Begawan Abioso
(Volume 17, Nomor 2, December 2026)
P-ISSN: 1858-2990 I E-ISSN: 2810-0727
Open access at: https://ejournal.hukumunkris.id/index.php/abioso

The Evolution of Cyber Mercenaries: Criminal Responsibility of Non-State Actors under International Law in Cyber Conflict

[a] **Femas Adi Saputra***, [a] **Vieta Imelda Cornelis**, [a] **Dedi Wardana Nasoetion**, [a] **Muhammad Yustino Aribawa**

[a] Faculty of Law, Universitas Dr. Soetomo, Surabaya, Indonesia

Submitted: 05-06-2026 | Revised: 26-06-2026 | Accepted: 06-08-2026 | Published: 10-08-2026

Abstract

The rapid advancement of information and communication technologies in the twenty-first century has fundamentally transformed the landscape of modern warfare, giving rise to cyberspace as a new domain of armed conflict. This transformation has been accompanied by the emergence of non-state actors operating as proxies or cyber mercenaries employed by states to conduct strategic cyber operations while maintaining plausible deniability. This study aims to examine the evolution and legal status of cyber mercenaries within the framework of International Humanitarian Law and to evaluate the extent to which criminal responsibility may be imposed on these actors under the Rome Statute and the jurisdiction of the International Criminal Court (ICC). Employing a normative juridical legal research method with conceptual and statutory approaches, the study finds that the traditional definition of mercenaries under Article 47 of Additional Protocol I to the Geneva Conventions is ill-suited to the realities of cyber proxies, thereby creating a significant legal gap. Furthermore, although the Rome Statute does not explicitly regulate cyber crimes, a consequences-based approach allows cyber mercenaries to be prosecuted for war crimes when cyber operations result in physical damage or the loss of functionality of civilian infrastructure equivalent to that caused by kinetic attacks. The study concludes that existing legal instruments require reinterpretation, supported by proactive policies within the framework of international criminal justice, to ensure individual criminal accountability for increasingly destructive contemporary cyber conflicts.

Keywords: Attribution; Cyber Mercenaries; Cyber Conflict; International Criminal Law; Rome Statute.

A. Introduction

The twenty-first century, often conceptualized as the era of the Fourth Industrial Revolution and digital transformation, has brought about profound structural changes across virtually every aspect of human civilization, including significant shifts in national security doctrines, geopolitical strategies, and the conduct of modern warfare (Tampubolon, 2019). The increasing dependence of states' critical infrastructure—including banking and macroeconomic systems, electricity distribution networks, satellite communication systems, and healthcare facilities—on computer networks and the Internet of Things (IoT) has created a new spectrum of vulnerabilities that has been extensively exploited by international threat actors (Sukmawan & Setyawan, 2023). Within this historical and strategic context, warfare is no longer dominated exclusively by kinetic confrontations involving the deployment of conventional military forces across territorial borders. Instead, the battlefield has expanded into cyberspace, which is now widely recognized by military strategists and international legal scholars as the fifth domain of warfare, complementing the traditional domains of land, sea, air, and outer space (Taufiqurrohman et al., 2022). This operational

* **Corresponding author:** waliarifin22@gmail.com

paradigm shift has given rise to what is commonly referred to as cyber warfare, in which computer code, malicious software (malware), ransomware, and network intrusions are employed as strategic instruments to undermine the economic capacity, military logistics, and political stability of adversary states (Suharto & Apriyani, 2021).

One of the greatest challenges to the rule of law confronting the international legal community today is the widening gap between *das Sollen*—the normative framework encompassing the legal rules, principles, doctrines, and ideals that prescribe how the law ought to operate—and *das Sein*, namely the empirical reality of how contemporary conflicts are actually conducted. From a normative perspective (*das Sollen*), the traditional architecture of international legal instruments, including the Charter of the United Nations governing the use of force and the 1949 Geneva Conventions together with their Additional Protocols regulating International Humanitarian Law (IHL), was developed on the fundamental premise that armed conflicts are initiated and conducted exclusively by the regular armed forces of sovereign states (Thurnher, 2020). In practice (*das Sein*), however, contemporary warfare increasingly reflects an alarming trend toward asymmetric conflict, in which states delegate offensive capabilities to, and rely upon, non-state actors—including independent hackers, underground hacktivist collectives, state-sponsored transnational organized criminal groups, and even corporate entities operating as Private Military and Security Companies (PMSCs)—to carry out clandestine cyber operations (Rodio, 2024).

This trend toward the privatization of warfare has given rise to a new category of actors commonly described in academic and professional discourse as cyber mercenaries (Mannan, 2019). The deployment of cyber mercenaries does not occur in a strategic vacuum; rather, it provides sponsoring states with substantial tactical and strategic advantages, particularly by enabling them to maintain plausible deniability. Through this form of proxy warfare, states can more readily deny their involvement in destructive cyber operations by exploiting the technical complexity of cyberspace and the anonymity inherent in the Internet's architecture, thereby complicating the attribution of responsibility under international law (Mannan, 2019). The growing disconnect between classical legal frameworks—which were developed on the assumption of physical battlefields—and the reality of digital proxies operating beyond conventional territorial and jurisdictional boundaries has created a profound accountability crisis. This challenge is further exacerbated by the fact that international legal scholarship and treaty codification have often struggled to keep pace with the rapid evolution of offensive cyber technologies (Taufiqurrohman et al., 2022).

Existing legal frameworks, including the definition of mercenaries codified in Article 47 of Additional Protocol I (1977), were specifically designed to regulate physically deployed combatants operating on conventional territorial battlefields—such as those encountered during the decolonization conflicts in Africa—rather than software specialists who launch remote cyber intrusions from jurisdictions located thousands of kilometers away from the targeted critical infrastructure (Rodio, 2024). As a consequence of this typological mismatch, contemporary cyber mercenary groups such as the Lazarus Group, which is widely believed to be affiliated with North Korea's intelligence apparatus, and Sandworm, which has been closely linked to Russia's Main Intelligence Directorate (GRU), frequently operate within the international legal grey zone, where no clear or binding framework of legal accountability exists (Antoniuk, 2024). When cyber operations orchestrated by these groups succeed in disabling critical civilian infrastructure—as demonstrated by the nationwide power outages in Ukraine and the destructive cyberattack against the country's largest telecommunications provider, Kyivstar—serious jurisdictional uncertainty arises. The international community

continues to debate whether such conduct should be characterized as ordinary domestic cybercrime, cross-border cyber espionage, or whether its scale and consequences exceed the threshold of armed violence, thereby warranting classification as war crimes subject to the regime of international criminal law (Abdullah, 2022).

This study is particularly important in addressing the international community's urgent need for a definitive legal accountability framework capable of ending the prevailing impunity of non-state actors in cyberspace. To clarify the position and contribution of this study, it is necessary to undertake a comparative mapping of the existing literature. Most previous studies have focused predominantly on the dimension of state responsibility and the geopolitical dynamics of proxy warfare (Biller & Maurer, 2018), for example, concentrate on how states delegate coercive capabilities to private actors in order to avoid direct military confrontation. Similarly (Jensen, 2017), analysis of the Tallinn Manual 2.0, which represents an important effort to systematize the application of international law to cyberspace, remains largely state-centric and is constrained by the persistent difficulties associated with attributing cyber operations to sponsoring states. Meanwhile, studies addressing the jurisdiction of the International Criminal Court (ICC) remain comparatively limited. For instance (Trahan, 2025), confines the analysis specifically to the challenges surrounding the application of the crime of aggression. (Al-Bayati, 2024) examines the threat posed by cyber mercenaries to national security without providing a sufficiently developed framework for resolving questions of international criminal jurisdiction. This review therefore reveals a clear research gap: existing scholarship has yet to comprehensively connect the limitations of International Humanitarian Law in addressing cyber mercenaries with the corresponding gaps in international criminal prosecution and individual criminal accountability for hackers involved in cyber conflicts.

To address this gap in the literature, the novelty of this study can be articulated at both the conceptual and practical levels. Conceptually, this study proposes a new doctrinal nexus. Rather than centering on the often-inconclusive debate over state responsibility and the attribution of cyber operations, it examines the incompatibility between the legal status of cyber mercenaries under International Humanitarian Law, particularly Article 47 of Additional Protocol I, and the existing legal framework governing such actors. It further connects this doctrinal deficiency to the potential application of individual criminal responsibility under the 1998 Rome Statute. Practically, this study develops a prospective and pragmatic framework for prosecution before international criminal justice institutions. Drawing upon recent prosecutorial developments at the International Criminal Court (ICC), it advances forensic and legal arguments for holding hackers, proxy operators, and malware architects individually accountable for war crimes through a consequences-based approach, thereby circumventing the protective shield of plausible deniability relied upon by sponsoring states.

Against this background, this study seeks to address two interrelated questions. First, it examines the conceptual evolution and legal status of cyber mercenaries as non-state proxy actors, together with the attribution challenges arising under International Humanitarian Law and the doctrine of state responsibility. Second, it analyzes the prospects for, and doctrinal challenges associated with, imposing individual criminal responsibility on cyber mercenaries for alleged war crimes under the jurisdictional framework of the 1998 Rome Statute.

Accordingly, this study primarily aims to critically examine the historical evolution, ambiguous legal status, and attribution controversies surrounding cyber mercenaries under

the framework of International Humanitarian Law and the principles of state responsibility. It further seeks to doctrinally assess the extent to which destructive cyber operations may be classified as core international crimes under the 1998 Rome Statute, with particular reference to recent developments in enforcement strategy and prosecutorial policy associated with the ICC Prosecutor, Karim Khan. These developments have marked a significant shift toward bringing cyber-enabled conduct within the investigative scope of the Court where such conduct is directly connected to war crimes or crimes against humanity (Verbruggen, 2023). By examining the constituent elements of mens rea and actus reus in the digital domain, this study seeks to provide a robust academic foundation for bridging and addressing the existing legal gap concerning the individual criminal responsibility of non-state actors in contemporary asymmetric warfare.

B. Research Methodology

This study employs normative legal research, understood as a method of legal inquiry that uses systematic legal reasoning to identify coherent legal rules, principles, and doctrines for addressing the legal issues under examination (Marzuki, 2021). In the context of this study, the central issue concerns the criminal responsibility of cyber mercenaries in cyber conflicts. The primary objects of analysis comprise international legal norms governing the law of armed conflict (LOAC), international human rights instruments, doctrines of international criminal law, and the legal relationship between states as principals and hackers operating as proxies.

The study relies exclusively on library research and draws upon secondary data classified into primary, secondary, and tertiary legal materials. The primary legal materials consist of binding international legal instruments, including the 1949 Geneva Conventions, Additional Protocol I of 1977, the 1998 Rome Statute of the International Criminal Court, and the 2001 Articles on Responsibility of States for Internationally Wrongful Acts (ARSIWA), together with relevant draft resolutions and other official international legal documents. The secondary legal materials comprise contemporary scholarly literature, predominantly from international sources, including the Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations, reports of the United Nations Open-Ended Working Group (OEWG), articles published in reputable international academic journals, and strategic policy documents issued by the International Criminal Court (ICC) during the 2023–2026 period.

The analytical framework adopted in this study combines a statutory approach, a conceptual approach, and a case approach. The statutory approach is employed to examine the hierarchy, scope, and textual interpretation of relevant international conventions. The conceptual approach is used to explore the doctrines of attribution and technological neutrality, while the case approach focuses specifically on major cyber operations affecting critical infrastructure, including the attack on Kyivstar in Ukraine and the deployment of NotPetya. The materials collected are subsequently reduced, synthesized, and analyzed using a descriptive qualitative method, followed by deductive legal reasoning. This method is intended to construct a coherent, logical, and prescriptive legal argument concerning the evolution of the legal status of cyber mercenaries and the prospects for enforcing international law against non-state actors operating in cyberspace.

Given that written international legal instruments do not yet explicitly regulate cyberspace in a comprehensive manner, thereby giving rise to a normative lacuna, this study applies three methods of legal interpretation in a gradual and integrated manner. The analysis

begins with literal interpretation to examine the strict textual meaning of traditional legal terminology, including the term "mercenary" under Article 47 of Additional Protocol I (AP I) of 1977 and the term "attack," defined as "acts of violence," under Article 49 of AP I. This literal interpretation provides the initial basis for identifying the doctrinal incompatibility that arises when conventional laws of war are applied to the non-kinetic characteristics of digital operations.

To address these limitations, the analysis proceeds through systemic interpretation, which is used to reconstruct the structural relationship between the law of armed conflict, or International Humanitarian Law, and International Criminal Law. Under this approach, the inability of cyber mercenaries to satisfy the conventional legal requirements for combatant or mercenary status under International Humanitarian Law is not treated as a juridical dead end. Rather, it is connected to the regime of individual criminal responsibility under the 1998 Rome Statute within the broader framework of the unity of the international legal system.

Finally, teleological and evolutive interpretation serves as the principal analytical lens for construing the concepts of "acts of violence" and "civilian objects" dynamically in the digital age. By interpreting these concepts in light of the fundamental purposes of humanitarian protection and the prevention of impunity, this approach provides the normative basis for applying a consequences-based approach, under which the loss of functionality of software or digital systems may be treated as legally equivalent to physical destruction or damage in the kinetic domain.

All legal materials and data structured through the foregoing methods of interpretation are subsequently analyzed qualitatively through a normative legal analysis grounded in deductive reasoning and legal syllogism. The process of deriving prescriptive conclusions is structured around three interrelated components of legal syllogistic reasoning. The major premise establishes the applicable theoretical framework and abstract legal norms, including the elements of war crimes under Article 8 of the 1998 Rome Statute, the doctrine of technological neutrality, and the relevant parameters of command responsibility and *mens rea*, particularly direct intent (*dolus directus*).

The minor premise incorporates the legally relevant facts and operational characteristics of cyber mercenary activities, including digital forensic evidence relating to the reconnaissance phase, the deployment of wiper malware, the authorization and execution of malicious code, and the concrete consequences of such operations, including the large-scale disruption of civilian communications infrastructure in the Kyivstar cyberattack in Ukraine and the NotPetya incident. By testing these digital facts against the abstract legal norms identified in the major premise, the analysis generates a prescriptive legal conclusion.

The resulting conclusion assesses the jurisdictional basis for international criminal accountability, identifies potential prosecutorial pathways, and examines the doctrinal challenges involved in bringing individual non-state proxy actors within the jurisdiction of the International Criminal Court (ICC).

C. Results

Investigations into destructive cyber operations that disable critical infrastructure are no longer confined to the theoretical domain; they have increasingly entered the realm of practical international criminal enforcement. The policy direction adopted by the Office of the Prosecutor (OTP) of the International Criminal Court (ICC) under Karim Khan during

the 2023–2026 period reflects an important development in the Court's approach to the scope of its ratione materiae jurisdiction. The OTP has indicated that criminal conduct carried out or facilitated through cyberspace may fall within the jurisdiction of the 1998 Rome Statute where its consequences satisfy the applicable gravity threshold and the constituent elements of crimes within the Court's jurisdiction, particularly war crimes under Article 8.

To provide a comprehensive understanding of the framework for the international criminal prosecution of cyber mercenaries and other cyber proxy groups, the analysis focuses on two principal dimensions. The first concerns the doctrinal transition reflected in the OTP Policy Paper and its adoption of the principle of technological neutrality.

Under the prosecutorial approach advanced by the Office of the Prosecutor (OTP), criminal liability is determined not by the technological means or medium employed—such as software or malicious code—but by the consequences of the operation and the perpetrator's *mens rea*. This consequences-based approach substantially reduces the normative gap arising from the absence of an explicit reference to cyberspace in the 1998 Rome Statute. Accordingly, cyber operations that result in the permanent or long-term loss of functionality of critical civilian infrastructure may, depending on their consequences, be treated as legally equivalent to kinetic attacks for the purposes of international criminal law.

The large-scale disruption of Kyivstar, Ukraine's major telecommunications provider, in December 2023, reportedly carried out by the Sandworm cyber group and associated actors linked to Russia's GRU, provides an important case study for examining the evidentiary construction of international criminal responsibility in the cyber domain. The submission of investigative materials to the International Criminal Court (ICC) through an Article 15 communication by the Security Service of Ukraine (SBU), together with the Human Rights Center at the University of California, Berkeley, illustrates how digital forensic evidence may be used to establish the constituent elements of international crimes arising from cyber operations.

With respect to *actus reus*, the operation involved the infiltration of internal systems followed by the coordinated deployment of destructive wiper malware, resulting in the disabling of Kyivstar's core civilian communications network. The consequences reportedly extended beyond the telecommunications system itself, producing wider reverberating effects that disrupted air-raid warning systems, banking services, and access to emergency medical communications for millions of civilians. From the perspective adopted in this study, such cascading consequences may support the characterization of the operation as an attack against civilian objects within the meaning of Article 8(2) (b) (ii) of the 1998 Rome Statute, provided that the remaining contextual and material elements of the offence are satisfied.

The construction of *mens rea* rests primarily on digital forensic indicators demonstrating the deliberate and sustained nature of the operation. Evidence of prolonged unauthorized access to Kyivstar's internal systems during the reconnaissance phase, the selection and deployment of specifically destructive malware, and the preparation of carefully designed execution scripts may collectively support an inference of direct intent (*dolus directus*). In the digital forensic context, source code, intrusion logs, malware configurations, command histories, and other technical artifacts may provide important evidence from which the perpetrator's subjective intent can be inferred, particularly when assessing whether the resulting disruption was deliberately engineered rather than caused by accidental malfunction.

The analysis of the Kyivstar case therefore demonstrates how the plausible deniability traditionally relied upon by cyber mercenaries and state-sponsored proxy actors may be challenged through the combined use of sophisticated digital forensic investigation and the emerging application of individual criminal responsibility within the jurisdictional framework of the ICC. In this respect, forensic attribution does not merely assist in identifying the technical source of a cyber operation; it may also provide the evidentiary bridge necessary to connect particular individuals to the material and mental elements of crimes falling within the Rome Statute.

D. Discussion

D.1 Conceptual Dynamics, the Evolution of Legal Status, and the Attribution Challenges of Non-State Actors under International Law

The transformation of the battlefield from a territorially defined physical domain measured in geographic terms to the logical infrastructure and information layers of cyberspace has fundamentally disrupted one of the defining foundations of the modern state: the monopoly on the legitimate use of physical force. In the digital era, non-state actors have gained increasingly accessible and widespread access to commercial cyber capabilities whose destructive potential may equal or even exceed that of certain conventional military capabilities. In previous decades, comparable offensive tools were largely concentrated within the intelligence and security apparatuses of major powers (Hakmeh et al., 2026).

Tim Maurer argues that, in pursuing geopolitical influence in cyberspace, states have increasingly adopted an entrepreneurial approach to the exercise of power. Rather than relying exclusively on their formal security institutions, states may deliberately sponsor, train, mobilize, and exploit independent hackers as cyber proxies through which they project coercive power beyond their territorial borders (Maurer, 2018). The growing reliance on such proxies has produced a broad and increasingly complex ecosystem of non-state cyber actors. This spectrum ranges from conventional private security contractors that have expanded their services into hack-for-hire operations to underground hacktivist collectives, state-sponsored transnational cybercrime groups and Advanced Persistent Threat (APT) actors, as well as cyber mercenaries whose activities are primarily driven by the pursuit of financial gain (Rodio, 2024).

Historically, the terminology and conceptualization of mercenaries in international law developed significantly in the aftermath of the Congo Crisis and the Katanga secessionist movement in the early 1960s. The involvement of foreign mercenaries in these conflicts generated widespread international condemnation, particularly because their activities were frequently associated with brutality and serious human rights violations. These developments ultimately prompted a series of United Nations Security Council resolutions and initiatives by the Organization of African Unity (OAU) aimed at criminalizing and eliminating mercenary activities (IHL Databases, 1987).

International Humanitarian Law (IHL) responded to these developments through the adoption of Article 47(2) of Additional Protocol I (AP I) of 1977, which establishes a legal definition of the persons who may be classified as mercenaries. However, applying and interpreting this nearly half-century-old provision in the context of contemporary cyber mercenaries reveals a fundamental and increasingly problematic legal incompatibility. Under the legal framework established by Article 47 of AP I, an individual may be classified as a

mercenary and denied the entitlement to combatant or prisoner-of-war status only if all six cumulative criteria are satisfied simultaneously (Rodio, 2024).

These cumulative requirements have consistently been regarded by international legal scholars and security experts, as well as reflected in the work of the United Nations Working Group on the Use of Mercenaries, as imposing an exceptionally demanding evidentiary threshold that is difficult to satisfy in practice. The problem becomes even more pronounced when these traditional criteria are transposed to cyber operations, where physical presence on the battlefield, nationality, formal recruitment, direct participation, and financial motivation may assume forms substantially different from those contemplated when Article 47 was originally drafted (Rodio, 2024).

To illustrate this legal adaptation problem more systematically, Table 1 provides a comparative analysis of the application of the six criteria under Article 47 of AP I to the traditional model of physically deployed mercenaries and the structural obstacles that arise when those same criteria are applied to contemporary cyber proxies or cyber mercenaries.

Table 1. Critical Analysis of the Obstacles to Applying the Mercenary Criteria under Article 47 of Additional Protocol I to Cyber Operations

Article 47 AP I Criterion (Cumulative Requirement)	Application to Conventional/Physical Mercenaries	Legal and Technical Obstacles in Applying the Criterion to Cyber Mercenaries
a) Specially recruited locally or abroad in order to fight in an armed conflict	Relatively straightforward to establish through contractual documents, recruitment records, intermediary arrangements, or corporate records of Private Military and Security Companies (PMSCs).	Cyber recruitment frequently occurs anonymously through closed online forums, dark-web platforms, intermediaries, or cryptocurrency-based transactions. These practices may obscure the recruitment chain and make it difficult to establish who recruited the operator, for what purpose, and in connection with which armed conflict.
b) Takes a direct part in the hostilities	Physical participation is generally more readily identifiable where an individual carries weapons, joins combat units, or operates directly on the front lines of a territorial armed conflict.	Cyber operations are commonly conducted remotely, with the operator physically located outside the conventional conflict zone. The use of distributed infrastructure, intermediary servers, compromised systems, and cross-border routing may complicate the legal determination of whether a particular cyber activity constitutes direct participation in hostilities.
c) Is motivated to take part in the hostilities essentially by the desire for private gain and is promised	The requirement may be assessed through salary records, contractual remuneration, or other evidence showing that the individual receives	The motives of cyber operators are frequently mixed. Financial incentives may coexist with nationalism, political ideology, hacktivism, personal loyalty, or

	material compensation substantially in excess of that paid or promised to combatants of similar rank and function	compensation materially exceeding that of comparable members of the regular armed forces.	pressure from state security institutions. Such overlapping motivations make it difficult to establish that the operator was motivated essentially by the desire for private gain, as required by Article 47.
d)	Is neither a national of a Party to the conflict nor a resident of territory controlled by a Party to the conflict	The criterion traditionally distinguishes foreign mercenaries from nationals or residents directly connected to the parties to an armed conflict.	Cyber proxies are frequently recruited, tolerated, or supported by their own governments. Where an operator is a national of the sponsoring state and that state is a Party to the conflict, the nationality requirement may prevent classification as a mercenary under Article 47, irrespective of the operator's functional role.
e)	Is not a member of the armed forces of a Party to the conflict	The criterion is generally applicable to genuinely private actors whose command structures remain institutionally separate from the regular armed forces.	In cyber operations, the distinction between private actors and state personnel may be blurred by covert functional integration, informal coordination, intelligence relationships, or personnel moving between governmental and private roles. Groups publicly presented as independent may therefore maintain operational links with military or intelligence structures, complicating their legal classification.
f)	Has not been sent by a State which is not a Party to the conflict on official duty as a member of its armed forces	Documentary evidence such as deployment orders, official mandates, military records, or diplomatic documentation may assist in determining whether an individual was officially dispatched by a third state.	Cyber proxy arrangements are often deliberately structured to avoid formal state authorization. Sponsoring states may rely on informal relationships, covert tasking, or intermediaries precisely to preserve plausible deniability, making it difficult to determine whether the cyber operator was acting independently or pursuant to an official state mission.

Source: Author's synthesis based on Additional Protocol I to the Geneva Conventions, the Tallinn Manual 2.0, and reports of the UN Working Group on the Use of Mercenaries, 2026.

Based on the doctrinal comparison presented in the table, it becomes evident that contemporary international law continues to rely on a restrictive terminology that is

increasingly disconnected from technological developments. Because the requirements under Article 47 of Additional Protocol I are cumulative, failure to satisfy even one element—such as criterion (d), where the hacker is a national of the Party conducting the relevant military operation—precludes the individual from being legally classified as a mercenary (Rodio, 2024).

Where a cyber actor does not qualify as a combatant or mercenary, the binary structure of International Humanitarian Law generally places that individual within the category of civilians under the civilian–combatant distinction (Rodio, 2024). This default classification creates an operational paradox within the law of armed conflict. As civilians, hackers affiliated with military proxy networks are, in principle, protected against direct attack unless and for such time as they directly participate in hostilities (DPH) (Wilmshurst et al., 2026).

To address this gap, contemporary interpretations of IHL have relied on the concept of continuous combat function (CCF). Under this approach, a civilian member of an organized armed group who assumes a continuous function involving the direct participation in hostilities may lose protection against direct attack for the duration of that function. Applied to cyber operations, hackers who continuously and systematically perform offensive cyber activities as part of an organized armed group supporting a Party to the conflict may therefore become lawful targets under IHL, subject to the applicable requirements of distinction, proportionality, and precautions in attack (Wilmshurst et al., 2026).

A further critical and equally complex dimension of the debate concerns attribution and state responsibility. As a general principle of international law, cyber operations conducted by individuals, corporate entities, or other non-state actors cannot automatically be attributed to a sovereign state for the purposes of international responsibility. An important exception, however, is reflected in Article 8 of the 2001 Articles on Responsibility of States for Internationally Wrongful Acts (ARSIWA), a principle also reflected in Rule 17 of the Tallinn Manual 2.0. Under Article 8, the conduct of a person or group may be attributable to a state where that person or group is, in fact, acting on the instructions of, or under the direction or control of, that state in carrying out the conduct in question (Özdemir, 2022).

The degree of control required to establish legal attribution has long been the subject of significant jurisprudential debate, particularly between the effective control and overall control standards. In Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America), the International Court of Justice (ICJ) articulated the effective control standard, which requires a particularly close connection between the state and the specific conduct alleged to constitute an internationally wrongful act (Özdemir, 2022). Under this test, evidence that a state financed a cyber proxy, supplied offensive software or exploit capabilities, organized supporting infrastructure, or provided technical training would not necessarily be sufficient, standing alone, to attribute every operation conducted by that proxy to the state. Attribution requires evidence that the state exercised the requisite direction or control over the particular operation or conduct at issue (Rodio, 2024).

This demanding attribution threshold is especially difficult to apply in cyberspace. Cyber proxies and cyber mercenaries may be afforded substantial operational discretion and autonomy, while portions of the cyber kill chain may be automated through software and algorithmic processes. At the same time, operational links between a sponsoring state and a proxy may be deliberately obscured through technical measures such as IP spoofing, layered infrastructure, compromised intermediary systems, and third-party command-and-control (C2) servers (Sakib & Nayeem, 2024). These characteristics enable states to maintain varying

degrees of strategic influence over cyber proxies while simultaneously reducing the availability of direct evidence capable of satisfying the legal threshold for attribution.

In response to the rigidity of the ICJ's approach, an alternative line of jurisprudence emerged from the International Criminal Tribunal for the former Yugoslavia (ICTY) in the landmark Tadić decision, which introduced the overall control standard (Rodio, 2024). Under this approach, the conduct of an organized armed group—or, by extension, an organized cyber proxy group—may be attributable to a state where the evidence demonstrates that the state exercised overall coordination, financing, and strategic direction over the group's activities. Unlike the effective control test, this standard does not require proof that every individual operation or tactical action was specifically instructed or directed by state officials (Rodio, 2024).

Contemporary cyber intelligence practice suggests that many Advanced Persistent Threat (APT) operations exhibit characteristics more consistent with this model of overall control. Frequently cited examples include the Sandworm group, which has been associated with the destructive NotPetya malware campaign targeting Ukraine, and the Lazarus Group, which has been linked to the cyberattack against the Bangladesh Bank and the large-scale theft of financial assets allegedly used to support North Korea's nuclear and ballistic missile programs (Hakmeh et al., 2026). These operations appear to involve sustained strategic coordination with state institutions while allowing participating cyber operators considerable operational autonomy, rather than requiring direct tactical supervision over each individual command or keystroke.

Nevertheless, much of the existing framework governing state responsibility, including the interpretation of Article 8 of ARSIWA by reference to the ICJ's effective control jurisprudence, continues to impose a demanding evidentiary threshold for legal attribution. As many commentators have observed, this threshold is particularly difficult to satisfy in cyberspace, where digital forensic evidence is often fragmented, anonymized, or deliberately concealed (Özdemir, 2022). Consequently, sponsoring states may exploit these evidentiary limitations to deny legal responsibility while avoiding the potential consequences of state responsibility under international law, including countermeasures, acts of retorsion, and coordinated multilateral sanctions.

The persistent difficulties associated with attribution at the state-to-state level raise a broader concern: the existing international legal framework may, in practice, create incentives for the continued privatization and proxy-based conduct of cyber conflict by allowing sponsoring states to benefit from uncertainty surrounding legal attribution. This accountability gap has therefore strengthened the case for shifting the focus of international legal enforcement toward an alternative and more direct level of responsibility. Rather than relying exclusively on the attribution of cyber operations to states, greater attention may be directed toward identifying and prosecuting the individual actors who plan, authorize, facilitate, or execute unlawful cyber operations through proxy structures. In this respect, the framework of international criminal law, particularly the jurisdictional mechanisms of the International Criminal Court (ICC), provides a potential avenue for pursuing individual criminal responsibility independently of the unresolved question of state attribution.

D.2 Paradigmatic Shift and the Implementation of Individual Criminal Responsibility under the Jurisdiction of the Rome Statute

Given the substantial structural obstacles to pursuing state responsibility, particularly the evidentiary difficulties associated with establishing effective control and the limited functional applicability of the mercenary definition under Additional Protocol I of 1977, the framework of individual criminal responsibility before the International Criminal Court (ICC) offers a potentially more direct avenue for accountability (Rodio, 2024). The 1998 Rome Statute, as the constitutive treaty establishing the ICC, was designed to combat impunity for the most serious crimes of concern to the international community as a whole. Article 25 of the Rome Statute provides the central legal basis for this individualized model of accountability by establishing the Court's jurisdiction over natural persons.

This legal architecture creates an important pathway for the direct prosecution of cyber mercenaries, contract hackers, malware developers operating within corporate structures, and commanders responsible for cyber operations, provided that their conduct satisfies the applicable jurisdictional and substantive requirements of the Rome Statute. Crucially, the determination of an individual's criminal responsibility does not necessarily depend upon the prior attribution of the underlying cyber operation to a sponsoring state. Accordingly, the ICC framework may enable accountability to focus directly on the individuals who plan, order, facilitate, or execute conduct amounting to crimes within the Court's jurisdiction, notwithstanding unresolved difficulties in establishing state responsibility (Wilmshurst et al., 2026).

A principal objection frequently advanced by skeptics is that the Rome Statute contains no explicit reference to terms such as "cyber attacks," "cyberspace," or "digital infrastructure." Nevertheless, contemporary scholars of international criminal law argue that criminal law is fundamentally guided by the principle of technological neutrality, under which criminal liability depends on the consequences of the prohibited conduct rather than the technological medium through which it is carried out. From this perspective, the absence of cyber-specific terminology from the Rome Statute does not, in itself, preclude the application of its substantive provisions to cyber-enabled conduct.

This interpretative approach has been increasingly reflected in the policy direction of the Office of the Prosecutor (OTP) under the leadership of ICC Prosecutor Karim Khan. Through policy initiatives and public statements issued between late 2023 and the projected policy framework extending to 2026, the Prosecutor has emphasized that the digital domain is capable of producing destruction and humanitarian suffering comparable to those that the drafters of the Rome Statute sought to prevent in the aftermath of the Second World War (Verbruggen, 2023).

Consistent with this policy direction, the OTP has indicated that destructive cyber conduct may, depending on its factual circumstances and consequences, satisfy the constituent elements of crimes within the jurisdiction of the Court. This includes war crimes, crimes against humanity, genocide, and the crime of aggression, provided that the conduct meets the applicable jurisdictional requirements, including the Court's gravity threshold (Trahan, 2025).

The most direct application of international criminal law to cyber mercenaries arises in the context of war crimes under Article 8 of the Rome Statute. A central doctrinal controversy in the application of war-crimes law to the digital domain concerns the interpretation of the term "attack," whose underlying definition in International

Humanitarian Law is derived from Article 49 of Additional Protocol I of 1977 as "acts of violence against the adversary" (Wilmshurst et al., 2026). The critical question is whether a purely digital operation—such as the manipulation of internet data flows or the alteration of software instructions controlling industrial systems, including Supervisory Control and Data Acquisition (SCADA) systems—may legally constitute an act of "violence" even in the absence of an immediately observable kinetic effect.

Scholarly debate on this issue remains divided. A narrower interpretation maintains that a cyber operation qualifies as an attack only where it causes, or is reasonably expected to cause, physical damage, destruction, injury, or death. By contrast, proponents of a broader consequences-based approach, reflected in the position of the International Committee of the Red Cross (ICRC) and in parts of the expert commentary associated with the Tallinn Manual, argue that the legal characterization of a cyber operation should focus primarily on its consequences rather than on the technological means employed. Under this approach, a cyber operation that causes a significant loss of functionality of civilian infrastructure may, depending on the nature, severity, and duration of the impairment, fall within the legal concept of an attack even where no immediate physical destruction occurs (Wilmshurst et al., 2026).

To further assess the potential application of international criminal law to cyber operations conducted by non-state actors, Table 2 provides a comparative synthesis of the circumstances under which such conduct may potentially fall within the four core crimes under the Rome Statute.

Table 2. Comprehensive Mapping of the Potential Classification of Cyber Conduct under the Core-Crime Jurisdiction of the Rome Statute

Crime Category under the Rome Statute	Potential Manifestation of Cyber Conduct by Cyber Mercenaries	Applicable Threshold and Evidentiary Requirements
War Crimes (Article 8)	Directing cyber operations against protected civilian infrastructure, such as hospitals, water-treatment facilities, and electricity networks, including operations comparable to the Industroyer2 campaign and attacks affecting Ukraine's electricity grid.	The operation must satisfy the applicable elements of a war crime and possess the required nexus to an armed conflict. Under a consequences-based approach, significant and sustained loss of functionality of civilian objects may support the characterization of a cyber operation as an "attack," even in the absence of immediate physical destruction, although this interpretation remains doctrinally contested.
Crimes against Humanity (Article 7)	Conducting widespread or systematic cyber operations against civilian populations, such as attacks on medical-record systems, disruption of essential civilian services, or cyber-enabled conduct that contributes to severe	No nexus to an armed conflict is required. However, the conduct must form part of a widespread or systematic attack directed against a civilian population, with knowledge of the attack, and must satisfy the relevant underlying act under Article 7. Where applicable, the

	deprivation, starvation, serious injury, or death.	attack must also be linked to a State or organizational policy.
Genocide (Article 6)	Deliberately using cyber operations to facilitate conduct directed against a protected national, ethnical, racial, or religious group, for example by disabling life-sustaining systems, obstructing evacuation mechanisms, or manipulating critical data in a manner connected to prohibited genocidal acts.	The principal challenge is establishing the required specific intent (*dolus specialis*) to destroy, in whole or in part, a protected group as such. Digital forensic evidence, including source code and operational logs, may contribute to proving intent, but technical evidence alone will rarely be sufficient to establish genocidal intent.
Crime of Aggression (Article 8 bis)	Cyber operations forming part of a broader state use of armed force, particularly where a large-scale cyber operation causes consequences potentially comparable in character, gravity, and scale to conventional uses of force or accompanies a kinetic military campaign.	Liability is subject to the leadership clause. Only a person in a position effectively to exercise control over, or direct, the political or military action of a State may incur individual criminal responsibility for the crime of aggression. Operational-level private cyber mercenaries will therefore generally fall outside Article 8 bis unless they satisfy this leadership requirement.

Source: Author's synthesis based on the provisions of the Rome Statute and contemporary doctrinal debates in international criminal law, 2026.

Based on the comparative analysis presented in Table 2, the war-crimes framework under Article 8 of the Rome Statute appears to provide the most plausible and doctrinally developed pathway for pursuing individual criminal responsibility in relation to destructive cyber operations. This possibility is no longer purely theoretical; it is increasingly being tested against real-world cyber incidents, most notably the December 2023 attack on Kyivstar, one of Ukraine's largest telecommunications providers. The incident caused extensive disruption to Kyivstar's core infrastructure and has been attributed by Ukrainian authorities and cybersecurity sources to actors associated with Sandworm, a group widely linked to Russia's GRU, together with the affiliated group Solntsepek (Antoniuk, 2024).

Unlike conventional cyber intrusions primarily aimed at data theft or financial gain, the Kyivstar operation exhibited characteristics of a destructive sabotage campaign. The attackers reportedly obtained covert access to internal systems through compromised privileged credentials and subsequently deployed destructive wiper malware in a coordinated manner against critical components of the civilian telecommunications network (Institute of Mass Information, 2024). The resulting disruption affected millions of Ukrainian users and reportedly interfered with access to essential communications, including emergency services, banking-related services, and certain air-raid warning functions.

From the perspective of International Humanitarian Law and international criminal law, these cascading consequences illustrate the significance of reverberating effects in assessing cyber operations against civilian infrastructure. Even where a cyber operation does not immediately produce visible kinetic destruction, the prolonged loss of functionality of

essential civilian systems may generate humanitarian consequences comparable in practical severity to those caused by conventional attacks (Antoniuk, 2024).

The response to these large-scale cyber incidents represents a potentially significant development in the evolving landscape of international criminal justice. The Security Service of Ukraine (SBU), through its cyber division under the leadership of Illia Vitiuk, has collaborated with leading academic and research institutions, including the Human Rights Center at the University of California, Berkeley, associated with Lindsay Freeman. This collaboration has contributed to efforts to prepare an Article 15 communication, a procedural mechanism under the Rome Statute through which individuals, organizations, and other entities may submit information concerning alleged crimes to the Office of the Prosecutor (OTP) of the International Criminal Court.

The materials reportedly submitted or prepared for submission encompass evidence relating to several major cyber operations affecting Ukraine, including the attack on Kyivstar, the cyber operations against Ukraine's electricity grid in 2015 and 2016, the NotPetya incident of 2017—which reportedly caused approximately USD 10 billion in global economic losses—and the disruption of the Viasat satellite communications network (Antoniuk, 2024). Taken together, these incidents provide a factual basis for examining whether destructive cyber operations conducted in the context of an armed conflict may fall within the existing jurisdictional framework of the Rome Statute.

The significance of this initiative should not be understated. If the OTP were ultimately to pursue an investigation and subsequent prosecution on the basis of such conduct, it could contribute to the development of an unprecedented body of international criminal law concerning cyber operations. In particular, proceedings involving individuals associated with Sandworm could provide one of the earliest opportunities for the ICC to assess whether cyber operations themselves may constitute, facilitate, or form part of war crimes or other crimes within the Court's jurisdiction (Antoniuk, 2024).

The initiative undertaken by the SBU and human rights experts at UC Berkeley to bring these incidents within the analytical framework of Article 8 of the Rome Statute also highlights an important development in the evidentiary assessment of *mens rea* in cyber-related war crimes. A central argument emerging from this prosecutorial discourse is that establishing intent in destructive cyber operations—often assumed to be inherently difficult—may in certain circumstances be more susceptible to forensic reconstruction than intent in conventional kinetic attacks (Verbruggen, 2023).

In a kinetic attack affecting a civilian area, for example, the accused military command may argue that the resulting civilian harm was unintended and arose from navigational error, mechanical failure, defective targeting information, or other operational circumstances. Such explanations may complicate the assessment of whether the requisite mental element for a particular war crime was present. By contrast, destructive cyber operations frequently generate extensive and persistent digital traces that may assist investigators in reconstructing the sequence of conduct preceding the attack.

Digital forensic evidence may reveal, for instance, that cyber operators maintained unauthorized access to an internal network for an extended period during a reconnaissance phase, selected or developed malware specifically designed to disable particular systems, and prepared execution scripts to activate at a predetermined operational moment. Evidence of prolonged access to the internal architecture of telecommunications or critical-infrastructure networks, together with the deployment of purpose-built destructive malware such as

Industroyer2, may therefore provide a detailed evidentiary record of preparation, target selection, and execution (Verbruggen, 2023).

These technical features can constitute powerful circumstantial evidence from which direct intent (*dolus directus*) may be inferred when considered together with the broader operational context. Source code, malware configurations, intrusion logs, execution commands, persistence mechanisms, and target-specific scripts may substantially undermine claims that the resulting loss of functionality occurred accidentally or through unintended system malfunction. In this sense, the architecture and sequencing of a destructive cyber operation may provide investigators with a comparatively granular record from which the perpetrator's subjective intent can be assessed (Antoniuk, 2024).

Where the framework of state responsibility encounters persistent evidentiary obstacles, particularly in establishing attribution through digital forensic evidence, international legal accountability increasingly shifts toward the paradigm of individual criminal responsibility under the 1998 Rome Statute. This development has gained particular momentum through the policy direction of the Office of the Prosecutor (OTP) of the International Criminal Court (ICC) under Prosecutor Karim Khan during the 2023–2026 period, which has emphasized that cyber operations producing grave consequences for critical civilian infrastructure may fall within the Court's ratione materiae jurisdiction. The cyberattack against Ukraine's telecommunications provider Kyivstar, allegedly conducted by actors associated with Sandworm, provides an important factual context for assessing the application of the consequences-based approach to destructive cyber operations.

Nevertheless, this emerging legal framework continues to face substantial doctrinal opposition from proponents of a more restrictive interpretation of international criminal law. According to this view, Article 8 of the Rome Statute, read together with the definition of "attack" in Article 49 of Additional Protocol I, confines the concept of an attack to acts of violence traditionally understood as involving physical force or physical destruction. On this reasoning, digital data constitute intangible property, and the deletion, corruption, or encryption of data alone cannot readily be characterized as damage to a protected civilian object for the purposes of international humanitarian law.

However, this restrictive interpretation becomes less persuasive when assessed through teleological and systemic methods of treaty interpretation. The disruption of the Kyivstar network did not merely interrupt the transmission of digital data; it also reportedly impaired essential civilian functions, including elements of Ukraine's air-raid warning capability and access to critical financial and communications services. The increasing interdependence of modern critical infrastructure has therefore blurred the traditional distinction between physical and non-physical harm. From this perspective, a significant and sustained loss of functionality may, in appropriate circumstances, warrant legal treatment comparable to physical damage where its foreseeable consequences directly endanger civilian life or the operation of indispensable civilian infrastructure.

Accordingly, there is a credible legal basis for examining whether individuals responsible for designing, directing, or deploying destructive malware could incur criminal responsibility under Article 8(2) (b) (ii) or, where the factual circumstances satisfy the relevant legal elements, Article 8(2) (b) (iv) of the Rome Statute. Pursuing individual accountability in such cases is not merely a symbolic or political response to emerging forms of cyber conflict. Rather, it represents an important doctrinal avenue for addressing the accountability gap that

has enabled cyber mercenaries and other proxy cyber actors to operate behind the shield of plausible deniability traditionally relied upon by sponsoring states.

Furthermore, as technological innovation continues to advance, the integration of autonomous technologies such as artificial intelligence (AI) into offensive cyber systems creates additional challenges for the doctrine of command or superior responsibility under Article 28 of the Rome Statute. As cyber mercenaries increasingly deploy hacking architectures enhanced by machine-learning capabilities to identify vulnerabilities, conduct zero-day discovery, and autonomously propagate malware across conflict environments without requiring continuous human-in-the-loop authorization at every operational stage, the traditional structure of individual criminal responsibility becomes increasingly difficult to apply (Carpenter, 2025).

In particular, the growing degree of machine autonomy may complicate the evidentiary relationship between the subjective mental state of the person who authorizes or supervises the operation (*mens rea*) and the ultimate harmful conduct generated by an autonomous system (*actus reus*). Where an AI-enabled cyber system independently selects technical pathways, adapts its behavior, or produces consequences beyond those specifically anticipated by its human operator, defendants may seek to argue that the resulting harm fell outside the scope of their knowledge, intent, or effective control (Carpenter, 2025). Such developments may therefore create significant challenges in determining when commanders, superiors, developers, or operators possess the degree of knowledge and control required for responsibility under Article 28.

Addressing this emerging accountability gap requires international criminal law to develop interpretative criteria capable of assessing the functional consequences of cyber operations rather than relying exclusively on visible physical destruction. Existing ICC jurisprudence concerning the concept of an "attack," including judicial reasoning in cases involving physical destruction such as Ntaganda, was developed primarily in the context of conventional kinetic violence (Orr, 2024). The growing use of cyber operations against critical civilian infrastructure consequently presents the Office of the Prosecutor and, ultimately, the Court with an opportunity to consider how established legal concepts should apply to technologically mediated forms of harm.

From this perspective, an evolutive interpretation of the Rome Statute should account for serious intangible or functional effects where those effects deprive protected civilian infrastructure of its essential operational capacity. The protection afforded to civilian objects should therefore not depend solely on whether tangible hardware—such as physical servers, communications equipment, or industrial machinery—has been visibly destroyed. It should also encompass, where the applicable legal elements are satisfied, severe and sustained loss of functionality caused by the destruction, corruption, or disabling of software and digital systems upon which essential civilian services depend (Wilmshurst et al., 2026).

At the final level of analysis, the crime of aggression presents a markedly different jurisdictional structure. The legal regime established through the Kampala Amendments and incorporated into Article 8 bis of the Rome Statute contains a specific leadership clause that significantly restricts the category of persons who may incur individual criminal responsibility for this crime (Wilmshurst et al., 2026). Under this clause, responsibility for the crime of aggression is limited to a person "in a position effectively to exercise control over or to direct the political or military action of a State." The provision therefore confines liability to

individuals occupying sufficiently senior positions within the political or military decision-making structure of the state.

This hierarchical limitation has important consequences for cyber mercenaries and other non-state cyber operators. Independent hackers, hacktivist groups, private cyber contractors, and operational-level proxy actors will generally fall outside the personal scope of the crime of aggression where they merely execute technical tasks without exercising effective control over, or directing, the political or military action of a state. In this sense, the Rome Statute treats such actors primarily as operational implementers rather than as the senior political or military decision-makers contemplated by the leadership requirement (Permanent Mission of Liechtenstein to the United Nations, 2021).

By contrast, potential responsibility for the crime of aggression is directed toward those senior officials who possess the authority required by Article 8 bis, including, depending on the factual circumstances, high-ranking political leaders, senior military commanders, or other state officials capable of effectively controlling or directing state political or military action (Ambos, 2022). Where such individuals design, authorize, coordinate, or provide strategic support for cyber operations forming part of an act of aggression by a State, their conduct may potentially fall within the scope of the crime of aggression, provided that all substantive and jurisdictional requirements are satisfied.

Any such prosecution, however, remains subject to the particularly complex jurisdictional regime governing the crime of aggression. The ICC may exercise jurisdiction only where the relevant conditions established by the Kampala Amendments and the Rome Statute are fulfilled, including the applicable rules concerning ratification or acceptance of the amendments and the status of the states concerned. In addition, a referral by the United Nations Security Council acting under Chapter VII of the UN Charter may provide an alternative jurisdictional pathway under the Statute (Trahan, 2025). Accordingly, even where senior state officials exercise the requisite leadership authority over cyber operations, the Court's ability to prosecute the crime of aggression remains dependent upon these distinct and highly restrictive jurisdictional prerequisites.

E. Conclusion

Based on the foregoing analysis of the evolving legal status, attribution challenges, and doctrinal ambiguities surrounding non-state proxy actors in contemporary cyber warfare, this study concludes that the existing framework of international humanitarian law (IHL) faces significant limitations when applied to the digital battlespace. In particular, the cumulative criteria contained in Article 47 of Additional Protocol I (1977) remain firmly rooted in the paradigm of conventional kinetic warfare and therefore prove difficult to reconcile with the operational realities of cyber mercenaries. The legal classification of cyber mercenaries under Article 47 is especially problematic because the provision's cumulative requirements—including direct participation in hostilities and the traditional conception of material involvement—were formulated with physically deployed mercenary forces in mind rather than remotely operating cyber actors.

This conclusion, however, is not without controversy. A number of international legal scholars, including Schmitt and Jensen, argue that relaxing the cumulative requirements of Article 47 in order to accommodate cyber actors could undermine the principle of nullum crimen sine lege and blur the fundamental distinction between civilians and combatants. According to this restrictive interpretation, private hackers who remain outside the formal

structure of a state's armed forces cannot readily be equated with traditional mercenaries because they lack physical presence on the battlefield and do not satisfy the legal criteria established by the Protocol. Nevertheless, such an approach risks isolating the law from the operational realities of contemporary armed conflict, where malicious code may produce functional consequences comparable in gravity to those caused by conventional kinetic attacks against civilian infrastructure.

These interpretative limitations are compounded by the continuing uncertainty surrounding state responsibility under the Articles on Responsibility of States for Internationally Wrongful Acts (ARSIWA) (2001). The traditional attribution standard derived from the ICJ's Nicaragua judgment, based on effective control, requires proof that the sponsoring state exercised specific direction or control over the particular cyber operation conducted by a proxy actor. Several states adopting a more conservative position—including the approaches reflected in the cyber sovereignty policies of the United States and the United Kingdom—have argued that lowering the attribution threshold to the overall control standard articulated by the ICTY in Tadić would generate legal uncertainty and expose states to disproportionate responsibility for the conduct of non-state actors operating from their territory.

Nevertheless, this objection has been increasingly qualified by the evolution of the due diligence principle reflected in the Tallinn Manual 2.0. Although the Tallinn Manual does not constitute a legally binding instrument, it represents an influential restatement of expert opinion and reflects an emerging doctrinal trend in the interpretation of international cyber law. Under this approach, a state's failure to take reasonable measures to prevent its territory or digital infrastructure from being used by cyber mercenaries to launch harmful cyber operations against another state may itself constitute a breach of an independent international obligation. Consequently, the attribution difficulties associated with the ARSIWA framework should no longer be understood as providing sponsoring states with absolute immunity from international legal responsibility.

Against the backdrop of persistent doctrinal difficulties in attributing cyber operations to states, the jurisdictional framework of the International Criminal Court (ICC) offers an increasingly significant alternative avenue for accountability through the doctrine of individual criminal responsibility. This potential has become particularly evident in the growing attention devoted to the destructive cyber operations affecting Kyivstar and allegedly associated with Sandworm, alongside the evolving prosecutorial policy of the ICC's Office of the Prosecutor (OTP). These developments suggest that the Rome Statute possesses a degree of interpretative flexibility capable of accommodating technological change through the principle of technological neutrality.

Under a consequences-based approach, destructive cyber operations conducted by non-state cyber actors may fall within the framework of war crimes under Article 8 of the Rome Statute where their consequences satisfy the applicable legal elements. In particular, cyber operations that cause severe and sustained loss of functionality to essential civilian infrastructure may warrant treatment comparable to attacks producing physical damage where the resulting disruption generates serious humanitarian consequences. The legal characterization of such conduct should therefore not depend exclusively on the occurrence of immediate physical destruction, injury, or loss of life. Rather, the assessment should account for the functional and humanitarian consequences produced by the operation, while

remaining subject to the requisite nexus with an armed conflict, the protected status of the targeted object, and the relevant elements of *actus reus* and *mens rea*.

Important limitations nevertheless remain. In the context of the crime of aggression, the leadership clause substantially restricts individual criminal responsibility to persons capable of effectively exercising control over or directing the political or military action of a state. Operational-level cyber mercenaries and technical cyber operators will therefore generally fall outside the personal scope of Article 8 bis. At the same time, the increasing use of autonomous and AI-enabled cyber technologies may create new evidentiary challenges concerning *mens rea*, effective control, and superior responsibility. In this respect, digital forensics—including source code, intrusion logs, malware configurations, and other technical artifacts—will assume an increasingly important role in reconstructing individual conduct and determining the mental elements required under international criminal law.

As an academic and policy recommendation, the international legal community, including the Assembly of States Parties to the Rome Statute, should encourage the progressive clarification of how the concept of an "attack" applies to cyber operations. Rather than relying exclusively on formal treaty amendment as the immediate means of addressing technological change, greater emphasis may be placed on principled interpretation, state practice, prosecutorial policy, and the gradual development of jurisprudence capable of clarifying when severe functional disruption of civilian infrastructure falls within existing rules of international humanitarian and criminal law. Any such interpretative development must, however, remain consistent with the principle of legality and should not extend criminal liability beyond the permissible interpretation of the applicable legal provisions.

At the domestic level, states should likewise clarify the legal consequences of civilian participation in cyber operations connected with armed conflicts. National legal frameworks, military manuals, and judicial practice should provide clearer guidance on the circumstances in which sustained cyber activities may constitute direct participation in hostilities (DPH) or, where the relevant organizational requirements are satisfied, indicate a continuous combat function (CCF). Such clarification is necessary not only for enforcement purposes but also to provide meaningful legal notice to civilian cybersecurity professionals, independent hackers, and private contractors regarding the consequences of participating in offensive cyber operations.

Ultimately, clearer legal standards may reduce the vulnerability of civilian cyber specialists to recruitment and exploitation by state intelligence services and other conflict actors. Strengthening legal awareness, clarifying the boundaries of civilian protection, and developing credible mechanisms of individual accountability are therefore essential to addressing the increasingly blurred relationship between civilian technological expertise and participation in twenty-first-century asymmetric cyber warfare.

References

Abdullah, N. A. (2022). *Urgensi Pengaturan Cyber Espionage dalam Masa Damai Ditinjau dari Hukum Internasional* [Universitas Brawijaya]. https://repository.ub.ac.id/id/eprint/199253/

Al-Bayati, M. (2024). Regulatory Framework for the Use of Cyber Mercenaries under International Law. *Journal of Anbar University for Law and Political Sciences*, *14*(Issue: 1 part 1), 1052–1080. https://doi.org/10.37651/aujlps.2023.144035.1097

Ambos, K. (Ed.). (2022). *Rome Statute of the International Criminal Court: Article-by-Article Commentary*. Verlag C.H.BECK oHG. https://doi.org/10.17104/9783406779268

Antoniuk, D. (2024). Ukraine gathers evidence to prosecute hackers behind Kyivstar attack in Hague. *The Record*. https://therecord.media/kyivstar-cyberattack-war-crimes-prosecution-ukraine

Biller, J., & Maurer, T. (2018). Cyber Mercenaries: The State, Hackers, and Power. *Naval War College Review*, *71*(4). https://digital-commons.usnwc.edu/nwc-review/vol71/iss4/16/

Carpenter, C. (2025). Whose [Crime] is it Anyway? Adapting the Crime of Aggression to Grapple with AI and the Future of International Crimes. *Journal of International Criminal Justice*, *23*(1), 69–86. https://doi.org/10.1093/jicj/mqae055

Hakmeh, J., Moynihan, H., & Prakash, N. (2026). *Holding state-sponsored hackers and other cyber proxies to account: Lessons from Russia's war on Ukraine*. Royal Institute of International Affairs. https://doi.org/10.55317/9781784136772

IHL Databases. (1987). Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977. *International Humanitarian Law Databases*. https://ihl-databases.icrc.org/en/ihl-treaties/api-1977/article-47/commentary/1987

Institute of Mass Information. (2024). *SBU identifies the Russian hackers who attacked Kyivstar*. https://imi.org.ua/en/news/sbu-identifies-the-russian-hackers-who-attacked-kyivstar-i60440

Jensen, E. T. (2017). The Tallinn Manual 2.0: Highlights and Insights. *BYU Law Research Paper*, *17–10*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2932110

Mannan, S. H. (2019). *Projecting Power: How States Use Proxies in Cyberspace*. https://nationalsecurity.law.georgetown.edu/journal/2019/12/11/projecting-power-how-states-use-proxies-in-cyberspace/

Marzuki, P. M. (2021). *Penelitian Hukum Normatif*. Kencana Prenada Media Group.

Maurer, T. (2018). *Cyber Mercenaries: The State, Hackers, and Power* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781316422724

Orr, Z. (2024). *Cyber Attack War Crimes and Ntaganda*. https://lieber.westpoint.edu/cyber-attack-war-crimes-ntaganda/

Özdemir, J. Ö. (2022). *Attribution of cyber operations in International Law* [Lund University]. https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=9081150&fileOId=9087494

Permanent Mission of Liechtenstein to the United Nations. (2021). *The Council of Advisers' Report on the Application of the Rome Statute of the International Criminal Court to Cyberwarfare*. https://www.ila-americanbranch.org/wp-content/uploads/2022/10/The-Council-of-Advisers-Report-on-the-Application-of-the-Rome-Statute-of-the-International-Criminal-Court-to-Cyberwarfare.pdf

Rodio, C. (2024). *Regulating Military Force Series—Justice and Accountability in the Era of Modern Mercenarism*. https://lieber.westpoint.edu/justice-accountability-era-modern-mercenarism/

Sakib, M. S. A., & Nayeem, J. (2024). Cybersecurity and International Law: Defining State Responsibility for Cross-Border Cyberattacks. *Journal of Business, IT, and Social Science*, *3*(1), 10–16. https://doi.org/10.51470/BITS.2024.03.01.10

Suharto, M. A., & Apriyani, M. N. (2021). Konsep Cyber Attack, Cyber Crime, Dan Cyber Warfare Dalam Aspek Hukum Internasional. *Risalah Hukum*, 98–107. https://doi.org/10.30872/risalah.v17i2.705

Sukmawan, D. I., & Setyawan, D. P. (2023). Hacker, Fear, and Harm: Data Breaches and National Security. *Jurnal Global & Strategis*, *17*(1), 153–182. https://doi.org/10.20473/jgs.17.1.2023.153-182

Tampubolon, K. E. A. (2019). Perbedaan Cyber Attack, Cybercrime, dan Cyber Warfare. *Jurist-Diction*, *2*(2), 539. https://doi.org/10.20473/jd.v2i2.14250

Taufiqurrohman, Moch. M., Fahri, M. T., Wijaya, R. K., & Wiranata, I. G. P. (2022). Meninjau Perang Siber: Dapatkah Konvensi-Konvensi Hukum Humaniter Internasional Meninjau Fenomena Ini? *Jurnal Kawruh Abiyasa*, *1*(2), 145–165. https://doi.org/10.59301/jka.v1i2.21

Thurnher, J. S. (2020). *Mercenaries on the Battlefield: What Legal Advisors Must Know*. https://lieber.westpoint.edu/mercenaries-battlefield-legal-advisors/

Trahan, J. (2025). Cyber Operations and the Crime of Aggression. *Case Western Reserve Journal of International Law*, *57*(1). https://scholarlycommons.law.case.edu/cgi/viewcontent.cgi?article=2698&context=jil

Verbruggen, Y. (2023). *Cybercrimes under consideration by the ICC*. https://www.ibanet.org/cybercrimes-under-consideration-by-the-ICC

Wilmshurst, E., Moynihan, H., & Van Benthem, T. (2026). *Securing justice for cyber-enabled international crimes: Legal foundations and practical routes to prosecution*. Royal Institute of International Affairs. https://doi.org/10.55317/9781784136635