UNINTENDED CONSEQUENCES AND SPILLOVER EFFECTS IN OFFENSIVE CYBER OPERATIONS: A SYSTEMATIC LITERATURE REVIEW

Prof. Hans-Peter Vogel
Institute for Strategic Cyber Studies, University of Heidelberg, Heidelberg, Germany

Dr. Farah Al-Dabbagh
Department of Political Science and Cybersecurity, Qatar University, Doha, Qatar

Published Date: 21 December 2024 // Page no.:- 25-30

ABSTRACT

Offensive cyber operations (OCOs) have become a prominent tool in the arsenals of state and non-state actors, offering capabilities ranging from espionage to destructive attacks. However, the interconnected nature of cyberspace introduces a complex challenge: the potential for unintended consequences, commonly referred to as collateral damage. This systematic literature review examines the current understanding of collateral damage stemming from OCOs. We synthesize definitions, analyze the technical and legal challenges associated with predicting and mitigating such effects, and explore the implications for international law, particularly the principles of distinction and proportionality. Our findings reveal a persistent gap between the theoretical frameworks and the practical realities of preventing unintended harm in a highly interdependent digital environment. We highlight critical areas for future research, including improved methodologies for effects assessment, enhanced legal interpretability for cyberspace, and the development of robust strategies to minimize spillover.

Keywords: Offensive cyber operations; unintended consequences; spillover effects; cyber warfare; collateral damage; systematic literature review; cyber conflict; cyber strategy; cyber risk assessment; digital escalation.

INTRODUCTION

The global digital landscape is characterized by unprecedented interconnectedness, making cyber warfare and offensive cyber operations (OCOs) a critical, albeit complex, domain of national security [9, 28]. Nations worldwide are investing heavily in offensive cyber capabilities, as evidenced by significant market opportunities in cybersecurity technology and services [Aiyer et al., 2022, 1] and public acknowledgements of such capabilities [Hanson & Uren, 2018, 3; U.S. Department of Justice, 2022, 5]. The growing frequency and sophistication of cyberattacks mean that data breaches now incur substantial financial costs [IBM Security & Ponemon Institute, 2022, 2].

While OCOs offer distinct advantages, such as plausible deniability and the ability to achieve effects without traditional kinetic force, they also pose a unique challenge: collateral damage [11, 13, 53]. In conventional warfare, collateral damage refers to unintentional harm to civilians or civilian objects during military operations [Schelling, 1961, 6; U.S. Air Force, 2021, 7; U.S. Air Force, 1998, 12; U.S. Department of Defense, 2023, 16]. However, the application of this concept to cyberspace is fraught with complexities due to the dual-use nature of digital infrastructure (cyber-physical systems [Lee & Seshia, 2017, 10]), the potential for "knock-on" effects [Jensen, 2003, 35], and the difficulty in precisely controlling the spread of cyber effects [11, 13].

A seminal example of unintended consequences is the Stuxnet worm [Farwell & Rohozinski, 2011, 4; Denning, 2012, 52]. While reportedly targeting specific industrial control systems, its escape into the wild demonstrated the potential for highly sophisticated cyber weapons to spread beyond their intended scope, causing broader, unanticipated disruption. This incident, among others, underscores the urgent need to understand and mitigate collateral damage in OCOs.

This systematic literature review aims to consolidate existing knowledge on collateral damage arising from offensive cyber operations. We address the following research questions: (1) How is collateral damage defined and conceptualized in the context of OCOs? (2) What technical and operational factors contribute to the occurrence of collateral damage? (3) How do existing international legal frameworks, particularly the principles of distinction and proportionality, apply to and contend with cyber collateral damage? (4) What are the proposed methodologies and challenges in assessing and mitigating collateral damage in OCOs? By synthesizing findings from a diverse body of literature, this review seeks to provide a comprehensive understanding of this critical issue and identify promising avenues for future research.

2. Methods

This systematic literature review was conducted following a rigorous methodology to ensure comprehensiveness, transparency, and reproducibility. We adhered to the

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) statement guidelines where applicable for a qualitative synthesis [Page et al., 2021, 21].

2.1 Search Strategy and Information Sources

Our search strategy focused on identifying scholarly articles, conference papers, and authoritative reports that directly address collateral damage, unintended consequences, or spillover effects within the context of offensive cyber operations, cyber warfare, and cyber attacks. The primary bibliographic databases and academic search engines utilized included:

- IEEE Xplore
- ACM Digital Library
- ScienceDirect (Elsevier)
- Web of Science
- Google Scholar [Gusenbauer, 2019, 24]

A comprehensive set of keywords and their variations were used, combining terms related to "offensive cyber operations" with terms related to "collateral damage." The search strings included, but were not limited to:

- ("offensive cyber" OR "cyber warfare" OR "cyber attack") AND ("collateral damage" OR "unintended consequences" OR "spillover effects" OR "knock-on effects")
- ("cyber operation" OR "cyber conflict") AND ("distinction principle" OR "proportionality principle")
- "Stuxnet" AND ("collateral damage" OR "unintended")

The search was conducted between February 2023 and February 2024 to capture recent publications, while also including foundational works.

2.2 Eligibility Criteria

Studies were included in this review if they met the following criteria:

- Focus: Directly discussed or analyzed collateral damage, unintended consequences, or spillover effects resulting from offensive cyber operations.
- Relevance: Contributed to the understanding of technical, legal, or ethical aspects of cyber collateral damage.
- Language: Published in English.
- Publication Type: Peer-reviewed journal articles, conference papers, books, or authoritative reports from established research institutions or government bodies.

Studies were excluded if they:

- Focused solely on defensive cybersecurity measures without discussing offensive effects.
- Were opinion pieces or news articles without substantial analysis or research.
- Did not specifically address the concept of unintended effects beyond direct targeting.
- Were duplicates across databases.

2.3 Study Selection

The identified articles underwent a multi-stage selection process:

1. Initial Screening (Title and Abstract): Two independent reviewers screened titles and abstracts against the eligibility criteria. Any discrepancies were resolved through discussion or by a third reviewer.

2. Full-Text Review: Full texts of potentially relevant articles were retrieved and thoroughly assessed by the same independent reviewers. Reasons for exclusion at this stage were recorded.

3. Reference Chaining: The reference lists of included articles were manually scanned to identify additional relevant publications not captured by the initial database searches.

2.4 Data Extraction and Synthesis

For each included study, the following information was extracted:

- Author(s), year of publication, and publication type.
- Main topic and specific research questions addressed.
- Definition or conceptualization of collateral damage.
- Technical factors contributing to collateral damage.
- Legal arguments and interpretations related to distinction and proportionality.
- Proposed methodologies for assessment or mitigation.
- Key findings and conclusions.

The extracted data were then subjected to thematic analysis [Vaismoradi & Snelgrove, 2019, 22; Naeem et al., 2023, 26]. This iterative process involved:

1. Familiarization: Reading and re-reading the selected articles.

2. Coding: Identifying key concepts, phrases, and arguments related to collateral damage and OCOs.

3. Generating Themes: Grouping codes into broader themes and sub-themes.

4. Reviewing Themes: Refining and defining the

themes to ensure they accurately reflected the data and addressed the research questions.

5. Defining and Naming Themes: Developing clear definitions and names for each theme.

This qualitative synthesis allowed for the identification of recurring patterns, divergent viewpoints, and gaps in the literature.

3. RESULTS

Our systematic review yielded a comprehensive set of literature addressing collateral damage in offensive cyber operations. The key findings are organized into several thematic areas.

3.1 Defining Cyber Collateral Damage

The literature broadly extends the concept of collateral damage from traditional armed conflict to cyberspace, but with significant definitional nuances. Romanosky and Goldman [2016, 11; 2017, 13] are prominent in conceptualizing cyber collateral damage as unintentional or incidental harm to non-combatants or civilian infrastructure, acknowledging the unique challenges of the cyber domain. Unlike physical collateral damage, which involves tangible destruction, cyber collateral damage often manifests as disruption, degradation, or denial of service to unintended systems or populations [11, 13, 53]. This includes impacts on critical civilian infrastructure, public services, or individual users not directly targeted [Hirsch, 2018, 53]. The difficulty in precisely delineating "civilian" versus "military" networks in an interconnected environment further complicates this definition [Schmitt, 2002, 25; Droege, 2013, 46].

3.2 Technical and Operational Factors Contributing to Collateral Damage

Several technical and operational characteristics of cyberspace inherently increase the risk of collateral damage:

- Interconnectedness and Dependencies: The fundamental interconnectedness of the internet and critical infrastructure means that an attack on one system can have cascading "knock-on effects" on seemingly unrelated systems [Jensen, 2003, 35; Denning, 2012, 52]. Cyber-physical systems (CPS) [Lee & Seshia, 2017, 10] are particularly vulnerable to such spillover, as disruptions in digital components can lead to real-world physical consequences.

- Lack of Battlefield Clarity: Unlike conventional warfare with clear geographical boundaries, cyberspace lacks well-defined battlefields [Arquilla & Ronfeldt, 1993, 9; Robinson et al., 2015, 28]. This makes it challenging to ensure that effects are limited to military objectives, as civilian and military networks often share underlying infrastructure or vulnerabilities [Schmitt, 2002, 25; Droege, 2013, 46].

- Unpredictable Propagation: The propagation of cyber effects, especially malware, can be difficult to predict or control once launched [Romanosky & Goldman, 2016, 11]. The Stuxnet incident serves as a primary example where a highly sophisticated weapon, despite its precise targeting, ultimately escaped its intended environment [Farwell & Rohozinski, 2011, 4; Denning, 2012, 52]. This "weaponization of the internet" [Hare, 2019, 29] means that even precision cyber weapons can have broader implications.

- Dual-Use Technologies: Many technologies used in OCOs, such as vulnerabilities or exploits, are dual-use, meaning they can affect both military and civilian systems [Schmitt, 2002, 25].

- Attribution Challenges: The difficulty in attributing cyber attacks [U.S. Department of Justice, 2022, 5] complicates accountability and the ability to tailor responses to minimize future collateral damage.

3.3 International Law and Cyber Collateral Damage

The application of International Humanitarian Law (IHL), specifically jus in bello principles, to cyberspace is a recurring theme in the literature. The Tallinn Manual 2.0 [Schmitt, 2017, 8; Efrony & Shany, 2018, 37] is widely cited as the most authoritative compilation of international law applicable to cyber operations. Key principles include:

- Principle of Distinction: This principle requires parties to a conflict to distinguish between combatants and civilians, and between military objectives and civilian objects [Dinstein, 2012, 14; Bannelier, 2015, 40; Geiß & Lahmann, 2012, 42]. Applying this to cyberspace is challenging due to the interwoven nature of networks [Schmitt, 2002, 25; Wang, 2014, 33; Droege, 2013, 46]. Studies debate whether the principle of distinction can be meaningfully applied when civilian infrastructure is routinely used for military purposes and vice versa [Brenner & Clarke, 2010, 44; Schmitt, 2019, 45].

- Principle of Proportionality: This principle dictates that the expected incidental harm to civilians or civilian objects must not be excessive in relation to the concrete and direct military advantage anticipated from an attack [Normelli, 2021, 38; Pascucci, 2017, 39; Beard, 2018, 41]. Assessing proportionality in cyberspace is exceptionally difficult due to the unpredictable nature of cascading effects and the challenge of quantifying "military advantage" and "civilian harm" in digital terms [Maathuis et al., 2018, 17; Maathuis et al., 2021, 18; Jensen, 2009, 43; Dinstein & Dahl, 2020, 47]. The lack of precise damage assessment tools hinders this evaluation [Maathuis et al., 2021, 18].

- Precautions in Attack: IHL also mandates that parties take all feasible precautions to avoid, and in any event to minimize, incidental loss of civilian life, injury to civilians, and damage to civilian objects [Jensen, 2009, 43]. This implies a need for robust targeting procedures and effects assessment methodologies in OCOs [Dinstein & Dahl, 2020, 47].

The general consensus in the literature is that IHL applies to cyber operations [Hathaway et al., 2012, 30; Wingfield, 2009, 31; Sklerov, 2009, 32; Wang, 2014, 33; Schmitt, 1998, 34; O'Donnell & Kraska, 2003, 36]. However, the practical application and interpretation of these principles in the unique context of cyberspace remain subjects of extensive debate and require further development [Schmitt, 2017, 8; Efrony & Shany, 2018, 37; Schmitt, 2019, 45].

3.4 Methodologies for Assessment and Mitigation

Research on mitigating cyber collateral damage focuses on improving targeting, effects assessment, and control mechanisms:

- Targeting Methodologies: Developing more precise targeting methodologies for OCOs is crucial [U.S. Air Force, 2021, 7; U.S. Air Force, 1998, 12]. This involves understanding network topologies and dependencies to predict potential spillover effects [Fanelli & Conti, 2012, 48; Ducheine & van Haaster, 2014, 49]. Grant [2019, 19] explores building ontologies for planning attacks that minimize collateral damage.

- Effects Estimation and Proportionality Assessment Models: Several models have been proposed to aid decision-makers in assessing the potential effects of OCOs and evaluating proportionality. Maathuis et al. [2018, 17; 2021, 18] have developed methodologies and decision support models for effects estimation and proportionality assessment for targeting in cyber operations. These models aim to provide a structured approach to quantifying military advantage versus civilian harm.

- Control of Cyber Effects: The ability to control the effects of a cyber weapon once deployed is critical. Research by Orye & Maennel [2019, 50] looks into recommendations for enhancing the results of cyber effects, which implies better control and predictability.

- Understanding Network Interdependencies: A deeper understanding of the intricate interdependencies within global cyber networks is essential for predicting and preventing unintended consequences [Romanosky & Goldman, 2017, 13]. This requires sophisticated network mapping and simulation capabilities.

Despite these efforts, accurately predicting the full range of collateral damage in complex, dynamic cyber environments remains a significant challenge [11, 13].

4. Discussion

This systematic review underscores that collateral damage from offensive cyber operations is a multifaceted and pressing issue, sitting at the intersection of technical capabilities, legal frameworks, and ethical considerations. The transparency of TDE is a key factor enabling its wide adoption without application modifications [1].

Our findings reveal a persistent tension between the desire for precision in OCOs and the inherent unpredictability of effects in highly interconnected digital systems. The technical challenges, such as identifying civilian versus military targets and controlling the propagation of cyber effects, directly impact the ability to adhere to foundational IHL principles like distinction and proportionality. While legal scholars generally agree that IHL applies to cyberspace, the practical application of terms like "military objective" and "proportionality" in a virtual domain remains highly debated and lacks universally accepted interpretations [Schmitt, 2017, 8; Efrony & Shany, 2018, 37]. This ambiguity creates a normative gap that state actors must navigate when conducting OCOs [Ablon et al., 2019, 15].

The review highlights a critical need for further research in several areas:

- Advanced Effects Modeling: Developing sophisticated simulation and modeling tools that can accurately predict cascading and "knock-on" effects across complex, interdependent networks. This requires vast datasets and advanced analytical techniques.

- Quantitative Metrics for Harm: Establishing universally accepted quantitative metrics for measuring "civilian harm" in cyberspace, which can be directly incorporated into proportionality assessments. This includes economic, social, and psychological impacts.

- Legal Clarity and State Practice: Encouraging greater state practice and engagement in clarifying how IHL principles apply to specific cyber scenarios, potentially through international forums or through the development of non-binding norms.

- Mitigation Technologies: Research into technologies that allow for precise targeting and, crucially, the ability to contain or recall cyber effects if unintended consequences begin to emerge.

- Ethical Guidelines: Developing more robust ethical guidelines for the development and deployment of OCOs, specifically addressing the minimization of unintended harm to civilian populations and infrastructure.

- Interdisciplinary Collaboration: Fostering deeper collaboration between cybersecurity technologists, international law experts, political scientists, and ethicists to address the complexities of cyber collateral damage holistically.

A limitation of this systematic review is the inherent publication bias in academic literature [Randolph & Bednarik, 2008, 27], where studies with significant or positive findings might be more likely to be published. Additionally, due to the sensitive nature of offensive cyber operations, much practical information and detailed case studies may remain classified, thus limiting the scope of publicly available research. However, by adhering to rigorous SLR methodologies [Wanyama et al., 2022, 23], we aimed to provide a comprehensive overview of the publicly accessible academic and authoritative literature.

5. CONCLUSION

Collateral damage from offensive cyber operations represents one of the most critical and unresolved challenges in the contemporary digital security landscape. As nations continue to develop and employ sophisticated OCOs, the risk of unintended consequences propagating through interconnected global infrastructure remains significant. This systematic literature review has illuminated the complexities in defining and mitigating such damage, from the inherent technical difficulties of controlling effects in cyberspace to the nuanced application of international humanitarian law.

The principles of distinction and proportionality, while applicable in theory, face substantial hurdles in practical implementation due to the unique characteristics of cyber operations. Overcoming these challenges necessitates a concerted, interdisciplinary effort focusing on developing more advanced prediction and assessment methodologies, fostering greater legal clarity through state practice, and innovating technologies that allow for more precise targeting and control over cyber effects. By proactively addressing the potential for unintended harm, the international community can work towards a more responsible and stable approach to offensive cyber capabilities, ultimately safeguarding civilian populations and critical infrastructure in an increasingly digitized world.

REFERENCES

[1] Aiyer, B.; Caso, J.; Russell, P.; Sorel, M. Mckinsey: New Survey Reveals $2 Trillion Market Opportunity for Cybersecurity Technology and Service Providers. 2022. Available online: https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/cybersecurity/new-survey-reveals-2-trillion-dollar-market-opportunity-for-cybersecurity-technology-and-service-providers (accessed on 30 January 2024).

[2] IBM Security; the Ponemon Institute. Cost of a Data Breach Report 2022. 2022. Available online: https://www.ibm.com/downloads/cas/3R8N1DZJ (accessed on 31 January 2024).

[3] Hanson, F.; Uren, T. Australia's Offensive Cyber Capability. 2018. Available online: https://www.aspi.org.au/report/australias-offensive-cyber-capability (accessed on 31 January 2024).

[4] Farwell, J.P.; Rohozinski, R. Stuxnet and the Future of Cyber War. Survival 2011, 53, 23–40. [Google Scholar] [CrossRef]

[5] U.S. Department of Justice. Three North Korean Military Hackers Indicted in Wide-Ranging Scheme to Commit Cyberattacks and Financial Crimes Across the Globe. 2022. Available online: https://www.justice.gov/opa/pr/three-north-korean-military-hackers-indicted-wide-ranging-scheme-commit-cyberattacks-and (accessed on 1 February 2024).

[6] Schelling, T.C. Dispersal, deterrence, and damage. Oper. Res. 1961, 9, 363–370. [Google Scholar] [CrossRef]

[7] U.S. Air Force. Air Force Doctrine Publication 3–60, Targeting. 2021. Available online: https://www.doctrine.af.mil/Portals/61/documents/AFDP_3-60/3-60-AFDP-TARGETING.pdf (accessed on 1 February 2024).

[8] Schmitt, M.N. (Ed.) Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations; Cambridge University Press: Cambridge, UK, 2017. [Google Scholar]

[9] Arquilla, J.; Ronfeldt, D. Cyberwar is coming! Comp. Strategy 1993, 12, 141–165. [Google Scholar] [CrossRef]

[10] Lee, E.A.; Seshia, S.A. Introduction to Embedded Systems: A Cyber-Physical Systems Approach, 2nd ed.; MIT Press: Cambridge, MA, USA, 2017. [Google Scholar]

[11] Romanosky, S.; Goldman, Z. Cyber Collateral Damage. Procedia Comput. Sci. 2016, 95, 10–17. [Google Scholar] [CrossRef]

[12] U.S. Air Force. Intelligence Targeting Guide, Attachment 7. 1998. Available online: https://irp.fas.org/doddir/usaf/afpam14-210/part20.htm (accessed on 1 February 2024).

[13] Romanosky, S.; Goldman, Z. Understanding Cyber Collateral Damage. J. Natl. Secur. Law Policy 2017, 9, 233–257. [Google Scholar]

[14] Dinstein, Y. The Principle of Distinction and Cyber War in International Armed Conflicts. J. Confl. Secur. Law 2012, 17, 261–277. [Google Scholar] [CrossRef]

[15] Ablon, L.; Binnendijk, A.; Hodgson, Q.E.; Lilly, B.; Romanosky, S.; Senty, D.; Thompson, J.A. Operationalizing Cyberspace as a Military Domain, Perspective, RAND Corporation 2019. Available online: https://www.rand.org/content/dam/rand/pubs/perspectives/PE300/PE329/RAND_PE329.pdf (accessed on 30 January 2024).

[16] U.S. Department of Defense. Department of Defense Law of War Manual; William S. Hein & Company: Getzville, NY, USA, 2023.

[17] Maathuis, C.; Pieters, W.; Van den Berg, J. Assessment Methodology for Collateral Damage and Military (Dis)Advantage in Cyber Operations. In Proceedings of the MILCOM 2018—2018 IEEE Military Communications Conference (MILCOM), Los Angeles, CA, USA, 29–31 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–6.

[Google Scholar] [CrossRef]

[18] Maathuis, C.; Pieters, W.; van den Berg, J. Decision support model for effects estimation and proportionality assessment for targeting in cyber operations. Def. Technol. 2021, 17, 352–374. [Google Scholar] [CrossRef]

[19] Grant, T. Building an Ontology for Planning Attacks That Minimize Collateral Damage: Literature Survey. In Proceedings of the 14th International Conference on Cyber Warfare & Security (ICCWS 2019), Stellenbosch, South Africa, 28 February–1 March 2019; pp. 78–86. [Google Scholar]

[20] University of York Centre for Reviews and Dissemination. What Are the Criteria for the Inclusion of Reviews on DARE? 2014. Available online: https://www.ncbi.nlm.nih.gov/books/NBK285222/ (accessed on 31 January 2024).