International Review of Law, Computers & Technology

ISSN: 1360-0869 (Print) 1364-6885 (Online) Journal homepage: https://www.tandfonline.com/loi/cirl20

Vigilantism and cooperative criminal justice: is there a place for cybersecurity vigilantes in cybercrime fighting?

K. K. e Silva

To cite this article: K. K. e Silva (2018) Vigilantism and cooperative criminal justice: is there a place for cybersecurity vigilantes in cybercrime fighting?, International Review of Law, Computers & Technology, 32:1, 21-36, DOI: 10.1080/13600869.2018.1418142

To link to this article: https://doi.org/10.1080/13600869.2018.1418142

Published online: 27 Dec 2017.

INTERNATIONAL REVIEW OF LAW, COMPUTERS & TECHNOLOGY, 2018
VOL. 32, NO. 1, 21–36
https://doi.org/10.1080/13600869.2018.1418142

Routledge
Taylor & Francis Group

Vigilantism and cooperative criminal justice: is there a place for cybersecurity vigilantes in cybercrime fighting?

K. K. e Silva

Tilburg Institute for Law, Technology, and Society, Tilburg University, Tilburg, Netherlands

ABSTRACT

Are cybersecurity vigilantes at odds with criminal justice? Perhaps. In general terms, vigilantism could be understood as an act of retaliation launched by private agents in response to a perceived criminal conduct and targeting alleged perpetrators of a crime. This form of unofficial crime control has flourished on the Internet, where non-State actors have enforced informal means of justice to counter criminal behavior. Recently, the actions of cybersecurity vigilantes have become a recurrent (and sometimes disruptive) element in the fight against cybercrime. In this paper, I shall argue that individuals who make use of force in response to criminal activities online could pose a serious threat to cybercrime investigations – but also that by acting upon a presumable moral duty to counter crime, cybersecurity vigilantes are turning the tables on how law enforcement is effectuated online and shaping the future of cooperative criminal justice.

KEYWORDS

Vigilantism; cybercrime; botnets

1. Introduction

Vigilantes' responses to perceived malicious activity have reportedly caused the loss of digital evidence, thereby obstructing law enforcement's effort in ascertaining attribution and jurisdiction over cybercrime offences. Acknowledging the fleeting existence of digital evidence and its importance for criminal prosecution, the Council of Europe Convention on Cybercrime empowered law enforcement agencies with orders to preserve evidence at peril and to produce data that is soon to be erased (Arts. 16 to 18). However, if evidence is altered or destroyed as the result of a countermeasure launched by informal justice makers, law enforcement may not have the opportunity to safeguard crucial data. Ultimately, if the authorities are unable to successfully prosecute cybercriminals, perpetrators will remain at large and attacks are bound to persist. Even if unknowingly, vigilantes may be tampering with evidence and hampering criminal investigations, a criminal offence under the national law of various EU Member States (e.g. Section 200 of the Dutch Code of Criminal Procedure, Art. 434–4 of the French Criminal Code).

On the other hand, actors may have a right and a duty to respond to perceived cybercrime offences that, if not countered timely, could damage private and public interests. Moreover, certain categories of expert vigilantes (e.g. computer scientists), may have

CONTACT K. K. e Silva k.k.esilva@uvt.nl

valuable insights that could prevent, halt, and mitigate cybercrime. Built on the premise that the dichotomy between *law enforcement* and *private enforcement* shall be surpassed, for the challenges regarding cybercrime calls for a model by which State and non-State actors may join forces, this paper investigates online vigilantism against cybercrime, and its opportunities and challenges within a framework for cooperative justice.

2. Defining vigilantism

The starting point of a discussion on vigilantism is conceptual. Although described as a longstanding social phenomenon, a criminological definition of the term was offered by Johnston in 1996 (1996). This concept is at the root of vigilantism studies and continues valuable to this date. By Johnston (1996, 221), vigilantism is a 'social movement giving rise to premeditated acts of force – or threatened force – by autonomous citizens. It arises as a reaction to the transgression of institutionalized norms by individuals or groups – or to their potential or imputed transgression'. The model of vigilantism advanced by Johnston is a compound concept built upon six key elements: (1) Planning, premeditation, and organization, (2) Private voluntary agency, (3) Autonomous citizenship, (4) The use or threatened use of force, (5) Reaction to crime and social deviance, and (6) Personal and collective security. The following paragraphs expose the scope of these elements as imagined by Johnston and the reason why they are essential to the idea of vigilantism.

According to Johnston, vigilantism is a *calculated, willful activity* that is preceded by at least minimal planning. Notably, the very origin of the word vigilantism relates to staying awake and watchful (late fifteenth century: from Latin vigilant- 'keeping awake', from the verb *vigilare*, and from vigil, Latin *vigilia*, 'awake').[1] As such, vigilantes – or those who engage in vigilantism – would not act on a whim: there is a purposeful intention and prior considerations about the actions to be undertaken. Corroborating this view, Dumsday (2009, 51) argues that actions of spontaneous vigilantism 'would not be instances of vigilantism; they are simply instances of altruistic defense in which the rescuers exceed the bounds of what is strictly necessary for protecting the victim'.

Acts of vigilantism are *launched by private agents*, meaning State actors and public authorities are out of the scope of the term, by default. Johnston (1996, 224) refuses the idea that vigilantism by off-duty officers would contradict this requirement, given that the acts committed by such persons may not always be distinguishable from the competences of their tasks, represent their own views and not those of their public organizations. Dumsday (2009, 51) rejects the idea that off-duty officers fall into blurry lines whenever the officer acts anonymously and not at the behest of higher government officials, what seems reasonable. Additionally, while much of the discussion is focused on individual citizens and State actors, the debate on vigilantism pays little attention to whether vigilantism can be a product of collective groups or whether legal persons may also act as vigilantes.

Autonomous citizenship refers to the fact that vigilantes act without a mandate from the State and therefore are not to be recognized as agents deploying delegated powers. Vigilantes are 'active citizens' (Johnston 1996, 226) who tale matters onto their own hands, perceiving themselves as moral actors in society who must intervene in face of a perceived threat. Thus, unofficial justice initiatives that are openly supported by public policies or public authorities are outside the scope of vigilantism (Dumsday 2009, 51).

Vigilantes respond to an event with the *use of or a threaten to use force*. Vigilantism and harm (or a promise of harm) are correlated: vigilantes exploit violence as retaliation. To that end, vigilantism goes beyond the boundaries of self-defense and should be differentiated from it. Vigilantes may use physical force, weapons, naming and shaming, among other forms of harm to inflict pain on the alleged 'assailant'. What other forms of interference could be classified as 'force', especially in a digital environment, remains uncertain.

Interestingly, vigilantism is not a preventive, but mostly a reaction following a 'social disturbance' in the view of vigilantes. Johnston recognizes the difference between vigilantism as 'crime control' and 'social control' (1996, 228). This division has faced disagreement in scholarship. Haas contends that only vigilantes that *responds to a (perceived) criminal act* are true to the notion of vigilantism (2010, 34). Haas argues, and I agree, that extending the notion of vigilantism to reactions launched against non-criminal acts would risk including anti-democratic behaviors and violations of fundamental rights as vigilantism, what should be avoided (2010, 34). The perception over what is the social norm and what is socially desirable cannot be defined by a narrow group of active citizens, but must be in line with the virtues of the democratic state and public interest. Therefore, this study also adopts the notion that vigilantism is only to be regarded as such when in response to (perceived) criminal behavior as defined by the law.

Finally, vigilantism is a social *response to a perceived lack of security*. Consequently, vigilantism arises as an initiative aimed at restoring the levels of personal and social security desired by vigilantes. The vigilante is thus an 'upholder of the established rules' (Dumsday 2009, 55). As noted by Johnston (1996, 228), 'vigilantism arises when some established order is perceived to be under threat from the transgression (or potential transgression) of institutionalized norms'.

Emblematic of the early understanding of vigilantism, the conceptualization of Johnston is a landmark in this field of study and offers a systematic approach to the identification of vigilantism and its affordances. A contemporary development has been suggested by Haas (2010), which focuses on the line of events surrounding the activities of vigilantes and provides an interesting insight on vigilantism as a process. Haas (2010) discusses five elements of vigilantism, namely: (1) precipitating event, (2) formal response to the precipitating event, (3) vigilantism, (4) vigilante, and (5) victim. Although the typology of Haas is possibly less clear or consistent than the model advanced by Johnston, it is relevant to reflect on the notion of 'vigilantism event sequence', formed by the first three elements.

The concept of precipitating event framed by Haas (2010, 37) refers to the incident that triggers the reaction of vigilantes. In the terminology of Johnston, this would be the equivalent of element n. 5 (reaction to a criminal action or social disturbance), except that, as abovementioned, Haas defends only criminal acts can classify as authentic precipitating events. Next, a formal response to the precipitating event, or the answer issued by the criminal justice system. Haas (2010, 37) emphasizes that 'this formal response can vary from no action at all, to for instance police arrest or a judge's verdict'. The perception of vigilantism as a social response is widened by Haas to comprise nonviolent acts, including those with physical, psychological, and economic consequences. This is strictly relevant to our examination of cybersecurity vigilantism.

3. Cybersecurity vigilantes

When applied to the context of information systems, the conceptualization of vigilantism as proposed by Johnston requires a certain level of adaptation. Smallridge, Wagner, and Crowl (2016, 59) argue that the compound concept crafted by Johnston must be broaden to apply to the specific circumstances of 'cyber vigilantism'. Cyber vigilantism, perceived as the variation of vigilantism taking place in the Internet and through information systems, is often associated with mass retaliation and collective efforts (Smallridge, Wagner, and Crowl 2016, 59). Cyber vigilantes act in response to a perceived and repercussive criminal act and, although often geographically dispersed, are united by connectivity and the opportunities of digital communication. In these cases, the use of force is seldom a product of physical harm but mostly a form of harm that can be executed digitally or inflicted upon the digital identity and reputation of the alleged assailant. Moreover, groups of vigilantes online have frequently convened under umbrella organizations and, although individual efforts remain common, unofficial justice has gained new nuances in the Internet era.

Cyber vigilantism takes varying forms: hacktivism, scam baiting, crowdsourcing, and citizen-led intervention (Smallridge, Wagner, and Crowl 2016, 59) are examples of how vigilantism is transmuting. As Dizon (2016, 25) explains, hacktivism is a form of hacking for overtly socio-political purposes, where the act of hacking is both the form and substance of their activism. Initiatives launched by groups such as Anonymous and LulzSec against governmental agencies and corporations are examples of how hacktivists express their social and political dissatisfaction with the actions of their targets (Dizon 2016, 25). Scambaiters turn fraudulent campaigns around to prey on online scammers, in a form of cybervengeance (Rosenbaum 2007). Crowdsourcing vigilantism is enabled by citizens amassing efforts towards a shared objective (Smallridge, Wagner, and Crowl 2016, 59–60), and has become popular in China through the 'human-flesh search engine' practice. This particular form of retaliation involves releasing one's personal details and private communications publicly, with the intent of ridiculing and harassing the subject of the search.[2]

In this study, however, we are concerned with cyber vigilantism that is moved by the objective of restoring cybersecurity and trust in digital environments. Borrowing from Haas's typology of vigilantism (Haas 2010, 34), the precipitating event to which cybersecurity vigilantes respond is thus an act that disrupts or endangers the security and reliability of the Internet. This sub-group, here referred to as cybersecurity vigilantes, can be defined as active citizens who, voluntarily and without the sanction of the State, launch attacks against cyber threats and cybercriminals with the goal of reestablishing justice and cybersecurity. In sum, cybersecurity vigilantism is *a social movement composed by individuals or collective groups who respond via technical means to a perceived and repercussive criminal act against the security of the Internet and information systems*.

At this point, it is important to highlight that the form of vigilantism discussed here is different from the one characterized by Trottier as *digital vigilantism* (2017, 56). By Trottier (2017, 55), digital vigilantism

> is a process where citizens are collectively offended by other citizen activity, and respond through coordinated retaliation on digital media, including mobile devices and social media platforms. The offending acts range from mild breaches of social protocol (bad parking; not removing dog feces) to terrorist acts and participation in riots. These offensive

> acts are typically not meant to generate large-scale recognition. Therefore, the targets of DV are initially unaware of the conflict in which they have been enrolled.

The social movement described as cybersecurity vigilantism is not restricted to social media. Plus, it has a very specific motivation: restoring and sometimes increasing levels of cybersecurity. Therefore, although a certain overlap between the concepts exists, they are significantly different in that cybersecurity vigilantes are mostly a group of skilled individuals who use their technical knowledge of computer science and engineering to retaliate criminal acts committed against information systems and remediate cybersecurity. Paraphrasing Dumstay, *cybersecurity vigilantes are upholders of the functioning and the future of the Internet*.

4. Defense of other

From a legal standpoint, cybersecurity vigilantism poses a series of questions regarding legitimacy and legality of bottom-down contributions to criminal justice. A tenet of the criminal justice system is the State monopoly of the use of violence. By this tenet, only the State, in the form of its institutions and official bodies, is permitted to lawfully deploy force – to the extent that is justifiable for the purpose of ensuring respect for the rule of law. Exceptionally, individuals will not be punished for using force if this classifies as a measure of public or private defense. For instance, individuals are authorized to counteract with reasonable force if such use of force is necessary, proportional, and well-timed for stopping or preventing a crime against themselves (self-defense) or another (defense of other). Aside from these narrow circumstances, the use of force by individuals is punishable and the actors who engage in such use are criminally liable for any criminal offences ensuing from it.

Defense of other and self-defense are closely linked concepts and share similar elements. The idea of defense of other could be understood as a variety of self-defense which is transferred to another actor who has the conditions and opportunities to act in a moment where the victim may not or will fail to respond to the aggression. Consequently, in order to escape criminal liability under the auspices of defense of other, the general theory of defense of other requires the actions of the agent be preceded by the genuine belief that the rights of another are at risk or in imminent danger. In addition, the response of the agent must be timely, proportional, and necessary to prevent or halt the criminal act. Every excess on the use of force escapes the shield of defense of other and brings back the shadow of criminal liability.

The question that arises is whether cybersecurity vigilantism could be covered under the umbrella of defense of other. Because defense of other excludes the criminal punishment of the activities launched by one or more citizens in the protection of another individual (or group of individuals), it is possible to consider if the elements of cybersecurity vigilantism could actually translate into an action that is covered by this private defense – or at least analyze which requirements would have to be present in order to justify the activities of cybersecurity vigilantism as a non-criminal act. To answer that question, I shall consider two specific cases highly publicized in the media: the actions of 'Janit0r' through Brickerbot (the anti-Mirai botnet) and the ones of Marcus Hutchins following the spread of Wannacry ransomware.

4.1. BrickerBot

The spread of Mirai, the first large-scale IoT botnet publicized, spawned a turmoil in the cybersecurity community. Although IoT devices had long been reported as a ticking bomb: the level of security embodied in the technology was questionable and a spur of threats against IoT was envisioned. In 2016, Mirai emerged as a powerful, remote network affecting cameras and routers, causing massive disruptions worldwide. The insufficient security standards deployed in the development of these mass IoT devices were captured by botherders and used to perpetrate potent DDoS attacks. The attacks emerging from Mirai have paralyzed more than 900,000 Deutsche Telekom customers (Krebs 2016b), a prominent cybersecurity website,[3] and the telecommunications infrastructure of Liberia (Leyden 2016). Several developments made these outages possible, including the leaking of the Mirai source code, what allowed other cybercriminals to create powerful and resilient versions of the original botnet.[4] According to Flashpoint, by November 2016 Mirai had already compromised a total of 5 million devices (Costello et al. 2016).

In April 2017, an interesting development was spotted. Security researchers at Radware identified a PDoS (permanent denial of service attack) aimed at corrupting the storage of specific IoT devices and therefore incapacitating their functioning permanently (Radware 2017). Four versions of BrickerBot were released in the wild (BricketBot.1, BrickerBot.2, BrickerBot.3, BrickerBot.4) in a short interval, disabling close to 2 million devices according to their creator (Cimpanu 2017). Described as a gray hat (Radware 2017), vigilante hacker, Janit0r was allegedly motivated by the idea that unsafe IoT devices should be treated as any other product whose security standards are so poor as to undermine the safety of others, failing to offer minimum state-of-the-art security features (Millman 2017). By the vigilante (Millman 2017), various IoT manufacturers placed dangerous devices on the market which could be easily exploited by the sorts of Mirai. But the response from regulators was yet to come and the widely known vulnerabilities were still running large by the time BrickerBot surfaced.

The events that follow are a de facto technical remedy to the issue of product unsafety: the unleash of BrickerBot disabled the use of poorly secured IoT devices prone to botnet infections, permanently. In sum, new cyberattacks were prevented by the spread of what would legally constitute another cybercrime offence, with the difference that the latter was motivated to protect citizens and public interest. Such forms of interference are normally a privilege of the State, whose delegated authorities can use force to protect society from danger. From a result-oriented angle, however, the actions of BrickerBot are not much different from what regulators enforce in product liability situations. When a product placed on the market fails to comply with the safety and security standards part of the state-of-the-art, a recall is often issued with the purpose of removing the unsafe devices from the market and making them unavailable to customers until the defect is patched. However, there is a large gap between the activities of actors such as Janit0r and companies promoting a recall, as the second are grounded on legislative instruments that justify the removal of unsafe products from the market and legitimately authorized by the law to do so.

But I will continue to analyze the event from a criminal law, though from a private defense perspective. From what has been reported on the media, the actions of Janit0r could fulfill the requirements specified for the use of defense of other. The actions are

based on the presumption that a grave attack is imminent and poses a serious risk – Mirai and its varieties were at large and growing, preying the exact vulnerabilities present in the devices targeted by Janit0r and BrickerBot. The release of BrickerBot is timely in that it was disseminated opportunely to prevent new infections to take place and avoid new DDoS attacks from being launched. However, the assessment of the element of imminence is based on the accounts reported on the media (Krebs 2016b) and grounded on the likelihood that the devices disrupted by BrickerBot would be soon targeted by Mirai, given the fact that they presented the same vulnerability being exploited in mass. But the element of imminence cannot be sufficiently proven by factual inferences alone, since the assessment of the imminence of the attack is both factual and subjective. Therefore, for the element of imminence to be sufficiently present, the author of BrickerBot should be expected to substantiate the reasons why, from her perspective, the attack was imminent and therefore justified the launch of the malware.

The elements of necessity and proportionality of this case are obscure and failure to abide to these elements would thwart an excuse based on defense of other. The main problem behind BrickerBot is its permanent character, in that it simply made targeted IoT devices unusable. Was irreversibly disabling the unsafe devices absolutely necessary for preventing the spread of Mirai? That is a question that only security experts would be capable to firmly answer. In using force, less aggravating and interfering means are always preferable as an alternative to disproportionately strong interventions. From a non-technical lens, the dissemination of a patch could have been equally effective in shutting down the vulnerability and therefore preventing Mirai-like infections without permanently disrupting the targeted devices. To that extent, there seems to be an excess in the use of force by Janit0r that would escape the shield of defense of other and subject the agent to liability in relation to the excess. Furthermore, the actions of Janit0r show elements of a cybercrime offence, more specifically of illegal interference with information systems (Art. 4, CoE Convention on Cybercrime, Art. 4, Directive 2013/40/EU).

4.2. WannaCry

Ransomware is a popular means of monetizing cybercrime, as this form of malware demands a monetary payment from the user of the victimized device. Ransomware usually locks in valuable content stored in a device and prevents its access by the user in exchange for a ransom. WannaCry ransomware was identified in May 2017, following its spread over Windows operating systems and an upheaval in the British health care systems, where 16 hospitals had their systems compromised. The malware upsurge took many by surprise as contamination was achieved without user interaction and easily carried out through interconnected networks, what explains its paralyzing effects over the British National Health System. The lock-in was achieved through encryption and the encryption key offered for a ransom of 300USD in bitcoins (Symantec Security Response 2017), in an effort to cash in on the malware as fast as possible within a short period of time.

Shortly after the news spread about WannaCry infections hitting the UK health care system, Marcus Hutchins, known under the alias of MalwareTech, found a built-in kill switch that allowed one to shut down WannaCry from within. The malware expert employed by an LA-based security firm analyzed a sample of the ransomware and

identified it was connecting back to an unregistered domain. Hutchins then decided to buy the domain to test the reaction of the malware, what in turn activated the kill switch and disabled its intended global spread. The fact that the registration killed the malware was a surprise to the cybersecurity expert, who only realized the profound impact of the action afterward (Khomami and Solon 2017). According to Hutchins, the purchase of the domain had been made to facilitate the study of the botnet afterward, which is a common practice at the company where he worked and would give them further insight into the functioning of the malware (Khomami and Solon 2017). Even if unknowingly, the actions of Hutchins were fundamental in hampering the dissemination of WannaCry and gave the authorities and businesses the necessary edge to alert users and minimize the impact of the infections.

The case of WannaCry is distinctively interesting because the actions of the vigilante were not necessarily intended at halting the attack but, as described by Hutchins himself, to allow him and his colleagues to later investigate the operations of the malware. It is difficult to place this particular event under the umbrella of defense of other, since the causality link is not certain. In the case of WannaCry, the reaction of the vigilante is proportional, necessary, and timely, though not moved by the purpose of halting or preventing the infliction of damage to another. Although these reaction could be construed as an accidental use of defense of other, the accidental nature of the intervention preempts the use of the exception, since the actions grounded on defense of other must be preceded by the genuine belief that the interference is aimed at preventing harm to another who is believed to be at risk or in imminent danger. Because the interference was non-intentional, the disruption of WannaCry cannot be classified as a legitimate use of defense of other.

Even though exculpation on the grounds of defense of other would not hold, it is possible to claim that, in responding to the malware, Hutchins did not use force – and therefore did not violate the legal system in that regard. The actions undertaken by Hutchins were arguably within the limits of the legal framework and did not overstep regulatory boundaries, as a *prima facie* analysis of the events suggests. However, the action of registering the domain and activating the kill switch could still be interpreted as a digital act of violence against the perpetrator of the attack. By interfering with the operation of the malware and diverting its communications, Hutchins could have made use of what in digital environments is a violent behavior. Yet, in doing so, Hutchins did not infringe any specific legislation, in that his response to the malware was as simple as registering the domain hardcoded in Wannacry and therefore could not be regarded as a form of illegal interference.

The analysis run by Hutchins does not seem to indicate further offences could have been committed, since the agent examined a sample which was publicly available and could be easily acquired by any user operating an unpatched Windows OS. Inquiring whether the specific techniques deployed by Hutchins to find the kill switch would be classified with a hacking offence, however, seems innocuous. Hacking is the product of an unauthorized and unlawful interference; once in possession of the malware sample acquired via the Internet, the Hutchins had subsumed the property rights over the malware and had thus freedom to study the code.

Considering the above, the actions of Hutchins should not be interpreted as a form of defense of other. Furthermore, although the interference itself might not be regarded as a

criminal offence, it does not necessarily lead to the interpretation that the interference is by no means an act of digital violence. Hence, the disruption of WannaCry by Hutchins could still be regarded as a form of vigilantism. On an important note, even if the agent would have used force without the clear understanding that his actions would ultimately lead to the disruption of the malware – what would prevent exculpation of the use of force on the grounds of defense of other – the actor should have not been punished for his non-intentional act. It is my interpretation that it would have been reasonable to offer him the same level of protection as it would have been afforded in case his actions had been motivated by the will to interfere with the criminal act – what would call for the shield of defense of other. Understanding otherwise would be of no value to the criminal justice system and severely punish the actions of an agent who ultimately contributed to defeating a serious threat to public security.

5. Data protection concerns

While vigilantism is a typical expression of the use of force, the results of vigilantes' activities in cybersecurity may fall out of the scope of criminal law and inside the realm of civil liability. One clear example is the potential interference with the right to data protection, seen that various cybersecurity techniques imply the observation and analysis of information that can be regarded as personal data under the GDPR (EU General Data Protection Regulation).[5] IP and email addresses, as well as other sources of data that can serve to single-out users fall within the notion of personal data and their use in cybersecurity vigilantism calls for the application of the GDPR, therefore limiting the lawful processing that these forms of data may undergo in the EU.

For the discussion at hand, the most important aspects of the GDPR revolve around the provision of lawfulness of processing, namely art. 6(1), which defines the grounds in which personal data can be deployed in compliance with EU data protection law. The appropriation by cybersecurity vigilantes of information concerning the whereabouts and identity of cybercriminals, as well as victims of a cybercrime, may be a problem from a data protection perspective. To avoid incurring in a data protection violation, cybersecurity vigilantes must ensure their activities are covered by at least one of the circumstances where processing of personal data is regarded as lawful under the GDPR.

This is particularly complicated, since the actions of cybersecurity vigilantes are unbeknownst to data subjects. One possible solution for this issue is exploring whether subparagraph (e) of art. 6(1), which authorizes the processing of personal data for the carrying out of a task in the public interest, could serve as a justifiable ground for the actions of cybersecurity vigilantes. Looking back at the cases examined before, if Janit0r and Hutchins are capable of demonstrating to a reasonable degree that the actions undertaken have been motivated by the purpose of protecting the integrity of information systems, the functioning of the Internet, and users themselves, their activities could be presumed as a processing conducted in the public interest. However, the main obstacle to the use of subparagraph (e) is the lack of legal certainty around the application of this provision to the realm of cybercrime and, more specifically, vigilantism by ethically motivated third parties. Recital 45 of the GDPR determines that the scope and reach of art. 6(1)(e) is a task for the national or EU legislator, what creates substantial blur over the application of the provision EU-wide – and the opportunity for uneven levels of protection and

regulatory standards at national level, what was one of the issues that supposedly motivated the repeal of the Data Protection Directive. In conclusion, while there is room to believe the activities of cybersecurity vigilantes could be in line with data protection regulation, the case is to be further analyzed following national legislation or future EU law regulating the matter. Failure to comply with the GDPR subjects the agent to penalties established at the national level – which are often established in the form of fines.

Aside the potential threat to the right to data protection, the actions of vigilantes could also be examined from a substantive criminal law perspective. It could be of relevance to consider whether the actions of vigilantes in fact mounted to crimes against the information systems and the data they intended to protect. These considerations could lead to a deeper investigation of whether vigilantism in the cases of Brickerbot and WannaCry constituted illegal access to information systems (Art. 2, CoE Convention on Cybercrime, Art. 3, Directive 2013/40/EU), illegal interference with information systems (Art. 4, CoE Convention on Cybercrime, Art. 4, Directive 2013/40/EU) or illegal interference with data (Art. 5, CoE Convention on Cybercrime, Art. 5, Directive 2013/40/EU). However, the focus of this paper is on the matter of vigilantism as obstruction of justice, as shall be discussed further. Future research on the potential criminal offences committed by cybersecurity vigilantes is welcome and encouraged, as the phenomenon is becoming more prevalent.

6. Tampering with evidence

Unofficial crime control in the form of cybersecurity vigilantism has flourished on the Internet. As discussed above, non-State actors have enforced informal means of justice to counter criminal behavior and have often succeeded at it. As a result, the actions of cybersecurity vigilantes have become a recurrent (and sometimes disruptive) element in the fight against cybercrime outside the course of an investigation. But individuals who make use of force in response to criminal activities online could be posing a serious threat to cybercrime investigations. Vigilantes' responses to perceived malicious activity could cause the loss of digital evidence, thereby obstructing law enforcement's effort in ascertaining attribution and jurisdiction over cybercrime offences. As a result, cybersecurity vigilantism could have collateral effects on the long-term results of criminal justice.

Criminal law has a distinctive interest in protecting the success and integrity of investigations. For that reason, obstructing justice and corrupting criminal evidence are regarded as criminal offences – offences against the administration of justice and the criminal justice system. The debate that arises is whether cybersecurity vigilantes, in trying to fill in the gaps left by the authorities, could also be hampering criminal investigations by unknowingly corrupting evidence. Tampering with evidence is a fundamental criminal offence in civil and common law regimes, the reason why I chose to examine the criminal codes of the Netherlands and France to further understand the elements of this criminal type and whether the actions of cybersecurity vigilantes run the risk of obstructing justice.

Art. 434–4 of the French Criminal Code[6] is situated under Section 1 – Obstacles to the pursuit of justice, Chapter III, Title III, Book IV, and punishes with imprisonment and a fine whoever (1) modifies the place of a crime or misdemeanor by altering, falsifying or erasing traces or clues by the taking, removing, moving, or suppressing objects, or (2) destroys, subtracts, receives, or alters a public or private document or an object with the intention of facilitating the discovery of crime or misdemeanor, search for evidence, or conviction of

suspects. A quick analysis of the provision suggests that paragraph 1 of the offence is applicable both in its intentional and negligent form, but paragraph 2 requires the conscience of the criminal act, which is committed with a specific purpose in mind. For the purpose of cybersecurity vigilantism, only paragraph 2 is relevant, since the physical place and objects of a crime are rarely affected by the actions of vigilantes. The actions of cybersecurity vigilantes take place in the realm of data flows and systems operations, which is related to the tampering of private documents (data). Nonetheless, even if cybersecurity vigilantes interfere with evidence in the course of their activities, French criminal law only makes punishable those actions which are intendedly launched with the goal of interfering the pursuit of criminal justice. Since cybersecurity vigilantes intervene to fill in the gaps of State response, it is hard to see how the threshold of intent would have been demonstrated when no such intention exists. The requirement that the tampering with evidence be intentional is sufficient to exclude the actions of various cybersecurity vigilantes, including those of Janit0r and Hutchins, from criminal liability under French law.

In the Netherlands, tampering with evidence is criminalized under art. 200 of the criminal code.[7] Here again the criminal offence is marked by the presence of intent. Article 200 punishes with up to 3 years of imprisonment or a fine whoever intentionally destroys, damages, renders unusable, or disposes of objects intended to serve the competent authorities as evidence or proof, instruments, documents, or registers permanently or temporarily held for safekeeping by order of the authorities, or which have been handed over either to a civil servant or any other person in the interest of public service.[8] By the wording of the Dutch criminal code, cybersecurity vigilantes would most likely fall out of the scope of Art. 200. However, since Dutch criminal law operates under a low threshold of intent, which regards as intentional the actions of an agent who assumes the risk of a foreseeable result, cybersecurity vigilantes who anticipate that their actions may have a negative impact on future cybercrime investigations by law enforcement could be held criminally liable under Art. 200. Nevertheless, given the discretionary powers afforded to the criminal prosecutors in the Netherlands, who make an evaluative assessment of the cases that require the attention of the State, it is unlikely that public prosecutors would find reasonable to prosecute cybersecurity vigilantes under the low threshold of intent, especially when the actions under scrutiny have contributed to larger achievements in criminal justice.

In sum, cybersecurity vigilantes are improbably incurring into offences against the administration and pursuit of justice when their actions are not intended to disrupt the success of subsequent or ongoing criminal investigations. This scenario is ever more unlikely when there is public support for the actions of vigilantes and concrete public interest gains arising from their active citizenship.

7. Cooperative criminal justice

The theory of cybersecurity is marked by the concept of shared responsibility. In cybersecurity literature, multistakeholderism is a grounding precept of successful and holistic operations. This concept is grounded on the fact that the architecture of the Internet and information systems permeates public and private infrastructures managed by a diversity of actors. These actors, each at their own stance, can influence the outcome of a security threat based on the decisions made at their control level. In the realm of

regulation, the concepts of network models and nodal regulation have been largely discussed, exposing the intricate relationship between operational control and regulatory power exercised by the broad spectrum of agents involved in the functioning of the Internet and information systems. In digital environments, multiple actors perform a dual regulator-regulatee role, as the decisions made at given control level reverberate through other circles of the web. Harnessing this unique interplay seems not only strategic but absolutely necessary for regulatory success.

Following the premise of multistakeholderism, States, businesses, and citizens alike are called upon to exercise their fair share of responsibility and will be held accountable (legally or morally) for failing to meet these standards. In the EU, States carry out the responsibility for implementing cybersecurity legislation and monitoring its application. Businesses are bound by national and EU regulation determining the technical standards of cybersecurity to be observed in the development of products, as well as procedures for reporting security incidents and principles for processing personal data in a lawful manner. Individuals are reasonably expected to keep their devices up to date and away from engaging in activities that undermine public security, which includes the security of information systems. Cybersecurity thus emerges as a collaborative effort, and a goal which is only achievable as long as all actors involved take ownership of their fair share (of the problem and the solution).

Considering that minimizing cybercrime is a core aspect of cybersecurity, the perspective of multistakeholderism should remain valid here. The context of cybercrime fighting is one where the interplay between private and public actors is ever more blurred in view of the flow of relevant information undergoing privately held channels. Although the same is observable in other types of crime – such as white-collar crimes, where the evidence is often in the hands of those being investigated – cybercrime is aided by the fleeting nature of data as criminal evidence and the abundance of anonymization and security tools that may conceal one's identity and remove traces that could lead to pivotal discoveries. Furthermore, these same obstacles intensify the challenges of international cooperation and mutual assistance in criminal matters, where insufficiencies and conflicting differences between legal systems may hamper the success of investigations. Additionally, lack of political will and technical resources in a country may increase its attractiveness to cybercriminals and create a shield for criminal activities originating in these territories. As a result, prosecution of cybercrime often falls short of expectation, what could be interpreted as a probable reason for the thriving of cybersecurity vigilantism.

Cybersecurity vigilantism has gained distinguished attention because of the nature of expert knowledge in the hands of this particular group of vigilantes. The experiences of BrickerBot and Wannacry ransomware reveal the valuable input citizens may have in halting cybercrime and supporting the activities of law enforcement. In the mentioned cases, intervention arising from society, namely from individuals ethically imbued with the sense of participative citizenship, delivered a positive, desirable result which was influential in promoting and preserving cybersecurity. The main lesson drawn from these recent episodes of vigilantism is that the goal of enforcing the law in digital environments can be partly assisted by the help of cybersecurity vigilantes. A purist would rule out the participation of non-State agents in law enforcement, based on the tenet of the monopoly of the use of force. A pragmatic would consider what the potential perks of allowing for this form of contributory justice are. I embrace the second viewpoint and defend that social movements should be understood as legitimate expressions of society for criminal

justice and that a path to legality of cybersecurity vigilantism should be considered by the legislator.

8. A path to legality

A few regulatory recommendations come to mind after running this legal analysis. First, the activities of vigilantes are seldom a concern of regulatory activity and have largely been governed by criminal law applicable to non-vigilantism. The lack of political will in regulating vigilantism as a *sui generis* form of contributive citizenship fortifies the principle that the State controls the monopoly on the use of force. Second, the activities of cybersecurity vigilantes have important repercussions in the field of data protection, as they touch upon data identifiers which may only be processed under lawful grounds. While art. 6(1)(e) of the GDPR opens doors for the processing of personal data by vigilantes, the absence of further regulation specifying the scope and exact interpretation that can be given to the expression 'task in the public interest' frustrates this expectation. Only Union and Member State law may define the precise terms in which art. 6(1)(e) can be used and until this further specification is passed into law or extended to the realm of cybersecurity, any attempt to justify processing of personal data as a task in the public interest is to no avail.

A regulatory framework that confronts the legal void of cooperative criminal justice would be greatly welcome. By setting permissions and boundaries to the activities of vigilantes, the State would acknowledge the role of vigilantism that, under specific circumstances, is welcomed by society. Ignoring the actions of vigilantes and the growing phenomenon of cybersecurity vigilantism will not prevent citizens from attempting to filling in the gaps in the response of official authorities. In effect, it may as well foster a parallel social culture of cybersecurity where individuals feel they are to take matters onto their own hands, and a legal culture where cybercriminals and active citizens are penalized as equals.

Offering a path to legality would support the activities of cybersecurity researchers, offer legal certainty in the research of cybercrime, and guide active citizens towards a responsible and ethical way of contributing to criminal justice. Provisions illuminating the scope of defense of other within the realm of cybersecurity and cybercrime, aside from potential shields from criminal and civil liability covering the activities of cybersecurity vigilantes should be seriously considered by the EU and national regulators. Finally, it would be opportune for the Article 29 Working Party to issue an opinion on the scope and uses of art. 6(1)(e) of the GDPR, similarly to Opinion 06/2014 on the 'Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC' (Article 29 Data Protection Working Party 2014). This could expose the exact concept of public interest under art. 6(1)(e), and whether and under what circumstances the activities of cybersecurity vigilantes should be covered by that provision.

9. Conclusion

Cybersecurity vigilantism, a concept advanced in this paper, adds to the pressure cook that cybercrime justice has become. Here, anonymity tools, encryption, limited State resources, insufficient technical knowledge by law enforcement, and jurisdictional boundaries

grounded on territoriality contribute to a complex scenario where investigations are regarded as of humble or delayed effect on preserving Internet security. The perception that the official response to cybercrime is inadequate or tardy, and that cybercriminals flourish and profit from our insufficient response, form the arguable motive driving the upsurge of cybersecurity vigilantism.

In the previous sections, it was argued that cybersecurity vigilantism is an emerging phenomenon with important implications for criminal justice. The challenges of furthering cybersecurity vigilantism, however, are many, flowing from the sword of criminal liability (cybercrime offences and tampering with evidence) to data protection violations. Ultimately, the question that this paper poses revolves around society's willingness to regulate the engagement of cybersecurity vigilantes in an environment of emerging cooperative criminal justice. I conclude with the idea that cybersecurity vigilantes have a role to play in the advancement of cybersecurity and should not be excluded from this process. Finally, I contend that further regulatory clarifications that could facilitate the activities of these agents and provide a path to legality, as well as delineate which undertakings are bordering criminal and civil liability, would be a great addition to the debate and help illuminate the gray zone where active citizenship in cybersecurity is currently trapped.

Notes

1. Oxford Dictionary of English, terms 'vigilant' and 'vigil'.
2. Human-flesh search engines – renrou sousuo yinqing – have become a Chinese phenomenon: they are a form of online vigilante justice in which Internet users hunt down and punish people who have attracted their wrath. The goal is to get the targets of a search fired from their jobs, shamed in front of their neighbors, run out of town. It's crowd-sourced detective work, pursued online – with offline results (See Downey 2010).
3. The website KrebsonSecurity was hit by an unusually powerful DDoS attack in September 2016, as a result of a Mirai attack (See Krebs, KrebsOnSecurity Hit With Record DDoS 2016a).
4. As noted in a recent report from Flashpoint and Level 3 Threat Research Labs, the threat from IoT-based botnets is powered by malware that goes by many names, including 'Lizkebab', 'BASHLITE', 'Torlus' and 'gafgyt'. According to that report, the source code for this malware was leaked in early 2015 and has been spun off into more than a dozen variants (See Krebs, KrebsOnSecurity Hit With Record DDoS 2016).
5. At the time of writing, the GDPR was only months away from coming into force, what justified the focus of the analysis on the first rather than on the Data Protection Directive 95/46/EC.
6. Article 434-4
   Est puni de trois ans d'emprisonnement et de 45 000 euros d'amende le fait, en vue de faire obstacle à la manifestation de la vérité:

   (1) De modifier l'état des lieux d'un crime ou d'un délit soit par l'altération, la falsification ou l'effacement des traces ou indices, soit par l'apport, le déplacement ou la suppression d'objets quelconques;
   (2) De détruire, soustraire, receler ou altérer un document public ou privé ou un objet de nature à faciliter la découverte d'un crime ou d'un délit, la recherche des preuves ou la condamnation des coupables.

   Lorsque les faits prévus au présent article sont commis par une personne qui, par ses fonctions, est appelée à concourir à la manifestation de la vérité, la peine est portée à cinq ans d'emprisonnement et à 75 000 euros d'amende.

7. Artikel 200

   (1) Hij die opzettelijk zaken, bestemd om voor de bevoegde macht tot overtuiging of bewijs te dienen, akten, bescheiden of registers die voortdurend of tijdelijk op openbaar gezag bewaard worden, of hetzij aan een ambtenaar, hetzij aan een ander in het belang van de openbare dienst zijn ter hand gesteld, vernielt, beschadigt, onbruikbaar maakt of wegmaakt, wordt gestraft met gevangenisstraf van ten hoogste drie jaren of geldboete van de vierde categorie.

   (2) Onder bevoegde macht wordt mede verstaan: een internationaal gerecht dat zijn rechtsmacht ontleent aan een verdrag waarbij het Koninkrijk partij is.

8. Translation retrieved from http://www.ejtn.eu/PageFiles/6533/2014%20seminars/Omsenie/WetboekvanStrafrecht_ENG_PV.pdf.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This article is part of the research developed within the project "Public-private actions against botnets: establishing the legal boundaries" ('BotLeg') funded by the Netherlands Organisation for Scientific Research (NWO). The BotLeg consortium is a collaboration between Tilburg University (TILT), SURFnet, The Dutch National High Tech Crime Unit, SIDN, Abuse Information Exchange, and Leaseweb.

References

Article 29 Data Protection Working Party. 2014. *Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller Under Article 7 of Directive 95/46/EC*. Brussels: European Commission.

Cimpanu, Catalin. 2017. "BrickerBot Author Claims He Bricked Two Million Devices." April 21. Accessed August 31, 2017. https://www.bleepingcomputer.com/news/security/brickerbot-author-claims-he-bricked-two-million-devices/.

Costello, John, Allison Nixon, Brian Hein, Ronnie Tokazowski, and Zach Wilkhom. 2016. "New Mirai Variant Leaves 5 Million Devices Worldwide Vulnerable." November 29. Accessed August 31, 2017. https://www.flashpoint-intel.com/blog/cybercrime/new-mirai-variant-involved-latest-deutsche-telekom-outage/.

Dizon, Michael. 2016. *Breaking and Remarking the Law*. Tilburg: Tilburg University.

Downey, Tom. 2010. "China's Cyberposse." March 3. Accessed August 31, 2017. http://www.nytimes.com/2010/03/07/magazine/07Human-t.html?pagewanted=all&mcubz=3.

Dumsday, Travis. 2009. "On Cheering Charles Bronson: The Ethics of Vigilantism." *The Southern Journal of Philosophy* 47: 49–67.

Haas, Nicole Evelin. 2010. *Public Support for Vigilantism*. Leiden: Leiden University.

Johnston, Les. 1996. "What is Vigilantism." *The British Journal of Criminology* 36 (2): 220–236.

Khomami, Nadia, and Olivia Solon. 2017. "'Accidental Hero' Halts Ransomware Attack and Warns: This Is Not Over." May 13. Accessed August 31, 2017. https://www.theguardian.com/technology/2017/may/13/accidental-hero-finds-kill-switch-to-stop-spread-of-ransomware-cyber-attack.

Krebs, Brian. 2016a. "KrebsOnSecurity Hit With Record DDoS." September 21. https://krebsonsecurity.com/2016/09/krebsonsecurity-hit-with-record-ddos/.

Krebs, Brian. 2016b. "New Mirai Worm Knocks 900K Germans Offline." November 30. Accessed August 31, 2017. https://krebsonsecurity.com/2016/11/new-mirai-worm-knocks-900k-germans-offline/.

Leyden, John. 2016. "Mirai IoT Botnet Blamed for 'Smashing Liberia Off the Internet'." November 4. https://www.theregister.co.uk/2016/11/04/liberia_ddos/.

Millman, Rene. 2017. "BrickerBot 'Creator' Claims Two Million IoT Devices have been Destroyed." April 25. Accessed August 31, 2017. https://internetofbusiness.com/brickerbot-iot-devices-destroyed/.

Radware. 2017. "BrickerBot PDoS Attack: Back With a Vengeance." April 21. Accessed August 31, 2017. https://security.radware.com/ddos-threats-attacks/brickerbot-pdos-back-with-vengeance/.

Rosenbaum, Ron. 2007. "How to Trick an Online Scammer into Carving a Computer Out of Wood." June. Accessed August 31, 2017. https://www.theatlantic.com/magazine/archive/2007/06/how-to-trick-an-online-scammer-into-carving-a-computer-out-of-wood/305903/.

Smallridge, J., P. Wagner, and J. Crowl. 2016. "Understanding Cyber Vigilantism." *Journal of Theoretical & Philosophical Criminology* 8 (1): 57–70.

Symantec Security Response. 2017. "What You Need to Know About the WannaCry Ransomware." May 23. https://www.symantec.com/connect/blogs/what-you-need-know-about-wannacry-ransomware.

Trottier, Daniel. 2017. "Digital Vigilantism as Weaponisation of Visibility." *Philosophy and Technology* 30: 55–72.