# 攻防制衡与国际网络冲突*

张　耀　许开轶

**【内容提要】** 网络空间攻防理论的传统解释认为，网络空间的攻防关系是进攻占优，即在国际网络空间中，网络攻击的发起国占据先发优势，目标国则处于被动地位，不利于网络空间的战略稳定。然而，这一理论既无法解释国家间没有爆发大规模网络战争的原因，也难以解释为何网络攻击的目标国往往不选择妥协和让步，以及网络攻击产生的实际效用也仍然有限的现象。对此，网络攻防平衡理论虽然给予了修正，但仍没有摆脱进攻占优的陷阱。本文将在网络冲突的攻防结果维度对主体间的攻防关系进行重构，提出网络冲突的攻防关系模型，认为国际网络冲突主体间存在“攻防制衡”的关系模式。其次，本文通过双边网络事件和争端数据库(DCID，Version 1.1)，对“攻防制衡”模式的生成及其影响因素进行数据检验，得出初步结论。另外，本研究将进一步根据伊朗对沙特的“沙蒙”行动与美国对伊朗的“震网”行动两个案例，对以上假设分别进行证实与证伪。结果表明，网络冲突的严重程度对“攻防制衡”模式的生成具有显著的负向影响；然而，正是“攻防制衡”模式所具有的低烈度特征，才使其成为当前国际网络冲突攻防互动的常态。因此，准确把握国际网络冲突攻防互动的态势，对于网络空间安全与国际体系的战略稳定具有重要意义。

**【关键词】** 网络安全　网络冲突　网络攻防理论　攻防制衡　战略稳定

---

* 本文系国家社会科学基金重点项目“政治安全视阈下的网络边疆治理研究”(项目编号：16AZZ008)以及2018年江苏省研究生科研创新计划“网络强国战略视阈下的制网权问题研究”(项目编号：KYCX18_1143)的阶段性研究成果。作者感谢漆海霞、刘杨钺、姜鹏、杜鑫、孟维瞻等老师对于本文的指导或启发，感谢《国际政治科学》编辑部及匿名审稿专家对本文提出的极具建设性的修改意见，文责自负。

---

《国际政治科学》2019年第4卷第3期(总第15期)，第90—124页。
*Quarterly Journal of International Politics*

【作者简介】 张耀，南京师范大学江苏高校东亚国际问题研究中心硕士研究生。
电子邮箱：alfie1994@163.com
许开轶，南京师范大学公共管理学院副院长、教授、博士生导师，江苏高校东亚国际问题研究中心研究员。
电子邮箱：xukaiyi7384@163.com

## 一、问题的提出

网络安全议题的兴起源于频发的国家间网络冲突事件，其成为国际冲突的新形式，并对国际网络空间的战略稳定构成了挑战。例如，美国前国防部长利昂·帕内塔（Leon Panetta）表示："下一个'珍珠港事件'很可能是网络攻击……将对美国产生巨大影响。这是我们必须担心和防范的问题。"① 美国前总统网络安全问题特别顾问理查德·克拉克（Richard Clarke）则认为，网络冲突有可能改变世界军事平衡，从而在根本上改变政治和经济关系。② 2018年9月20日，美国特朗普政府发布的首份《美国国家网络战略报告》指出，美国将继续加强面对网络空间威胁的防御能力，并提出了"防御前置"（defend forward）的新理念。③ 防御前置将大概率导致一些先发制人的行为，但理论上防御前置也可包括一些主动防御活动，例如在目标网络系统中种植木马程序、安插网络暗门或"逻辑炸弹"等积极防御行为。无论是防御前置还是积极防御，都令网络空间安全滋生了不稳定因素。因此，国家间

① Karne Parrish, "Panetta Warns of Cyber Threat Growing Quickly," *Department of Defense News*, February 6, 2013, http://archive.defense.gov/news/newsarticle.aspx?id=119214, 访问时间：2018年12月4日。

② Richard A. Clarke and Robert K. Knake, *Cyber War: The Next Threat to National Security and What to Do About It* (New York: HarperCollins Publishers, 2012), p. 32.

③ The White House, "National Cyber Strategy of the United States of America," September 20, 2018, https://www.whitehouse.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf, 访问时间：2018年12月4日。

在网络冲突中的攻防态势为研究网络空间安全与大国间的战略稳定提供了一种可能的视角。

在既有的国内外研究中,通过探究网络进攻与防御之间谁更占据有利地位来探讨网络冲突频发的观点较为常见。纵观国际网络冲突动态,我们可以发现,相比于政治矛盾与贸易冲突,网络冲突的发起国在成功完成网络入侵之后,目标国既没有轻易选择妥协和让步,也没有做出过激的报复性行为,而是对网络入侵事件淡化处理。源自经典的进攻—防御理论的网络攻防理论成为解释上述现象背后生成逻辑的理论基础。

然而,无论是传统的进攻占优论(Offense Dominance Theory)还是新兴的攻防平衡论(Offense-defense Balance Theory),均未能有效解释网络冲突的目标国面临网络攻击之后的政策选择问题。一方面,网络空间的攻防平衡理论是对网络进攻占优的有力改进,并有效解释了网络冲突维持低烈度态势的过程机理,为国际网络空间的战略稳定提供了理论依据;另一方面,网络攻防平衡存在矫枉过正的问题,即我们无法回避网络空间仍存在攻击的先发优势,网络攻击成功的案例大量存在,并给攻击目标造成了切实的损失。这就自然引出了一系列问题:为什么网络冲突的发起国在成功完成网络攻击之后却难以从根本上实现其战略意图?如何规避网络空间的进攻优势对国际体系战略稳定的负面影响?这些问题成为网络攻防理论的若干短板,也是本文探讨的核心。基于此,本研究将从国际网络冲突处理结果的维度对攻防关系进行重构,提出新的网络攻防关系模型作为分析框架,并通过定量与案例分析相结合的方法进行实证检验,得出相应结论。

## 二、网络空间的进攻—防御理论及其批判

本文首先将对网络攻防理论相关文献进行回顾和梳理。进攻占优论作为网络攻防理论的重要范式之一,存在若干不足。网络攻防平衡理论则进行了部分修正,成为国内外研究的新趋势。本文则试图从已有的理论出发,探究其解释现实问题的局限性,为本研究提出的新框架奠定理论基础。

## (一) 网络空间的进攻—防御理论及其基本要义

网络空间的攻防理论可以追溯到传统国际体系中的进攻—防御理论。[①] 以肯尼斯·华尔兹(Kenneth Waltz)为代表的新现实主义认为,在国际无政府状态下,国家追求的终极目标不是权力,而是安全。[②] 由于缺少"利维坦"式的权力来源,国际无政府状态下所产生的安全困境[③],使国家为追求安全所采取的行动自然会被他国视为威胁。即使国家没有表现出恶意,但国家之间的恐惧与不安全感仍然是普遍存在的。攻防理论正是在国际无政府状态下的安全困境中孕育出来的。其前提假设是:国家所经历的安全困境的强度可能会有所不同,这取决于军事技术的发展状况和某些情境因素。当军事进攻占上风时,合作就很困难,战争就更有可能发生;当防御占上风时,战争就容易避免,合作便更容易实现。[④] 罗伯特·杰维斯(Robert Jervis)认为,当进攻比防守有优势时,进攻是维护国家利益的最佳途径。任何一个国家,如果不设法扩大其规模和影响力,都将难以维护其既有的国家利益。[⑤]

---

① 一些学者曾对传统的进攻—防御理论进行了详细综述与阐释,参见:Tang Shiping, "Offense-defense Theory: Towards a Definitive Understanding," *Chinese Journal of International Politics*, Vol. 3, No. 2, 2010, pp. 213-260; 王伟光:《攻防平衡理论及其批判》,《国际政治科学》2012 年第 3 期,第 84—120 页。

② Kenneth N. Waltz, "The Origins of War in Neorealist Theory," *The Journal of Interdisciplinary History*, Vol. 18, No. 4, 1988, pp. 615-628.

③ John Herz, "Idealist Internationalism and the Security Dilemma," *World Politics*, Vol. 2, No. 2, 1950, pp. 157-180; Robert Jervis, "Cooperation under the Security Dilemma," *World Politics*, Vol. 30, No. 2, 1978, pp. 167-214; Randall L. Schweller, "Neorealism's Status-Quo Bias: What Security Dilemma?" *Security Studies*, Vol. 5, No. 3, 1996, pp. 90-121; Charles L. Glaser, "The Security Dilemma Revisited," *World Politics*, Vol. 50, No. 1, 1997, pp. 171-201.

④ 也有一些学者认为,在防御占优的情况下,战争的风险会增加。参见:James D. Fearon, "The Offense-defense Balance and War since 1648," paper delivered to *the Annual Meetings of the International Studies Association*, Chicago, February 21-25, 1995, pp. 379-414; Peter Liberman, "The Offense-defense Balance, Interdependence, and War," *Security Studies*, Vol. 9, No. 1-2, 1999, pp. 59-91。

⑤ Robert Jervis, "Cooperation under the Security Dilemma," *World Politics*, Vol. 30, No. 2, 1978, pp. 167-214.

斯蒂芬·埃弗拉(Stephen Van Evera)则持相同的观点,认为当征服变得很容易时,战争发生的可能性要大得多,攻防平衡的变化对战争的风险有很大影响。[①] 攻防理论认为影响攻防平衡的两个变量是:(1)防御性武器和政策是否可以区别于进攻性武器和政策;(2)防御或进攻在准备和结果上是否占据优势。[②]因此,进攻优先的国家战略成为主权国家维护和争取国家利益的重要选择,而这会成为国家间发生冲突的催化剂。随着时间的推移,无论是进攻占优还是攻防平衡,都逐渐成为攻防理论的核心变量,并广泛用于分析国家行为体的国际互动行为中。

在无政府的国际体系中,国家将在威胁面前保持平衡。斯蒂芬·沃尔特(Stephen Walt)认为,对国家来说,威胁是综合实力、地理邻近性、进攻性、能力和侵略意图的结果。[③] 网络攻击的匿名性与归因难题更是加剧了国际网络空间的无政府特性与安全困境,给国家带来严重威胁。因此,攻防理论与网络安全领域的结合便存在了可能。基于网络空间技术特征所带来的未知恐惧与不安全感,国家往往会持续加强网络攻防能力建设,甚至不惜采取先发制人的策略以保自身安全。对于一些网络大国来说,国家网络战略的制定成为一种例行举措,"网络部队"发展为继陆军、海军、空军之后的又一新军种,成为维护国家网络空间安全与担负网络攻防对抗任务的重要力量。

网络空间的进攻占优论成为学界的重要理论解释。在网络冲突的攻防问题上,多数学者认为进攻占优,因此预计会有更多的攻击者和破坏者。网络攻击增加了发起国进攻并对目标造成损害的机会,同时降低了发起国承

---

① Stephen Van Evera, "Offense, Defense, and the Causes of War," *International Security*, Vol. 22, No. 4, 1998, pp. 5-43.

② Robert Jervis, "Cooperation under the Security Dilemma," *World Politics*, Vol. 30, No. 2, 1978, pp. 167-214.

③ Stephen M. Walt, *The Origins of Alliances* (Ithaca: Cornell University Press, 1987), pp. 17-33.

担的风险，因为发送特殊的程序代码比派遣特种部队更为容易。[①] 进攻占优的理论依据源于斯蒂芬·埃弗拉所认为的进攻容易导致战争的论点，现实依据则体现在网络空间的技术特性上。一方面，对于网络冲突的发起国来说，网络空间的技术特征极大地提高了进攻的效率和胜率。网络战的准入门槛非常低，即便是大国也很容易受到不间断的虚拟攻击。具体而言，一是网络空间超越了时空的限制，对于进攻方来说具有出其不意的优势；二是网络攻击的匿名性与隐蔽性使得网络攻击的发起国可以减少自身被暴露的风险，从而增加敌方归因的困难，进而减少被报复的概率；三是发动网络攻击的成本较低，因为网络武器主要是基于计算机程序代码。肯尼斯·纳普(Kenneth Knapp)和威廉·博尔顿(William Boulton)指出，大范围强大的网络武器已变得更加便宜和可用，攻击者只需花400美元就能制造出一枚电磁炸弹。[②] 网络攻击方式主要包括破坏、分布式拒绝服务、入侵与渗透。其中，

---

① 学界关于网络空间进攻占优论的有关论述，参见：William F. Lynn III, "Defending a New Domain-the Pentagon's Cyberstrategy," *Foreign Affairs*, Vol. 89, No. 5, 2010, p. 97; Lucas Kello, "The Meaning of the Cyber Revolution: Perils to Theory and Statecraft," *International Security*, Vol. 38, No. 2, 2013, pp. 7-40; Kenneth Lieberthal and Peter Warren Singer, *Cybersecurity and US-China Relations* (Washington, D. C: Brookings, 2012); Martin C. Libicki, *Cyberdeterrence and Cyberwar* (Santa Monica: Rand Corporation, 2009); Joseph Nye, *Cyber Power* (Cambridge, Massachusetts: Harvard Univ Cambridge Mabelfer Center for Science and International Affairs, 2010); Adam P. Liff, "Cyberwar: A New 'Absolute Weapon'? The Proliferation of Cyberwarfare Capabilities and Interstate War," *Journal of Strategic Studies*, Vol. 35, No. 3, 2012, pp. 401-428; Chris C. Demchak, *Wars of Disruption and Resilience: Cybered Conflict, Power, and National Security* (Athens: University of Georgia Press, 2011); Nazli Choucri, *Cyberpolitics in International Relations* (Cambridge, Massachusetts: MIT Press, 2012); Keir Lieber, "The Offense-defense Balance and Cyber Warfare," *Cyber Analogies*, 2014, pp. 96-107; Timothy J. Junio, "How Probable is Cyber War? Bringing IR Theory back in to the Cyber Conflict Debate," *Journal of Strategic Studies*, Vol. 36, No. 1, 2013, pp. 125-133; John B. Sheldon, "Deciphering Cyberpower: Strategic Purpose in Peace and War," *Strategic Studies Quarterly*, Vol. 5, No. 2, 2011, pp. 95-112。

② Kenneth Knapp and William Boulton, "Cyber-Warfare Threatens Corporations: Expansion into Commercial Environments," *Information Systems Management Journal*, Vol. 23, No. 2, 2006, p. 76.

网络渗透可以细分为逻辑炸弹、病毒、蠕虫、监听等方式。① 这些网络攻击的方式对于国家来说,成本微不足道,国家更无须承担人员伤亡的风险。即使网络攻击的成本再高,使用网络武器的门槛也会低于其他种类的武器。②

另一方面,对于网络冲突的目标国来说,网络空间的技术特征使得网络防御的成本相对昂贵,并使目标国处于被动地位。具体而言,一是实时归因问题难以解决,网络攻击可以在技术上对攻击源头进行伪装和分散,这会使网络冲突的目标国难以确定攻击来源。此外,一旦目标国人为归因失误,将会产生新的敌人,并引发新的冲突,归因的不确定性将使战争爆发的风险增大,从而对网络空间的战略稳定构成挑战。③ 二是国际网络冲突中目标国网络系统的脆弱性,会加大其遭受网络攻击后所承受的损失。现代网络系统的脆弱性加速了国际社会将不可避免地滑向网络战争。④事实上,网络应用较为发达的国家可能会成为网络攻击的重要目标,因为它们对普遍应用的网络系统依赖程度更高,网络攻击所带来的损失也可能较大。而网络技术并不发达的国家在遭受网络攻击时所产生的负面影响可能有限。而在国际网络冲突中,国家网络基础设施的脆弱性更加成为网络防御的弱点。

总之,网络攻防理论的传统解释以进攻占优论为特点,并将经典的攻防理论与网络空间的技术特征有机结合了起来,使国际政治中网络空间安全理论的完善与发展迈出了重要一步。

---

① 关于网络攻击方式的介绍与分类,参见:沈逸、江天骄:《网络空间的攻防平衡与网络威慑的构建》,《世界经济与政治》2018 年第 2 期,第 49—70 页;Ryan C. Maness and Brandon Valeriano, "The Impact of Cyber Conflict on International Interactions," *Armed Forces & Society*, Vol. 42, No. 2, 2016, pp. 301-323; Stephen M. Walt, "Is the Cyber Threat Overblown?" *Foreign Policy*, March 30, 2010, https://foreignpolicy.com/2010/03/30/is-the-cyber-threat-overblown/, 访问时间:2018 年 12 月 13 日。

② Timothy J. Junio, "How Probable is Cyber War? Bringing IR Theory back into the Cyber Conflict Debate," *Journal of Strategic Studies*, Vol. 36, No. 1, 2013, p. 130.

③ 刘杨钺:《网络空间国际冲突与战略稳定性》,《外交评论》2016 年第 4 期,第 106—129 页;Timothy J. Junio, "How Probable is Cyber War? Bringing IR Theory back into the Cyber Conflict Debate," *Journal of Strategic Studies*, Vol. 36, No. 1, 2013, pp. 125-133。

④ Gary McGraw, "Cyber War is Inevitable (unless We Build Security in)," *Journal of Strategic Studies*, Vol. 36, No. 1, 2013, p. 109.

## （二）网络空间的进攻—防御理论的局限

进攻占优理论作为网络空间攻防理论的重要解释，存在以下两个问题。第一，进攻占优理论在理论层面上过于聚焦于网络冲突的攻防过程而忽视网络冲突后的攻防结果。网络空间的进攻与防御体现了网络冲突从网络攻击的发起到给目标国造成实质上的物理损害为止这一攻防过程。在这一过程中，网络冲突的发起国同样存在无法达到网络攻击效果的可能。虽然发起国成功完成了网络空间的攻击，但这并不意味着目标国必然会选择妥协和屈服。如果目标国选择积极防御甚至采取报复行为，那么网络冲突的攻防态势可能会像日本偷袭美国珍珠港一样出现反转，并进一步扭转网络冲突的攻防态势。退一步讲，如果网络冲突的目标国在修复网络漏洞或挽救损失之后，延续先前的政策选择，则意味着网络冲突发起国并没有在根本上实现其战略目标。例如，2007 年 4 月，爱沙尼亚因苏俄纪念铜像事件引发俄罗斯政府资助下的黑客组织的大规模 DDoS（分布式拒绝服务）攻击，导致爱沙尼亚多个政府网站全线瘫痪。然而，这使得爱沙尼亚加速倒向北约，成为北约对抗俄罗斯地缘政治压力的桥头堡，俄爱矛盾进一步恶化。

第二，进攻占优论在经验层面上已被相关学者证伪，且难以得到事实和数据的支持。正如詹姆斯·费伦（James D. Fearon）认为的，如果战争爆发，攻防理论还需要区分从争端中获得的优势和从积极行动中获得的优势。[①] 费伦指出，进攻占优意味着国家更倾向于进攻而不是防御，而不管力量的平衡如何，弱国在任何情况下都很少占上风。网络攻防平衡理论也对进攻占优的观点进行了修正：丽贝卡·斯莱顿（Rebecca Slayton）从网络攻防的成本—效益角度出发，认为在攻防双方开展网络行动所获得的实际效用相等时，攻防平衡是可以实现的。衡量这种效用的方法就是双方通过进攻—防御所获得的绝对效用减去攻防成本得出的净效用。[②] 更多的网络攻防平衡

---

① James D. Fearon, "The Offense-defense Balance and War since 1648," paper delivered to *the Annual Meetings of the International Studies Association*, Chicago, February 21-25, 1995, pp. 379-414.

② Rebecca Slayton, "What is the Cyber Offense-defense Balance? Conceptions, Causes, and Assessment," *International Security*, Vol. 41, No. 3, 2017, pp. 72-109.

理论视角从攻防过程出发,认为网络攻防态势存在一种平衡关系,网络攻击的效用被过分夸大,且防御的能力被人为低估,这有利于国际网络空间形成新的战略稳定。[①] 此外,发起者寻找机会并成功实施网络破坏行动的复杂程度正在上升,复杂系统的保护和防御设置越好,发起者所需资源越多、技术越复杂、设计越具体、需要的组织越多。如托马斯·里德(Thomas Rid)认为,只有很少的精密的战略行动者才有可能完成类似于"震网"(Stuxnet)那样顶级的计算机破坏行动。[②] 在归因问题的解决方面,攻防平衡理论认为,随着技术的发展,归因变得愈发容易,即入侵检测系统可以更好地实时识别入侵,更快地利用更多的数据。更具适应性的网络可能会提高进攻行动的成本,从而消除混乱,释放资源,更好地识别高调的入侵行为。[③] 事实上,纵观国际网络冲突动态,国家间并没有公开爆发大规模持续激烈的网络战争,而是维持着一种长期低烈度的态势,网络攻防行动大多是在暗中进行的。而且多数网络攻击的目标国并没有采取过度的回应,甚至不惜淡化处理网络入侵事件。

通过文献梳理发现,首先,网络空间的攻防平衡理论是对网络进攻占优理论的有力改进,并有效解释了网络冲突维持低烈度态势的过程机理。该理论为本文核心问题的解释提供了理论挖掘的基础,也为国际网络空间的战略稳定提供了技术逻辑。然而,聚焦于冲突过程的网络攻防能力并非衡量攻防平衡的唯一因素[④],因此,攻防平衡理论仍未完全规避进攻占优的陷阱。正如费伦所言,学界更倾向于将不断变化的攻防平衡与实力平衡混为

---

① Ilai Saltzman, "Cyber Posturing and the Offense-defense Balance," *Contemporary Security Policy*, Vol. 34, No. 1, 2013, pp. 40-63; Salma Shaheen, "Offense-defense Balance in Cyber Warfare," *Cyberspace and International Relations* (Heidelberg, Berlin: Springer, 2014), pp. 77-93; 沈逸、江天骄:《网络空间的攻防平衡与网络威慑的构建》,《世界经济与政治》2018 年第 2 期,第 49—70 页;左鲁亦:《国家安全视域下的网络安全——从攻守平衡的角度切入》,《华东政法大学学报》2018 年第 1 期,第 148—157 页。

② Thomas Rid, "Cyber War will not Take Place," *Journal of Strategic Studies*, Vol. 35, No. 1, 2012, p. 28.

③ Thomas Rid and Ben Buchanan, "Attributing Cyber Attacks," *Journal of Strategic Studies*, Vol. 38, No. 1-2, 2015, p. 29.

④ Erik Gartzke, "The Myth of Cyberwar: Bringing War in Cyberspace Back Down to Earth," *International Security*, Vol. 38, No. 2, 2013, p. 66.

一谈。[①] 其次，成本—效用模式作为一种理论解释存在网络攻防损益难以衡量的问题，因为网络攻击的成本大小具有相对性，且这种成本也会因为发起国的网络战略定位、经济实力以及国防投入能力的不同而展现出不同的承受能力。此外，网络冲突攻防双方的净效用也难以衡量。成本—效用平衡的理论模式以双方互动产生效用的差值来评判网络攻防是否平衡，而更多的攻防平衡论侧重于论述网络冲突中发起者发动攻击的难度在加大，且目标防御行为的能力在增强。其实这仅是对进攻绝对占优论的一种修正，提醒我们认清进攻优势与网络空间威胁被过分夸大。但是这并不能改变在实际的网络冲突过程中进攻者仍然占据一定的先发优势的事实，只是这种攻防失衡的程度在逐渐减轻，但仍未达到绝对的平衡态势。最后，当前进攻与防御的界限已变得非常模糊，现有的聚焦于网络冲突过程的攻防理论将难以区分孰为进攻、孰为防御。例如，特朗普政府提出的“防御前置”理念为美国的积极防御行为赋予了合法性。防御前置最后必然会导致一些先发制人的行为，难以避免地带上了进攻的色彩。

综上所述，现有的网络空间进攻占优理论与攻防平衡理论均难以对本文的核心问题作出有效的解释。本研究将在现有的网络攻防平衡理论的基础上，对国际网络冲突的攻防关系进行重构，提出一种新的理论模式来作深入阐释。

## 三、国际网络冲突“攻防制衡”(offense-defense counterbalance)模式的理论阐释

在建立“攻防制衡”的核心理论假设之前，首先需对国际网络冲突进行清晰的界定。其次，由于“攻防制衡”模式生成有其特有的内在逻辑，这种逻辑是建立在网络攻防关系模型的基础之上的，而“攻防制衡”理论模式的建立既需要分析其生成逻辑的自洽性，同时也需要提出相应的理论假设，以便

---

① James D. Fearon, “The Offense-defense Balance and War since 1648,” paper delivered to *the Annual Meetings of the International Studies Association*, Chicago, February 21-25, 1995, pp. 6-7.

检验理论模式的可行性。因此,本文继而提出符合本文研究问题的基本假定与理论假设。

### (一) 国际网络冲突的概念界定

本文的研究对象为国家间的网络冲突,即国际网络冲突。[①] 关于网络冲突的界定,学界一直存在争议,并出现网络冲突与网络战争混淆使用的情况。[②] 郎平对网络冲突进行了全面科学的分类,包括技术、社会公共政策以及经济和国家安全三个层面。[③] 里德认为网络战争不会发生,他引用克劳塞维茨(Carl Von Clausewitz)在《战争论》中对战争的界定标准,认为进攻性行为只有符合致命性、工具性、政治性等三个标准才构成战争行为,然而事实上并没有符合这三个标准的网络攻击。[④] 约翰·斯通(John Stone)对此进行了反驳,他认为网络战争是存在的,战争行为通过使用武力以产生暴力影响。这些暴力影响在本质上不一定是致命的,但仍可产生巨大的暴力影响,依然属于战争范畴。[⑤] 加里·麦格劳(Gary McGraw)持相近观点,认为控制关键基础设施的信息系统非常容易受到网络攻击,除非改善网络防御与所依赖系统中的系统性安全漏洞,否则网络战争是不可避免的。[⑥] 对此,亚当·里夫(Adam P. Liff)则持较为中立的立场,他拒绝把自己的观点与里德

---

① 本文将国家作为国际网络冲突重点关注的行为主体,虽然出现了各种各样的其他非国家行为体,包括个人、跨国公司和非政府组织,但在无政府状态的国际体系中,拥有主权的国家是最高等级的行为体,也是网络空间安全互动的主要参与者,这种国家的独特属性适用于网络空间。此外,为达到控制变量的目的,本文控制各个国家的内部因素这一变量,以排除对国际互动行为的影响差异。

② 刘杨钺:《国际政治中的网络安全:理论视角与观点争鸣》,《外交评论》2015 年第 5 期,第 122 页。

③ 郎平:《网络空间国际秩序的形成机制》,《国际政治科学》2018 年第 1 期,第 25—54 页。

④ Thomas Rid, "Cyber War will not Take Place," *Journal of Strategic Studies*, Vol. 35, No. 1, 2012, pp. 5-32.

⑤ John Stone, "Cyber War will Take Place!" *Journal of Strategic Studies*, Vol. 36, No. 1, 2013, pp. 101-108.

⑥ Gary McGraw, "Cyber War is Inevitable (unless We Build Security in)," *Journal of Strategic Studies*, Vol. 36, No. 1, 2013, pp. 109-119.

混为一谈，他认为网络攻击能力的扩散虽然可能会增加战争爆发的频率，但这种影响很可能是微弱的，而且会因情况而异，即它对战争可能性的边际影响是正面的，但却很小。①

由此可见，学界对于网络冲突的界定争论十分激烈，对于网络战争是否会发生也没有统一的定论，但这另一方面也说明网络冲突已然成为国际网络安全领域中的重要问题，值得深入探讨。为避免产生概念分歧，本文将研究对象界定为外延较广的国际网络冲突，并定义为国家行为体利用网络信息技术在网络空间对目标国家的网络系统进行攻击并试图影响或改变对方行为的国际互动形式。国际网络冲突体现为一种国家为实现本国利益在网络空间对目标国发起的施压过程。这种冲突形式在网络空间是普遍存在的，但网络冲突并不等同于网络战争，而是将其包含于内。就本研究而言，电磁脉冲（electromagnetic pulses）、雷达对抗（radar countermeasure）、光电对抗（optoelectrionic countermeasure）以及传统上被认定是电子战（electronic warfare）的其他攻防对抗，并不将其定义为国际网络冲突。②

## （二）“攻防制衡”的前提假定

核心概念界定清晰之后，本文需确定三条前提假定，分别概括为“理性国家行为体假定”“敌对国家入侵成功假定”和“攻防结果维度假定”。只有满足这三条假定，“攻防制衡”模式才具备生成的环境与条件。

**假定 1：** 理性国家行为体假定。参与网络冲突的行为主体必须为主权国家或有国家的支持和参与的其他行为体。由于国家是理性行为体，因此参与网络冲突的行为体默认偏向于作出符合本国利益最大化的政策选择。

**假定 2：** 敌对国家入侵成功假定。国际网络冲突中对目标国的攻击是

---

① Adam P. Liff, “The Proliferation of Cyberwarfare Capabilities and Interstate War, Redux: Liff Responds to Junio,” *Journal of Strategic Studies*, Vol. 36, No. 1, 2013, pp. 134-138.

② 为使国际网络冲突“攻防制衡”模式能在实证检验过程中获得数据分析的支撑，本研究将国际网络冲突区别于传统的电子对抗，即网络冲突事件需要出于恶意目的的操纵计算机程序代码。电子操纵则通过电子（如无线电波）、定向能损坏或破坏电路等方式进行干扰和破坏。相关数据分析将在下文详述。

发起国的主观意图,且发起国成功完成了网络入侵。如果网络冲突的发起国不仅突破了目标国的“网络边疆”,同时对目标国的网络系统完成信息窃取、窜改和破坏等行为,即视为发起国在网络空间层面中攻击成功,反之则为不成功。如果网络冲突的目标国主动在实际反应中表现出妥协和让步,或者被迫中断、推迟现实行动,从而实现发起国的进攻意图,则视为目标国作出了妥协,反之则为不妥协。

**假定 3**:攻防结果维度假定。国家间网络冲突攻防关系的重构必须站在网络冲突攻防结果的维度上作出评判。这一结果维度独立于传统的攻防过程维度。过程维度重视网络攻防对抗过程来考察攻防互动的优劣态势;结果维度则跳出网络冲突攻防互动的过程,立足发起国网络攻击之后,目标国依据网络冲突的损害程度等因素而作出的政策选择这一时间节点。因此,这一假定既能说明网络攻防态势对冲突结果不产生决定性影响,也可较好地避免进攻与防御界限模糊的问题。

### (三)“攻防制衡”模式生成的内在逻辑

从均势理论来看,国家通常会在权力失衡的状态下采取制衡行为。均势关注的是实力和状态,制衡关注的是政策和行为。[①] 网络冲突的发起与目标国的回应在客观上也可以形成一种权力均衡的态势。既有的网络攻防平衡理论更像是一种对未来趋势的展望。

当前网络空间的攻防关系仍然体现出一种进攻占优的态势,网络攻防平衡的实现不能因网络攻击被有效遏制且防御能力得到增强而直接判断。通过对 2000 年至 2014 年发生的国际网络冲突数据统计结果[②]可以发现:在绝大多数的网络冲突中,防御一方都处于被动局面,即网络攻击成功的案例仍占绝大多数。这显然对现有的网络攻防平衡理论构成了挑战。事实上,

---

① 刘丰:《大国制衡行为:争论与发展》,《外交评论》2010 年第 1 期,第 113 页。

② 该双边网络事件和争端数据库(2000—2014,Version1.1)及编码信息可从以下地址获取:https://drryanmaness.wixsite.com/cyberconflcit/cyber-conflict-dataset,访问时间:2018 年 12 月 15 日;Ryan C. Maness and Brandon Valeriano, “The Impact of Cyber Conflict on International Interactions,” *Armed Forces & Society*, Vol. 42, No. 2, 2016, pp. 301-323.

更多的网络冲突均以发起国完成网络空间层面的攻击目标且目标国并没有作出发起国想要的让步行为而收场。这反映了当下的国际网络冲突在攻防结果维度上客观实现了相互制约与平衡的态势，也就是本文所说的“攻防制衡”模式。

为建立全面的网络冲突攻防关系理论模型，本文在假定 1 与假定 3 的条件下对国家间网络冲突的攻防关系进行重构。表 1 展示了网络冲突的发起国与目标国之间存在的四种可能的攻防关系模式。

**表 1　国际网络冲突的攻防关系模型**

| 发起国 \ 目标国 | 妥协 | 不妥协 |
|---|---|---|
| 成功 | 进攻制胜 | 攻防制衡 |
| 不成功 | 攻防错位 | 防御制胜 |

如表 1 所示，如果一国为网络冲突的发起国，必然以进攻成功为目的。在进攻成功的条件下，目标国将随之作出两种反应：妥协与不妥协。目标国作为理性行为体，一般不会在损害本国利益的条件下轻易选择妥协政策。因此，当且仅当目标国作出妥协的成本小于不妥协的成本时，才会做出妥协决定。如果目标国作出妥协的政策选择，那么网络冲突的发起国就成功实现了其攻击的战略意图，至少在形式上攻防双方形成“进攻制胜”的态势。在目标国不妥协的情况下，发起国虽然成功完成了网络空间的进攻，但却无法实现其战略意图，攻防双方在网络冲突结果维度上实现了客观的“攻防制衡”态势。另一方面，如果发起国没有成功完成网络攻击，这说明目标国成功防御了发起国的网络进攻。此时，目标国同样可以做出是否妥协的决定：如果目标国选择妥协，那么此时的攻防关系就表现为一种错位关系；如果目标国选择不妥协，那么攻防双方则实现了“防御制胜”的态势，显然这是一种经验性的判断和选择。

根据国际网络冲突的发起国与目标国之间的攻防关系模型，可以引申出四种可能的攻防关系模式，分别是“进攻制胜”“攻防制衡”“攻防错位”和“防御制胜”。然而，在理性国家行为体的前提假定下，“攻防错位”缺少实际意义，网络冲突的目标国作为理性行为体不可能在发起国发起网络攻击失

败后仍然作出让步和妥协。“防御制胜”的攻防关系模式不符合本文的敌对国入侵成功假定,防御成功的目标国自然不会主动选择让步或屈服。因此,无论是从符合前提假定还是从符合既有的研究意义出发,“攻防错位”和“防御制胜”的攻防关系模式均不作为本文探讨的核心。本文将对“进攻制胜”与“攻防制衡”两个攻防关系模式进行检验和讨论。此外,在满足假定 2 的条件下,“进攻制胜”与“攻防制衡”存在互斥关系,即国际网络冲突中的发起国进攻成功后,目标国无法同时作出两个相反的选择。如果国际网络冲突“攻防制衡”的关系模式假设检验成功,那么意味着国际网络冲突“进攻制胜”的关系模式被成功证伪。

因此,基于网络空间进攻占优及日益夸大的网络威胁,本研究将重点对国际网络冲突“攻防制衡”的关系模式及其影响因素进行实证检验,力求探索出当前国际网络冲突攻防互动的规律。

### (四)“攻防制衡”模式的基本假设

**假设 1**:当前国际网络冲突的攻防关系普遍表现为“攻防制衡”的关系模式,即发起国进攻成功,目标国则往往不选择妥协和让步。

**假设 2**:国际网络冲突的严重程度对事件结果具有负向影响,严重程度越低,“攻防制衡”模式生成的概率越高。

## 四、国际网络冲突“攻防制衡”模式的实证检验

本部分首先通过网络冲突领域数据库对“攻防制衡”关系模式的相关假设进行数据检验,并试图分析影响“攻防制衡”模式生成的因素;进一步根据相关案例分别进行证实与证伪,验证网络空间的“攻防制衡”理论模式的可靠性与生成的内在逻辑。

### (一)网络“攻防制衡”模式的数据检验

在数据检验部分,笔者通过统计描述的方式对“攻防制衡”模式的存在性与普遍性进行检验,同时立足国际网络冲突的演化过程,选取相关变量探

讨影响“攻防制衡”模式生成的相关因素，并重点探讨国际网络冲突的严重程度对“攻防制衡”模式的影响。

1. 数据来源

本文使用的数据来自美国海军研究生院国防分析部的瑞恩·曼尼斯(Ryan C. Maness)所参与整理的双边网络事件和争端数据库(DCID, Version 1.1)2000—2014年的数据。[①] 该数据库包含2000年至2014年的192个由国家行为体发起的双边网络冲突事件，编码方法与战争相关指数数据库(COW)进行了关联。网络攻击的目标必须是政府实体或作为目标国家的国家安全机构(例如电网、国防承包商和安全公司)的部分私人实体、重要媒体组织(第四产业)或重要公司。该数据库不包括多边网络冲突事件，因此，该数据库符合本文关于理性国家行为体和攻防结果维度两个假定。由于网络归因存在不确定性，这难免会降低坐实发起国攻击身份的准确性，从而会影响研究结果。但DCID数据库完美地避开了归因问题，即发起网络攻击的国家必须相当明确。如果归因指向存在严重疑问，那么该网络冲突事件将会被排除在样本之外。此外，该数据库的所有事实判断必须通过政府声明、政策报告、世界著名网络安全机构的白皮书等消息来源作为验证，从而进一步保证了信息的真实性和准确性。最后，该数据库还通过15位具有相关军事教育背景的专家协助编码关键变量来进行卡帕检验(Kappa Test)[②]，以验证编码的变量是否可靠，并得出了0.646的卡帕系数。所以，

---

① 该双边网络事件和争端数据库(2000—2014，Version1.1)及编码信息可从以下地址获取：https://drryanmaness.wixsite.com/cyberconflcit/cyber-conflict-dataset，访问时间：2018年12月15日；Ryan C. Maness and Brandon Valeriano，“The Impact of Cyber Conflict on International Interactions，” *Armed Forces & Society*，Vol. 42，No. 2，2016，pp. 301-323.

② 卡帕检验(Kappa Test)是一种统计方法，用于评估固定数量的评分者在对一些项目进行分类评分或对项目进行分类时达成一致意见的可靠性。kappa计算结果为－1～1，但通常kappa是落在0～1，可分为五组来表示不同级别的一致性：0.0～0.20极低的一致性(slight)、0.21～0.40一般的一致性(fair)、0.41～0.60中等的一致性(moderate)、0.61～0.80高度的一致性(substantial)和0.81～1几乎完全一致(almost perfect)。Anthony J. Viera and Joanne M. Garrett，“Understanding Interobserver Agreement：The Kappa Statistic，” *Fam Med*，Vol. 37，No. 5，2005，pp. 360-363.

该数据库无论是样本、变量选择还是编码的一致性都具有较强的可靠性,同时较好地契合了本文的研究设计,成为数据检验的基础。

2. 变量选取

DCID 数据库所统计的网络冲突数据及包含的相关变量为检验以上假设与全面呈现国际网络冲突演化过程提供了有力支持。

(1) 因变量的选取。对于网络冲突的严重程度是否会影响该模式的生成假设,本文将网络攻击是否成功(objective success)与目标国是否妥协(concession)两个变量进行分类组合。排除不符合前提假定的样本后,如果发起国在网络空间成功实现攻击目标且目标国拒绝作出妥协,则冲突攻防双方形成"攻防制衡"的关系模式(假设 1),反之,则为"进攻制胜"的关系模式。本文将因变量确定为是否符合"攻防制衡"模式,其中"是"赋值为 1,"否"赋值为 0。由于"攻防制衡"与"进攻制胜"存在非此即彼的互斥关系,因此当因变量取值为"0"时,表示"进攻制胜"的攻防关系模式。最后,本文确定数据库符合条件的样本量为 169 个。

(2) 自变量的选取。关于国际网络冲突中的概念操作化问题,已有若干国内外学者进行了有益尝试。首先,针对网络攻防能力的测量,有学者对不同国家的网络攻击能力、网络防御能力以及网络依赖能力进行了评估,并将得分的总和作为其网络对抗能力的总体指标。① 然而,网络空间存在其特有的组成方式,"网络空间的虚拟性意味着很难判断国家间的实力格局、使用实力的意志及对两者的感知"②。因此,难以寻找合适的变量与指标来对国家行为体的网络实力进行有效测量。斯莱顿则认为攻防关系态势并非由技术决定,而是效用。③ 网络攻击的效用是攻击目标的价值(例

---

① Richard A. Clarke and Robert K. Knake, *Cyber War: The Next Threat to National Security and What to Do About It* (New York: Harper Collins Publishers, 2012).

② 任琳:《网络空间战略互动与决策逻辑》,《世界经济与政治》2014 年第 11 期,第 76 页。

③ 这一概念源于经济学,在战争的博弈论和讨价还价理论中都得到了应用,参见:Rebecca Slayton, "What is the Cyber Offense-defense Balance? Conceptions, Causes, and Assessment," *International Security*, Vol. 41, No. 3, 2017, pp. 72-109.

如夺取地盘、窃取秘密或控制计算机）减去实现目标的最低成本；防御的效用是防御目标的价值（例如控制领土、保守机密、控制计算机）减去最低的防御成本。[①] 然而，正如前文所述，网络攻击的成本会因其所在国家的网络战略定位、经济实力以及国防投入能力的不同而不同。即使在发起网络攻击成本相同的情况下，网络冲突给双方带来的相对效益的衡量仍缺少统一的标准。

网络空间的相互博弈使得信息的非对称程度得以加剧，造成干扰变量过多，影响理论模型的建构与检验的假设。因此，本文跳出还原主义的视角，运用分析性方法[②]理解网络冲突的演化过程与攻防关系。如图 1 所示，

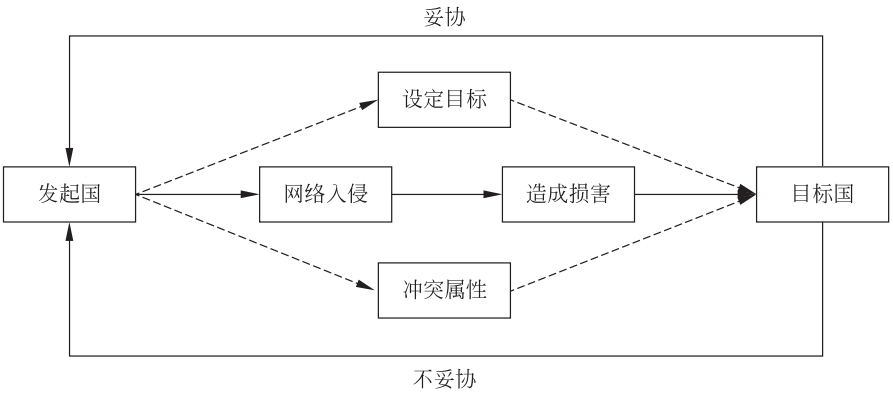


**图 1　国际网络冲突演化过程图**

① Rebecca Slayton, "What is the Cyber Offense-defense Balance? Conceptions, Causes, and Assessment," *International Security*, Vol. 41, No. 3, 2017, p. 80. 这种操作化方法曾被斯蒂芬·比德尔所批评，他认为攻防所需的最低成本是不可观察到的，而胜率是更好的指标之一。相关论述参见：Stephen Biddle, "Rebuilding the Foundations of Offense-defense Theory," *The Journal of Politics*, Vol. 63, No. 3, 2001, p. 749.

② 分析性方法是将整体还原为分离的个体，然后检验各部分的性质和彼此间的联系。对整体的理解是通过对处于相对简单状态的各要素的研究以及对其联系的观察得来的。它在一定程度上是有效的，尤其是当某些因素之间的联系可以分解为成对变量之间的关系，而其他因素则保持不变，而且可以假设未包括在变量中的干扰性因素的影响很小的情况下，该方法对本文探讨国际网络冲突的演化过程十分有帮助。参见：肯尼斯·华尔兹：《国际政治理论》，信强译，上海人民出版社，2017 年，第 41 页。

本文从国际互动形式的主体—过程—结果三个维度来呈现国际网络冲突的演化过程。冲突的主体包含发起国与目标国。在发起国对目标网络空间的入侵过程中,锁定攻击目标、使用的攻击方式以及对目标网络所造成的损害程度成为完成网络攻击的必要环节和因素。结果维度体现了目标国在发起国完成网络空间的攻击之后,需要作出的政策选择(是否妥协)。这一分析框架的优势是它从复杂的网络冲突互动中抽象出若干独立的变量,形成逻辑关系,以达到控制变量的目的。

曼尼斯等人在分析网络冲突对国际互动的影响时则通过将网络冲突事件的关键构成要素进行概念化操作,从而起到简化识别网络冲突复杂过程的作用。DCID 数据库成为学界通过系统的定量方法研究网络冲突的首次尝试。[①] 依照国际网络冲突演化过程的分析框架与 DCID 数据库的变量情况,设定目标通过"目标类型"(target type)来表示,网络入侵通过"攻击方式"(cyber method)与"是否 APT"(Advanced Persistent Threat,高级持续性威胁)来表示,造成损害通过"损害类型"(damage type)、"持续时间"(time FL)与"严重程度"(severity)来表示。此外,网络冲突属性的不同也会对目标国的政策选择产生可能的影响,该属性通过"冲突类型"(interaction type)来表示。其中,"严重程度"是核心自变量,"目标类型""攻击方式""是否 APT""损害类型""持续时间"为控制变量。控制变量也是解释变量,可能会对因变量的生成产生影响,但不作为本研究关注的重点。以上所有自变量的分类与取值情况见表 2。

3. 数据检验与讨论

在前提假定与变量选取的基础上,检验假设 1 与假设 2 是否成立成为下文的主要任务。通过对整理后样本量进行统计,发现符合"进攻制胜"模式的样本数量为 11 个,占样本总量的 6.5%;而符合"攻防制衡"模式的样本数量为 158 个,占样本总量的 93.5%。由此可见,"攻防制衡"关系模式数量占绝大多数,成为国际网络冲突攻防互动的常态,假设 1 被初步证实。

---

① Ryan C. Maness and Brandon Valeriano, "The Impact of Cyber Conflict on International Interactions," *Armed Forces & Society*, Vol. 42, No. 2, 2016, pp. 301-323

表 2 自变量设置、赋值与描述性统计

| 变量名称 | 变量设置 | 攻防制衡 | | |
|---|---|---|---|---|
| | | 赋值 | 频数(次) | 频率(%) |
| 目标类型 | 私人或非政府机构 | 1 | 34 | 21.52 |
| | 政府非军事机构 | 2 | 99 | 62.66 |
| | 政府军事机构 | 3 | 25 | 15.82 |
| 攻击方式 | 破坏 | 1 | 30 | 18.99 |
| | 拒绝服务(DDoS) | 2 | 25 | 15.82 |
| | 入侵 | 3 | 75 | 47.47 |
| | 渗透 | 4 | 28 | 17.72 |
| 是否 APT | 是 | 1 | 63 | 39.87 |
| | 否 | 0 | 95 | 60.13 |
| 损害类型 | 直接意图与即时攻击 | 1 | 91 | 57.59 |
| | 直接意图与延迟攻击 | 2 | 65 | 41.14 |
| | 间接意图与即时攻击 | 3 | 1 | 0.63 |
| | 间接意图与延迟攻击 | 4 | 1 | 0.63 |
| 持续时间 | 三天之内 | — | 53 | 33.54 |
| | 两周之内 | — | 32 | 20.26 |
| | 两周以上 | — | 73 | 46.2 |
| 严重程度 | 缺少动态网络的探测 | 1 | 4 | 2.53 |
| | 出现骚扰、扰乱治安等现象 | 2 | 73 | 46.20 |
| | 目标关键信息被窃取 | 3 | 47 | 29.75 |
| | 广泛的重要信息被窃取 | 4 | 32 | 20.25 |
| | 出现单个关键网络破坏的企图 | 5 | 0 | 0.00 |
| | 单个关键网络大面积破坏 | 6 | 2 | 1.27 |
| | 直接造成的最小人员伤亡 | 7 | 0 | 0.00 |
| | 对国民经济造成严重影响 | 8 | 0 | 0.00 |
| | 对国家基础设施造成严重破坏 | 9 | 0 | 0.00 |
| | 大量人员伤亡 | 10 | 0 | 0.00 |
| 冲突类型 | 妨害性行为 | 1 | 79 | 50.00 |
| | 防御性行为 | 2 | 0 | 0.00 |
| | 进攻性行为 | 3 | 79 | 50.00 |

注:(1)“持续时间”作为连续型变量,为便于呈现变量特征,我们在描述性统计中虽不做赋值处理,但划分了若干区间以呈现样本分布情况;(2)本文将“严重程度”也作为取值区间为 1～10 的连续型变量处理。

在检验严重程度对“攻防制衡”形成的影响之前,本文利用 Stata 14 统计分析软件对影响“攻防制衡”模式的自变量进行描述性统计,并形成自变量与因变量的交互表,从而更直观地了解变量赋值与在“攻防制衡”模式下的变量取值分布情况。

通过上述描述性统计,可以发现三条重要规律。首先,从损害类型来看,国际网络冲突“攻防制衡”模式中,网络入侵并给目标国造成损害是发起国的意图,但网络攻击对目标国产生的损害却不一定会即时产生效果。例如,“震网”蠕虫病毒的目的是通过代码操作慢慢影响和破坏伊朗纳坦兹工厂的离心机来阻挠核试验。第二,观察持续时间变量,可以发现 33.54%的网络冲突事件在 3 天之内结束,20.26%的网络冲突事件在 2 周之内结束,随着持续时间的延长,事件发生频次减少,呈现出瞬时性与逐时递减的特点。第三,从网络冲突结果的严重程度来看,网络攻击的破坏能力并没有我们想象的那么严重,而是维持在一种低烈度的态势,更缺少因为网络攻击而造成人员伤亡的数据支持。尤其是在“攻防制衡”模式下,网络冲突的严重程度占比最高的指标值仅为 2,但却占据样本总量的 46.20%。这就意味着数量相当的网络攻击给目标国带来的损害较为轻微。

那么,严重程度是否会影响“攻防制衡”模式的生成呢?为检验假设 2,本文根据变量特征依次建立 Probit、Logit 以及 Cloglog 模型,并进行多模型比较分析。因为“是否生成攻防制衡模式”是一个典型的二分变量,所以本文构建二元选择模型如式(1)所示

$$y_i^* = \beta_0 + \beta_1 \cdot severity_i + x_i'\beta + \mu_i, \quad i = 1,2,\cdots,N$$
$$y_i = \begin{cases} 1, & y_i^* > 0 \\ 0, & y_i^* \leqslant 0 \end{cases} \tag{1}$$

在式(1)中,下标 $i$ 表示冲突事件序号,$y$ 表示可观测的因变量,即国际网络冲突事件中“攻防制衡”模式是否生成:当该模式生成时,$y=1$;当该模式没有生成时,$y=0$。$y^*$ 表示与 $y$ 对应的不可观测的潜变量,并假设该潜变量可以被描述为一个由冲突事件特征 $x_i$(包括事件严重程度、持续时间等)定义的线性函数。其中,本文重点关注的冲突事件特征即严重程度记为 $severity$。据此,在冲突事件 $i$ 中生成“攻防制衡”模式的概率为

$$\begin{aligned} P(y_i = 1 \mid severity_i, x_i) &= P(y_i^* > 0 \mid severity_i, x_i) \\ &= P(\mu_i > -(\beta_0 + \beta_1 \cdot severity_i + x_i'\beta)) \\ &= P(\mu_i < \beta_0 + \beta_1 \cdot severity_i + x_i'\beta) \\ &= F(\beta_0 + \beta_1 \cdot severity_i + x_i'\beta) \end{aligned} \tag{2}$$

在式(2)中,误差项 $\mu_i$ 被假设服从对称分布,且其累积分布函数记为 $F(\cdot)$。于是,在冲突事件 $i$ 中没有生成“攻防制衡”模式的概率为

$$P(y_i = 0 \mid severity_i, x_i) = 1 - F(\beta_0 + \beta_1 \cdot severity_i + x_i'\beta) \tag{3}$$

根据式(2)和式(3)在全样本 $N$ 的范围内采用最大似然法进行估计,即可得到待估参数 $\beta$ 的估计量。分别假设 $\mu_i$ 服从不同的分布函数,则式(1)所示的二分选择模型可以被定义为 Probit、Logit 和 Cloglog 模型。本文首先基于 Probit 模型在不同的控制变量下对式(1)进行估计,结果如表 3 所示。

**表 3 “攻防制衡”模式生成的 Probit 模型估计结果**

| 变量 | 模型 1 | | 模型 2 | | 模型 3 | |
|---|---|---|---|---|---|---|
| | 系数 | 边际效应 | 系数 | 边际效应 | 系数 | 边际效应 |
| 严重程度 | −0.726*** | −0.071*** | −1.166*** | −0.069*** | −1.170*** | −0.069*** |
| | (−3.23) | (−3.21) | (−3.13) | (−3.13) | (−3.14) | (−3.19) |
| 攻击方式(DDoS) | | | −1.644 | −0.097* | −1.608 | −0.095* |
| | | | (−1.60) | (−1.81) | (−1.60) | (−1.85) |
| 直接延迟 | | | −0.571 | −0.028 | −0.341 | −0.018 |
| | | | (−0.81) | (−0.93) | (−0.53) | (−0.56) |
| 间接即时 | | | −3.098*** | −0.516*** | −3.045*** | −0.506*** |
| | | | (−3.91) | (−2.93) | (−3.66) | (−2.87) |
| 间接延迟 | | | −1.825* | −0.196 | −1.668* | −0.182 |
| | | | (−1.66) | (−1.03) | (−1.66) | (−1.11) |
| 持续时间 | | | | | −0.034 | −0.002 |
| | | | | | (−0.30) | (−0.30) |
| 政府非军事机构 | | | | | 0.450 | 0.027 |
| | | | | | (0.70) | (0.66) |

续表

| 变量 | 模型 1 | | 模型 2 | | 模型 3 | |
|---|---|---|---|---|---|---|
| | 系数 | 边际效应 | 系数 | 边际效应 | 系数 | 边际效应 |
| 政府军事机构 | | | | | −0.103 | −0.008 |
| | | | | | (−0.18) | (−0.18) |
| 常数项 | 3.929*** | | 6.488*** | | 6.293*** | |
| | (4.76) | | (3.60) | | (3.54) | |
| 样本量 | 169 | | 169 | | 169 | |
| 拟 R 平方 | 0.239 | | 0.532 | | 0.547 | |
| Chi2 统计量 | 10.414 | | 35.582 | | 38.526 | |
| $p$ 值 | 0.001 | | 0.000 | | 0.000 | |

注:(1)括号内表示估计系数对应的 $t$ 值;(2) * 、** 和 *** 分别表示在 0.1、0.05 和 0.001 的统计水平上显著;(3)所有边际效应都取样本平均值。

如表 3 所示,基于三种模型设定形式估计的 Probit 模型整体拟合效果都表现良好,所有模型的 Chi2 统计量都在 1%的统计水平上高度显著。模型 1 中仅控制了“严重程度”变量,此时模型的拟 R 平方为 0.239,说明冲突事件的严重程度可以在 23.9%的程度上解释是否生成“攻防制衡”模式。从估计系数来看,事件严重程度对“攻防制衡”模式生成的概率在 1%的统计水平上会产生显著的负向影响。结合边际效应的估计结果,事件严重程度每提高 1 个等级,冲突双方实现“攻防制衡”模式的概率会下降 7.1%。

考虑到单变量模型中可能由遗漏变量引起内生性问题,导致严重程度的估计结果有偏且不一致,本文补充估计了模型 2 和模型 3,分别引入了更多的事件特征作为控制变量。随着在模型中引入更多的控制变量,模型的拟 R 平方得到了显著的提高,而严重程度仍然被证实在 1%的统计水平上会对“攻防制衡”的生成产生显著的抑制作用。估计结果显示,事件严重程度对“攻防制衡”的边际影响效果仅从−7.1%下降到了−6.9%。据此可以认为,事件严重程度对“攻防制衡”的负向影响在不同模型设定下表现得十分稳健,且应该可以排除模型中存在内生性问题的可能性。

此外,从控制变量的估计结果来看,本文选择“是否 DDoS”的虚拟变量

代表“攻击方式”与“是否 APT”变量放入模型中①。通过模型估计,攻击方式中采取 DDoS 攻击在 10%的统计水平上会对“攻防制衡”模式的生成产生显著的负向影响,即相对于其他方法,在 DDoS 攻击下,冲突双方实现“攻防制衡”态势的概率会下降约 9.5%。相对于直接即时的攻击损害类型,直接延迟在 1%的统计水平上也会对“攻防制衡”模式的生成产生显著的负向影响,且在间接即时损害类型下实现“攻防制衡”态势的概率较直接即时低了 50%以上。然而,通过损害类型不同分类下的样本分布来看,间接即时与间接延迟下的冲突事件样本各仅为 1 个,考虑到样本数量极少,并不能说明损害类型对因变量的生成具有因果关系。而基于共线问题,导致模型无法控制“冲突类型”变量,因此只能把它排除在控制变量之外。

接下来,本文进而考虑采用 Logit 和 Cloglog 模型对式(1)再次进行估计,结果如表 4 所示。结合表 3 和表 4 的估计结果,可以认为包括“严重程度”在内的变量对“攻防制衡”模式生成的影响效果在不同的模型设定下仍十分稳健。

通过 Logit 与 Cloglog 模型的比较可以发现,两种模型的整体拟合效果表现依然良好,所有模型的 Chi2 统计量都在 1%的统计水平上高度显著,且估计结果高度一致。因此,我们可以得出以下三点结论:一是国际网络冲突的“攻防制衡”模式成为当前国际网络冲突攻防互动的常态,假设 1 被证实。二是从影响“攻防制衡”模式生成的可能自变量描述性统计分析中,可看出国际网络冲突的损害程度与现实国家之间的军事冲突相比,还不可同日而语,网络安全给国家安全带来的威胁存在人为夸大的成分。造成网络安全议题被夸大的原因可从媒体、决策者和专家三个主体层面分析,包括并不仅

① 本研究将“是否 DDoS”作为重要参考值代替其他攻击方式,一方面是为了规避分类变量过多而在实际统计建模过程中所导致的共线性问题,另一方面是因为 DDoS 攻击作为一种公开式的网络攻击,会对目标国的网络系统与社会民众的心理产生较大的负面影响。换句话说,虽然 DDoS 攻击所带来的破坏力并不强大,但造成的社会影响却不容小觑。相似观点可参考:Ryan C. Maness and Brandon Valeriano, “The Impact of Cyber Conflict on International Interactions,” *Armed Forces & Society*, Vol. 42, No. 2, 2016, pp. 301-323.

表 4 “攻防制衡”模式的 Logit 与 Cloglog 模型比较分析

| 变量 | Logit 1 | | Logit 2 | | Cloglog 1 | | Cloglog 2 | |
|---|---|---|---|---|---|---|---|---|
| | 系数 | 边际效应 | 系数 | 边际效应 | 系数 | 边际效应 | 系数 | 边际效应 |
| 严重程度 | −1.439*** | −0.072*** | −2.323*** | −0.071*** | −0.540*** | −0.069*** | −1.196* | −0.075** |
| | (−2.89) | (−2.67) | (−2.82) | (−3.39) | (−2.98) | (−3.49) | (−1.82) | (−2.12) |
| 攻击方式(DDoS) | | | −3.483* | −0.106** | | | −1.513 | −0.095 |
| | | | (−1.70) | (−2.27) | | | (−1.26) | (−1.43) |
| 直接延迟 | | | −0.716 | −0.019 | | | | −0.312 |
| | | | (−0.49) | (−0.54) | | | (−0.69) | (−0.71) |
| 间接即时 | | | −5.974*** | −0.487*** | | | | −2.949*** |
| | | | (−2.94) | (−4.19) | | | (−2.63) | (−1.78) |
| 间接延迟 | | | −3.423* | −0.192 | | | | −1.286 |
| | | | (−1.74) | (−1.41) | | | (−1.28) | (−0.76) |
| 持续时间 | | | −0.179 | −0.005 | | | 0.052 | 0.003 |
| | | | (−0.64) | (−0.69) | | | (0.45) | (0.46) |
| 政府非军事机构 | | | 1.275 | 0.040 | | | | 0.311 |
| | | | (0.79) | (0.77) | | | (0.70) | (0.64) |
| 政府军事机构 | | | −0.144 | −0.006 | | | | 0.105 |
| | | | (−0.12) | (−0.12) | | | (0.20) | (0.21) |
| 常数项 | 7.523*** | | 12.780*** | | | | 5.399** | |
| | (4.16) | | (2.93) | | | | (2.35) | |
| 样本量 | 169 | 169 | 169 | 169 | 169 | 169 | 169 | 169 |
| 拟 R 平方 | 0.234 | | 0.551 | | | | | |
| Chi2 统计量 | 8.356 | | 29.300 | | 8.869 | | 28.652 | |
| *p* 值 | 0.004 | | 0.000 | | 0.003 | | 0.000 | |

注:(1)括号内表示估计系数对应的 *t* 值;(2)*、** 和 *** 分别表示在 0.1、0.05 和 0.01 的统计水平上显著;(3)所有边际效应都取样本平均值。

限于恐惧、权力以及经济利益的存在。[①] 三是从严重程度对“攻防制衡”模式

① 刘建伟:《恐惧、权力与网络安全议题的兴起》,《世界经济与政治》2013 年第 12 期,第 43—59 页;蔡翠红、杰古:《网络战叙事的结构分析:主题和动因》,《情报杂志》2014 年第 8 期,第 25—30 页。

生成影响情况的回归模型中可看出，冲突事件的严重程度是影响“攻防制衡”模式生成的关键因素，且影响效果非常显著，假设 2 被证实。

因此，一方面，可以证实假设 1 是成立的，即国际网络冲突的攻防关系常态表现为“攻防制衡”的关系模式。虽然国际网络冲突在冲突的过程中是有利于攻方的，但网络攻击的效用在网络冲突结果维度中得到了约束与制衡，这有利于网络空间的相对安全与国际体系的战略稳定。另一方面，通过多模型比较，可以证实网络攻击给目标国造成的严重程度与“攻防制衡”模式的生成构成非常显著的负向关系，故假设 2 得到证实。接下来，本文将通过证实、证伪两个现实案例进一步验证之前的假设是否成立。

## （二）网络“攻防制衡”模式的案例考察

为进一步验证“攻防制衡”的关系模式是当前国际网络冲突攻防互动的常态，本部分选取了伊朗对沙特的“沙蒙”（Shamoon）行动作为证实案例，选取了美国对伊朗的“震网”行动作为证伪案例。[①] 这两则案例的选取依据是：“沙蒙”与“震网”都是以国家行为体作为主要参与者发起的网络冲突事件，且两次行动均成功完成了网络空间的攻击。但作为证实案例，“沙蒙”的目标国沙特阿拉伯并没有作出任何妥协举动；而“震网”的目标国伊朗的核项目因为该事件而遭到重创，并被迫做出一定程度的示弱行为，因此可作为证

① 为避免案例中可能存在的归因不确定问题，本部分的案例选取仍然源于双边网络事件和争端数据库（DCID）。原因在于该数据库中的案例基于高度一致性的卡帕系数以及多方权威机构和专家的认证，否则案例将被数据库排除在样本之外。从案例的因果推断来看，西方学者的观点占据已知的众多公开资料。因此，我们仍然不能绝对认定“沙蒙”与“震网”的幕后主使，但作为学术研究者，基于条件的限制只能力求通过多方公开资料还原事件真相。因此，案例分析也不失为一种可取的检验方式。可以确定的是，无论是“沙蒙”行动还是“震网”行动，仅凭一般个人或非政府组织的能力是无法独立发动如此大规模网络攻击的。此外，主动宣称对网络攻击事件负责往往更符合独立黑客或非政府组织的心理，因为非政府组织可能会存在名声、金钱利益等其他目的，而不纯粹是为了发起攻击。所以，“沙蒙”与“震网”归因不确定的问题可以得到较好的解决。关于非政府组织发起网络攻击的动机问题，可参考：Erik Gartzke，“The Myth of Cyberwar：Bringing War in Cyberspace Back Down to Earth，” *International Security*，Vol. 38，No. 2，2013，pp. 41-73.

伪案例使用。

1. 对证实案例的考察:伊朗对沙特的“沙蒙”行动

2012 年 8 月 15 日始,沙特阿美石油公司[①]遭受了严重的黑客攻击。在几个小时内,超过 35000 台计算机的内部信息被部分擦除或计算机被完全摧毁。这种病毒抹杀了沙特石油公司 85%的计算机数据并以燃烧的美国国旗图像取代。由于无法供给石油,沙特石油公司供应全球 10%石油的能力突然受到威胁。随后,计算机专家发现了这种新型病毒——“沙蒙”,也被称为“W32. DisTrack”。这个恶意软件有一个逻辑炸弹,触发了主引导记录和数据清除有效负载并使联网计算机无法使用。公司前安全顾问克里斯·库贝卡(Chris Kubecka)说,如果计算损失的成本,索尼影业遇袭事件与这次网络攻击事件相比都显得微不足道。[②] 不过,该公司在一份声明中称,这次袭击对石油生产作业没有任何影响,也没有造成任何基础设施的破坏。[③] 国际社会普遍认为伊朗的嫌疑最大。著名网络安保公司托菲诺安全(Tofino Security)认为,这次袭击是由一位不满沙特政府的内部人士参与发起的。但仅凭借其个人能力是无法发起如此大规模的网络攻击的。托菲诺认为这位内部人士与伊朗政府展开了密切的合作。[④] 美国中情局同样认为此次“沙蒙”攻击是由伊朗网军和沙特什叶派教徒合作完成的一次成功的网络

---

① 沙特阿美石油公司,又称沙特阿拉伯石油公司。据一些金融杂志估计,这家石油巨头的市值高达 10 万亿美元,政府持股达到 100%,是世界上市值最高的公司。因此,对沙特阿美石油公司的威胁可能危及沙特阿拉伯的国家安全。故沙特王国投入了 33000 名士兵和 5000 名警卫的武装部队,以确保沙特阿美石油公司设施的安全。相关介绍请参考:Anthony H. Cordesman, *Saudi Arabia: National Security in a Troubled Region* (Santa Barbara, California: ABC-CLIO, 2009)。

② CNN Business, Jose Pagliery, “The Inside Story of the Biggest Hack in History,” August 5, 2015, https://money.cnn.com/2015/08/05/technology/aramco-hack/index.html, 访问时间:2018 年 12 月 10 日。

③ BBC News, “Shamoon Virus Targets Energy Sector Infrastructure,” August 17, 2012, 访问时间,2018 年 12 月 10 日。

④ Heather MacKenzie, “Shamoon Malware and SCADA Security-What are the Impacts?” Tofino Security, October 25, 2012, https://www.tofinosecurity.com/blog/shamoon-malware-and-scada-security---what-are-impacts, 访问时间:2018 年 12 月10 日。

袭击。①

那么沙特政府是如何回应的？从现有的公开信息来看，沙特石油公司在遭受网络攻击后即刻请出数名美国网络安全专家对病毒进行监控。在袭击发生后数小时内，著名安全公司赛门铁克(Symantec)的研究人员开始分析病毒样本。② 实际上，沙特的网络防御能力仍较为落后，但其有着充足的资金来聘请外援帮助其解决网络安全问题。在“沙蒙”行动之后，沙特并没作出任何过激反应，而是加紧修复网络安全漏洞，并在石油供应停摆之后恢复了供给。可以说，这场历史性的网络攻击并未给沙特的石油生产和供给造成严重影响。从长远发展战略规划上的反应来看，沙特先后制定了一系列改革方案来增强本国的综合实力。以上事实足以说明，沙特在遭受“沙蒙”袭击之后并没作出任何形式的妥协和让步。而分析伊朗发起“沙蒙”行动的原因，一种观点认为，行动的原因在于沙特政府对叙利亚和巴林逊尼派的援助；另一种观点认为，这次网络攻击是伊朗对美国对其核设施发动的“震网”行动的报复。“沙蒙”行动没有动摇地区安全结构，沙特与伊朗在中东地区的对立格局并没有发生任何改变，更没有撼动中东地区大国的战略稳定，而是地缘政治博弈在网络空间的一次反映，符合国际网络冲突中“攻防制衡”关系模式的理论假设。

2. 对证伪案例的考察：美国对伊朗的“震网”行动

如上所述，本文不仅关注支持“攻防制衡”模式的网络冲突事件，也对看似不支持假设的案例进行重点分析。本文的核心假设认为，当前国际网络冲突普遍表现为“攻防制衡”的关系模式，国际入侵虽有利于进攻，但网络攻击的效用得到制衡，有利于网络空间的相对安全和大国的战略稳定。为什么震惊世界的“震网”事件却实现了发起国的战略意图——破坏伊朗核计

① 转引自：Richard Sale，“Iran behind Shamoon Attack，” Industrial Safety and Security Source，October 15，2012，http://www.isssource.com/iran-behind-shamoon-attack/，访问时间：2018 年 12 月 10 日。

② Nicole Perlroth，“In Cyberattack on Saudi Firm，U.S. Sees Iran Firing Back，” The New York Times，October 23，2012，https://www.nytimes.com/2012/10/24/business/global/cyberattack-on-saudi-oil-firm-disquiets-us.html，访问时间：2018 年 12 月 13 日。

划？上述案例在网络冲突的结果维度上看似与“攻防制衡”模式相悖，只有证明该逆向案例在本质上仍属于“攻防制衡”的理论范畴，才能够确认“攻防制衡”模式在解释力和预测力上的有效性。

“震网”是一种计算机蠕虫病毒，于 2010 年 6 月被发现。该病毒最主要的攻击目标是伊朗纳坦兹铀浓缩厂的核设施。“震网”专门被用于针对可编程逻辑控制器(Programmable Logic Controllers，PLCs)，这种控制器允许机电过程自动化，例如用于控制工厂装配线上的机械或用于分离核材料的离心机。“震网”病毒利用了 4 个零日漏洞(Zero-day Exploit)①，通过瞄准使用微软窗口操作系统(Windows)和网络机器来发挥作用。据报道，“震网”病毒破坏了伊朗的可编程逻辑控制系统，收集工业系统的信息，导致快速旋转的离心机分裂。② “震网”病毒毁坏了伊朗近 1/5 的离心机，感染了 20 多万台计算机，导致 1000 台机器物理退化，并使得伊朗核计划倒退了两年。③ 赛门铁克在 2010 年 8 月指出，全球 60％的受感染计算机在伊朗。④ 俄罗斯网络安全公司卡巴斯基实验室进一步得出结论，如此复杂的攻击只能在“国家支持下”才可进行。⑤ 这进一步证实了发起“震网”攻击的幕后主使有伊朗的宿敌美国。针对这起网络入侵，伊朗时任总统内贾德虽证实了攻击事实，

---

① Ryan Naraine, “Stuxnet Attackers Used 4 Windows Zero-day Exploits,” ZDNet, September 14, 2010, https://www.zdnet.com/article/stuxnet-attackers-used-4-windows-zero-day-exploits/, 访问时间：2018 年 12 月 13 日。

② David Kushner, “The Real Story of Stuxnet,” IEEE Spectrum, February 26, 2013, https://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet/, 访问时间：2018 年 12 月 3 日。

③ Yaakov Katz, “Stuxnet Virus Set back Iran's Nuclear Program by 2 Years,” *The Jerusalem Post*, December 15, 2010, https://www.jpost.com/Iranian-Threat/News/Stuxnet-virus-set-back-Irans-nuclear-program-by-2-years, 访问时间：2018 年 12 月 13 日。

④ William MacLean, “UPDATE 2-Cyber Attack Appears to Target Iran-tech Firms,” Reuters, September 24, 2010, https://www.reuters.com/article/security-cyber-iran/update-2-cyber-attack-appears-to-target-iran-tech-firms-idUSLDE68N1OI20100924, 访问时间：2018 年 12 月 13 日。

⑤ Kaspersky Lab, “Kaspersky Lab's Experts Believe that Stuxnet Manifests the Beginning of the New Age of Cyber-warfare,” September 24, 2010, https://www.kaspersky.com/about/press-releases/2010_kaspersky-lab-provides-its-insights-on-stuxnet-worm, 访问时间：2018 年 12 月 13 日。

但并未明确描述更多细节，且伊朗政府宣称该病毒并未对其核项目产生较大的影响，但鉴于“震网”病毒的扩散程度，要清除铀浓缩过程中涉及的所有计算机设备的病毒是非常困难的。① 也许正是这些忧虑使得伊朗在 2010 年 11 月全面暂停了纳坦兹的铀浓缩生产。②

该案例显示了在这次国际网络冲突中，发起国美国成功完成网络空间的攻击，并给目标国伊朗的核设施带来重大打击，且伊朗在一定程度上被迫选择暂停核武器重要原料铀浓缩生产。这在形式上符合国际网络冲突攻防关系模型中的“进攻制胜”假设。因此，该案例完全可以作为“攻防制衡”模式的证伪案例。

本文认为，纵使美国针对伊朗的“震网”行动在表面上符合“进攻制胜”假设的逻辑，但实际上，“震网”仍然符合“攻防制衡”理论模式的内在逻辑。第一，纳坦兹铀浓缩生产的暂停是伊朗为防止面临更大的损失而被迫选择的结果，并不反映伊朗政府主观意愿上的屈服和让步。第二，伊朗被破坏的网络系统并没有被彻底毁坏，且具有恢复的可能。虽然“震网”在短时间内给伊朗核武的研制带来了冲击，但并没有从根本上毁灭伊朗核武的研制能力。第三，美伊两国常年敌对的关系并不会使姿态强硬的伊朗单方面作出实质性的妥协和让步，反而更加刺激伊朗不断加强网络空间的攻防能力建设，加紧修复网络攻击给伊朗核设施带来的损失，并加快推动伊朗核计划的实施，伊核问题也没有得到实质性解决。

国际网络冲突并不能使网络冲突的目标国轻易选择妥协和让步，也没有升级为激烈程度更高的网络战争，更难以轻易改变国际体系的战略稳定与国家政策选择。因此，无论是证实案例还是所谓的证伪案例从正反两面均证实了国际网络冲突的攻防互动态势基本符合“攻防制衡”理论模式的内在逻辑。依据以上案例分析冲突严重程度对“攻防制衡”模式生成的影响，可以看出严重程度对“攻防制衡”模式的生成产生显著的负向影响。首先，

---

① 保罗·沙克瑞恩等，《网络战：信息空间攻防历史、案例与未来》，吴奕俊译，北京：金城出版社，2016 年，第 273—274 页。

② David Albright et al., *Did Stuxnet Take Out 1,000 Centrifuges at the Natanz Enrichment Plant?* (Washington, DC: Institute for Science and International Security, 2010), p. 6.

网络攻击严重程度较高不利于“攻防制衡”模式的生成。“震网”事件不可不谓一次成功的网络入侵事件,虽然伊朗政府并没有主动作出妥协和让步,但该病毒成功破坏和推迟了伊朗核计划的实施,这在一定程度上实现了美国的部分战略意图。其次,网络攻击的严重程度仍然有限,难以催使目标国主动作出妥协和让步的政策选择。从“沙蒙”行动可以看出,该网络入侵也是一次严重程度较高的事件,并对沙特石油公司的网络信息造成了严重破坏,但是依然没有促使沙特放弃对伊朗逊尼派势力的援助和对本国什叶派群体的打压。伊朗的“沙蒙”行动虽然成功对沙特的国有资产造成了一定的损失,但这种损失还没有达到促使沙特作出让步的程度。综上,本文认为,网络冲突的严重程度对“攻防制衡”模式的生成具有一定的负面影响,但不具备决定性作用。

## 五、“攻防制衡”模式与网络空间的战略稳定

本文实证检验表明,“攻防制衡”模式是当前国际网络冲突攻防互动的常态。网络空间攻防关系理论模型的构建有助于从根本上把握网络冲突攻防双方的互动态势,进而有利于国家依据互动规律积极应对网络入侵,抑制网络冲突的战略效果,维护网络空间的战略稳定。对于发起国来说,赢得网络冲突并不会给自身带来更多的政治权力,输掉网络冲突也不会真正损害一个在网络空间已经很强大的行动者的权威。① 从国际稳定的视角来看,网络空间的进攻占优论暗含着单极稳定论的假设。威廉·沃尔弗斯(William Wohlforth)在首创单极稳定论时曾指出:多极世界中,大国之间进行着霸权竞争和安全竞争,从而导致国际体系的不稳定。② 网络空间是全球性虚拟空间,任何国家都有入侵他国“网络边疆”的可能性,进攻占优的技术优势使得这种客观的多极体系增添了不稳定的因素,不利于网络空间的相对安全与国际体系的战略稳定。而“攻防制衡”模式下的网络空间则更偏向于多极稳

① Mariarosaria Taddeo, “Cyber Conflicts and Political Power in Information Societies,” *Minds and Machines*, Vol. 27, No. 2, 2017, pp. 265-268.

② William C. Wohlforth, “The Stability of a Unipolar World,” *International Security*, Vol. 24, No. 1, 1999, pp. 5-41.

定论的理论假设。正如卡尔·多伊奇(Carl W. Deutsch)与戴维·辛格(David J. Singer)所认为的:随着国际体系从两极转向多极化,国家之间的战争倾向度较低,而国家间的互动关系不仅是竞争性的,合作性也会凸显。[①] 一方面,"攻防制衡"模式下的国家在孕育着多极环境下的网络冲突中存在一种相互的制衡关系,这种客观上的制衡对网络冲突的升级和国际体系的失序起到稳定作用。在传统的国际社会,战争或冲突的解决往往作为政治权力转移的标志,比如19世纪初的滑铁卢战役。约瑟夫·奈(Joseph S. Nye)认为,信息时代的权力已实现在国家和非国家行为体之间扩散,而非转移[②]。另一方面,尽管世界各国纷纷加强其网络攻防能力的建设,但合作仍然是网络空间互动的重要方式,如网络安全对话作为中美两国四个高级别对话机制之一,对加强两国之间的战略稳定具有重要意义。全球网络安全以敌对或合作的方式为特征,可能取决于网络大国在制定这些负责任国家行为准则方面所做努力的短期成败。[③] 而国际网络冲突的"攻防制衡"会对网络空间的冲突态势产生抑制作用,这有助于网络空间的战略稳定。

如图2所示,"攻防制衡"模式的生成与网络冲突的严重程度存在较强的负向关系,但不具备决定性作用。在"攻防制衡"模式下,网络冲突的目标国弱化了发起国网络攻击的实际效果,使得发起国的战略意图难以实现。因此,网络冲突对国际体系的战略稳定难以构成挑战。"进攻制胜"模式下的网络冲突虽表现为目标国做出对预谋活动的暂停、放弃等行为,但这并不能说明是目标国主动作出了妥协和让步。所以,"进攻制胜"模式下的网络冲突的内在逻辑在较大程度上仍存在符合"攻防制衡"理论模式的可能。当然,如果发起国对目标国的网络入侵行为造成了切实严重的损害,则会在一定程度上改变目标国的战略行动,并令国际体系的战略稳定滋生不稳定因素。

---

① Carl W. Deutsch and J. David Singer, "Multipolar Power Systems and International Stability," *World Politics*, Vol. 16, No. 3, 1964, pp. 390-406.

② Joseph S. Nye, *The Future of Power* (New York: PublicAffairs, 2011), p. 15.

③ Paul Meyer, "Seizing the Diplomatic Initiative to Control Cyber Conflict," *The Washington Quarterly*, Vol. 38, No. 2, 2015, pp. 50-51.

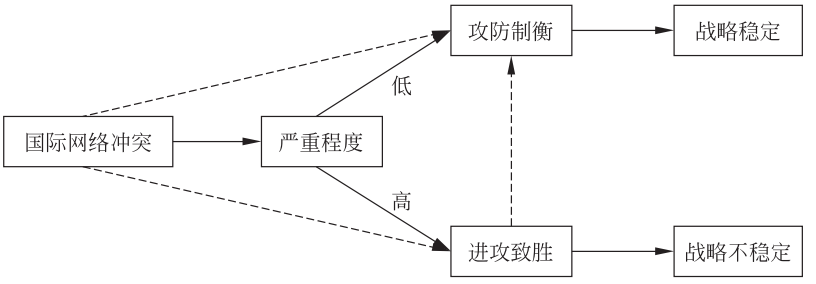


**图 2 "攻防制衡"模式对战略稳定的影响**

事实上,网络攻击不太可能被证明具有特别强大的战略意义,除非它们能对对手造成实质性的、持久的伤害。"由于网络攻击涉及对目标军事能力和民用基础设施的临时软杀伤,如果不能同时进行旨在对目标的恢复能力造成永久性损害的地面攻击,那么攻击的价值就基本上失效了。"[①]里夫认为,网络战似乎更像是一种工具,只有在非常有限的情况下,才能以相对较低的成本追求政治(战略)或军事(战术)目标。[②] 在大多数情况下,只有在网络攻击伴随地面军事力量(或其他旨在利用互联网取得的任何使目标国暂时丧失能力的行动)时,这种情况才会发生。例如,2008 年俄罗斯通过对格鲁吉亚发动网络攻击以配合地面军队的跨境行动,使得网络攻击为军事进攻提供了辅助作用。从这个角度来看,网络攻击并没有给军事战争带来颠覆性影响,相反,它只是承诺要扩大现有的国家间权力与影响力的差距。[③]仅仅通过互联网造成危害的能力并不能预测网络战将取代陆地战争,甚至不能预测网络战将成为未来战争的一个重要独立领域。[④] 从曼尼斯在早先整理的 2001 年至 2011 年的 DCID 数据库也可发现这样的规律:在 126 起网

① Erik Gartzke, "The Myth of Cyberwar: Bringing War in Cyberspace Back Down to Earth," *International Security*, Vol. 38, No. 2, 2013, p. 62.

② Adam P. Liff, "The Proliferation of Cyberwarfare Capabilities and Interstate War, Redux: Liff Responds to Junio," *Journal of Strategic Studies*, Vol. 36, No. 1, 2013, pp. 135-136.

③ Erik Gartzke, "The Myth of Cyberwar: Bringing War in Cyberspace Back Down to Earth," *International Security*, Vol. 38, No. 2, 2013, p. 43.

④ Ibid, p. 57.

络冲突事件中，仅有 20 对（约 16%）事件存在目标国报复与反击的行为，而绝大多数（80%）网络冲突事件仅仅是一种由发起国发起的单向网络攻击，而目标国并没有做出反击和报复的行为。[①] 一方面，发起国难以通过网络攻击对目标国造成较大的损害；另一方面，即便发起国成功完成网络空间的攻击行为，绝大多数的目标国也并没有做出较为激烈的报复行为，这也在侧面上证明了网络攻击带来的损害并不能达到发起报复性行动的阈值。相反，这种损害的后果被证明是效用甚微的，因为在地区或全球事务中，改变权力平衡的能力只适用于已具有相当国际影响力的国家行为体。因此，国际网络冲突事件虽然频发，但却始终维持在低烈度的态势，在“攻防制衡”的攻防关系模式下，并不能对国际网络空间的战略稳定构成挑战。

## 六、结　论

综上所述，“攻防制衡”模式被证实为当前国际网络冲突攻防互动的普遍态势。该模式较好地解释了国际网络冲突中为何发起国常常进攻成功，而目标国却选择不妥协的问题。其背后理论模式的构建也进一步弥补了传统的进攻占优理论的不足，并有助于增强网络空间攻防平衡理论的解释效力。其生成的内在逻辑在理论上源于网络空间的攻防关系模型的构建；在现实层面源于网络冲突给目标国造成的损害程度仍然较低，且大概率远低于目标国作出妥协的成本，即使严重程度较高会增加目标国作出妥协的概率。事实上，网络冲突带来的现实危害难以达到发起国的目标阈值，正是其低烈度特征使得“攻防制衡”模式成为当前国际网络冲突攻防互动的常态。这既是“攻防制衡”模式生成的现实逻辑，也为网络空间的相对安全和国际体系的战略稳定提供了可能。

本研究构建的“攻防制衡”理论模式也存在一定的局限性，但并不会影响理论构建和检验的可靠性。首先，从理论模式的构建上来看，“攻防制衡”

---

① 刘杨钺：《国际政治中的网络安全：理论视角与观点争鸣》，《外交评论》2015 年第 5 期，第 131 页；Brandon Valeriano and Ryan C. Maness, “The Dynamics of Cyber Conflict between Rival Antagonists, *2001—11*,” *Social Science Electronic Publishing*, Vol. 51, No. 3, 2014, pp. 347-360。

模式必须建立在一定的前提假定下才具备生成的环境与条件。其次,由于DCID 数据库仍在建设中,理论检验的样本量限制难免会导致变量之间产生共线性突出等问题,同时数据的搜集和整理仍无法避免存在一定程度的遗漏和偏差。这看似对本文数据检验的科学性构成了挑战,但从编码过程的严谨性和统计建模的显著度来看,该数据库具备较高的契合性与可靠性,且难以影响本文核心理论假设的验证结果。最后,关于本文案例选择,其参考资料大多源于多方公开信息,由于网络安全领域涉及诸多敏感信息,甚至被国家有意保密和隐瞒,所以可能存在一定偏差。作为学术研究者,基于研究条件的限制,只能力求通过多方公开的信息还原事实原貌。本文将案例的选取限定在 DCID 数据库中,基于其对归因问题的专业化处理,故而极大降低了归因失误带来的影响。因此,该局限之处同样难以对本研究的理论检验构成挑战。

本研究为国家间网络冲突行为的研究提供了一个新的解释维度。这种维度将跳出传统的网络攻防过程,而聚焦于网络攻防背后的冲突结果。这既有利于研究者揭开网络攻防技术对国际网络冲突过程施加的“伪装”,从根本上认知国际网络冲突在国际互动过程中带来的直接影响;也成功规避了进攻占优的“魔咒”,从而有力反驳了网络威胁的过分夸大的观点。事实证明,国际网络冲突客观存在的“攻防制衡”模式,极大地削弱了网络冲突给国际体系的战略稳定所带来的威胁和挑战,并为主权国家参与网络空间安全的治理与合作提供了理论依据和现实根基。