Is There Real Hacktivism? A Method To Distinguish False-Flag Operations From Genuine Hacktivists

Há hacktivismo real? Um Método para Distinguir Operações de Falsa Bandeira de Hacktivistas Genuínos

Rev. Bras. Est. Def. v. 12, e025020, 2025, p. 1–43
ISSN 2358-3932

EDUARDO ARTHUR IZYCKI
LUIZ GUSTAVO LAVANDOSKI DA SILVA

INTRODUCTION

Hacktivism is a phenomenon as old as the public-access internet. The term emerged as an amalgamation of the words "hacking" and "activism" in the 1990s, as part of a subculture linked to hacker communities and the internet. The perception of hacktivism is often imbued with romanticism, evoking ideals such as those in John Perry Barlow's Declaration of the Independence of Cyberspace (Barlow 2016), which extolled the internet as a space of freedom and resistance against oppression.

The early hacktivist actions were inconsequential. The most referenced examples included virtual sit-ins (denial-of-service attacks), malicious artifacts with non-aggressive messages (e.g., the "I Love You" virus), and irreverent actions (such as Worms Against Nuclear Killers — WAKN Worm) carried out through digital means (Dreyfuss 1997).

However, during the 2000s, some hacktivist actions gained notoriety. Among them were operations by the Anonymous collective and information leaks publicized by WikiLeaks. This period marked a transition in

Eduardo Arthur Izycki is a doctoral candidate in International Relations at the University of Brasília (UnB). He is a professor at IDP, Brasília, DF, Brazil. He leads a research group on Cybersecurity at the Group for Studies and Research in International Security of the Institute of International Relations at the University of Brasília — Gepsi UnB. Orcid.org/0000-0001-9514-7723. Email: eduardo.izycki@aluno.unb.br.

Luiz Gustavo Lavandoski da Silva is a PhD candidate in Global Political Economy at the Federal University of ABC (UFABC). He works as a cybersecurity analyst and researcher in the private sector in São Paulo, SP, Brazil. He is a member of the Study and Research Group on International Security at the Institute of International Relations of the University of Brasília (Gepsi-UnB) and the Cyber Defense Institute (IDCIBER).

which these groups' actions transcended mere nuisances and warranted formal reactions from governments and major private organizations affected by the campaigns.

This attention was not limited to the victims or the audience of these campaigns. State-sponsored advanced persistent threats (APT) also recognized hacktivism as an opportunity.

APT groups are a kind of apex predator in the cyber threat ecosystem due to their advanced nature (technical capabilities) and persistence (abundant resources). Records of APT activities date back to the early 2000s. However, as hacktivist actions gained notoriety in the 2010s, some APT groups saw an opportunity to simulate hacktivism.

False-flag hacktivism served to legitimize state-sponsored actions, and even when technical evidence attributed the attack to a state actor, it provided sponsoring states with a degree of plausible deniability. It falls within a broader strategy of evading attribution. As an example, state actors mimic the behaviour of cybercriminals through a technique known as living-off-the-land. By leveraging widely available tools and software already present on targeted devices, APTs disguise their identity as regular criminals.

It is worth noting that not all APT actions benefit from false-flag tactics. Conventional information-gathering operations — such as traditional and industrial espionage — as well as acts of sabotage, do not seek publicity and therefore do not benefit from hacktivist simulations. On the contrary, exposing such operations could reduce their effectiveness.

It is within this context — where APTs engage in false-flag operations — that this article is framed. The distinction between genuine hacktivism and state-actor simulations is relevant to public opinion (the audience of these actions), governments and private entities (the targets of such operations), and the epistemic community that follows this issue.

For policymakers, misidentifying false-flag operations as genuine grassroots activism can distort perceptions of public opinion and ultimately lead to ineffective responses. Such misinterpretations may result in engaging in dialogue with a rogue actor while missing the opportunity to properly attribute an ongoing information-operation campaign sponsored by a state actor. Likewise, misplaced attribution by law enforcement can compromise legitimate criminal prosecution.

The private sector also benefits from a tested methodology, as cybersecurity analysts rely on threat modelling for cyber risk management. Having clear steps to distinguish hacktivists from APTs optimizes their workflow. Academia also benefits, as most research on information operations is conducted by multidisciplinary teams from security studies and in-

ternational relations, lacking a strong technical background to distinguish hacktivists from APTs and open-source information without this kind of methodological support.

The current literature considers that nations leverage a broad range of non-state actors to conduct cyberattacks and project power in the digital domain, including criminals, hacktivists, patriot hackers, and cybermilitia (Egloff 2015; Sigholm 2016; Maurer 2018). Given the research gap on the existence of non-state actors' campaigns independent from state sponsors, this paper develops a methodology based on 24 hacktivist operations. It identifies common characteristics across the cases, highlighting indicators that differentiate genuine hacktivist actions from simulated operations.

To achieve these objectives, the article follows this structure: The first section presents a brief history of hacktivism and defines the term. Next, it examines 24 hacktivist actions, extracting common elements to establish clear indicators that distinguish authentic hacktivism from state-sponsored simulations. The third section analyses the collected data. Finally, the article concludes with findings on the proposed methodology, its limitations, and potential directions for future research.

HACKTIVISM HISTORY

The internet's architecture was not designed with security as a primary requirement but rather with resilience as its main virtue. In the early days of the internet, information remained publicly accessible because applications followed standards that did not yet incorporate security considerations (Autenrieth and Kirstädter 2000). As a result, individuals with technical skills could operate freely and even advance political agendas through offensive cyber actions. It is within this context that the concept of hacktivism emerges.

One of the earliest organized examples of this practice was the Electronic Disturbance Theatre (EDT) in the 1980s. The group used rudimentary tools, such as 'virtual sit-ins,' to overload servers and disrupt the operation of government and corporate websites (Taylor 2005).

From the 2000s onward, with the rise of global connectivity, hacktivism evolved into collective actions known as 'Ops' (Operations), involving multiple individuals. Hacktivism gained greater visibility with the emergence of decentralized groups such as Anonymous. These collectives, characterized by their informal structure, became associated with the use of the Guy Fawkes mask as a symbol (*V for Vendetta* — 2005) and the dark-mode terminal aesthetics (*The Matrix* — 1999) (Taylor 2005).

The global impact of data breaches from the 2010s served as a pivotal moment in solidifying hacktivism. The publication of the Afghanistan War Logs and Cablegate, publicized by WikiLeaks, exemplifies this shift. The disclosures relied on individuals such as Chelsea Manning and Edward Snowden. Such cases reinforced the idea that hacking could serve as a form of digital accountability, using the exposure of classified information to reveal abuses of power.

Against this backdrop, state-sponsored actors' APTs reflect the fact that hacktivist groups had gained legitimacy in their messaging. By simulating these groups, APTs could enhance the credibility of their information operations, leveraging the perception of hacktivism as an act with a 'pure' political cause rather than a maneuver within geopolitical rivalries. Gross et al also argue that hacktivism allows "foreign governments to conduct offensive cyber operations by proxy" (Gross et al. 2017).

Moreover, in an era of a growing cybersecurity industry focused on attributing cyber operations to state actors, the ability to simulate hacktivism provided an additional layer of plausible deniability for sponsoring states.

It was at this crossroads, starting in the mid-2010s, that APT groups began to take an interest in hacktivism.

HACKTIVISM AS A CYBER THREAT

The group *Cult of the Dead Cow* likely was the first to use the term 'hacktivism' — a simple combination of hacking and activism — (Samuel 2004), without any doctrinal intentions. However, this terminology remains in use today.

As the visibility of this phenomenon increased, hacktivism became a subject of academic interest throughout the 2000s. Due to academia's persistent focus on definitions, the concept gradually gained clearer contours. Almost unanimously, academic definitions begin by merging the concepts of *'hacking'* and activism as foundational elements of the term.

Denning defines hacktivism as the use of digital tools — often illegal or ambiguous — for political or social purposes (Denning 2001). Her approach emphasizes the underlying intent: advancing values such as freedom of expression, human rights, and transparency. Similar to Karagiannopoulos' emphasis on the intended social impact, asserting that hacktivism encompasses *'any use of digital technologies for political reasons.'* (Karagiannopoulos 2021; 2018) Authors like Taylor use even broader definitions, describing hacktivism as *'an umbrella term that includes a disparate range of activities'* (Taylor 2005). Goldstein also notes that hacktivism *'involves the subversive use of computers and/or computer networks to further so-*

cietal/political change' (Goldstein 2018). Similarly, Mike Milone states that *'a hacktivist, therefore, uses the same tools and techniques as a hacker but does so in order to bring attention to a larger political or social goal'* (Milone 2003). Finally, Romagna also describes hacktivism as *'the use of hacking techniques to promote a political agenda on the Internet.'* (Romagna 2020)

The most widely recognized definition comes from Gabriella Coleman, author of *Hacker, Hoaxer, Whistleblower, Spy: The Many Faces of Anonymous.* Based on her extensive ethnographic research on the Anonymous collective, the Canadian scholar frames hacktivism as the intersection of hacking — the creative use of digital technologies — and political activism to promote ideals such as freedom of expression, social justice, and resistance to power abuses (Coleman 2014a).

Coleman identifies four key elements as essential to hacktivism:

- Technological subversion (use of digital tools);
- Anonymity and collectivism (anonymity as protection and as a means of constructing a non-identitarian collective identity);
- Political motivation (rooted in an ethics of resistance against power imbalances, censorship, and surveillance);
- Performative techniques (engagement in symbolic actions). (Coleman 2014a).

The aforementioned authors based their definitions on the truthfulness of hacktivists' statements and reported facts. It means that the hacktivists' goals and intentions were accepted as legitimate because of the open-source evidence they provided during the campaigns.

A different academic strain posits that the cyber threat landscape can be blurry and the threat actor taxonomy can intermingle.

Cristano et al. (2023) question the distinction between cybercrime and hacktivism: 'Should the distinction between cyber offence and defence be questioned? Should the state's level of democracy/authoritarianism decide whether its cyber operations are justified? Should the lines between hacktivism and criminality be challenged?' (Cristiano et al. 2023). Others, such as Chertoff, highlight that many of the noble motivations behind hacktivism do not hold up under legal scrutiny, asserting that 'hacktivism is another issue that is not black and white. While the aims of some hacktivists may be debatable, their methods are often offensive and, more importantly, illegal.' (Chertoff 2017). That position is endorsed by other authors as well (Rader and Wash 2015) but few people receive explicit computer security training. Despite this lack of formal education, users regularly make many important security decisions, such as "Should I click on this potentially

shady link?" or "Should I enter my password into this form?" For these decisions, much knowledge comes from incidental and informal learning. To better understand differences in the security-related information available to users for such learning, we compared three informal sources of computer security information: news articles, web pages containing computer security advice, and stories about the experiences of friends and family. Using a Latent Dirichlet Allocation topic model, we found that security information from peers usually focuses on who conducts attacks, information containing expertise focuses instead on how attacks are conducted, and information from the news focuses on the consequences of attacks. These differences may prevent users from understanding the persistence and frequency of seemingly mundane threats (viruses, phishing. Their point is the political effect of classifying an actor as a hacktivist rather than a criminal, in a similar vein to the terrorist vs. freedom fighter dilemma, without proposing a methodology to distinguish the two categories.

Conversely, authors rooted in a state-centric approach acknowledge that non-state actors (including hacktivists) can act as a proxy for states. Egloff, Maurer, and Sigholm consider that a broad range of non-state actors are leveraged by nations to conduct cyberattacks and project power in the digital domain, including criminals, hacktivists, patriot hackers, and cybermilitias. Sigholm analyzes how and when non-state actors' objectives coincide with those of nation-states, leading to cooperation (Sigholm 2016). Egloff suggests the relationship between states and non-state actors in cyberspace can be understood through historical analogies to mercantile companies, privateers, and pirates (Egloff 2015). Maurer's work emphasizes the relationships between states and non-state actors in cyberspace, particularly focusing on how states sponsor, deploy, and exploit hackers as proxies to project power in cyberspace (Maurer 2018).

These works focus on how states may rely on non-state actors to obscure attribution and achieve political objectives through the cyber domain. Therefore, they substantially differ from the structured, comparative framework to separate authentic hacktivist activity from covert influence operations proposed by this paper.

This article adopts Coleman's definition, conceptualizing hacktivism as a voluntary phenomenon (not motivated by financial gain or coercion) in which individuals with technical skills (*hacking*) engage in political action (*activism*) aimed at a specific audience.

Voluntary participation is a crucial criterion for distinguishing genuine hacktivists. Hacktivists typically have other professional occupations (or, in the stereotype, are teenagers living with their parents), limiting their

resources. This makes genuine hacktivists less persistent and technically capable than government-backed APT groups because of their inherent constraints, as they lack expertise in all areas necessary for an operation. In contrast, state actors can hire specialists to address knowledge gaps, a point also subscribed by Egloff and Maurer (Egloff 2015; Maurer 2018). Consequently, genuine hacktivism tend to be erratic and less sophisticated, targeting victims within their offensive capabilities (low-hanging fruits) rather than the best suited to achieving their political goals.

Another key difference is that the political objective precedes the hacktivist. The individual voluntarily applies technical skills in pursuit of an existing political stance. When political motivation emerges after the formation of a hacktivist group, it becomes more likely that the operation is a false-flag attack justified by a newly adopted political narrative.

Finally, in line with Coleman's argument, hacktivist operations seek to engage an audience, and they might use "colluding accounts" in social media to achieve this (Ikwu et al. 2023). This contrasts with state-sponsored actions, which serve geopolitical agendas without targeting a specific audience.

This definition — combining voluntary initiative, technical capabilities, political motivation, and audience engagement — will be instrumental in the next section. There, the presence or absence of these defining elements will serve as criteria for distinguishing genuine hacktivism from false-flag operations.

METHODOLOGY

This article employs empirical verifiable evidence as criteria to distinguish between genuine hacktivism and state-sponsored false-flag operations, contributing to a clearer understanding of this phenomenon in the contemporary context.

This section presents 24 historical cases of hacktivist actions. Each event is concisely described, considering known characteristics of the perpetrators, the actions carried out, the political motivation, and technical details.

The 24 cases were selected from open sources based on the incidents' prominence, the existence of academic references framing the groups as hacktivists (regardless of being tagged as false-flag), and the availability of technical analyses, preferably conducted by cybersecurity firms or government CSIRTs. Traditional media outlets were deprioritized as sources due to their potential bias — whether sympathetic to or antagonistic toward the political agenda behind the incidents.

For instance, daily web defacements are consistent with hacktivists groups in Brazil, such as CyberTeam and Protowave Reloaded; however, there are no academic references or technical analyses on the groups. Thus, the selection of the 24 cases was source-oriented — availability — as well as the clear academic and industry perception that the group behaved as hacktivists.

The preference for academic references and technically focused sources introduced an involuntary bias in the cases. Given the overwhelming presence of Global North companies in the cybersecurity sector, many of these analyses focus on hacktivist actions aligned with political agendas relevant to that region, as their clientele is also concentrated there.

CASE STUDIES

Genuine hacktivism

Operation Payback

Operation Payback was a campaign launched by the Anonymous collective in 2010, initially in response to anti-piracy efforts by the Motion Picture Association of America (MPAA) and the Recording Industry Association of America (RIAA). These organizations had launched denial-of-service (DDoS) attacks against websites that facilitated file sharing via torrents, such as The Pirate Bay. In retaliation, Anonymous targeted these entities and soon expanded the campaign to a global scale (Roberts 2010).

Other targets included PayPal, Visa, and Mastercard, which had blocked donations to WikiLeaks following the Cablegate disclosures, restricting its ability to fund operations. Anonymous carried out DDoS attacks against these companies, accusing them of censorship (Sauter 2014; Coleman 2014b).

The operation employed tools such as the Low Orbit Ion Cannon (LOIC), enabling anyone with internet access to participate in the DDoS attacks, even without technical expertise. The complexity of these actions aligns with the technical sophistication expected from unstructured collectives (Olson 2013).

Operation Payback exemplified digital mobilization, with thousands of participants worldwide contributing to the takedown of targeted websites. The campaign was organized in a decentralized manner by collectives associated with Anonymous, consistent with the concept of genuine hacktivism (Streams 2012).

LulzSec

LulzSec (short for Lulz Security) was a group that gained prominence in 2011 through attacks on technology companies (such as Bethesda), media outlets, and government institutions (such as the CIA). The group's name derives from the slang term "lulz," a variation of "laughs," reflecting its irreverent approach (Leyden 2010).

LulzSec compromised Sony's systems, exposing customers' personal information, and infiltrated the American public television network PBS in response to a documentary that criticized WikiLeaks. Additionally, the group targeted the cybersecurity firm HBGary, leaking internal emails that exposed controversial practices, including the use of digital espionage tools against activists (Olson 2013; Goldstein 2018).

The complexity of their intrusion and DDoS attacks aligned with the technical capabilities expected from collectives with some skilled members. LulzSec formally disbanded after approximately 50 days of intense activity (Beraldo 2022).

LulzSec did not adhere to a unified political agenda, and its actions addressed issues such as privacy in a loosely organized manner. Nevertheless, the group's dismantling, which resulted in the arrest of several members, reinforced its status as a case of genuine hacktivism (Streams 2012).

Operation Tunisia

The Arab Spring had multiple causes, one of which was the role of social media in humanizing political oppression, exemplified by the self-immolation of Mohamed Bouazizi (Howard et al. 2011). The events in Tunisia were also influenced by the *Cablegate* disclosures, which exposed cases of government corruption (Jordan 2015).

Hacktivist groups mobilized to launch DDoS attacks against Tunisian government websites (Klein 2015) and to publish content calling for an end to oppression in the country (Howard et al. 2011). Additionally, they facilitated information leaks and provided technological support to citizens subjected to state surveillance (Coleman 2013). This engagement extended to similar actions during the Arab Spring in other countries throughout 2010 and 2011 (Coleman 2013).

Although Anonymous participants were not necessarily from Tunisia, their involvement in the operation was voluntary. It was neither orchestrated nor technically subordinated to Western actors but rather an act of solidarity with local movements (Howard et al. 2011).

Phineas Phisher

The hacktivist known as Phineas Phisher (PP) carried out attacks against private companies, including Gamma Group, the creator of FinFisher (2014), and HackingTeam (2015), the provider of Galileo (Franceschi-Bicchierai 2017).

In both cases, PP leaked internal company documents, exposing how their spyware was used by governments to repress dissidents and political activists (Marczak et al. 2015; Marczak and Scott-Railton 2016; Privacy International 2018). PP made these leaks widely available to the public (WikiLeaks 2014; 2015).

Beyond these actions, PP also targeted banks, oil companies, and other entities that, according to him, contributed to inequality and unethical practices (Coleman 2017). In his manifesto, PP outlined his motivations against institutional oppression and the structural inequalities of the capitalist world (Fisher 2014; 2016; 2017). The hacktivist — who had become a 'celebrity' in the dark web — gave interviews explaining his motivations (Fisher and CrimethInc 2018).

PP demonstrated a highly sophisticated level of operational security, yet the intrusion methods used did not suggest external sponsorship (Wes 2024).

PP is widely considered an example of authentic hacktivism, characterized by clear ideological motivations and consistently oriented actions. His capabilities align with those of an individual or a small, politically motivated group. His efforts to engage a specific audience without ties to explicit geopolitical interests further support his classification as a case of genuine hacktivism.

Belarusian Cyber Partisans

The *Belarusian Cyber Partisans* (BCP) emerged in 2020 following the Belarusian presidential elections, which were marred by allegations of fraud and violent repression. Presenting themselves as pro-democracy hacktivists, BCP has conducted cyberattacks against government targets, companies, and individuals affiliated with Alexander Lukashenko's regime (CyberPartisans 2025).

Their actions include leaking government documents, exposing the identities of security force members (Naprys 2024), and disrupting state infrastructure systems (Greenberg 2024). These targets align with BCP's stated political motivations (Antoniuk 2024).

The group claims to include former officials and dissidents from Belarusian security forces. This collaboration reinforces its credibility as a hacktivist entity and justifies its operational capabilities without external sponsorship. The involvement of insider dissidents is consistent with a local group possessing the relevant knowledge to cause greater damage.

BCP publicly shares leaked materials without restrictions, aligning with hacktivists seeking popular support rather than advancing a specific narrative.

The group does not exhibit signs of state sponsorship. While the sophistication of certain operations targeting critical infrastructure warrants further analysis, BCP appears to be a genuine hacktivist collective formed by local dissidents.

AgainstTheWest

AgainstTheWest (ATW) gained notoriety for its actions against the Chinese government on platforms such as RaidForums. The group claims to fight against authoritarian regimes and promote global transparency. ATW's operations exposed connections between private organizations and the Chinese government in surveillance activities deemed unjust (DataBreaches.Net 2022).

ATW's attacks were almost exclusively directed at China, suggesting an alignment with the strategic interests of the United States and other Global North actors (Backchannel and DeVera 2021).

The group's focus on Chinese targets coincides with a geopolitical agenda that raises suspicions of state sponsorship. Additionally, ATW sought to maximize political impact by selectively presenting leaks that reinforced negative perceptions of the Chinese government and its use of surveillance tools. On the other hand, ATW shared materials on publicly accessible platforms, maintaining transparency. The group also engaged directly with the public, a characteristic typical of genuine hacktivists.

In 2023, *Global Times*, a Chinese state-affiliated publication, reported that ATW members had been identified as French and Canadian nationals, naming one individual born in Switzerland (Siqi 2023). This evidence supports the group's authenticity, as its members appeared to be an independent collective without clear state sponsorship.

ATW exemplifies the complexities involved in assessing the authenticity of hacktivist groups. In this case, the group's technical capabilities and skill in shaping political narratives appear to have been genuine attributes of a small collective rather than the result of state support.

False Flag Operations

Syrian Electronic Army

The *Syrian Electronic Army* (SEA) is a hacktivist group that gained prominence during the Syrian civil war. It conducted attacks against Western media outlets (*New York Times*, *Washington Post*, *BBC*), humanitarian organizations (*Human Rights Watch*), and dissidents of the Syrian regime, aiming to discredit sources critical of Assad's government (Deibert 2011).

SEA presented itself as a defender of the Syrian government against "Western propaganda." However, its strong political alignment weakened the hypothesis that it was a genuine hacktivist group, even though its tactics — phishing, defacement, and DDoS — were consistent with hacktivist methods (Keller 2011).

Furthermore, as evidence of government support for SEA emerged, the group came to be regarded as an extension of Syria's information operations and digital repression efforts. Its attacks amplified state narratives in a temporally correlated manner, suggesting a controlled agenda. Forensic analyses revealed the use of state resources, including infrastructure and coordination with other government propaganda operations (Shehabat 2012).

SEA's operations from within Syrian territory, coupled with public statements from then-President Bashar al-Assad encouraging the group, strongly indicate its classification as a false-flag operation (Deibert 2011).

Shamoon

Shamoon was a cyber-campaign targeting *Aramco*, Saudi Arabia's state-owned oil company, in August 2012. The *Shamoon* malware, a wiper, overwrote data on approximately 35,000 company computers, significantly disrupting its operations. Beyond Aramco, the malware was also used against other energy sector companies in the Middle East (Bronk and Tikk-Ringas 2013).

The attack was claimed by the group *Cutting Sword of Justice*, which identified itself as hacktivists and justified the operation as a response to Saudi Arabia's *"tyrannical government"*.

However, several factors cast doubt on the hacktivist nature of the attack. Shamoon's high level of sophistication — as a destructive wiper capable of completely erasing hard drives and preventing data recovery — along with its resemblance to *Duqu* (a malware previously used

against Iran), suggested an advanced actor beyond the typical capabilities of hacktivist groups. Additionally, the use of internal Aramco credentials to spread the malware within the company's network pointed to a degree of planning that exceeded usual hacktivist operations (Great 2012).

The attack targeted critical infrastructure with strategic importance to Saudi Arabia's economy, a key regional rival of Iran. The group had no known history before the attack and did not conduct any subsequent operations that reinforced its identity as hacktivists opposing the Saudi regime (Cylance 2014).

Further analyses attributed the attack to Iranian actors, reinforcing the conclusion that *Cutting Sword of Justice* served as a front to obscure the true origin of the operation (BBC 2012; Threat Hunter Team 2018). Beckerman considered the campaign as part of a possible escalation between Iran and Saudi Arabia (Beckerman 2022).

CyberBerkut

CyberBerkut presented itself as a pro-Russia hacktivist group formed by Ukrainian separatists during the Crimean conflict (2014). The group's name referenced the Ukrainian police unit *Berkut*, which was disbanded after the Maidan Revolution, emphasizing its opposition to the Ukrainian government.

CyberBerkut operations included DDoS attacks, website defacements, and leaks of confidential documents — tactics consistent with hacktivism. However, its selection of high-profile targets, such as the German Parliament, NATO, and Ukrainian government websites, demonstrated strategic alignment with Russian geopolitical interests and a level of technical sophistication incompatible with typical hacktivist groups (RecordedFuture 2015).

Burton and Lain (2020) highlight that hacker collectives have successfully demonstrated to both the U.S. and NATO the severity and impact of their operations. According to the authors, hacktivists not only pose a cyber threat but also influence kinetic warfare by shaping the social and psychological aspects of conflict, expanding the reach and consequences of their activities (Burton and Lain 2020).

The sophistication of CyberBerkut actions — including the coordinated execution of disinformation campaigns based on *Hack-and-Leak-Operations (HLO)* — suggests assistance from Russian state actors (Hulcoop et al. 2017). The overlap of tactics and technical infrastructure with known Russian cyber-espionage groups further reinforced this hypothesis (RecordedFuture 2015; Kostyuk and Zhukov 2019).

Despite its public rhetoric suggesting spontaneity, the group functioned as an extension of Russian narratives, operating under the technical guidance of Russian APTs (Defense Intelligence Agency 2017). As a result, CyberBerkut has been classified as a false-flag operation.

CyberCaliphate

CyberCaliphate declared itself a hacktivist group aligned with the Islamic State (ISIS), conducting jihadist propaganda in cyberspace. Its operations began in 2014, targeting entities such as the French television channel *TV5Monde* (disrupting its broadcasts) and high-profile social media accounts, including those of the *United States Central Command (USCENTCOM)* (Franceschi-Bicchierai 2017).

The group's operations employed customized malware and a command-and-control (C2) infrastructure, suggesting technical capabilities that were unusually advanced for an independent jihadist group (Suiche 2017).

The techniques and infrastructure used were later linked to tools associated with *APT28*, a group connected to Russian military intelligence (GRU), reinforcing the hypothesis that CyberCaliphate operated with Russian state support (ANSSI 2015; 2017).

The selection of high-profile targets among Russia's geopolitical rivals, combined with its technical sophistication, disqualified CyberCaliphate as a case of genuine hacktivism.

Guardians of Peace

The 2014 attack on Sony was carried out by the group *Guardians of Peace (GOP)*. The GOP claimed to be acting in protest against the release of *The Interview*, a comedy depicting an assassination attempt on North Korean leader Kim Jong-un. The escalation of their actions culminated in threats of physical attacks against cinemas screening the film — an unprecedented form of intimidation in genuine hacktivism (Zetter 2014).

The incident led to the exposure of Sony's confidential data, including corporate emails, employee information, and unreleased films. The use of advanced techniques, such as customized malware, indicated a level of organization and resources uncommon for hacktivists. Additionally, the GOP had no known history of hacktivist activity prior to the attack (Novetta 2016).

The attack was later attributed to the *Lazarus Group*, a North Korean state-sponsored actor, as several of its characteristics aligned with previous campaigns linked to the group (United States Department of Justice 2018a).

The group's lack of prior activity, its motivations, and its choice of target — a film deemed offensive to the North Korean regime — reinforced its classification as a false-flag operation. Furthermore, the use of sophisticated techniques strengthened the hypothesis that the attack simulated hacktivism to legitimize its actions (Haggard and Lindsay 2015).

Yemen Cyber Army

The *Yemen Cyber Army* (YCA) emerged in 2015 as a pro-Houthi hacktivist group targeting Saudi entities (Yemen Cyber Army 2015). YCA gained notoriety following attacks on the *Al Hayat* newspaper and the Saudi Ministry of Foreign Affairs, during which they leaked classified documents (Frenkel 2015). The attacks were framed as part of the struggle against the Saudi government in support of the Houthis in the Yemeni war (Cordesman 2016).

YCA had no record of activity before 2015, when the Yemeni war had reached a critical stage. The group's sudden emergence and subsequent disappearance suggest an *ad hoc* formation for specific purposes (Franceschi-Bicchierai 2015).

The methods employed by YCA — including the extraction of large volumes of data with notable sophistication — exceeded the typical capabilities of hacktivist groups. Additionally, the domain used to publish the leaked documents (*wikisaleaks[.]com*) was linked to hosting providers previously associated with Iranian state-sponsored operations (Kausch 2017).

The leak of Saudi documents coincided with escalating regional tensions between Iran and Saudi Arabia. Furthermore, the technical similarities between Iranian APT groups and YCA, along with the group's lack of engagement with the political reality of its own country, strongly indicated a false-flag operation.

Anonymous Poland

Anonymous Poland leaked documents from the *World Anti-Doping Agency (WADA)* and the *Court of Arbitration for Sport (TAS-CAS)*, claiming to expose corruption and injustices in international sports (WADA 2016).

The group presented itself as part of the broader *Anonymous* collective, adopting its communication style and visual identity. However, Anonymous Poland had no prior history of activism or connections to traditional hacktivist agendas (FireEye 2017). The stated justification for their actions was to expose *"corruption and injustices in international sports"* (ThreatConnect 2016b).

Furthermore, the attacks attributed to the group demonstrated both sophistication and precise targeting. The coordination of document releases combined a specific narrative and timing strategy, signalling an intent to maximize the impact of the disclosures.

While the official justification centred around transparency in sports, the choice of targets, the strategic nature of the attacks, and the orchestration of the leaks strongly suggested alignment with Russia's geopolitical agenda (Rid 2017).

The U.S. government attributed the operation to Russian APT groups, linking individuals associated with these groups to the techniques used in previous cyber campaigns (United States Department of Justice 2018b).

The simultaneous release of WADA and TAS-CAS documents with Russian narratives aimed at discrediting international institutions, along with the use of advanced techniques resembling those of Russian APTs, indicated that Anonymous Poland was a false-flag operation.

Shadow Brokers

The *Shadow Brokers (SB)* presented themselves as hacktivists dissatisfied with the U.S. government, claiming to have access to hacking tools attributed to the *NSA*. These tools included zero-day exploits used against various systems. However, contradicting their purported activist motivation, SB initially attempted to auction access to these tools (Tripwire 2016).

After the failed auction, the group decided to publicly release advanced hacking tools, including *EternalRomance* and *EternalBlue*, which were later weaponized in global ransomware campaigns (*WannaCry* and *NotPetya*, respectively) (Groll 2016; Brown 2017).

The disclosure aimed to undermine the credibility of the U.S. government, portraying it as irresponsible for failing to patch critical vulnerabilities and instead exploiting them in secret. Additionally, technical analysis of the case reinforced the presence of Russian geopolitical objectives (Perlroth and Sanger 2017).

The fact that SB accessed highly sophisticated and presumably well-protected tools indicates an advanced espionage capability. These tools were reportedly obtained in 2013 but were not leaked until 2016, coinciding with the U.S. election period — demonstrating patience inconsistent with typical hacktivist groups (Schneier 2017). Moreover, the public release of these tools led to widespread ransomware attacks, which contradicted the group's initial justification for the leaks.

SB had no prior history of hacktivist activities and disappeared after the operation, further distancing them from genuine hacktivism. Ultimately, the technical ability required to breach the NSA's operational security strongly indicates that this was a false-flag operation.

Guccifer 2.0

Guccifer 2.0 became one of the most notorious cases associated with hacktivism. Claiming responsibility for leaking documents from the *Democratic National Committee (DNC)* during the U.S. presidential elections, he presented himself as a lone Romanian hacker operating through independent platforms to expose corruption within institutions (Nakashima and Harris 2018).

Although Guccifer 2.0 claimed to be Romanian, linguistic analyses revealed basic language errors, indicating that he was not a native speaker. Additionally, he had no known hacktivist history — factors that undermined his legitimacy. Another significant clue was that his activities consistently aligned with Moscow's business hours, further raising suspicions of a Russian operation (ThreatConnect 2016a).

Technical analyses showed that the servers used to operate the Guccifer 2.0 account and the malware deployed in the DNC breach shared characteristics with tools used in operations attributed to Russian APT groups, particularly *APT28* (CrowdStrike 2020).

The documents released by Guccifer 2.0 were intended to damage the Democratic Party and its presidential candidate, Hillary Clinton. Some of these documents were manipulated, and their selective disclosure — both in timing and target — reinforced the hypothesis of a coordinated influence campaign (Rid 2016).

Despite the Trump administration, the U.S. government launched legal action against Russian individuals linked to the GRU, identifying Guccifer 2.0 as part of a Russian state-orchestrated operation (United States Department of Justice 2018b; United States Department of Justice Mueller 2019).

IntrusionTruth

IntrusionTruth is a collective that uses independent platforms to disclose information about Chinese state-sponsored APTs (Intrusion Truth 2017). Active since 2017, the group described itself in an interview as *"a global network of anonymous contributors united by a common goal to expose Chinese APTs"* (Zetter 2022a).

The depth of the information they have released — including connections between individuals and Chinese government agencies — goes beyond what would typically be accessible to ordinary hacktivists relying on public sources (Cox 2018). However, the group claims that *"[readers who write in with tips and information] are a key part of our Intrusion Truth network"* (Zetter 2022b).

Their stated motivation is to counter Chinese global oppression, a geopolitical stance aligned with the interests of Global North countries (Natto Thoughts 2024).

The timing of their disclosures, which often coincides with periods of heightened tensions between China and the Global North, weakens the hypothesis that IntrusionTruth is a genuine hacktivist group. While the group exhibits traits of independent investigative work, its target selection and the sophistication of its disclosures — such as naming individuals linked to APTs — suggest that it operates as a false-flag operation.

SpiderZ

The SpiderZ group claimed responsibility for an attack against the Lebanese financial organization Al-Qard Al-Hassan (AQAH), which is associated with Hezbollah. The group had no prior record of cyber activities, raising doubts about its authenticity as a hacktivist entity.

AQAH, a credit cooperative, has been accused of serving as a financial conduit for Hezbollah to circumvent economic sanctions (Jofre 2020). SpiderZ leaked internal documents, including financial data and client records, exposing Hezbollah's financial practices (Seblani 2021).

The choice of AQAH — a high-value target for Israel — suggests that state sponsorship may have been involved in the hacktivist operation as part of efforts to weaken Hezbollah. Gaining access to internal systems, including financial databases and client records, also indicates a well-planned attack atypical for hacktivist groups (Blumenthal 2021).

Finally, SpiderZ had no prior history and did not conduct any subsequent operations. These factors suggest that the attack served to amplify Israeli narratives, potentially justifying assertive foreign policies against Hezbollah.

Lab Dookhtegan and GreenLeakers

The collectives Lab Dookhtegan and GreenLeakers were groups that leaked tools used by Iranian APTs, specifically OilRig (APT34) and MuddyWater (APT39), respectively. Both groups used Telegram as their

primary platform to disseminate leaks and conveyed a political message opposing the Iranian government.

The leaks included hacking tools, internal documentation, and sensitive information about operations attributed to Iranian APTs. Both groups presented themselves as entities aiming to expose the regime's cyber activities (Greenberg 2019).

The disclosed materials contained highly technical and operational details, requiring deep access to Iranian APT groups (Vijayan 2019). This level of access raises doubts about whether a hacktivist group could have executed these actions without external support. Additionally, neither group claimed to be composed of former APT insiders, which would have justified their access to such information. The individuals behind the groups were never identified.

The selection of high-profile targets — Iranian APT groups — and the orchestration of leaks coinciding with geopolitical tensions undermined the hypothesis that these were genuine hacktivist collectives. As a result, their activities have been classified as false-flag operations.

Meteor Express

The Meteor Express collective deployed a series of wipers starting in 2019 against critical infrastructure systems in Iran and Syria. These attacks were attributed to various groups identifying themselves as hacktivists, including *Indra*, *Predatory Sparrow*, and *Adalat Ali*.

The operations disrupted Iran's railway system in 2021, displaying digital messages instructing passengers to "call the Supreme Leader's office." Other attacks targeted automation systems in steel plants, causing significant operational disruptions.

These groups used social media accounts and Telegram to publicize their activities, including data leaks allegedly extracted from their targets, accompanied by images of the attacks. Their direct engagement with Western audiences — evidenced by posts in English — further shaped their outreach strategy (GonjeshkeDarande 2023).

Despite their hacktivist rhetoric, none of these three groups had a prior history or verifiable connections with legitimate Iranian or Syrian activist movements, weakening the hypothesis that they were genuine hacktivists.

Another striking inconsistency was *Predatory Sparrow's* Twitter posts, in which the group justified its actions as *"consistent with International Law"* (GonjeshkeDarande 2023). This concern for legal justification is entirely inconsistent with the ethos of hacktivism, making it more plausible that

state actors were framing their operations against critical infrastructure as legally compliant actions.

Additionally, the Meteor Express malware demonstrated advanced technical expertise in industrial automation systems, a level of sophistication inconsistent with the capabilities of genuine hacktivist groups.

Homeland justice

In July 2022, Albania suffered a large-scale denial-of-service attack (Halili 2022). The group *Homeland Justice (HJ)* claimed responsibility, presenting itself as a hacktivist collective.

The attack specifically targeted Albania due to its government's decision to provide shelter to *Mujahedin-e-Khalq (MEK)*, an Iranian opposition group. This unusual choice of target aligns exclusively with Iran's agenda to weaken international support for regime dissidents. Furthermore, the group's rhetoric closely mirrored Iran's official narrative, reinforcing the classification of its actions as a false-flag operation (Jenkins et al. 2022).

HJ operated with an organized structure, executing attacks in coordinated waves — an approach inconsistent with the typically decentralized nature of genuine hacktivist groups. Additionally, the group employed ransomware and advanced digital espionage tools, indicating access to state-level resources (CISA 2022a; Intelligence 2022).

The group's significant operational capabilities and its geopolitical motivations, which were entirely aligned with the Iranian government, strongly indicate a false-flag operation. Further technical analysis later linked the group to Iran's *Ministry of Intelligence and Security* (CISA 2022a).

XakNet

XakNet is a pro-Russia group that presents itself as a hacktivist collective, conducting cyberattacks and leaking data related to Ukrainian and Western organizations. Its operations are publicized through social media and *Telegram*, aligning with common hacktivist practices. *XakNet's* claims are explicitly aligned with defending Russian interests, a position the group openly endorses.

Technical analyses by cybersecurity firm *Mandiant* indicate that *XakNet's* activities are controlled by *APT28*, a group linked to Russia's *GRU* military intelligence agency (Mandiant 2022). The company observed that in one-third of the Ukrainian data leaks attributed to *XakNet*,

APT28 had carried out intrusions into the same networks within a 24-hour window before the leaks (Mandiant 2023).

While the group's technical characteristics and modus operandi resemble those of a typical hacktivist collective, the additional evidence strongly suggests that *XakNet* operates as a false-flag entity under the GRU's direction.

Killnet

Killnet is a pro-Russia group that gained notoriety during the 2022 invasion of Ukraine. The group primarily conducts DDoS attacks against government institutions and private companies in various countries.

Killnet's targets include European and Western nations that expressed support for Ukraine. In June 2022, the group claimed responsibility for DDoS attacks that disrupted internet services in Lithuania, in response to the country's decision to block the transit of certain goods to Kaliningrad (Goodin 2022b).

Killnet justifies its actions as a defence against "Western propaganda" and frequently uses public channels to recruit sympathizers and promote its activities.

The Cybersecurity and Infrastructure Security Agency (CISA) classifies Killnet as a cybercriminal group aligned with the Russian government (CISA 2022b). Mandiant recently reported an expansion of the group's technical capabilities, which could indicate state support (Mandiant 2023b).

Given its initial characteristics, the recent identification of one of its leaders, Nikolai Serafimov (known as Killmilk) (Antoniuk 2023), Killnet originally operated as a genuine hacktivist group. However, its growing alignment with Russian objectives and additional technical capabilities suggest that the group has since been co-opted by the government.

IT Army of Ukraine

The *IT Army of Ukraine* was established in February 2022, shortly after the Russian invasion, at the initiative of Ukraine's Vice Prime Minister and Minister of Digital Transformation, *Mykhailo Fedorov* (Ukrainian Ministry of Digital Transformation 2022).

The group gathers volunteers from around the world to conduct cyberattacks against Russian targets, including critical infrastructure, government websites, and private companies (Goodin 2022a). Its operations are primarily coordinated through a *Telegram* channel, where target lists

and attack instructions are publicly shared with participants (IT Army of Ukraine 2022).

The public call for participation by a Ukrainian government official and the stated motivation of defending Ukraine's sovereignty indicate that this is not an organic hacktivist group but rather a state-organized initiative (Soesanto 2022). While the public dissemination of target lists and the nature of the attacks align with typical hacktivist behaviour, the group operates under clear directives from the Ukrainian government.

DISCUSSION

The description and analysis of the 24 case studies provide a significant sample for the methodology to distinguish between genuine hacktivism and false-flag operations.

Among the 24 cases analysed, only six were classified as genuine hacktivism, representing a 25% margin. This finding suggests that the majority of hacktivist actions are not the product of authentic grass-roots groups.

Most of the analysed operations (22) resulted in Hack-and-Leak Operations (HLO), while nine involved denial-of-service (DoS) attacks, and two cases included destructive wiper malware. All genuine hacktivist actions performed HLO, with two also conducting DoS attacks, but none involved data destruction.

This indicates that the type of action alone — HLO or DoS — is insufficient as a criterion. However, the use of data-destruction techniques (wipers) suggests a higher level of sophistication, making it a plausible criterion for identifying sophistication.

Another critical consideration is the assumption that false-flag operations are solely a strategy of non-Western states, a perception influenced by the market incentives of the cybersecurity industry (Guerrero-Saade and Bartholomew 2016). While most of the documented operations originated in non-Western countries (11 cases), at least six instances of false-flag operations had motivations linked to the Global North. This proportion was likely influenced by data collection bias, as cybersecurity firms from the Global North predominantly conducted the analyses used for the analysis.

Based on this dataset, we selected four key indicators whose presence or absence serve as distinction criteria: (i) group history, (ii) targeting, (iii) sophistication level, and (iv) narrative control. Additional elements can also be considered to distinguish hacktivist operations; however, they were not systematically observed in the cases.

Group history and continuity

The first indicator whose presence is predictive of genuine hacktivist operation is the group's history and continuity. Although hacktivist groups can be loosely organized collectives, their discussions and themes tend to persist over time. Groups such as *Anonymous*, *LulzSec*, *Phineas Fisher*, *Belarusian Cyber Partisans*, and *AgainstTheWest* maintained a digital presence and remained available to justify their actions after executing them.

Some false-flag groups also participated in post-operations engagements, such as *SEA*, *IntrusionTruth*, *XakNet*, *Killnet*, and *IT Army of Ukraine*. But these are the groups that were found to have direct state support for their actions.

However, the absence of history and the disposable nature of groups' identity are key traits of false-flag operations. Several collectives with no prior history disappeared after achieving their apparent objectives, displaying the common trait of their disposable identity. Examples include *Shamoon*, *CyberBerkut*, *CyberCaliphate*, *GOP*, *YCA*, *Anonymous Poland*, *SB*, *Guccifer 2.0*, *SpiderZ*, *Lab Dookhtegan*, *GreenLeakers*, *MeteorExpress*, and *Homeland Justice*.

This strong correlation suggests that false-flag campaigns create disposable personas for a particular operation whilst genuine hacktivists carefully craft their groups' identities through different social media and direct engagements.

Target selection and operational focus

In false-flag operations, target selection is a decisive factor, with activist rhetoric following as a justification. Groups such as *Shamoon*, *CyberBerkut*, *CyberCaliphate*, *GOP*, *Anonymous Poland*, *The Shadow Brokers*, *Guccifer 2.0*, *IntrusionTruth*, *SpiderZ*, *Lab Dookhtegan*, *GreenLeakers*, *MeteorExpress*, and *Homeland Justice* focused their operations on very specific targets, directly aligned with the geopolitical interests of their likely state sponsors.

This contrasts with genuine hacktivist, which target broader categories rather than specific organizations or individuals. For example, *Phineas Fisher* discussed target selection not as an absolute mission but as a mix of opportunity and willingness to attack surveillance companies (Fisher 2014; 2016; 2017). Similarly, *BCP* engaged in internal discussions to determine viable and legitimate targets within their hacktivist motivation (CyberPartisans 2025). Similar reasoning applied to *Anonymous* and *LulzSec* regarding their choice of targets (Coleman 2014b).

In addition to target selection, orchestrated timing is an additional factor. *IntrusionTruth*, *SpiderZ*, *Lab Dookhtegan*, *GreenLeakers*, and *MeteorExpress* all focused on targets that resonated with the geopolitical competition during crisis periods, evidence that they had a long-term focus on the targets.

This contrasts with the hacktivist opportunistic nature, targeting systems loosely aligned with their political motivations. Due to their resource limitation — time and skills — hacktivists exploit generic targets that do not necessarily epitomize their political cause. For instance, this leads to huge breaches with a small portion of the data belonging to the actual targets.

Therefore, this criterion assesses how generic or broad the target was, where hacktivists are prone to conduct actions against available targets and false-flag operators tend to be highly specific to engage.

Level of sophistication

The third criteria regards the sophistication level required for the action. Although intrusion, data collection, and exfiltration might not require sophisticated technical expertise, false-flag operations target well-secured entities with multi-step operations, implying a higher level of technical expertise.

This criterion was extracted from *Shamoon*, *CyberBerkut*, *CyberCaliphate*, *Shadow Brokers*, *Guccifer 2.0*, *IntrusionTruth*, *SpiderZ*, *Lab Dookhtegan*, *GreenLeakers*, *MeteorExpress*, and *Homeland Justice*, all of which engaged in complex operations that involved several steps (phishing, persistence, and exfiltration) over several weeks.

The genuine hacktivist groups typically apply human resources to complex operations, preferring to target low-hanging fruits in low-level security environments, such as targeting personal devices or retrieving information with leaked credentials.

This criterion distinguishes the actor's sophistication based on the operations complexity, the resources leveraged, and the techniques used during execution. The hacktivists' operations are straightforward and do not require multiple steps or task specialization.

Narrative Control and Information Manipulation

The final key indicator identified regards narrative control. False-flag operations do not aim to "reveal" information in the public interest but rather to steer discussions in a particular direction. This includes selec-

tive leaking, withholding information, and data manipulation to maximize public outcry.

In the cases of *CyberBerkut*, *CyberCaliphate*, *Anonymous Poland*, *Shadow Brokers*, and *Guccifer 2.0*, leaked content was manipulated to align with a predetermined agenda rather than merely exposing classified information. Similarly, *IntrusionTruth*, *SpiderZ*, *Lab Dookhtegan*, *GreenLeakers*, *MeteorExpress*, and *Homeland Justice* executed operations aimed at discrediting specific government organizations through selective leaks.

By contrast, *Belarusian Cyber Partisans*, *Phineas Phisher*, *Operation Payback*, *Operation Tunisia*, *ATW*, and *LulzSec* exposed whole datasets to the wider public and kept engaging in post-operation discussions defending their actions without content manipulation.

These four criteria are extracted from the comparison of 24 cases, based on the hacktivist definition provided in the introduction. They form the methodological core for distinguishing genuine hacktivist groups from false-flag operations orchestrated by state actors. Other possible criteria — such as the beneficiaries of the attacks or their unintended side effects — were also considered during the research process, but proved to be unreliable or insufficient for consistently differentiating between the two. No single criterion is decisive enough to override the others, as some groups exhibit contradictory characteristics.

CONCLUSION

This article provided a historical overview of hacktivism and, based on a defined concept of hacktivism, analysed 24 operations in search of factual indicators to distinguish genuine actions from mere simulations.

Through this empirical effort, the study identified four key criteria that differentiate authentic hacktivists from state-sponsored groups: (i) the history and continuity of the groups, (ii) the targeting of offensive actions, (iii) the sophistication of the operations, and (iv) narrative control.

The criteria proposed in this article should not be understood as absolute — meaning their presence or absence of a single criterion does not lead to a definitive conclusion about whether a group is authentic. It is natural for some overlap to exist, especially since one category (false-flag) intentionally seeks to mimic the characteristics of the other (genuine hacktivists). In essence, state-sponsored actors try to appear as genuine hacktivists.

As for applying the methodology, we recommend gathering open-source information to address the first criterion (group history). Genuine hacktivist groups typically maintain a digital presence and actively seek

visibility for their actions and causes. As a result, such information should be publicly accessible to any interested party, including independent researchers and academics.

The remaining three criteria rely on information that includes technical details of the actions performed by the group, preferably based on a larger sample of operations to increase the reliability of the assessment. These criteria typically depend on analyses and reports produced by governments, cybersecurity firms, or digital forensics teams, which often provide additional technical insights into the group's methods, targets, and infrastructure. It is crucial to consult more than one source, as relying on a single report may introduce origin bias. Cross-verifying information from multiple independent sources helps mitigate this risk and supports a more balanced and accurate classification.

Assuming a spectrum between false-flag operations and genuine hacktivism, where groups may exhibit conflicting characteristics, the methodology presented is effective in classifying self-proclaimed hacktivist groups as either authentic or false-flag actors.

It is important to acknowledge the temporal contingency of this methodology, as it relies on historically determined typologies and behaviours. As state actors refine their false-flag tactics, they learn from both failures and successes, improving their ability to simulate authentic hacktivist groups. In this sense, the very methodology developed here could potentially be subverted into a manual for simulating hacktivism.

There is a possible trend for future false-flag operations with the recruitment of genuine hacktivists as part of the campaign. *XakNet*, and *Killnet* appear to be boosted by government backing (and steering) to a particular goal, in both cases the group's behaviour changed significantly in terms of targeting and sophistication. If this becomes a trend, the methodology will require adjustments as existing hacktivists personas might be incorporated in larger state-sponsored operations.

Regarding the central question of this article, it is possible to affirm that genuine hacktivism exists, but it represents a minority of high-profile cases. This is because false-flag hacktivist operations tend to be more technically capable and target high-profile entities, making them more likely to be reported in the media than genuine hacktivist actions, which are often unable to breach well-defended targets. Hence, true hacktivism might not be making the headlines, but it still exists.

This article does not take a position on the legitimacy of hacktivism. Its objective was not to justify such actions in authoritarian contexts or even in democratic settings, where channels for reporting misconduct may not always function effectively. Instead, the goal was to assist in distin-

guishing between two types of actors that are often perceived as threats by their targets, despite their differing motivations.

One potential avenue for further research is the interaction between hacktivists — both genuine and false-flag — and the entities that aim to facilitate responsible disclosures of classified information. Following in the footsteps of WikiLeaks, various groups and journalist consortia now work to verify and freely distribute information deemed to be in the public interest (e.g., *DDoSecrets* and *Enlace Hacktivista*). This article did not examine the relationship between responsible disclosure organizations and hacktivists, nor did it explore the risks of vertical control by state actors, in which APT groups conduct HLOs and use responsible disclosure entities to legitimize the leaked content.

This article successfully establishes a methodology for analysing cases of authentic hacktivism. The expectation is that this tool will be useful in regions that are not global geopolitical hotspots and do not attract the attention of major cybersecurity firms analysing HLO operations. For instance, Latin America has seen cases of alleged hacktivism, such as *EterSec*, *Vaza Jato*, and *Guacamaya*, which have not been examined through this analytical framework. Future applications of this methodology may provide a clearer understanding of these cases, as well as other under-analysed hacktivist operations worldwide.

ANNEX

Case	Historical Background and Continuity	Target Selection	Sophisticated Methods	Narrative Control
Operation Payback	Interaction through messaging platforms Group identity prior and post campaign (until members' arrest)	Opportunistic behaviour, target selection through public debate among members	Utilization of publicly available DDoS tools	Members kept supporting a particular ideological view in open platforms
LulzSec	Open public participation Interaction through messaging platforms Group identity prior and post campaign (until members' arrest)	Financial companies, but without acute target selection	Utilization of publicly available DDoS tools	Members kept supporting a particular ideological view in open platforms
Operation Tunisia	Open public participation Interaction through messaging platforms Group identity prior and post campaign	Opportunistic behaviour, targets associated with the Tunisian government	Utilization of publicly available DDoS tools and leakage of generic data from governmental targets	Members kept supporting a particular ideological view in open platforms
Syrian Electronic Army (SEA)	Based in Syria, remained active until the maintenance of Assad's regime	Group operations in alignment with the Syrian government, mostly Western targets "opposing" the Syrian regime	Use of state infrastructure for attacks, but without innovative methods employed	Biased towards Assad's regime, with selective disclosure of information
Shamoon (Cutting Sword of Justice)	The group did not exist before the campaign and did not remain active after its conclusion	Highly specific target selection (Saudi oil company)	Conducts sophisticated actions against automation systems, based on complex software (Duqu)	The group presented a manifesto following the action without subsequent efforts to steer the narrative

Case	Historical Background and Continuity	Target Selection	Sophisticated Methods	Narrative Control
CyberBerkut	The group did not exist before the campaign and did not remain active after its conclusion	Target selection oriented by alignment with the Russian government	Overlap of techniques and methods utilized by Russian APT actors	Group effort to direct the narrative in favour of the Russian government, selective data disclosure
CyberCaliphate	The group did not exist before the campaign and did not remain active after its conclusion	Target selection oriented by alignment with the Russian government	Customized malware and a command-and-control (C2), infrastructure and TTPs linked to APT28	The group did not interact on digital platforms seeking narrative control
Guardians of Peace	The group did not exist before the campaign and did not remain active after its conclusion	Target selection oriented by alignment with the North Korean government	Use of sophisticated actions and voluminous data exfiltration	Continuous public threats, and attempt to prevent film release through physical threats
Phineas Fisher	Actor remained active in managed profiles with elevated OpSec level	Selection of specific targets in digital surveillance companies	Use of sophisticated actions and voluminous data exfiltration	Free data distribution, without directing repercussions, actor debated the campaign online
Yemen Cyber Army	The group did not exist before the campaign and did not remain active after its conclusion	Target selection oriented by alignment with the Iranian government	Use of sophisticated actions and voluminous data exfiltration, infrastructure shared with APT groups	Group effort to direct the narrative against the Saudi government, selective data disclosure
Anonymous Poland	The group did not exist before the campaign and did not remain active after its conclusion	Target selection oriented by alignment with the Russian government	Use of sophisticated actions and voluminous data exfiltration, infrastructure shared with APT groups	Group effort to direct the narrative against Olympic and sports authorities, selective data disclosure
The Shadow Brokers	The group did not exist before the campaign and did not remain active after its conclusion	Selection of targets with elevated security levels and geopolitical rivals of the Russian government	Use of sophisticated actions and voluminous data exfiltration	Group effort to direct the narrative against the US government

Case	Historical Background and Continuity	Target Selection	Sophisticated Methods	Narrative Control
Guccifer 2.0	The group did not exist before the campaign and did not remain active after its conclusion	Target selection oriented by alignment with the Russian government	Shared infrastructure and identification of individuals associated with APT groups	Group effort to direct the narrative against US presidential candidates, selective data disclosure
IntrusionTruth	Actor remained active in managed profiles with elevated OpSec level	Specific target selection against the Chinese government, with significant targeting	No evidence of sophisticated actions, the data presented appears to originate from open sources	The group is active in leak disclosure, but provides complete content
SpiderZ	The group did not exist before the campaign and did not remain active after its conclusion	Target selection oriented by alignment with the Israeli government	Use of sophisticated actions and voluminous data exfiltration	The group did not interact on digital platforms seeking narrative control
Lab Dookhtegan & GreenLeakers	Actor remained active in managed profiles with elevated OpSec level Continuation of group identity following operations	Specific target selection against the Iranian government, with significant targeting	Use of sophisticated actions and voluminous data exfiltration	Group effort to direct the narrative against the Iranian government, selective data disclosure
MeteorExpress	Actor remained active in managed profiles with elevated OpSec level and did not remain active after conclusion	Target selection oriented by alignment with the Israeli government	Use of sophisticated actions and voluminous data exfiltration	Group effort to direct the narrative against the Iranian government, selective data disclosure
Belarusian Cyber Partisans	Composed of dissidents from government security forces and patriotic hackers Continuation of group identity following operations	Target selection within the Belarusian government, aiming to maximize political and strategic damage	Conducts sophisticated actions against automation systems	The group is active in leak disclosure, but provides complete content

Case	Historical Background and Continuity	Target Selection	Sophisticated Methods	Narrative Control
AgainstTheWest	Members are information security professionals from different countries Elevated level of operational security, but with continuity following operations	Specific target selection against the Chinese government, with significant targeting	Methods were not disclosed in detail, but content suggests access to sensitive information (probable significant capability)	The group presents collected information in carefully constructed narratives
Homeland Justice	The group did not exist before the campaign and did not remain active after its conclusion	Specific target selection against the Iranian government, with exclusive focus on a group of Iranian dissidents	Use of sophisticated actions and voluminous data exfiltration, infrastructure shared with APT groups	The group did not interact on digital platforms seeking narrative control
Killnet	Group existed before the campaign and remained active following operations	Target selection oriented by alignment with the Russian government	Utilization of publicly available DDoS tools, data leakage, infrastructure shared with APT groups	Group effort to direct the narrative, selective data disclosure
Xaknet	Group existed before the campaign and remained active following operations	Target selection oriented by alignment with the Russian government	Utilization of publicly available DDoS tools, data leakage, infrastructure shared with APT groups	Group effort to direct the narrative, selective data disclosure
IT Army of Ukraine	Group did not exist before the campaign (Ukraine invasion), but remains active following operations	Target selection oriented by alignment with the Ukrainian government	Use of sophisticated actions and voluminous data exfiltration, the group has formal support from the Ukrainian government	Group effort to direct the narrative, selective data disclosure

REFERENCES

ANSSI. 2015. "Attaque Informatique Contre TV5 Monde: L'ANSSI Mobilisée". cyber.gouv.fr/publications/attaque-informatique-contre-tv5-monde-lanssi-mobilisee.

ANSSI. 2017. "Retour Technique de l'incident de TV5Monde". www.sstic.org/2017/presentation/2017_cloture/.

Antoniuk, Daryna. 2023. "Report Claims to Reveal Identity of Russian Hacktivist Leader". *The Record.* therecord.media/killmilk-identity-revealed-gazeta-ru-killnet-russia.

Antoniuk, Daryna. 2024. "How the Belarusian Cyber Partisans Are Fighting a Digital War against Two Dictators". *The Record.* therecord.media/belarusian-cyber-partisans-operations-politics-russia-ukraine.

Autenrieth, A. and A. Kirstädter. 2000. "Fault Tolerance and Resilience Issues in IP-Based Networks". www.semanticscholar.org/paper/Fault-Tolerance-and-Resilience-Issues-in-IP-Based-Autenrieth-Kirst%C3%A4dter/f44807ada8b61cdbd4293b79187721137c3233da.

Backchannel and Aaron DeVera. 2021. "AgainstTheWest, the Hacking Group Wreaking Havoc on Chinese Government and Corporate Targets". Substack newsletter. *Backchannel Blog* (December). backchannel.substack.com/p/against-thewest-the-hacking-group.

Barlow, John Perry. 2016. "A Declaration of the Independence of Cyberspace". Electronic Frontier Foundation (January). www.eff.org/cyberspace-independence.

BBC. 2012. "Shamoon Virus Targets Energy Sector Infrastructure". Technology. *BBC News* (August). www.bbc.com/news/technology-19293797.

Beckerman, Carly E. 2022. "Is There a Cyber Security Dilemma?" *Journal of Cybersecurity* 8, no. 1: tyac012. doi.org/10.1093/cybsec/tyac012.

Beraldo, Davide. 2022. "Movements as Multiplicities and Contentious Branding: Lessons from the Digital Exploration of #Occupy and #Anonymous". *Information, Communication & Society* 25, no. 8: 1098–114. doi.org/10.1080/1369118X.2020.1847164.

Blumenthal, Erin. 2021. "Hezbollah's al-Qard al-Hasan and Lebanon's Banking Sector". *FDD* (May). www.fdd.org/analysis/2021/05/11/hezbollahs-al-qard-al-hasan-and-lebanons-banking-sector/.

Bronk, Christopher and Eneken Tikk-Ringas. 2013. "The Cyber Attack on Saudi Aramco". *Survival* 55, no. 2: 81–96. doi.org/10.1080/00396338.2013.784468.

Brown, Rebekah. 2017. "The Shadow Brokers Leaked Exploits Explained". *Rapid7* (April). www.rapid7.com/blog/post/2017/04/18/the-shadow-brokers-leaked--exploits-faq/.

Burton, Joe and Clare Lain. 2020. "Desecuritising Cybersecurity: Towards a Societal Approach". *Journal of Cyber Policy* 5, no. 3: 449–70. doi.org/10.1080/23738871.2020.1856903.

Chertoff, Michael. 2017. "A public policy perspective of the Dark Web". *Journal of Cyber Policy* 2, no. 1: 26–38. doi.org/10.1080/23738871.2017.1298643.

CISA. 2022a. "Iranian State Actors Conduct Cyber Operations Against the Government of Albania" (September). www.cisa.gov/news-events/cybersecurity-advisories/aa22-264a.

CISA. 2022b. "Russian State-Sponsored and Criminal Cyber Threats to Critical Infrastructure" (May). www.cisa.gov/news-events/cybersecurity-advisories/aa-22-110a.

Coleman, E. Gabriella. 2017. "The Public Interest Hack". *Limn* 8 (May): 18.

Coleman, Gabriella. 2013. "Anonymous and the Politics of Leaking". In *Beyond WikiLeaks: Implications for the Future of Communications, Journalism and Society*, edited by Benedetta Brevini, Arne Hintz, and Patrick McCurdy. Palgrave Macmillan UK. doi.org/10.1057/9781137275745_13.

Coleman, Gabriella. 2014a. *Hacker, Hoaxer, Whistleblower, Spy: The Many Faces of Anonymous*. Verso. www.versobooks.com/en-gb/products/21-hacker-hoaxer--whistleblower-spy.

Coleman, Gabriella. 2014b. *Hacker, Hoaxer, Whistleblower, Spy: The Many Faces of Anonymous*. Verso. www.versobooks.com/en-gb/products/21-hacker-hoaxer--whistleblower-spy.

Cordesman, Anthony H. 2016. *The Changing Gulf Balance and the Iranian Threat* (August). www.csis.org/analysis/changing-gulf-balance-and-iranian-threat.

Cox, Joseph. 2018. "Meet 'Intrusion Truth', the Mysterious Group Doxing Chinese Intel Hackers". *VICE* (August). www.vice.com/en/article/intrusion-truth-group-doxing-hackers-chinese-intelligence/.

Cristiano, Fabio, Xymena Kurowska, Tim Stevens, et al. 2023. "Cybersecurity and the Politics of Knowledge Production: Towards a Reflexive Practice." *Journal of Cyber Policy* 8, no. 3: 331–64. doi.org/10.1080/23738871.2023.2287687.

CrowdStrike. 2020. "CrowdStrike's Work with the Democratic National Committee: Setting the Record Straight". www.crowdstrike.com/en-us/blog/bears-midst-intrusion-democratic-national-committee/.

CyberPartisans. 2025. "Belarusian Cyber Partisans". www.by.cpartisans.org/en.

Cylance. 2014. "Operation Cleaver". www.aclu.org/sites/default/files/field_document/Cylance-Operation-Cleaver-Report-1748-1833.pdf.

DataBreaches.Net. 2022. "An Interview with AgainstTheWest". databreaches.net/2022/04/03/an-interview-with-againstthewest/.

Defense Intelligence Agency. 2017. *Russia Military Power: Building a Military to Support Great Power Aspirations*." publicintelligence.net/dia-russia-military-power-2017/.

Deibert, Ronald. 2011. "Syrian Electronic Army: Disruptive Attacks and Hyped Targets" (June). deibert.citizenlab.ca/2011/06/syrian-electronic-army-disruptive-attacks-and-hyped-targets/.

Denning, D. 2001. "Activism, Hacktivism, and Cyberterrorism: The Internet As a Tool for Influencing Foreign Policy". www.semanticscholar.org/paper/Activism,-Hacktivism,-and-Cyberterrorism:-the-As-a-Denning/829b21633c51252429abcb1ac717ecc4efc64566.

Dreyfuss, Suelette. 1997. *Underground: Tales of Hacking, Madness, and Obsession on the Electronic Frontier*. Mandarin Australia.

Egloff, Florian J. 2015. "Cybersecurity and the Age of Privateering: A Historical Analogy". www.semanticscholar.org/paper/Cybersecurity-and-the-Age-of-Privateering%3A-A-Egloff/dd21875046c939b63afa704d6b4c5c26edf80401.

FireEye. 2017. "APT28 At the Center of the Storm".

Fisher, Phineas. 2014. "Hackback — A DIY GUIDE II". *The Anarchist Library*. theanarchistlibrary.org/library/phineas-fisher-hackback-a-diy-guide-ii.

Fisher, Phineas. 2016. "Hackback — A DIY GUIDE 1". *The Anarchist Library*. theanarchistlibrary.org/library/phineas-fisher-hackback-a-diy-guide-i.

Fisher, Phineas. 2017. "Hack Back — A DIY Guide (Hacking Team)". *The Anarchist Library* theanarchistlibrary.org/library/hack-back-subcowmandante--marcos-phineas-fisher-hack-back-a-diy-guide-hacking-team.

Fisher, Phineas, and CrimethInc. 2018. "HackBack! Talking with Phineas Fisher". *The Anarchist Library.* theanarchistlibrary.org/library/crimethinc-hackback-talking-with-phineas-fisher.

Franceschi-Bicchierai, Lorenzo. 2015. "There's Evidence the "Yemen Cyber Army" Is Actually Iranian". *VICE* (June). www.vice.com/en/article/theres-evidence-the-yemen-cyber-army-is-actually-iranian/.

Franceschi-Bicchierai, Lorenzo. 2017. "The Hack That Caused a Crisis in the Middle East Was Easy". *VICE* (June). www.vice.com/en/article/the-hack-that--caused-a-crisis-in-the-middle-east-was-easy/.

Frenkel, Sheera. 2015. "Meet The Mysterious New Hacker Army Freaking Out The Middle East". *BuzzFeed News* (June). www.buzzfeednews.com/article/sheerafrenkel/who-is-the-yemen-cyber-army.

Goldstein, Emmanuel. 2018. "Hacktivism and the Hacker Promise". *Blog Law Columbia.* blogs.law.columbia.edu/uprising1313/emmanuel-goldstein-hacktivism-and-the-hacker-promise/.

GonjeshkeDarande. 2023. "GonjeshkeDarandeOfficial". t.me/GonjeshkeDarande Official 3.

Goodin, Dan. 2022a. "After Ukraine Recruits an 'IT Army', Dozens of Russian Sites Go Dark". *Ars Technica* (February). arstechnica.com/information-technology/2022/02/after-ukraine-recruits-an-it-army-dozens-of-russian-sites-go--dark/.

Goodin, Dan. 2022b. "Pro-Russia Threat Group Killnet Is Pummeling Lithuania with DDoS Attacks". *Ars Technica* (June). arstechnica.com/information-technology/2022/06/pro-russia-threat-group-killnet-is-pummeling-lithuania-with-ddos-attacks/.

Great. 2012. "Shamoon the Wiper — Copycats at Work" (August). securelist.com/shamoon-the-wiper-copycats-at-work/57854/.

Greenberg, Andy. 2019. "A Mystery Agent Is Doxing Iran's Hackers and Dumping Their Code". *Wired.* www.wired.com/story/iran-hackers-oilrig-read-my-lips/.

Greenberg, Andy. 2024. "How a Group of Israel-Linked Hackers Has Pushed the Limits of Cyberwar". *Wired*. www.wired.com/story/predatory-sparrow-cyberattack-timeline/.

Groll, Elias. 2016. "The NSA Has a New Disclosure Policy: Getting Hacked". *Foreign Policy*. foreignpolicy.com/2016/08/18/the-nsa-has-a-new-disclosure-policy-getting-hacked/.

Gross, Michael L., Daphna Canetti, and Dana R. Vashdi. 2017. "Cyberterrorism: Its Effects on Psychological Well-Being, Public Confidence and Political Attitudes". *Journal of Cybersecurity* 3, no. 1: 49–58. doi.org/10.1093/cybsec/tyw018.

Guerrero-Saade, Juan Andrés and Brian Bartholomew. 2016. "Wave Your False Flags! Deception Tactics Muddying Attribution in Targeted Attacks". *Virus Bulletin*. www.virusbulletin.com/conference/vb2016/abstracts/wave-your-false-flags-deception-tactics-muddying-attribution-targeted-attacks.

Haggard, Stephan and Jon R. Lindsay. 2015. "North Korea and the Sony Hack: Exporting Instability Through Cyberspace" (May). www.eastwestcenter.org/publications/north-korea-and-the-sony-hack-exporting-instability-through-cyberspace.

Halili, Eduart. 2022. "Cyber Attacks Forces AKSHI Close Government Online Systems". *Albania Daily News*. albaniandailynews.com/news/cyber-attacks-forces-akshi-close-government-online-systems.

Howard, Philip N., Aiden Duffy, Deen Freelon, M. M. Hussain, Will Mari, and Marwa Maziad. 2011. "Opening Closed Regimes: What Was the Role of Social Media During the Arab Spring?" SSRN Scholarly Paper No. 2595096. *Social Science Research Network*. doi.org/10.2139/ssrn.2595096.

Hulcoop, Adam, John Scott-Railton, Peter Tanchak, Matt Brooks, and Ron Deibert. 2017. "Tainted Leaks: Disinformation and Phishing With a Russian Nexus". *Citizen Lab Research Report* no. 92. University of Toronto. citizenlab.ca/2017/05/tainted-leaks-disinformation-phish/.

Ikwu, Ruth, Luca Giommoni, Amir Javed, Pete Burnap, and Matthew Williams. 2023. "Digital Fingerprinting for Identifying Malicious Collusive Groups on Twitter". *Journal of Cybersecurity* 9, no. 1: tyad014. doi.org/10.1093/cybsec/tyad014.

Intelligence, Microsoft Threat. 2022. "Microsoft Investigates Iranian Attacks against the Albanian Government". *Microsoft Security Blog* (September). www.microsoft.com/en-us/security/blog/2022/09/08/microsoft-investigates-iranian-attacks-against-the-albanian-government/.

Intrusion Truth. 2017. intrusiontruth.wordpress.com/.

IT Army of Ukraine. 2022. t.me/itarmyofukraine2022.

Jenkins, Luke, Emiel Haeghebaert, Ben Read, and Alice Revelli. 2022. "Roadsweep Ransomware Targets the Albanian Government". cloud.google.com/blog/topics/threat-intelligence/likely-iranian-threat-actor-conducts-politically-motivated-disruptive-activity-against/.

Jofre, TZVI. 2020. "Hezbollah-Affiliated Financial Org Hacked, Information Leaked". *The Jerusalem Post* (December). www.jpost.com/breaking-news/hezbollah-affiliated-financial-org-hacked-information-leaked-653690.

Jordan, Tim. 2015. "Hacktivism: Operation Tunisia, Modular Tactics and Information Activism". In *Information Politics.* Liberation and Exploitation in the Digital Society. Pluto Press. doi.org/10.2307/j.ctt183p2xf.14.

Karagiannopoulos, Vasileios. 2018. *Living With Hacktivism.* Springer International Publishing. link.springer.com/10.1007/978-3-319-71758-6.

Karagiannopoulos, Vasileios. 2021. "A Short History of Hacktivism: Its Past and Present and What Can We Learn from It". In *Rethinking Cybercrime: Critical Debates,* edited by Tim Owen and Jessica Marshall. Springer International Publishing. doi.org/10.1007/978-3-030-55841-3_4.

Kausch, Kristina. 2017. "Cheap Havoc: How Cyber-Geopolitics Will Destabilize the Middle East". www.gmfus.org/news/cheap-havoc-how-cyber-geopolitics--will-destabilize-middle-east.

Keller, Max Fisher, Jared. 2011. *The Atlantic* (August). www.theatlantic.com/international/archive/2011/08/syrias-digital-counter-revolutionaries/244382/.

Klein, Adam G. 2015. "Vigilante Media: Unveiling Anonymous and the Hacktivist Persona in the Global Press". *Communication Monographs* 82, no. 33: 379–401. doi.org/10.1080/03637751.2015.1030682.

Kostyuk, Nadiya, and Yuri M. Zhukov. 2019. "Invisible Digital Front: Can Cyber Attacks Shape Battlefield Events?" *Journal of Conflict Resolution* 63, no. 2: 317–47. doi.org/10.1177/0022002717737138.

Leyden, John. 2010. "4chan Launches DDoS against Entertainment Industry". www.theregister.com/2010/09/20/4chan_ddos_mpaa_riaa/.

Mandiant. 2022. "Hacktivists Collaborate with GRU-Sponsored Threat Actors". *Google Cloud Blog*. cloud.google.com/blog/topics/threat-intelligence/gru-rise--telegram-minions.

Mandiant. 2023. "The GRU's Disruptive Playbook". *Google Cloud Blog*. cloud.google.com/blog/topics/threat-intelligence/gru-disruptive-playbook.

Mandiant. 2023b. "KillNet Showcases New Capabilities While Repeating Older Tactics". *Google Cloud Blog*. cloud.google.com/blog/topics/threat-intelligence/killnet-new-capabilities-older-tactics.

Marczak, Bill and John Scott-Railton. 2016. "The Million Dollar Dissident: NSO Group's iPhone Zero-Days Used against a UAE Human Rights Defender". *Citizen Lab Research Report* no. 78. University of Toronto. citizenlab.ca/2016/08/million-dollar-dissident-iphone-zero-day-nso-group-uae/.

Marczak, Bill, John Scott-Railton, Adam Senft, Irene Poetranto, and Sarah McKune. 2015. "Mapping FinFisher's Continuing Proliferation". *Citizen Lab Research Report* no. 64. University of Toronto. citizenlab.ca/2015/10/mapping--finfishers-continuing-proliferation/.

Maurer, Tim. 2018. *Cyber Mercenaries: The State, Hackers, and Power*. Cambridge University Press. www.cambridge.org/core/books/cyber-mercenaries/B685B7555E1C52FBE5DFE6F6594A1C00.

Milone, Mark. 2003. "Hacktivism: Securing the National Infrastructure". *Knowledge, Technology & Policy* 16, no. 1: 75–103. doi.org/10.1007/s12130-003-1017-5.

Mueller, Robert S. 2019. "Report on the Investigation into Russian Interference in the 2016 Presidential Election". www.justice.gov/archives/sco/file/1373816/dl.

Nakashima, Ellen and Shane Harris. 2018. "How the Russians Hacked the DNC and Passed Its Emails to WikiLeaks". *The Washington Post*. www.washingtonpost.com/world/national-security/how-the-russians-hacked-the-dnc-and-passed-its-emails-to-wikileaks/2018/07/13/af19a828-86c3-11e8-8553-a3ce89036c78_story.html.

Naprys, Ernestas. 2024. "Belarusian KGB Allegedly Breached, Hackers Dox over 8600 Agents". *Cybernews* (April). cybernews.com/news/belarusian-kgb-allegedly-breached-cyber-partisans/.

Natto Thoughts. 2024. "Intrusion Truth Methods: How Can They Get It Right Again and Again?" Substack newsletter. *Natto Thoughts* (April). nattothoughts.substack.com/p/intrusion-truth-methods-how-can-they.

Novetta. 2016. "Operation Blockbuster: Unraveling the Long Thread of the Sony Attack". *Docslib*. docslib.org/doc/5282771/operation-blockbuster-unraveling--the-long-thread-of-the-sony-attack-3-caveats.

Olson, Parmy. 2013. *We Are Anonymous: Inside the Hacker World of LulzSec, Anonymous, and the Global Cyber Insurgency*. Reprint edition. Back Bay Books.

Perlroth, Nicole, and David E. Sanger. 2017. "Hackers Hit Dozens of Countries Exploiting Stolen N.S.A. Tool". World. *The New York Times* (May). www.nytimes.com/2017/05/12/world/europe/uk-national-health-service-cyberattack.html.

Privacy International. 2018. *The Global Surveillance Industry*. privacyinternational.org/explainer/1632/global-surveillance-industry.

Rader and Wash. 2015. "Identifying patterns in informal sources of security information". *Journal of Cybersecurity* 1, no. 1: 121–44. doi.org/10.1093/cybsec/tyv008

RecordedFuture. 2015. *Cyber Berkut Graduates From DDoS Stunts to Purveyor of Cyber Attack Tools*. www.recordedfuture.com/blog/cyber-berkut-analysis.

Rid, Thomas. 2016. "All Signs Point to Russia Being Behind the DNC Hack". *VICE* (July). www.vice.com/en/article/all-signs-point-to-russia-being-behind--the-dnc-hack/.

Rid, Thomas. 2017. "Disinformation: A Primer In Russian Active Measures And Influence Campaigns Panel II". Select Committee on Intelligence. www.govinfo.gov/content/pkg/CHRG-115shrg25998/html/CHRG-115shrg25998.htm.

Roberts, Paul. 2010. "Attacks On MPAA's UK Law Firm Lead to Data Leaks, Lawsuit". Threatpost. threatpost.com/attacks-mpaas-uk-law-firm-lead-data-leaks-lawsuit-092710/74517/.

Romagna, Marco. 2020. "Hacktivism: Conceptualization, Techniques, and Historical View". In *The Palgrave Handbook of International Cybercrime and Cyberdeviance*, edited by Thomas J. Holt and Adam M. Bossler. Springer International Publishing. doi.org/10.1007/978-3-319-78440-3_34.

Samuel, Alexandra Whitney. 2004. *Hacktivism and the Future of Political Participation*. Harvard University.

Sauter, Molly. 2014. *The Coming Swarm : DDoS Actions, Hacktivism, and Civil Disobedience on the Internet*. Bloomsbury Academic.

Schneier, Bruce. 2017. "Who Are the Shadow Brokers?" *Schneier on Security* (May). www.schneier.com/blog/archives/2017/05/who_are_the_sha.html.

Seblani, Sana. 2021. *Lebanon: The Hacking of Hezbollah's Al-Qard Al-Hasan and US Sanctions.* daraj.media/en/lebanon-the-hacking-of-hezbollahs-al-qard-al-hasan-and-us-sanctions/.

Shehabat, Ahmad. 2012. *The Social Media Cyber-War: The Unfolding Events in the Syrian Revolution 2011*, no. 6: 2. www.hca.westernsydney.edu.au/gmjau/archive/v6_2012_2/ahmad_shehabat%20_RA.html.

Sigholm, Johan. 2016. "Non-State Actors in Cyberspace Operations". *Journal of Military Studies* 4, no. 1: 1–37. doi.org/10.1515/jms-2016-0184.

Siqi, Cao. 2023. *Exclusive: Hacker Group with Members from Europe, North America Found to Have Launched Cyberattacks against China.* www.globaltimes.cn/page/202302/1285744.shtml.

Soesanto, Stefan. 2022. "The IT Army of Ukraine". *Center for Security Studies* (June). css.ethz.ch/en/center/CSS-news/2022/06/the-it-army-of-ukraine.html.

Streams, Kimber. 2012. "Anonymous "Operation Payback" Hackers Convicted for Costly DDoS Attacks". *The Verge* (December). www.theverge.com/2012/12/6/3735622/anonymous-conviction-christopher-weatherhead--operation-payback.

Suiche, Matt. 2017. "Lessons from TV5Monde 2015 Hack". *Comae Technologies* (June). medium.com/comae/lessons-from-tv5monde-2015-hack-c4d62f07849d.

Taylor, Paul A. 2005. "From Hackers to Hacktivists: Speed Bumps on the Global Superhighway?" *New Media & Society* 7, no. 5: 625–46. doi.org/10.1177/1461444805056009.

Threat Hunter Team. 2018. *Shamoon: Destructive Threat Re-Emerges with New Sting in Its Tail.* www.security.com/threat-intelligence/shamoon-destructive--threat-re-emerges-new-sting-its-tail.

ThreatConnect. 2016a. "Guccifer 2.0: The Man, the Myth, the Legend?" *ThreatConnect* (July). threatconnect.com/blog/guccifer-2-0-the-man-the-myth--the-legend/.

ThreatConnect. 2016b. "Hacktivists vs Faketivists: Fancy Bears in Disguise". *ThreatConnect* (December). threatconnect.com/blog/faketivist-vs-hacktivist-how--they-differ-2/.

Tripwire. 2016. *Shadow Brokers Leaks Dilemma – History of Events Explained.* www.tripwire.com/state-of-security/shadow-brokers-leaks-dilemma-history-events-explained.

Ukrainian Ministry of Digiral Transformation. 2022. *Ukraine Ministry of Digiral Transformation: IT Army Blocks Russian Sites in a Few Minutes — the Main Victories of Ukraine on the Cyber Front.* www.kmu.gov.ua/en/news/mincifri-it-armiya-blokuye-rosijski-sajti-za-dekilka-hvilin-golovni-peremogi-ukrayini-na-kiberfronti.

United States Department of Justice. 2018a. *Park Jin Hyok — Complaint.* www.justice.gov/usao-cdca/press-release/file/1091951.

United States Department of Justice. 2018b. *U.S. Charges Russian GRU Officers with International Hacking and Related Influence and Disinformation Operations* (October). www.justice.gov/opa/pr/us-charges-russian-gru-officers-international-hacking-and-related-influence-and.

United States Department of Justice. 2019. *English: U.S. v. Viktor Borisovich Netyksho (Case 1:18-Cr-215) Indictment Filed by the Mueller Investigation* (July). www.justice.gov/file/1080281/download. commons.wikimedia.org/wiki/File:U.S._v._Viktor_Borisovich_Netyksho,_et_al,_Indictment.pdf.

Vijayan, Jai. 2019. *Data Dump Purportedly Reveals Details on Previously Unknown Iranian Threat Group.* www.darkreading.com/cyberattacks-data-breaches/data-dump-purportedly-reveals-details-on-previously-unknown-iranian-threat-group.

WADA. 2016. "Cyber Security Update: WADA's Incident Response". *World Anti Doping Agency* (October). www.wada-ama.org/en/news/cyber-security-update-wadas-incident-response.

Wes. 2024. "Phineas Fisher: The World's Hacktivist". *Medium* (November). medium.com/@phishfinding/phineas-fisher-the-worlds-hacktivist-483601802ca8.

WikiLeaks. 2014. *SpyFiles 4.* wikileaks.org/spyfiles4/.

WikiLeaks. 2015. *The Hackingteam Archives.* wikileaks.org/hackingteam/emails/.

Yemen Cyber Army. 2015. *Yemen Cyber Army Announces Target List.* cryptome.org/2015/07/yemen-cyber-army-targets.htm.

Zetter, Kim. 2014. "Sony Got Hacked Hard: What We Know and Don"t Know So Far". *Wired.* www.wired.com/2014/12/sony-hack-what-we-know/.

Zetter, Kim. 2022a. "Intrusion Truth — Five Years of Naming and Shaming China's Spies". *Zero Day* (March). www.zetter-zeroday.com/interview-with-intrusion-truth/.

Zetter, Kim. 2022b. "Unmasking China's State Hackers". *Zero Day* (March). www.zetter-zeroday.com/unmasking-chinas-state-hackers/.

IS THERE REAL HACKTIVISM? A METHOD TO DISTINGUISH FALSE-FLAG OPERATIONS FROM GENUINE HACKTIVISTS

ABSTRACT

Hacktivism has been a persistent yet evolving phenomenon since the advent of the internet. While early hacktivist actions were largely symbolic, the 2000s saw the emergence of impactful campaigns. However, as these activities gained visibility, state-sponsored actors began adopting false-flag tactics — mimicking hacktivism to obscure state involvement in cyber operations and leveraging audiences for their "pure" political goals. In a politically polarized world, distinguishing genuine political movements from foreign interference is relevant. This paper examines the distinction between genuine hacktivism and state-backed false-flag operations by analysing 24 notable cases. Through a comparative approach, it establishes a methodology to differentiate authentic hacktivist activities from strategically disguised state actions. The analysis identifies four key indicators: (i) the longevity and continuity of a group's activity, (ii) the selection and justification of targets, (iii) the sophistication of the attack methods employed, and (iv) the manipulation of leaked information to serve state interests. The findings reveal that authentic hacktivism represents a minority of high-profile cyber incidents, with most well-documented cases being linked to state-backed actors seeking plausible deniability for cyber operations.

Keywords: Hacktivism; False-Flag Operations; Cybersecurity; Hybrid Threats.

RESUMO

O hacktivismo tem sido um fenômeno persistente, embora em constante evolução, desde o advento da internet. Enquanto as ações hacktivistas iniciais eram amplamente simbólicas, os anos 2000 testemunharam o surgimento de campanhas mais impactantes. Contudo, à medida que essas atividades ganharam visibilidade, atores patrocinados pelo Estado começaram a adotar táticas de *false-flag* — imitando o hacktivismo para ocultar o envolvimento estatal em operações cibernéticas, explorando públicos devido ao seu objetivo político "puro". Em um mundo politicamente polarizado, distinguir movimentos políticos genuínos de interferência estrangeira torna-se relevante. Este artigo examina a distinção entre o hacktivismo genuíno e as operações de *false-flag* apoiadas por Estados, analisando 24 casos notáveis. Por meio de uma abordagem comparativa, estabelece uma metodologia para diferenciar atividades hacktivistas autênticas de ações estatais estrategicamente disfarçadas. A análise identifica quatro indicadores-chave: (i) a longevidade e continuidade das atividades de um grupo; (ii) a seleção e justificativa dos alvos; (iii) a sofisticação dos métodos de ataque empregados; e (iv) a manipulação de informações vazadas para servir a interesses estatais.

Palavras-chave: Hacktivismo; Operações de Falsa Bandeira; Segurança Cibernética; Ameaças Híbridas.

Recebido em 15/05/2025. Aceito para publicação em 01/10/2025.