实验的子博弈完美均衡显示于表4，反映了智能体寻求理性优化其收益的结果。混合攻击发生率较低，表明所选反混合措施侧重于威慑对手，而非减轻混合攻击的后果。鉴于假定对手的战略考量已知，防御方可策略性地选择最具成本效益、并能成功威慑对手不发动攻击的反混合措施。这解释了为何在两种措施均足以威慑对手时，情报共享优于市场限制。此外，当战略考量表明无论采取何种反混合措施，对手都可能继续实施混合行动时，子博弈完美均衡表明，防御方应致力于采用成本效益较高的反混合措施（如情报共享），以减轻攻击的影响。

表3：各反混合措施SHAP图中排名前三的重要特征

反混合措施	重要特征		
	#1	#2	#3
D1 情报共享	q_{11}	ω_{111}	ω_{113}
D2 提升网络韧性	γ_2	q_{12}	ω_{213}
D3 进攻性网络行动	q_{13}	ω_{311}	ω_{313}
D4 市场限制	q_{14}	γ_4	ω_{411}
D5 开放式通信	q_{15}	γ_5	θ_1
D6 不采取威慑措施	q_{16}	ω_{611}	θ_1

[8]完整的敏感性报告可在以下网址获取：https://github.com/HCSS-Data-Lab/Hybrid-Threat-Implementation

表4：子博弈完美均衡结果出现次数

反混合措施	混合行动	
	攻击	不攻击
D1 情报共享	15	697
D2 提升网络韧性	1	98
D3 进攻性网络行动	3	76
D4 市场限制	0	110
D5 开放式通信	0	0
D6 不采取威慑措施	0	0

5.1 结果验证

尽管由于混合威胁性质模糊以及被针对方不愿披露信息，对结果进行普遍验证仍然具有挑战性，但将我们的发现置于既有模型和先前研究的语境中，仍能提供有价值的见解。首先，我们的发现与其他模型的推论一致，即混合威胁能有效地在经济上耗尽防御方Balcaen等人（2022年），因为我们分析中最有效的反混合措施同时也是成本最高的。这凸显了优先配置稀缺资源以及战略性地构建跨领域威慑措施的必要性。我们的多主体模型的发现进一步印证了先前研究，这些研究强调防御策略需要在反混合措施的有效性与资源约束之间取得平衡Attiah等人（2018），并突出了在不同对抗策略下选择成本效益较高的方案的重要性。

此外，我们的分析支持关于民主威慑的现有研究，这些研究主张由国家协调、采取全社会方针 Tre- verton 等人 (2018)。具体而言，我们的研究结果凸显了市场限制和情报共享等措施的有效性，这些措施符合运用非军事手段、软实力和非对称战略来应对混合威胁的做法 Wigell(2021)。最后，我们的敏感性分析进一步印证了现有网络空间威慑建模研究的发现 Kostyuk(2021)，并强调此类威胁的有效性在很大程度上取决于对手对反制措施的易受影响程度。

6 讨论

本文提出了评估跨领域反混合措施有效性的新方法：在对这些措施的威慑效果和防御效果存在不确定性的背景下，权衡每项措施的成本、威慑能力和损害缓解能力，因此采用概率方法或博弈论方式对其进行建模。这些方法的实验涵盖了 1,000 种不同情景，涉及恶意网络行动，并纳入了多种反混合措施；这些措施源自广泛的文献综述，并辅以迷你德尔菲调查。

尽管博弈论方法在混合领域（Balcaen 等人，2022年）和网络领域（Attiah 等人，2018）日益受到重视，但单靠博弈论无法捕捉混合威胁动态中固有的不确定性所带来的细微复杂之处。另一方面，概率图模型在混合威胁领域中可能并不适用（Balaban 和 Mielniczek，2018），因为它们要求事先进行设定，且无法考虑战略互动。我们的研究首次将博弈论方法与概率图方法结合起来，将混合领域特有的不确定性编码为概率关系，并将不同的行为体建模为智能体。建模结果凸显了反混合措施的不同特征如何影响其在各种仿真中的有效性，并展示了由理性、最优决策通过博弈均衡所形成的博弈结果。

因此，政策影响有两方面。首先，我们的分析表明，在深度不确定的条件下，借鉴相关文献中的见解并通过专家引 elicitation，可以系统地估算惩罚、拒止和韧性等不同反混合措施的效果。多智能体方法的结果甚至揭示了面对采取战略行动的对手时，最优措施会如何变化。我们的模型为这一过程提供了原型，可加以仿效、扩展和完善，用于测试和设计反混合威胁政策，从而帮助政策制定者制定并确定反混合威胁政策的优先次序。

其次，基于引 elicitation 数据开展的大量仿真凸显了有利条件的重要性。这些条件会调节措施的效果，这意味着有效的反混合威胁战略需要以这些有利条件为目标。例如，惩罚性威慑措施取决于对手是否愿意接受，因为只有当对手对这类措施作出响应时，这些反混合威胁战略才会奏效。（Mazarr 等，2018）。建模工作表明，即使只是稍微加深对侵略者可能接受反混合措施的程度的了解，也能显著提升对措施有效性的评估。这意味着，用于预测对手对可能采取的反混合措施作何反应的资源投入，取决于反混合措施的效果。由于对手的响应也取决于沟通，因此，结果强调，实施反混合措施时，细致周密的沟通战略至关重要（Sweijs 和 Zilincik，2021）。

我们提出两个未来研究领域。首先，近期开发的方法侧重于条件概率表的引出（Hassall et al,2019; Barons et al, 2022年）。实施这些方法并让政策制定者加以检验，将提高模型输入数据的精确度。其次，应针对涉及超越网络领域的跨领域性质混合威胁的情景开展实验。这会注入额外的复杂性，但将更贴近地反映混合威胁的当代性质。

更根本地，我们的建模工作旨在发展知识库，包括可应用于以不确定性为特征的现实世界安全问题的方法、数据和技术。通过表明此类建模工作能够产生具有现实世界政策价值的洞见，我们希望这项工作能获得更广泛的追随。

资金支持

本研究由荷兰外交部和荷兰国防部通过 PROGRESS 研究框架协议提供经费支持。研究内容及所表达的观点完全由作者负责，不代表荷兰政府的认可，也不应被视为荷兰政府的认可。

披露声明

作者未报告任何潜在利益冲突。

参考文献

Andrew A, Spillard S, Collyer J 等（2022年）通过网络安全仿真开发最优因果网络防御智能体。网址 https://arxiv.org/abs/2207.12355，2207.12355

Arce D（2023）国防经济学家的网络安全。《国防与和平经济学》34(6):705–725

Attiah A、Chatterjee M、Zou CC（2018）一种用于建模网络攻击与防御策略的博弈论方法。载于：2018 IEEE 国际通信会议，IEEE，第 1–7 页

Balaban M、Mielniczek P（2018）混合冲突建模。载于：2018 年冬季仿真会议，IEEE，第 3709–3720 页

Balcaen P、Bois CD、Buts C（2022）对混合威胁的博弈论分析。《国防与和平经济学》33(1):26–41

Barons MJ、Mascaro S、Hanea AM（2022）在量化离散贝叶斯网络时，平衡专家征询负担与专家输入的丰富程度。《风险分析》42(6):1196–1234

Berejikian JD（2002）威慑认知理论。《和平研究杂志》39(2):165–183

Brantly AF（2018）网络威慑问题。载于：2018 年第十届国际网络冲突会议 CyCon，页码 31–54，https://doi.org/10.23919/CYCON.2018.8405009

DavisPK，O'Mahony A，Curriden C 等（2021）影响对手国家：平息完美风暴。技术报告，兰德公司，圣莫尼卡

Feledi D、Fenz S、Lechner L（2013）迈向基于网络的信息安全知识共享。《信息安全技术报告》17(4)：199–209。https://doi.org/https://doi.org/10.1016/j.istr.2013.03.004，网址 https://www.sciencedirect.com/science/article/pii/S1363412713000204，专题：ARES 2012 第七届国际可用性、可靠性与安全性会议

Freedman LD（2004）《威慑》。剑桥：Polity出版社

Gartzke E、Lindsay JR（2019）《跨域威慑：复杂时代的策略》。牛津大学出版社，https://doi.org/10.1093/oso/9780190908645.001.0001，网址 https://doi.org/10.1093/oso/9780190908645.001.0001

George AL、Smoke R（1974）《美国外交政策中的威慑：理论与实践》。哥伦比亚大学出版社

Gilad A, Tishler A (2023) 减轻高级网络攻击风险：网络武器的质量、隐蔽性与使用强度的作用。《国防与和平经济学》在线优先发表（在线优先发表）：1–21

Greenberg A (2019)《沙虫：网络战新时代，以及追捕克里姆林宫最危险黑客的行动》。Anchor

Hammond L, Fox J, Everitt T, 等 (2021) 多主体影响图的均衡精炼：理论与实践。arXiv 预印本 arXiv:210205008

Hassall KL, Dailey G, Zawadzka J, 等 (2019) 促进贝叶斯信念建模所用信念的 elicitation。环境建模与软件 122:104,539。https://doi.org/https://doi.org/10.1016/j.envsoft.2019.104539，网址 https://www.sciencedirect.com/science/article/pii/S136481521930221X

Howard RA, Matheson JE（2005）影响图。《决策分析》2(3):127–143

Johansson F, Falkman G（2008）一种用于威胁评估的贝叶斯网络方法，并应用于空中防御场景。载于：2008年第11届国际信息融合会议，IEEE，第1–7页

Kilcullen D（2019）非常规战争的演变。《斯堪的纳维亚军事研究期刊》2(1)

Koller D、Friedman N（2009年）《概率图模型：原则与技术》。MIT出版社

Koller D, Milch B (2003) 用多智能体影响图表示和求解博弈。《博弈与经济行为》45(1):181–221. https://doi.org/https://doi.org/10.1016/S0899-8256(02)00544-4，网址 https://www.sciencedirect.com/science/article/pii/S0899825602005444，博弈论学会首届世界大会

Kostyuk N (2021) 网络空间威慑：公共与私人网络能力。《国际研究季刊》65(4):1151–1162

Krolikowski A, Hall TH (2023) 华为 5G 困境中的非决策性决策：日本、英国和德国的政策。《日本政治科学期刊》页码 1–19

Lis P, Mendel J (2019) 对关键基础设施的网络攻击：经济学视角 1。《经济与商业评论》5(2):24–47

Lupovici A（2023）《通过施加成本实现威慑：介于惩罚性威慑与拒止威慑之间》。《国际研究评论》25(3)：viad036。https://doi.org/10.1093/isr/viad036，网址https://doi.org/10.1093/isr/viad036，https://arxiv.org/abs/https://academic.oup.com/isr/article-pdf/25/3/viad036/50957926/viad036.pdf

Mallory K（2018）跨域威慑的新挑战。技术报告，JSTOR

Mazarr MJ（2021）《理解威慑》，T.M.C. Asser出版社，海牙，第13–28页。https://doi.org/10.1007/978-94-6265-419-82，网址 https://doi.org/10.1007/978-94-6265-419-8_2

MazarrMJ、ChanA、Demus A等（2018）什么能够威慑以及为何：探索有效威慑国家间侵略的要求。 https://doi.org/10.7249/RR2451

McKenzie TM（2017年）网络威慑是否可能

Mikkola P、Martin OA、Chandramouli S等（2023）先验知识提取：过去、现在与未来。《贝叶斯分析》第1–33页。 https://doi.org/10.1214/23-BA1381，URL https://doi.org/10.1214/23-BA1381

Morato D，Berrueta E，Maga˜na E，等（2018）通过分析文件共享流量实现勒索软件早期检测。《网络与计算机应用杂志》124：14–32。https://doi.org/https://doi.org/10.1016/j.jnca.2018.09.013，网址 https://www.sciencedirect.com/science/article/pii/S108480451830300X

Morris LJ，Mazarr MJ，Hornung JW，等（2019）在灰色地带获得竞争优势：针对低于大规模战争门槛的胁迫性侵略的应对选项。技术报告，兰德国家国防研究所，美国圣莫尼卡

Nash Jr JF（1950）n人博弈中的均衡点。《美国国家科学院院刊》36（1）：48–49

Norgaard R，Killeen T（1980）期望效用与截断正态分布。《管理科学》26（9）：901–909

Nye Jr JS（2016）网络空间中的威慑与劝阻。《国际安全》41（3）：44–71

Nyemann DB，Sørensen H（2018）超越韧性：应对混合威胁的复兴方法。混合威胁卓越中心战略分析第13期

Paul C（2016）一名混合战争怀疑者的自白。《小战争期刊》

Pearl J（2009）《因果关系》。剑桥大学出版社

Roth F、Stirparo P、Bizeul D 等（2019）高级持续性威胁组织及其行动。《高级持续性威胁组织及其行动》

Schneider J（2019）网络空间内外的威慑。《跨域威慑：复杂时代的策略》，牛津大学出版社，牛津，第95–120页

Selten R、Bielefeld RS（1988）重新审视广义式博弈均衡点的完美性概念。施普林格

Spiegelhalter DJ、Lauritzen SL（1990）有向图结构上条件概率的序贯更新。《网络》20(5)：579–605

Stein Nv、Raponi E、Sadeghi Z 等（2022）可解释人工智能全局敏感性分析方法的比较及其在基因组预测中的应用。《IEEE开放获取》10：103,364–103,381。https://doi.org/10.1109/ACCESS.2022.3210175

Sullivan JE、Kamensky D（2017）乌克兰的网络攻击如何揭示美国电网的脆弱性。《电力期刊》30(3)：30–35。https://doi.org/https://doi.org/10.1016/j.tej.2017.02.006，网址 https://www.sciencedirect.com/science/article/pii/S1040619017300507

Sweijs T, Zilincik S（2019）《跨域威慑与混合冲突》。海牙战略研究中心

Sweijs T, Zilincik S（2021）《跨域威慑的本质》。NL ARMS《荷兰年度军事研究评论》2020：《21世纪的威慑——理论与实践的见解》，第129–158页。Tiirmaa-Klaar H（2016）《构建国家网络韧性并保护关键信息基础设施》。《网络政策杂志》1(1)：94–106。Tor U（2017）《“累积威慑”作为网络威慑的新范式》。《战略研究杂志》40(1-2)：92–117。Treverton GF, Thvedt A, Chen AR 等（2018）《应对混合威胁》。Valeriano B, Maness RC（2022）《对立对手之间网络冲突的动态，2001–11（数据集于2022年更新）》。《和平研究杂志》51(3)：347–360

Wagner TD、Mahbub K、Palomar E 等（2019）网络威胁情报共享：调查与研究方向。《计算机与安全》87:101,589。https://doi.org/https://doi.org/10.1016/j.cose.2019.101589, URLhttps://www.sciencedirect.com/science/article/pii/S016740481830467X

Wang Y、Sun Y、Li JY 等（2012），基于动态贝叶斯网络的防空威胁评估。载于：2012年国际系统与信息学会议（ICSAI2012），IEEE，第721–724页。Wigell M（2021），民主威慑：如何阻止混合干预。《华盛顿季刊》44(1)：49–67。Wilner A、Wenger A、Morgan PM 等（2021），拒止威慑：理论与实践。Cambria出版社。Yan G、Lee R、Kent A 等（2012），构建贝叶斯网络博弈框架，以评估分布式拒绝服务攻击与防御。载于：2012年ACM计算机与通信安全会议论文集，第553–566页。Zhu J、Peng G、Wang D（2020），结合条件变分自编码器与贝叶斯网络的基于双域的对抗防御。《IEEE工业信息学汇刊》17(1)：596–605

7 附录

8 网络威胁情景

在当前政治和经济关系面临挑战、危机加剧的时期，B国担心，修正主义国家（如A国）可能试图结合混合威胁与攻击来破坏稳定并影响B国。尤其是，鉴于A国在信息与网络空间领域已确立的实力，B国担心后者会对其关键基础设施（包括电网和港口）发动大规模网络攻击。

B国情报机构报告称，B国很可能成为针对关键基础设施的网络攻击的下一个目标，并将这一情况标记为可信威胁。

尽管一些与B国类似的国家已成为A国恶意网络行动的目标，但其中一些国家成功地通过跨领域方法威慑或减轻了这些攻击的影响：要么设定红线并明确传达反制措施（惩罚性威慑），要么增强韧性并提高攻击成本，以劝阻对手（拒止威慑）。

尽管目标国家采取了威慑行动，A国仍决定对其中一些国家实施进攻性网络行动。这些网络行动对各目标国家可能造成的影响差异很大，这也取决于目标国家的检测能力，以及其预防与恢复能力。如果检测能力一般，未能在事前识别威胁，对手就会更深入地利用漏洞，最终造成重大损害。

对手 A 根据潜在的收益和损失，决定实施网络行动。显然，A 会考虑到，在未被妥善检测到时，混合行动更容易成功。因此，针对设想中的网络攻击，A 会仔细考察防御方 B 的检测能力和归因能力。反过来，防御方 B 在增强自身韧性的同时，也会努力提升检测能力和归因能力。

9 实验设计数据

本节通过说明这些实验所依据的概率分布，详细介绍实验部分中的实验。

我们分别从截断正态分布 $f(x, \mu_1, \sigma_1, a, b)$ 和 $f(x, \mu_2, \sigma_2, a, b)$ 中抽取破坏性影响最大的值 θ_1 和显著影响值 θ_2，其中 $a = 0$，$\mu_1, \mu_2 = 1000, 100$ 和 $\sigma_1, \sigma_2 = 300, 50$ 分别取自 Norgaard 和 Killeen（1980）。因此，x 取自区间 $[0, \infty]$。

可忽略的破坏性影响 θ_3 取自基于正态分布 $\mathcal{N}(0,5)$ 的正半正态分布。所有数值均以百万计，表示破坏性成本。

这些措施的成本 γ_i 均从截断正态分布中抽取。 d_i 对手在采取反混合措施 q_{ij} 后实施混合行动 d_i 的概率取自贝塔分布，因为贝努利分布的共轭先验是贝塔分布（假设对手要么发起攻击，要么不发起攻击）。最后，基于反混合措施 w_{ijh} 和攻击d_i ，混合行为可能造成的影响 c_j 的概率取自狄利克雷分布，因为狄利克雷分布是分类分布的共轭先验，并且常用于影响图（Spiegelhalter 和 Lauritzen，1990）。

9.1 主动情报共享

共享情报时，不会直接将其传达给对手，因此对手在采取此措施后实施混合行动的概率 q_{11} 为 $Be(5,5)$。然而，这项措施显著提升了缓解混合行为破坏性影响的能力，因此相关概率取自狄利克雷分布：

$$\begin{pmatrix} w_{111} \\ w_{112} \\ w_{113} \end{pmatrix} \sim \text{Dir} \begin{pmatrix} 4 \\ 8 \\ 12 \end{pmatrix}.$$

这项措施的成本反映了机密信息泄露给无权获取者所造成的损失，其取自截断正态分布$f(x,\ \mu,\ \sigma,\ a,\ b)$，其中 $\mu = 150,\ \sigma = 50$和 $a = 0$。

如果对手不采取混合行动，破坏性影响可忽略不计的概率始终为 1。

9.2 在更广泛的社会层面提升网络韧性

提升网络韧性通过拒止威慑发挥作用，使该措施阻止对手实施混合行为的概率q_{21} 取自 $Be(4,\ 8)$，而潜在影响的缓解程度则取自以下狄利克雷分布：

$$\begin{pmatrix} w_{211} \\ w_{212} \\ w_{213} \end{pmatrix} \sim \text{Dir} \begin{pmatrix} 3 \\ 5 \\ 8 \end{pmatrix}.$$

这项措施的成本主要由私营部门承担，并服从截断正态分布 $f(x,\mu,\sigma,a,b)$，其中 $\mu = 300, \sigma = 50$和$a = 0$。

9.3 进攻性网络行动

进攻性网络行动包含拒止威慑和惩罚性威慑两种组成部分。为补偿这项措施也可能适得其反这一事实，我们从 $Be(1,1.2)$ 中抽取该措施成功阻止对手实施混合行为的概率 q_{31} ，并从以下狄利克雷分布中抽取损害缓解潜力：

$$\begin{pmatrix} w_{311} \\ w_{312} \\ w_{313} \end{pmatrix} \sim \text{Dir} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

这项措施的成本包括组建进攻性网络部队的成本，以及相关攻击所产生的成本。这些成本服从截断正态分布 $f(x, \mu, \sigma, a, b)$，其中 $\mu = 250, \sigma = 30$ 和 $a = 0$。

9.4 市场限制

实施市场限制通过拒止威慑发挥作用，因此，这项反混合措施阻止对手实施混合行为的概率 q_{41} 服从 $Be(2,8)$。潜在影响的概率服从狄利克雷分布

$$\begin{pmatrix} w_{411} \\ w_{412} \\ w_{413} \end{pmatrix} \sim \text{Dir} \begin{pmatrix} 2 \\ 2 \\ 15 \end{pmatrix}.$$

最后，这项措施的成本包括将某些私营组织排除在市场之外的成本，服从截断正态分布$f(x, \mu, \sigma, a, b)$，其中 $\mu = 400, \sigma = 50$ 和 $a = 0$。

9.5 通过战略传播开展公开威慑宣传

假设该信息涉及某种惩罚性威慑，且威胁过程中存在不确定性，我们将该威慑措施成功阻止对手实施混合行为的概率 q_{51} 从 $Be(0.4,2)$中抽取。这里不涉及损害缓解，因此，破坏性影响的概率与未采取任何措施时一样，均从同一狄利克雷分布中抽取：

$$\begin{pmatrix} w_{511} \\ w_{512} \\ w_{513} \end{pmatrix} \sim \text{Dir} \begin{pmatrix} 12 \\ 6 \\ 2 \end{pmatrix}.$$

最后，该措施的成本还包括所采取的高风险升级所产生的成本，该成本取自截断正态分布 $f(x,\ \mu,\ \sigma,\ a,\ b)$，其中 $\mu = 500,\ \sigma = 250$ 和 $a = 0$。

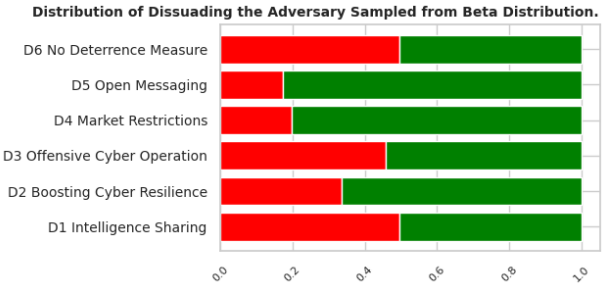

图5：各项反混合措施成功威慑对手的概率。绿色表示对手被成功劝阻的概率，红色表示对手仍实施网络行动的概率。

9.6 无威慑措施

假设未采取威慑措施 d_6，我们从 $Be(5,\ 5)$ 中抽取对手实施混合行动的概率 q_{61}。同样，如果防御方未采取威慑行动，而对手实施了混合行动，则从狄利克雷分布中抽取三种影响各自的概率：

$$\begin{pmatrix} w_{611} \\ w_{612} \\ w_{613} \end{pmatrix} \sim \mathrm{Dir} \begin{pmatrix} 12 \\ 6 \\ 2 \end{pmatrix}.$$

若对手不实施混合行动，则破坏性影响可忽略不计的概率始终为1。

图5 和6 分别展示了成功威慑的分布以及恶意网络攻击潜在影响的分布。

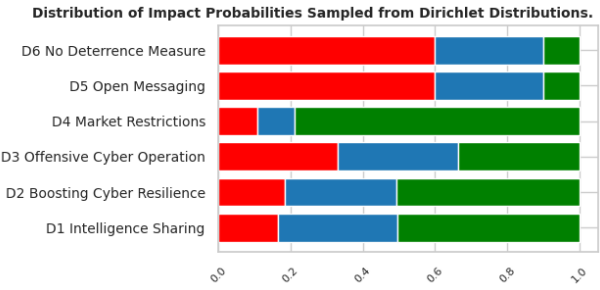

图 6：各项反混合措施成功减轻混合行动影响的概率。红色表示混合行动造成严重影响的概率，蓝色表示造成中等影响的概率，绿色表示造成轻微影响的概率。