Processing of botnet tracking data under the GDPR

Böck, Leon; Andersen, Martin Fejrskov; Demetzou, Katerina; Karuppayah, Shankar ; Mühlhäuser, Max; Vasilomanolakis, Emmanouil

Published in:
Computer Law & Security Review

DOI (link to publication from Publisher):
10.1016/j.clsr.2021.105652

Creative Commons License
CC BY 4.0

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Böck, L., Andersen, M. F., Demetzou, K., Karuppayah, S., Mühlhäuser, M., & Vasilomanolakis, E. (2022). Processing of botnet tracking data under the GDPR. *Computer Law & Security Review*, *45*, Article 105652. https://doi.org/10.1016/j.clsr.2021.105652

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/CLSR

Computer Law & Security Review

Processing of botnet tracking data under the GDPR

Leon Böck[a,*], Martin Fejrskov[b], Katerina Demetzou[c], Shankar Karuppayah[d], Max Mühlhäuser[a], Emmanouil Vasilomanolakis[e]

[a] Telecooperation Lab, Technische Universität Darmstadt, Darmstadt, Germany
[b] Telenor, Aalborg, Denmark
[c] Radboud Business Law Institute, Radboud University, Nijmegen, The Netherlands
[d] National Advanced IPv6 Centre, Universiti Sains Malaysia, Malaysia
[e] Cyber Security Group, Aalborg University, Denmark

ARTICLE INFO

Article history:

Keywords:
Botnets
GDPR
Legal ground
Legitimate interest
Research

ABSTRACT

Botnet research is one of the many research areas affected by the coming into force of the General Data Protection Regulation (GDPR). This article aims to identify the most appropriate legal bases that would legitimise data processing in the context of botnet tracking and to give an overview of the practical implications for practitioners. First, we give a technical introduction to botnet tracking techniques and the types of processed data. Afterward, we argue that botnet tracking qualifies as ”processing of personal data” and falls under the material scope of the GDPR. We then present three scenarios where these botnet tracking techniques apply: botnet tracking research in the public interest, botnet tracking in the commercial interest and botnet tracking conducted by Internet service providers. For each scenario, we discuss the differing goals, identify the appropriate legal bases, and elaborate on the practical implications. This article concludes that the legal implications are very different for each of the three scenarios, highlighting the importance of carefully considering the legal bases before engaging in botnet tracking.

1. Introduction

Botnets (which is short for robot networks) are networks of interconnected malware-infected devices, commonly referred to as bots. A botnet is an extension of traditional malware, i.e., malicious software[1] and belongs to the broad category of 'high-tech cybercrimes'.[2] Botnets have been characterised as

[1] A malware 'infiltrates and gains control over a computer system or a mobile device to steal valuable information or damage data' (EUROPOL, https://www.europol.europa.eu/crime-areas-and-trends/crime-areas/cybercrime), last accessed 28/10/2021.

[2] EUROPOL, https://www.europol.europa.eu/crime-areas-and-trends/crime-areas/cybercrime, last accessed 28/10/2021.

* Corresponding author.
E-mail addresses: boeck@tk.tu-darmstadt.de (L. Böck), mfea@telenor.dk (M. Fejrskov), k.demetzou@cs.ru.nl (K. Demetzou), karuppayah@tk.tu-darmstadt.de (S. Karuppayah), max@informatik.tu-darmstadt.de (M. Mühlhäuser), emv@es.aau.dk (E. Vasilomanolakis).

https://doi.org/10.1016/j.clsr.2021.105652

the 'backbone of modern criminality'.[3] The infected devices, the bots, are remotely controlled by the so-called botmaster, without the user's knowledge.

The botmaster can freely instruct the bots to carry out cyber-attacks such as Distributed Denial of Service (DDoS), information theft, ransomware extortion or phishing, while preserving their anonymity at the same time. Due to the feature of anonymity, botnets have become a preferred tool for cyber criminals.[4] 'They provide low-cost, high-profit opportunities with only a minuscule risk of identification, sanction, punishment or arrest.'[5]

Often, the defence against botnets is reactive, disorganized, and responsive against a particular attack that has taken place.[6] To address these problems, highly accurate botnet tracking data is necessary. This article aims to investigate the legal circumstances surrounding the collection and processing of data necessary to facilitate strategic disruption of botnets.[7]. 'Disruption stands for the broad range of actions that purposefully disturb an ongoing botnet infection or attack.'[8]

Disrupting botnets is an ongoing effort among researchers, security companies, Internet Service Providers (ISPs) and law enforcement agencies. The success of such disruption efforts relies greatly on detecting and taking down the Command and Control (C2) infrastructure employed by the botmasters. To effectively disrupt a modern botnet, one has to either cut the communication to the botmasters or disinfect all bots. Regardless of the technical challenges to carry out a disruption, one must first identify and track the infected machines. This process of collecting and processing information is called *tracking*. Defenders achieve this by infiltrating the botnets with specifically designed software or monitoring network (e.g. Domain Name System (DNS)) traffic. This allows the defenders to enumerate the infected machines, identify their network location, track commands issued by botmasters and retrieve information about the connections between individual bot infections. This information can then be used to track botmaster activities and infections, to eventually disrupt the botnet or capture the botmasters themselves.

Tracking botnets, which is critical to any disruption efforts, requires processing of personal information such as IP addresses. Any processing of information that qualifies as personal data falls under the protective scope of the General Data Protection Regulation (GDPR). As it will be argued in this article, tracking activities that aim at disrupting botnets fall under the definition of *processing of personal data* and for this reason they have to comply with the legal principles and rules set out in the GDPR. This article aims to investigate to what extent and under what circumstances botnet tracking can be carried out by three different actors who are (or should be) highly involved in any disruption efforts against botnets. The three application scenarios examined in the article are tracking activities carried out for research in the public interest, commercial interest, and those by Internet Service Providers (ISPs).

The role of law enforcement agencies and similar authorities is relevant to mention in this context. Law enforcement agencies often collaborate with and rely on the aforementioned parties for botnet discovery and data collection[9] and are therefore not typically using botnet tracking techniques or performing botnet research themselves. Also, such authorities are outside the GDPR's material scope[10] and are instead regulated by the Law Enforcement Directive[11] as transposed in national law by the EU Member States. For these reasons, we consider law enforcement agencies to be out of the scope of this article.

Our goal is to answer the following research question: *What are the practical implications for researchers, private companies, and ISPs when engaging in botnet tracking activities so as to comply with the European general legal framework on data protection (GDPR)?* By answering this question, we hope to shed light on how botnet tracking can be conducted lawfully under the GDPR.

The remainder of this article is structured as follows. Section 2 presents the technical context by introducing related work on the topic of botnet tracking. Section 3 presents the legal context and more specifically the general EU legal framework on data protection (i.e., the General Data Protection Regulation), and it explains the reasons why the GDPR becomes relevant in the discussion on botnet tracking. Sections 4, 5 and 6 particularise Section 3 by discussing the applicability of the GDPR to the three scenarios, i.e., researchers, private companies and ISPs. Section 7 concludes this article with a discussion and outlook for future work.

[3] Karine K. e Silva (2017) How industry can help us fight against botnets: notes on regulating private-sector intervention, International Review of Law, Computers & Technology, 31:1, 105-130, 106 DOI: 10.1080/13600869.2017.1275274.

[4] ENISA (2011), 'Botnets: Measurement, Detection, Disinfection and Defence', http://www.enisa.europa.eu/activities/Resilience-and-CIIP/networks-and-servicesresilience/botnets/botnets-measurement-detection-disinfection-and-defence.

[5] ENISA (2011), 'Botnets: Measurement, Detection, Disinfection and Defence', http://www.enisa.europa.eu/activities/Resilience-and-CIIP/networks-and-servicesresilience/botnets/botnets-measurement-detection-disinfection-and-defence.

[6] Martin, A. K, Andrade, N. N. G.de, 'Battling Botnets with Digital Rights in Mind', European Journal for Law and Technology, Vol. 3, No. 2, 2012, 1.

[7] C. Rossow, D. Andriesse, T. Werner, B. Stone-Gross, D. Plohmann, C. J. Dietrich, H. Bos, Sok: P2pwned-modeling and evaluating the resilience of peer-to-peer botnets, in: IEEE Symposium on Security and Privacy, 2013, pp. 97-111.

[8] K.K Silva, How industry can help us fight against botnets, 112. Disruption being one of the four pillars according to Silva, i.e., (1) prevention (2) information exchange (3) disruption (4) disinfection efforts.

[9] U.S. Leads Multi-National Action Against "Gameover Zeus" Botnet and "Cryptolocker" Ransomware, Charges Botnet Administrator. http://www.justice.gov/opa/pr/us-leads-multi-national-action-against-gameover-zeus-botnet-and-cryptolocker-ransomware

[10] Article 2(2d) GDPR.

[11] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA.

2. Background on botnet tracking

This section introduces background information on botnet tracking. We start with a description of different botnet types based on their different network architectures and techniques for resilient and stealthy C2. Next, we discuss the tracking techniques that can be employed for the different botnet types. Finally, we discuss the Personally Identifiable Information (PII) collected by using the different tracking techniques.

2.1. *Botnets*

Botnets can have different network architectures describing how bots are interconnected to their botmaster. Traditionally, botmasters formed a centralized network, in a client-server fashion, where a single server was used to control and communicate all their bots. However, the centralized control server presents a single point of failure, the central server, that can be targeted to disrupt the botnet completely. To overcome this, botmasters have adopted advanced C2 infrastructures such as Domain Generation Algorithms (DGAs) or fast-flux Domain Name System (DNS) to conceal their C2 servers.[12]

DGAs provide a remedy for the single point of failure of centralized botnets by leveraging the Domain Name System (DNS). Instead of addressing a C2 server through a unique IP address or domain name, this type of malware can fall back to generate new domain names in a deterministic fashion. In the event a C2 server becomes unreachable, the bots start to compute and connect to these newly generated domain names. A botmaster only needs to register one of these domain names and point it to a new C2 server to restore control over the infected machines.

Moving beyond DGA techniques, botmasters fully distribute the C2 channel by leveraging Peer-to-Peer (P2P) networks.[13] The use of such P2P networks for C2 purposes completely negates the single point of failure. It makes every infected machine capable of sending botmaster commands to all other bots. Peer-to-Peer (P2P) botnets rely on a resilient and difficult to disrupt C2 channel leveraging P2P network technology. In a P2P network, there is no central server that controls the infected machines. Instead, every client (referred to as peer) acts as both client and server simultaneously, distributing the server functionality among all peers. Therefore, every peer can distribute malware updates, disseminate botmaster commands, retrieve stolen information, or infect other vulnerable machines.

To attain such functionality, the bots need to ensure that they remain connected to the P2P network. This is commonly achieved by maintaining a list of other bots, the so-called Neighborlist (NL). The NL is frequently checked and updated to ensure that the stored entries are still active and reachable. It stores other bot's IP addresses and ports alongside additional information such as timestamps when the bot was last seen. To remain connected to each other, bots maintain a NL of other active bots. This NL is frequently updated by replacing inactive bots with active bots. To replace a peer, a bot sends *NL-request* messages to active peers within their NL, asking them to share IP-address and port of active bots from their NL. While the P2P architecture makes these botnets very resistant to disruptions, the communication protocol is inherently open to allow new bots to join the network. Defenders leverage this circumstance to develop software that allows them to infiltrate and monitor P2P botnets.

Table 1 – Applicability of botnet tracking techniques' for different botnet types. Legend: √= applicable, (√) = sometimes applicable, ✗= not applicable.

Botnet tracking technique	Centralized	P2P	Fast flux	DGA
Targeted:				
Crawlers	(√)	√	(√)	✗
Sensors	✗	√	✗	✗
Domain takeover	√	✗	√	√
Non-Targeted:				
Network traffic monitoring	√	√	√	√
DNS traffic monitoring	(√)	✗	(√)	√

2.2. *Tracking techniques*

We broadly differentiate between defender techniques that are: *i)* active and targeted, i.e., specifically designed to obtain botnet tracking information only and *ii)* passive, non-targeted approaches that retrieve botnet tracking data from existing source (e.g., by analyzing general network traffic data). An overview of the capabilities of the different tracking techniques described in this section can be found in Table 1.

2.2.1. *Targeted techniques*

Crawlers A crawler generally describes software that iteratively contacts Internet-connected devices to discover information. For botnet tracking, it can be used to discover C2 servers, track P2P botnet infections or identify devices in a fast-flux botnet. Crawlers are most prominent for discovering P2P bot infections by continuously requesting NL-entries from bots.[14] However, crawlers in P2P botnets can not discover devices behind firewalls or Network Address Translation (NAT) devices, which are common for most private networks. Moreover, webcrawlers such as Shodan[15] can be used to identify botnet C2 servers if they provide public interfaces, e.g., HTTP or HTTPS pages. Lastly, crawlers can be used to frequently contact a fast-flux C2 server to obtain a list of all bots participating in the fast-flux network.

Sensors

Sensors are specific to P2P botnets and provide more accurate enumerations of botnets than crawlers. They overcome

[12] M. Singh, M. Singh, S. Kaur, Issues and challenges in DNS based botnet detection: A survey, Computers & Security 86 (2019) 28-52.

[13] L. Bock, E. Vasilomanolakis, M. Muhlhauser, S. Karuppayah, Next generation P2P botnets: monitoring under adverse conditions, in: International Symposium on Research in Attacks, Intrusions, and Defenses, Springer, 2018, pp. 511-531.

[14] C. Rossow, D. Andriesse, T. Werner, B. Stone-Gross, D. Plohmann, C. J. Dietrich, H. Bos, Sok: P2pwned-modeling and evaluating the resilience of peer-to-peer botnets, in: IEEE Symposium on Security and Privacy, 2013, pp. 97-111.

[15] https://www.shodan.io/, last accessed 28/10/2021.

the aforementioned limitation of crawlers by passively waiting for messages from bots behind firewall and NAT devices. A sensor imitates the behavior of a regular bot by responding to probe messages from other bots. By remaining active within the botnet for prolonged periods, sensors become popular within the botnet. They will eventually be contacted by the majority of the bot population, enabling tracking of the entire P2P botnet.

Domain takeover

A common approach to track the bot population of centralized, fast-flux, and DGA botnets is to take over one of the C2 domains. This can be achieved by registering a yet unregistered domain found in the botnet malware or by going through legal processes to obtain control over a C2 domain. Once a C2 domain is in the control of the defenders, all infections contacting that domain can be tracked.

2.2.2. Non-targeted techniques

Network traffic monitoring In contrast to the aforementioned targeted techniques, a passive approach can be used, where the Internet communication between two or more bots is observed as it transits various routers on the Internet. This approach requires access to the Internet infrastructure and is therefore typically performed by larger companies or ISPs.

The monitoring is performed by bulk collection or analysis of many or all Internet data packets, as it is typically hard to know which data packets are related to the botnet before further analysis. From a legal and technical perspective, this represents a vastly different approach than the targeted approaches. While this approach is more invasive than targeted techniques, the expanded view provides unique possibilities to detect and track botnet infections. As an example, Zhuang et al.[16] leverage this perspective to detect P2P botnets based on mutual contacts and destination diversity of Internet-connected devices.

The primary advantage of the network monitoring approach is that it can be used to reveal and map new botnets of both known and unknown families. Moreover, it represents the most reliable approach to track centralized, fast-flux, and DGA based botnets if the botnets traffic signature is known.

DNS traffic monitoring

As discussed earlier in this section, the DNS is used by DGA and fast-flux-based botnets to connect to the C2 server. Furthermore, even traditional centralized botnets often use static domains to address the C2 server. This allows defenders to leverage DNS traffic to detect and track bots based on their DNS queries. In order to do this, one has to collect and process all DNS traffic from either the network itself or by running a DNS server. This information makes it possible to detect bots based on their connection attempt to known malicious domains. Advanced approaches such as Pleiades,[17] can even automatically identify the usage of a DGA and differentiate between different DGAs used by different malware. This allows them to find new, track existing and distinguish various malware using a DGA to reach their C2 server.

2.3. Collection of PII

The PII collected in botnet tracking depends on both the applied technique and the specifics of the botnet. However, all tracking techniques discussed previously rely on the collection of PII, primarily in terms of IP addresses. While other PII may become available during or as a follow up to botnet tracking, we do not go into further detail for two reasons: 1) If and what type of additional PII is specific to a botnet and not generally applicable to all botnets, and 2) the applicability of the GDPR is not affected by the amount of PII, but the fact that any PII is processed (c.f. Section 3). Therefore, we focus on IP addresses as they are considered PII and collected in any botnet tracking activities.

Nevertheless, we want to point out, that additional types of information specific to the technique or botnet may be collected during the tracking process. For instance, the DNS monitoring technique will collect the requested domain name and the sensor technique may collect the software version of the probing bot. For specific botnet tracking campaigns previous research has reported the collection of email addresses and account information[18] or the identity of the botmaster.[19]

Therefore, this work mainly differentiates who's PII is collected. Targeted techniques only collect the IP addresses of infected computers. In contrast, the non-targeted approaches collect IP addresses of the bots and benign traffic. Within the next section, we discuss the legal implications of collecting PII data as part of botnet tracking activities.

3. Applicability of the general data protection regulation (GDPR)

In this section, we examine the material applicability of the GDPR in the context of botnet tracking activities as described in Section 2. Article 2(1) GDPR, entitled 'Material Scope', reads: '[t]his Regulation applies to the processing of personal data wholly or partly by automated means [...]'. There are two main conditions which need to be fulfilled in order for botnet tracking to fall under the scope of the General Data Protection Regulation (GDPR);[20] firstly, that the information gathered, ac-

[16] D. Zhuang, J. M. Chang, Peerhunter: Detecting peer-to-peer botnets through community behavior analysis, in: Dependable and Secure Computing, 2017 IEEE Conference on, IEEE, 2017, pp. 493-500.

[17] M. Antonakakis, R. Perdisci, Y. Nadji, N. Vasiloglou, S. Abu-Nimeh, W. Lee, D. Dagon, From throw-away traffic to bots: Detecting the rise of DGA-based malware., in: USENIX Security Symposium, Vol. 12, 2012.

[18] Brett Stone-Gross, Marco Cova, Lorenzo Cavallaro, Bob Gilbert, Martin Szydlowski, Richard A. Kemmerer, Christopher Kruegel, Giovanni Vigna: Your botnet is my botnet: analysis of a botnet takeover. CCS 2009: 635-647.

[19] Spanish police take down massive Mariposa botnet, https://www.reuters.com/article/urnidgns852573c400693880002576da007372f2/spanish-police-take-down-massive-mariposa-botnet-idUS378118728720100303

[20] What is also required is that the processing of personal data does not fall under Article 2(2) GDPR, whereby exceptions from the material scope are enumerated. This, however, does not apply in

cessed, used, etc., for the tracking qualifies as 'personal data' and secondly, that the activities performed upon this information qualify as 'processing'.

3.1. *Personal data*

Personal data is defined as 'any information relating to an identified or identifiable natural person'.[21] The legislature's intention is to give a broad notion to the concept of personal data 'so as to include all information concerning an identifiable individual'.[22]

In the case of botnet tracking, the following types of information will be collected: *Network packets, DNS queries*, and *auxiliary information* such as *geo-location and timestamps*. Before examining whether these types of information qualify as personal data, we first need to explain what each of these information is and why we need to collect it for botnet tracking purposes.

The purpose of collecting the aforementioned types of information is to uniquely identify and track a bot infection with utmost accuracy. What is important is identifying the machine (the bot which has been infected) and not the user of the machine (the natural person). That is not to suggest that, there is no possibility that users are also identified in the process of bot identification. However, it should be clear, that in bot tracking we are only interested in bot identification.

The minimal information necessary to track an infected bot is its IP address and port, i.e., the identifier for a specific connection on a machine and a series of timestamps. Given this information, defenders can implement countermeasures such as IP blacklists to defend their networks or execute clean-up operations on infected machines under their control. The timestamps are necessary due to the dynamic allocation of IP addresses in many consumer networks. Lacking these timestamps will lead to inaccurate information, given that a new non-infected machine may be assigned an IP address previously used by an infected machine. In addition to IP address and timestamps, there are other types of information that, if collected, will improve the defensive capabilities.

Two examples are collecting DNS queries and NL messages (see Section 2.2). Collecting DNS queries allows the identification of bots that make use of a DGA. While this mechanism makes botnets more robust, defenders can exploit it to identify and track bot infections. The second example relates to P2P botnets. By consecutively sending NL messages to all known bots, defenders can quickly discover new infections shared by the other bots. In both cases, additional information is essential to discover and monitor machines infected by the specific type of botnet malware. The information that needs to be processed to identify and track bots qualify as personal data. Even though the user of the infected machine is not directly identified, e.g., via their name, they could, however, be indirectly identified (i.e., identifiable) due to the combination of various bits of information.

3.2. *Processing*

Processing means 'any operation or set of operations which is performed on personal data or on sets of personal data whether or not by automated means, such as collection, recording, organisation [...] erasure or destruction'.[23] Processing is also a very broad term encompassing all possible activities performed upon personal data. Access, collection, storage, sharing, anonymisation, deletion, and any other activity fall under the definition of 'processing'. In the case of botnet tracking specifically, the exact activities depend on the actors who perform them, their purposes, and their possibilities (for example in terms of infrastructure). As we will see in more detail in Sections 4, 5, and 6, researchers, private companies, and ISPs are examples of actors who engage in botnet tracking and perform various activities that fall under the term 'processing'. That means that the specific purpose defines both the type of information and the activities that need to be performed to achieve this specific purpose.

3.3. *Material applicability of the GDPR: data protection principles and legal grounds*

After acknowledging the material applicability of the GDPR in the context of botnet tracking, the next step is to highlight the legal consequences.

As the Court of Justice of the European Union (CJEU) has clarified various times, '*all processing of personal data must comply, first, with the principles relating to data quality set out in [Article 5 of the GDPR] and, secondly, with one of the criteria for making data processing legitimate listed in Article [6 of the GDPR]*'.[24] Compliance with the data protection principles and identification and application of the appropriate legal ground is the responsibility of the data controller,[25] who is 'the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data'.[26] To give an example, a company and an ISP (in our second (Section 5) and third (Section 6) scenarios, respectively) are legal persons. These actors qualify as data controllers under the GDPR, given that they determine the why (purpose) and the how (means) of processing operations in the context of botnet tracking.

Article 5 GDPR requires that all data protection principles enumerated in the Article are respected. The data protection principles are the following: (a) lawfulness, fairness, transparency (b) purpose limitation (c) data minimisation (d) accuracy (e) storage limitation (f) integrity and confidentiality

the case of botnet tracking and for that reason, it is not discussed in this article.

[21] Article 4(1) GDPR.

[22] Amended proposal for a Council Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data, COM(92) 422final, 28.10.1992,10 (commentary on Art 2).

[23] Article 4(2) GDPR.

[24] CJEU TK Judgment Case C-708/18, ECLI:EU:C:2019:1064, para 36,CJEU Google Spain and Google, C-131/12, EU:C:2014:317, paragraph 71 and the case-law cited.

[25] Under the GDPR, the data controller is the actor who is primarily responsible for compliance with all the provisions set out in this law.

[26] Article 4(7) GDPR.

(g) accountability. Article 6 GDPR[27] requires that each and every processing of personal data has a legal ground (legal basis) that justifies the processing. Processing that does not have a legal basis is unlawful.

In the following sections, we will discuss the three scenarios of botnet tracking performed by researchers in the public interest (Section 4), researchers for the commercial interest of a private company (Section 5) and ISPs (Section 6). In this Section we concluded that the GDPR is applicable when botnet tracking is taking place. Therefore, in the following sections we will identify for each of the three actors the appropriate legal ground(s) (if any) to process personal data for botnet tracking and we will analyse the relevant data protection principles in more detail.

4. Research in the public interest

Our first scenario addresses data collection and analysis of active botnets by researchers at public universities or other public research institutions. We specifically differentiate between research conducted by the aforementioned entities and those conducted by private companies or other entities not acting in the immediate interest of the public.

This section will first describe the goals and purposes, followed by an analysis of the legal grounds for processing. We conclude the section by proposing technical measures for practitioners based on the identified legal grounds.

4.1. *Goals and purpose of data collection*

The primary goal of researchers is to investigate and understand botnets in order to develop techniques to detect, track and mitigate botnet infections.[28] [29] [30] [31] While activities of an individual researcher may not lead to immediate remediation of botnet threats, research provides an invaluable basis for the successful disruption of botnets. One prominent example of this has been the collaboration between researchers, law enforcement, and security companies to take down the Gameover Zeus botnet.[32]

In order to do research on botnets, real data is essential. While simulations or other artificial data sources can be used to a limited degree[33] real data is required for evaluation purposes or to provide a close to real-world approximation by the artificial data.[34]

The botnet tracking approaches addressed in this section's scenario focus on targeted data collection. The reason for this is twofold. First, researchers often have a specific goal in mind, allowing them to collect the data in a targeted fashion. Second, many researchers do not have the resources or infrastructure for non-targeted data collection. Nevertheless, several examples exist where researchers collaborate with ISPs or other companies to track botnets.[35] [36] Such scenarios have to take into account the legal obligations of all parties involved. Moreover, the non-targeted bulk data collection is commonly carried out by the other party and not the researchers themselves, as discussed in more detail in Sections 5 and 6.

The PII collected by researchers for the purpose of botnet tracking is limited to IP addresses and data immediately derived from IP addresses such as geographical location. IP addresses are the essential datum for researchers that is collected in virtually every scenario. In the case of botnet tracking, it is necessary to identify an infected machine. An IP address represents the go-to identifier to track and enumerate bot infections. Lastly, it is essential to pinpoint and identify infected machines and botnet control infrastructure for botnet mitigation. Apart from IP addresses, various metadata are collected by researchers in order to measure, analyze, develop and test new anti-botnet mechanisms.

The collected data is commonly processed automatically to perform tasks such as botnet detection, enumeration, or tracking. Moreover, it is in the research community's interest as a whole to make the data available to other researchers. This may either be fully public or upon request, possibly including non-disclosure agreements or other restrictions. This is crucial to facilitate reproducible results and allow other researchers to extend and improve existing work.

[27] Article 6 (1) GDPR: 'Processing shall be lawful only if and to the extent that at least one of the following applies: (a) [...] consent, (b) [...] necessary for the performance of a contract, (c) [...] necessary for compliance with a legal obligation, (d) [...] necessary to protect the vital interests, (e) [...] necessary for the performance of a task carried out in the public interest, (f) [...] necessary for the purposes of the legitimate interests [...].'

[28] D. Zhuang, J. M. Chang, Peerhunter: Detecting peer-to-peer botnets through community behavior analysis, in: Dependable and Secure Computing, 2017 IEEE Conference on, IEEE, 2017, pp. 493–500.

[29] C. Rossow, D. Andriesse, T. Werner, B. Stone-Gross, D. Plohmann, C. J. Dietrich, H. Bos, Sok: P2pwned-modeling and evaluating the resilience of peer-to-peer botnets, in: IEEE Symposium on Security and Privacy, 2013, pp. 97–111.

[30] M. Antonakakis, R. Perdisci, Y. Nadji, N. Vasiloglou, S. Abu-Nimeh, W. Lee, D. Dagon, From throw-away traffic to bots: Detecting the rise of dga-based malware., in: USENIX security symposium, Vol. 12, 2012.

[31] S. Greengard, The war against botnets, Commun. ACM 55 (2) (2012) 16–18. doi:10.1145/2076450.2076456. URL https://doi.org/10.1145/2076450.2076456

[32] U.S. Leads Multi-National Action Against "Gameover Zeus" Bot-net and "Cryptolocker" Ransomware, Charges Botnet Administrator, http://www.justice.gov/opa/pr/us-leads-multi-national-action-against-gameover-zeus-botnet-and-cryptolocker-ransomware

[33] L. Böck, E. Vasilomanolakis, J. H. Wolf, M. Mühlhäuser, Autonomously detecting sensors in fully distributed botnets, Computers & Security 83 (2019) 1–13. doi:10.1016/j.cose.2019.01.004. URL https://doi.org/10.1016/j.cose.2019.01.004

[34] L. Bock, E. Vasilomanolakis, M. Mühlhäuser, S. Karuppayah, Next generation P2P botnets: monitoring under adverse conditions, in: International Symposium on Research in Attacks, Intrusions, and Defenses, Springer, 2018, pp. 511–531.

[35] M. Antonakakis, R. Perdisci, Y. Nadji, N. Vasiloglou, S. Abu-Nimeh, W. Lee, D. Dagon, From throw-away traffic to bots: Detecting the rise of dga-based malware., in: USENIX security symposium, Vol. 12, 2012.

[36] L. Bilge, D. Balzarotti, W. Robertson, E. Kirda, C. Kruegel, Disclosure: detecting botnet command and control servers through large-scale netflow analysis, in: Proceedings of the 28th Annual Computer Security Applications Conference, ACM, 2012, pp. 129–138.

To summarize, researchers need to collect, process, share and store personal data obtained through targeted data collection methods. However, most research is not interested in the person behind an infected machine but in the infected machine itself and the interconnections within the botnet. Moreover, researchers themselves are not able to link IP addresses to a person without additional non-public data sources.

In the following, we first discuss the legal grounds and the data protection principles applicable to botnet research and then provide several technical guidelines for lawful and privacy-preserving botnet research.

4.2. *Appropriate legal grounds for research and article 89 GDPR*

s Researchers need to collect data and further process it for scientific research. As it has been mentioned already in Section 3, for any processing operation to be lawful, researchers need to have an appropriate legal ground as well as to conform to the data protection principles. Before analysing the legal ground and the data protection principles, we should first make some remarks about the special regime attributed to scientific research under the GDPR.

4.2.1. *Special regime on scientific research under the GDPR*

Article 89 GDPR is the applicable provision in the case of processing for scientific research purposes[37] and it reads,

1. Processing for [...] scientific or historical research purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner.
2. Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes and such derogations are necessary for the fulfilment of those purposes.'

This special regime that becomes particularly clear in Article 89(2), 'demonstrates that research occupies a privileged position within the GDPR'.[38],[39] According to Article 179(1) TFEU, one of the European Union's objectives is to achieve a European Research Area.[40] However, and as highlighted by the European Data Protection Board (hereafter, the EDPB), it should be clear that the special derogatory regime is not to be perceived as a general exception to the GDPR requirements in cases of scientific research.[41] On the contrary, the special regime allows for exceptions to specific requirements only in specific situations.[42] It has been noted that Article 89 derogatory system provides flexibility to research but not a detailed framework for research data protection.[43] This flexibility is accompanied by appropriate safeguards which do not relieve the controller from all other GDPR obligations.[44] The controller must thus respect all data protection principles and identify the most appropriate legal ground for processing.

What type of research?

The first issue raised about the scientific research regime is that 'the concept of research is very broad'.[45] The GDPR does not define 'scientific research' in its Articles. It is only Recital 159 that clarifies that for the purposes of the GDPR 'the processing of personal data for scientific research purposes should be interpreted in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately funded research'. In its Opinion on Consent, the EDPB clarifies that the notion 'may not be stretched beyond its common meaning and understands that 'scientific research' in this context means a research project set up in accordance with relevant sector-related methodological and ethical standards, in conformity with good practice'.[46]

[37] Article 89 is placed under Chapter IX entitled 'Provisions relating to specific processing situations'.

[38] European Data Protection Supervisor (EDPS), A Preliminary Opinion on data protection and scientific research, 6 January 2020, 18.

[39] In their article, Mourby et al, Governance of academic research data under the GDPR – lessons from the UK, International Data Protection Law, 2019, Vol. 9, No. 3.,203, the authors examined in detail the UK case and showed that the UK 'has exercised the Article 89 derogation to the fullest possible extent. This suggests a desire to provide researchers and research archivists, with the greatest possible degree of freedom which is, nonetheless, consistent with the GDPR.' They mention in the same article that Ireland and Malta have also exercised this derogation to the fullest possible extent.

[40] Article 179(1) TFEU: '(1) The Union shall have the objective of strengthening its scientific and technological bases by achieving a European research area in which researchers, scientific knowledge and technology circulate freely, and encouraging it to become more competitive, including in its industry, while promoting all the research activities deemed necessary by virtue of other Chapters of the Treaties.'

[41] European Data Protection Board (EDPB) Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research, Adopted on 2 February 2021, para.4.

[42] European Data Protection Board (EDPB) Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research, Adopted on 2 February 2021, para.4.

[43] Mourby et al., Governance of academic research data under the GDPR – lessons from the UK, International Data Protection Law, 2019, Vol. 9, No. 3., 202.

[44] European Data Protection Board (EDPB) Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research, Adopted on 2 February 2021, para.54.

[45] European Data Protection Supervisor (EDPS), A Preliminary Opinion on data protection and scientific research, 6 January 2020, 6.

[46] European Data Protection Board (EDPB) Guidelines 05/2020 on consent under Regulation 2016/679, Version 1.1, adopted on 4 May 2020, para 153.

In the present article, we focus on research conducted in the field of cyber security. Researchers engaged in botnet tracking perform research in the field of computer security. Researchers can be acting in various contexts. It makes a difference, in data protection terms, to conduct research within the healthcare domain where there could be large scale processing of sensitive data, from conducting research on computer security domain whereby only processing of IP addresses for the purposes of identifying infected machines may take place. Thus, it is important to identify the field of research, given that the risks to the rights and freedoms of data subjects might differ.[47]

Research for whom?

A second grey area for applying this special regime is answering the question ’research for whom?’. A distinction should be made between ‘on the one hand, genuine research for the common good and, on the other, research which serves primarily private or commercial ends’.[48] To say this simply, there is a difference between a researcher in the university and a researcher working for a multinational company.[49] The reasoning behind this differentiation could lie in a proportionality assessment of the derogations and thus limitations of the right to data protection (i.e. the special regime of Article 89) in light of the purpose of producing and promoting goods in the public interest that will eventually be publicly shared, i.e. knowledge and research. In his Preliminary Opinion on data protection and scientific research, the European Data Protection Supervisor (hereafter, the EDPS) enumerates the three conditions that should be met in order for the specific data protection regime to apply:

1. Personal data are processed;
2. Relevant sectoral standards of methodology and ethics apply, including the notion of informed consent, accountability and oversight;
3. The research is carried out with the aim of growing society’s collective knowledge and well-being, as opposed to serving primarily one or several private interests.[50]

Once these conditions are met, the particular regime that the GDPR has established (i.e. Article 89) applies. More specifically, Article 89 allows for derogations from specific data subjects rights, i.e. right of access by the data subject (Article 15), right to rectification (Article 16), right to restriction of processing (Article 18) and the right to object (Article 21). These derogations are allowed only in so far as such rights are ‘likely to render impossible or seriously impair the achievement of the specific purposes and such derogations are necessary for the fulfilment of those purposes’ (purposes here meaning: scientific research purposes) and only under the condition that the safeguards of ‘technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation’.[51] The GDPR places a great emphasis on ‘safeguards and accountability’ which need to be guaranteed. As the EDPS has pointed out, ‘the scope of the derogations remain limited to cases where the integrity of research would be compromised by the exercise of data subjects’ rights’.[52] What needs to be stressed is the two-fold aim of the GDPR being to facilitate the free flow of data within the EU while at the same time safeguarding fundamental rights of individuals.

4.2.2. *Identifying the appropriate legal ground*

As previously mentioned,[53] under the GDPR framework, most legal obligations fall on the shoulders of the data controller because they are the actor who determines the purposes and the means of the processing.[54] The data controller is of great importance because they are held (primarily) accountable and responsible for the data processing operations. The data controller needs to guarantee the ‘safeguards and accountability’ mentioned in the previous paragraph. In the case of a researcher employed by a university to perform research on botnet tracking, the university is the employer and therefore the data controller. This scheme works under the assumption that the researcher is treating personal data within reasonable expectations according to their employment contract and research proposal.

One of the primary obligations of a data controller is to identify an appropriate legal ground for the processing operations to be lawful. According to Article 6(1) ‘Processing shall be lawful only if and to the extent that at least one of the following applies [...]’. The GDPR enumerates six legal grounds that can justify the processing of personal data; consent, performance of a contract, compliance with a legal obligation, protection of vital interests of data subjects, performance of a task carried out in the public interest or the exercise of official authority and legitimate interest. At least one of these six grounds should be met for the processing operation to be lawful in the sense of Article 6 and according to the data protection principle of ’lawfulness’ required in Article 5.

The appropriateness of a legal ground depends on the specific circumstances, the purpose of the processing, and the relationship with the individual. It could be the case that a pro-

[47] Under the GDPR, ‘risk’ is an important concept that calibrates data controllers’ legal obligations (see Article 24, but also Article 35 etc). Hence, it is of value to recognise that research in the field of botnet tracking presents fewer risks than research in other fields.

[48] European Data Protection Supervisor (EDPS), A Preliminary Opinion on data protection and scientific research, 6 January 2020, 5.

[49] For research performed for commercial purposes, see Section 5.

[50] European Data Protection Supervisor (EDPS), A Preliminary Opinion on data protection and scientific research, 6 January 2020, 12.

[51] In its Opinion on transparency, the EDPB explained (in the context of Art 14(5)b) that a serious impairment of objectives means that the objectives of the processing are nullified. The use of Art 14(5)b exception (and equally the derogations mentioned in Art 89), ‘presupposes that the data processing satisfies all of the principles set out in Article 5 and that most importantly, in all of the circumstances, the processing of the personal data is fair and that it has a legal basis.’, Article 29WP, Guidelines on Transparency under Regulation 2016/679, WP260 rev.01, para 65.

[52] European Data Protection Supervisor (EDPS), A Preliminary Opinion on data protection and scientific research, 6 January 2020, 22.

[53] See Section 3 on the GDPR applicability.

[54] For more information on the concept and the role of the data controller, see EDPB Guidelines 07/2020 on the concepts of controller and processor in the GDPR, Version 2.0, Adopted on 07 July 2021.

cessing operation satisfies several grounds for legitimation.[55] Although there is no hierarchy among the legal grounds,[56] and no legal ground is better than the others, the data controller has to identify the basis which stands out as the primary and most appropriate ground for the processing at hand. It has been suggested that processing for research purposes could rely on the ground of consent, on the public interest ground, or the legitimate interest ground, but this is not entirely clear.[57]

In the case of performing research on botnet tracking, we shall begin by clarifying that there is no contractual relationship between the researcher and the users that would justify the processing. There is no legal obligation for the researcher that requires them to conduct the specific research and no vital interests of the users are at stake. Thus, three out of six legal grounds (ie. 6(1)(b),(c),(d)) are not appropriate for legitimizing the specific processing activity. With regard to consent, it is technically non-feasible to ask the consent from all the users of the infected machines, given that the researcher only knows the IP of an infected machine, but not who the users are. Moreover, even if it was technically possible to ask for consent, it is improbable to get a response or approval from all affected individuals. Lastly, asking for consent could seriously impair the research objectives, i.e., alert the botmasters of tracking activity. There are two last legal grounds to be examined, the public interest ground (Art 6(1)(e)) and the legitimate interest ground (Art 6(1)(f)).

4.2.3. *Possible legal grounds: public interest task and legitimate interest*

Article 6(1)(e): the Public Interest task ground. [58]

According to Article 6(1)(e) GDPR processing can be lawful if it is 'necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;'. Recital 10 clarifies that regarding the public interest task legal ground, '[…] Member States should be allowed to maintain or introduce national provisions to further specify the application of the rules'. Recital 45 further explains that processing that is necessary for the performance of a task carried out in the public interest or in the exercise of official authority should have a basis in Union or Member State law. Among other details, Union or Member State law should determine 'whether the controller performing a task carried out in the public interest or in the exercise of official authority should be a public authority or another natural or legal person governed by public law, or, where it is in the public interest to do so, including for health purposes such as public health and social protection and the management of health care services, by private law, such as a professional association.'[59]

The 'public interest' legal ground covers, inter alia, situations where the data controller performs a task of public interest and for that task to be correctly fulfilled, processing of personal data is necessary.[60] Two important remarks should be made: first, the controller should establish that it is strictly necessary to process personal data for the specific task they need to accomplish.[61] Secondly, what constitutes a 'task in the public interest' and what is an 'official authority' are further defined in national laws. As the ICO clarifies, while the data controller does not need a specific statutory power, the task they perform shall have a clear basis in law.[62] Conducting research projects in universities or research centres with a legal mandate to do research in the public interest, could justify the use of the 'public interest task' as the lawful basis.[63] The public interest could legitimize the processing of personal data in the research context. An important note is that the right to data portability (Article 20 GDPR) shall not apply in case the lawful ground of 'public interest task' is used[64] as is the case also for the right to erasure (Article 17 GDPR), as long as the processing is necessary.[65]

Article 6(1)(f): the Legitimate Interest ground.

What also seems to be an appropriate legal ground is the sixth legal basis, namely the legitimate interest ground (Article 6 (1)(f)) which reads:

> 'processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child'.

What is important to keep in mind is the last sentence of Article 6(1): 'Point (f) of the first subparagraph shall not apply to processing carried out by public authorities in the

[55] See for example in the CJEU Manni Case C-398/15, ECLI:EU:C:2017:197, para. 42 'In that regard, as the Advocate General pointed out in point 52 of his Opinion, it should be noted that the processing of personal data by the authority responsible for keeping the register pursuant to Article 2(1)(d) and (j) and Article 3 of Directive 68/151 satisfies several grounds for legitimation provided for in Article 7 of Directive 95/46, namely (…)'.

[56] Gil González, E., de Hert, P. Understanding the legal provisions that allow processing and profiling of personal data–an analysis of GDPR provisions and principles. ERA Forum 19, 597–621 (2019). https://doi.org/10.1007/s12027-018-0546-

[57] Kelli et al., Processing personal data without the consent of the data subject for the development and use of language resources, Selected papers from the CLARIN Annual Conference 2018. Linköping Electronic Conference Proceedings 159: 72–82, 75.

[58] In the recently released 'Data: a new direction', by the Department for Digital, Culture, Media & Sport, 10 September 2021, the UK Government acknowledged that 'Uncertainty around determining lawful grounds could hinder or discourage important research' (para 43) and suggested 'public interest' as an appropriate lawful ground for research conducted by universities. More specifically para 44(a) reads: 'At present, universities are identifying a legal basis to use for research in an unclear and inconsistent way. Uncertainty may be creating burdens or discouraging useful research. Defining when universities can rely on Article 6(1)(e) of the UK GDPR may reduce these burdens and increase transparency for data subjects on how universities use personal data.'

[59] Recital 45 GDPR.

[60] In case the controller has a legal obligation to process personal data, then they should use the legal obligation basis of Art 6(1)(c).

[61] For the case of public research on botnet monitoring, see Section 4.1

[62] Information Commissioner's Office (ICO) Guide to the General Data Protection Regulation (GDPR) 01 January 2021 - 1.1.157, 76.

[63] Kelli et al, Processing personal data without the consent of the data subject for the development and use of language resources, Selected papers from the CLARIN Annual Conference 2018. Linköping Electronic Conference Proceedings 159: 72–82, 75.

[64] Article 20(3) GDPR.

[65] Article 17(3)(b) GDPR.

performance of their tasks.’. The data controller has to examine whether they fall under the status of ‘public authority’ under their national laws. If they do, then the legitimate ground does not apply for the processing performed within their tasks. Instead, they should consider legitimizing their processing based on the ‘public interest task’ (Art 6(1)(e)). The legitimate interest ground requires three conditions to be met, namely that:

1. there exists an interest that is legally qualified as legitimate (purpose test),
2. the processing is necessary for the purposes of this legitimate interest (necessity test),
3. the legitimate interest is not overridden by the interests or fundamental rights and freedoms of the data subject (balancing test).

As to the first condition, we shall examine whether botnet tracking could be qualified as a ‘legitimate interest’ of the institution where the research takes place under the meaning of Article 6(1)(f). The legislature is in favor of a very broad interpretation of what counts as a ’legitimate interest’ and it is thus a relatively easy threshold to pass. It is interesting to note the choice of the word ‘legitimate’ instead of ‘legal interest’.[66] [67]According to Corbin, a legal interest reflects the ‘aggregate of the legal relations of a person with respect to some specific physical object or the physical relations of specific objects’.[68] A legitimate interest need not stem from a specific legal instrument, but it should be acceptable under the applicable EU and national law. It should additionally be real and present, clearly articulated and sufficiently specific.[69] In the scenario presented in this Section, the question is whether research conducted in a public research institution on botnet tracking constitutes a legitimate interest. The answer is yes. Not only is research a legitimate interest but the EDPB has included ‘IT and network security’ in the list of the ‘most common contexts in which the issue of legitimate interest in the meaning of Article [6(1)(f)] may arise’.[70]

Regarding the second condition, the researcher has to ask themselves when performing the necessity test: ’can I achieve the same result in a less intrusive way?’ The necessity test is a facts-based test[71] meaning that it should be performed taking into account the very specific circumstances of the case at hand. According to the CJEU, ‘derogations and limitations in relation to the protection of personal data must apply only in so far as is strictly necessary’.[72] Additionally this necessity test ’must be examined in conjunction with the data minimisation principle’[73] according to which, personal data shall be ’adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed’.[74] As recently ruled by the District Court Midden-Nederland, the burden lies on the data controller to be able to provide proper explanation that the processing they perform is strictly necessary.[75] In our scenario, it has already been explained[76] that the personal data that are being processed are absolutely necessary to achieve the purpose of botnet tracking. This means, the purpose of tracking down infected bots (machines) ’cannot reasonably be as effectively achieved by other means less restrictive for fundamental rights and freedoms of data subjects’.[77]

The third condition requires a balancing test between the legitimate interest and the fundamental rights and freedoms of the data subjects. It is a test that should be done on a case-by-case basis whereby the researcher needs to consider the impact of the specific processing on the rights and freedoms of the users and assess whether this impact overrides the legitimate interest of the researcher in performing botnet tracking research.[78] This test ‘may require a complex assessment taking into account a number of factors’,[79] among which are the safeguards that will be guaranteed by the data controller mentioned above in Article 89(2).[80] The balancing test is a safeguard of data subjects’ rights, freedoms, and interests in light of the very broad scope of the concept of ’legitimate interest’. In the case of botnet tracking in the context of research performed in a public university or other public institution, the following elements should be considered. First of all, academic research is protected by Article 13 of the EUCFR[81] which reads that *’The arts and scientific research shall be free of constraint. Aca-*

[66] This is a point made in Kamara, I., de Hert, P.: Understanding the balancing act behind the legitimate interest of the controller ground. A pragmatic approach. In: Selinger, E., Polonetsky, J., Tene, O. (eds.) The Cambridge Handbook of Consumer Privacy (2018), 330.

[67] See also, Rechtbank Midden-Nederland, VoetbalTV, ECLI:NL:RBMNE:2020:5111, para 16.

[68] Arthur Corbin, Legal Analysis and Terminology, 29, Yale Law Journal 163, 1919-1920, p.173.

[69] Article 29WP, Opinion on 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, WP217, 25.

[70] Article 29WP, Opinion on 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, WP217, 24.

[71] European Data Protection Supervisor ‘ Developing a “Toolkit” for Assessing the Necessity of Measures that Interfere with Fundamental Rights, Background paper’, June 2016, p.8.

[72] Rīgas satiksme, C 13/16, ECLI:EU:C:2017:336, para30 and the case law the CJEU refers to: judgments of 9 November 2010, Volker und Markus Schecke and Eifert, C 92/09 and C 93/09, EU:C:2010:662, paragraph 86; of 7 November 2013, IPI, C 473/12, EU:C:2013:715, paragraph 39; and of 11 December 2014, Ryneš, C 212/13, EU:C:2014:2428, paragraph 28.

[73] CJEU TK Judgment Case C-708/18, ECLI:EU:C:2019:1064, para 48.

[74] Article (5)(1)(c) GDPR.

[75] Rechtbank Midden-Nederland, VoetbalTV, ECLI:NL:RBMNE:2020:5111, para 20.

[76] See Section 3.1

[77] CJEU TK Judgment Case C-708/18, ECLI:EU:C:2019:1064, para 47.

[78] Information Commissioner’s Office (ICO) Guide to the General Data Protection Regulation (GDPR) 01 January 2021 - 1.1.157.

[79] Article 29WP, Opinion on 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, WP217, 22.

[80] There is a debate on whether ‘safeguards’ should be part of the balancing test. For more information on that topic, see Kamara, I., de Hert, P.: Understanding the balancing act behind the legitimate interest of the controller ground. A pragmatic approach. In: Selinger, E., Polonetsky, J., Tene, O. (eds.) The Cambridge Handbook of Consumer Privacy (2018), 333.

[81] EUCFR stands for Charter of Fundamental Rights of the EU.

demic freedom shall be respected.' Secondly, conducting cybersecurity research is an interest not strictly limited to the data controller but a wider, public interest. The advancement of knowledge and the development of tools that enhance computer system security are relevant to the broader community. Third, it is socially and culturally expected and does not go beyond the reasonable expectations of data subjects that Universities and public institutions perform research.The advancement of knowledge is a legitimate expectation, also recognised by the GDPR in Recital 113 which reads: '[...] For scientific or historical research purposes or statistical purposes, the legitimate expectations of society for an increase of knowledge should be taken into consideration.'

The above-mentioned elements add 'weight' to the legitimate interest of research and could tip the balance in favor of research in the balancing test. It has been explicitly mentioned by the EDPB that 'the more compelling the public interest or the interest of the wider community and the more clearly acknowledged and expected it is in the community and by data subjects that the controller can take action and process data in pursuit of these interests, the more heavily this legitimate interest weighs in the balance'.[82]

4.3. *Practical implications for researchers*

Within the previous section, we established that data collection and further processing for botnet research in the public interest is lawful if the legitimate interest withstands the balancing test against fundamental rights and freedoms of the data subjects.

In order to withstand this balancing test, it is crucial to, inter alia, implement appropriate safeguards, which 'shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation'.[83] Unfortunately, the GDPR provides little guidance as to the appropriate safeguards 'which is alarming considering the potential scope of the derogations.'[84] The vagueness of the term 'appropriate safeguards' has been criticised by scholars.[85] At the time of writing of this article, the EDPB has announced the preparation and upcoming publication of guidelines that will further develop the concept of safeguards in cases of processing for scientific research.[86] The EDPB has highlighted 'the special role that safeguards may play in reducing the undue impact on the data subjects and thereby changing the balance of rights and interests to the extent that the data controller's legitimate interests will not be overridden'.[87]

Data security is among the data protection principles which, as has already been mentioned, have to be complied with whenever the processing of personal data takes place. Article 5(1)(f) GDPR reads that personal data shall be 'processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').' Moreover, it is explicitly mentioned in Article 89 that, wherever possible, personal data should be pseudonymized if such a technical measure would not interfere with the purpose of the research itself.

Hence, we provide the following practical guidelines for researchers:

- **Anonymization and Pseudonymization:** Article 89 of the GDPR states that if the purpose of the processing can be achieved on anonymized or pseudonymized data, the data shall be processed in that form. It is important to mention that anonymization should be applied if possible. For botnet tracking research, an identifier is necessary to track devices over time. Therefore, anonymization of IP addresses is impractical for botnet research purposes. However, pseudonymization can and should be applied. Moreover, suppose the actual IP address is needed to derive additional information, e.g., geo-location. In that case, the IP address should be pseudonymized after the computation of the derived data. For specifics on pseudonymization of IP addresses, we recommend following the best practices provided by ENISA.[88]
- **Minimization:** Data that is not relevant for the research should not be collected. Moreover, if data becomes obsolete after processing, it should be deleted to reduce the stored data to the minimum necessary to carry out a specific research goal. For example, if a non-targeted approach, e.g., network traffic monitoring, is used for data collection, non-botnet-related traffic will be recorded. In this case, this traffic should be removed as soon as it is known to be benign.
- **Secure Storage:** All research data containing PII should be stored in a manner to prevent unauthorized access or modification of the personal data. Suppose PII cannot be anonymized or pseudonymized for the research, the collected information should be protected to ensure integrity and confidentiality. For example, one could encrypt the stored data and apply access control measures to prevent unauthorized parties from accessing the data.
- **Accountability:** One should be able to demonstrate that the aforementioned safeguards are implemented and complied with. This could be achieved through proper documentation of the safeguards and keeping records about data access.

[82] Article 29WP, Opinion on 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, WP217, 35.

[83] Article 89 GDPR.

[84] C Staunton et al, The GDPR and the research exception: considerations on the necessary safeguards for research biobanks, European Journal of Human Genetics (2019) 27:1159–1167, 1166.

[85] Kart Pormeister, 'Genetic Data and the Research Exemption: Is the GDPR Going too Far?' (2017) 7 International Data Privacy Law 137.

[86] European Data Protection Board (EDPB) Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research, Adopted on 2 February 2021.

[87] Article 29WP, Opinion on 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, WP217, 31.

[88] Pseudonymization techniques and best practices, European Union Agency for Cybersecurity ENISA, 2019.

It is important to remember that data protection is 'much more than a technical issue requiring technical solutions'.[89] Additional legal, organisational, and technical safeguards are needed, which should be 'dynamic and responsive to an evolving science'.[90] Transparent and publicly available policies (on issues such as the storage of data) as well as clear governance procedures that oversee the use of data constitute important organisational safeguards.[91]

We want to highlight that collecting and (publicly) sharing datasets is crucial for network and system security research. Hence, collecting a variety of data to facilitate experiments is the actual purpose of some research. The GDPR provides specific derogations related to ”processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes”.[92]

Within our scenario, the derogations include Articles 15, 16, 18, and 21 of the GDPR for scientific research purposes and additionally, Articles 19 and 20 for archiving purposes in the public interest. While all data processing shall fall under the former, publishing the data to facilitate further research by others falls under the latter. Specifically for the context of this scenario, derogations from the following data subject rights are possible:

- **Right of access by the data subject:** Article 15 allows a data subject to obtain access to the stored data and information about the processing. While informing the actual person affected by the botnet infection is in the interest of most botnet tracking research, the researchers do not want to disclose this information to the botmaster. This is a practical issue, as the identifier known to the researchers is often only the IP address and not the device owner's name. Therefore, as a botmaster has control over the infected device, they could impersonate the affected person in order to learn if they are compromised or not. Therefore, access to personal data should only be provided upon proof of ownership of that IP address.
- **Right to erasure:** Article 17 of the GDPR states that in the context of scientific research the data subject can not request the erasure of their data, if it ”is likely to render impossible or seriously impair the achievement of the objectives of that processing”.[93] In most cases the erasure of botnet tracking data will affect achieving the intended purpose. As an example, identifying the origin of a botmaster command within a P2P requires accurate information about the infected devices and their interconnections. Removing even small portions of this information may render this goal impossible, justifying a derogation for the right to erasure.
- **Right to object:** Article 21 of the GDPR provides the data subject with the right to object. For scientific research the Article states ”Where personal data are processed for scientific or historical research purposes or statistical purposes pursuant to Article 89(1), the data subject, on grounds relating to his or her particular situation, shall have the right to object to processing of personal data concerning him or her, unless the processing is necessary for the performance of a task carried out for reasons of public interest.” As for the right to erasure, processing of the complete data is often necessary to achieve the purpose of the research, underlining that the processing is necessary for the performance of the task, justifying a derogation.

We want to point out that these derogations should be justified on a case-by-case basis. Nevertheless, the derogations provided for scientific research are often crucial in order to carry out the intended purpose of research in the public interest.

5. Research for commercial interest

Our second scenario addresses the data collection and analysis of active botnets by companies with commercial interests. These companies are primarily motivated by collecting botnet tracking data to provide better customer service, e.g., security companies. An example of such a scenario would be the collection of a botnet's known C2 IP addresses by an anti-virus company to be included in a blacklist, i.e., a list of known malicious devices. This blacklist is then distributed to the endpoint devices or applications to protect their client systems and networks.

This section will first describe the goals and purposes, followed by an analysis of the legal grounds for processing. We conclude the section by proposing practical guidelines for practitioners based on the identified as most appropriate legal grounds.

5.1. *Goals and purpose of data collection*

One of the primary goals of a company with a commercial interest in investigating and understanding botnets is to stay abreast of the emerging botnet threats and develop suitable solutions to protect their clients. An example would be the effort taken by anti-virus companies to disseminate policies and updates to their anti-virus applications. In return, these applications protect the end-devices and networks of their customers.

The botnet-related information collected by these companies can be categorized into internal and external sources.[94] Internal sources provide data from within the organization or those collected by the providers themselves. For example, companies may leverage targeted techniques similar to those of researchers described in Section 4 to track bots participating in the botnet as part of their internal sources. In addition, companies may also collect additional data from appli-

[89] C Staunton et al, The GDPR and the research exception: considerations on the necessary safeguards for research biobanks, European Journal of Human Genetics (2019) 27:1159–1167, 1165.

[90] C Staunton et al, The GDPR and the research exception: considerations on the necessary safeguards for research biobanks, European Journal of Human Genetics (2019) 27:1159–1167, 1165.

[91] C Staunton et al, The GDPR and the research exception: considerations on the necessary safeguards for research biobanks, European Journal of Human Genetics (2019) 27:1159–1167, 1165.

[92] Article 89 GDPR.

[93] Article 17(3) GDPR.

[94] M. Bromiley, Threat intelligence: What it is, and how to use it effectively, SANS Institute InfoSec Reading Room 15.

cations or devices deployed in the premises of their clients, e.g.intrusion detection systems.

In both internal and external sources, companies perform targeted data collection to offer high-quality service to their clients as per the Service Level Agreement (SLA). After collecting the data, the companies utilize techniques such as machine learning and statistical evaluations, allowing past and future trends of cyber-attacks to be analysed and used in improving the solutions deployed to protect their clients. This also implies the necessity of storing the gathered data for a longer period, if required, depending on the type of analysis or intelligence needed. For instance, tracking data of IP addresses of bots within an emerging botnet can be used as a temporary blacklist on all end-point security solutions of the clients to prevent communication to or from infected machines. Without this collected information, the companies would not be able to meet the contractual agreement with their clients, e.g.protecting them from botnet and other cybersecurity threats.

To summarize, companies with commercial interest need to collect, process, and store data specific to their clients to provide a better service for them, e.g., threat intelligence. However, it is worth noting that information from different clients or organizations serviced by a company can be jointly aggregated and analysed to provide reliable intelligence that is useful for the greater good of all customers.

5.2. *Appropriate legal grounds: two purposes, two legal processing grounds*

As explained in Section 3, the processing of IP addresses, as well as of any other personal data for the aforementioned purposes, needs to have a legal basis that legitimizes the processing. The appropriate legal basis depends on the particular context of the processing operation. There are two primary purposes for processing personal data in the context of security companies–first, the provision of security services to their customers. Second, the performance of research to improve the quality and update their services. Each purpose requires a separate legal ground that will legitimise the processing. Again, the legal ground of consent is not practically feasible, for the same reasons as explained previously in Section 4.2.2.

5.2.1. *Contract: provision of services*

To begin with, a security company will process the personal data of a customer to provide individualised security services. In the context of the provision of services, the most appropriate legal ground is 'performance of a contract with the data subject' (Article 6(1)(b) GDPR). In the case of the contractual legal basis, the personal data processed must be 'genuinely necessary'[95] for the performance –the normal execution- of the contract. For that, 'the exact rationale of the contract, i.e. its substance and fundamental objective'[96] has to be accurately understood. The link between the processing of the data and the purpose of the execution of the contract has to be direct and objective.[97] The question to be answered by the controller is: 'can the requested service be provided without the specific processing taking place?'.[98] If the answer is no and the necessity test is satisfied, then the processing is lawful under the legal basis of the contract (Article 6(1)(b)). It is a processing operation that is considered 'a priori legitimate and therefore only subject to compliance with other applicable provisions of the law. There is in other words a presumption that the balance between the different rights and interests at stake [...] is satisfied.'[99] There is no need to perform an additional balancing test, as is the case in the legitimate interest legal ground. The three requirements which need to be fulfilled are:

- A contract between the controller and the data subject exists.
- The contract is valid according to applicable national contract laws.
- The processing is objectively necessary for the performance of the contract.[100]

5.2.2. *Legitimate interest: improvement of services*

The second purpose of processing is to conduct research to improve the quality of the services the security company offers and to be updated with regard to the ever-changing landscape in botnets and cyber-crime. As discussed earlier in Section 4, processing for research enjoys a special regime under the GDPR. However, it is important to clarify that not every type of research enjoys this regime. It should be research conducted in the public interest. Even though the security services of big companies are relevant for the broader community as they protect against cyber-crime, the primary goal of these companies is driven by profit. The interest is primarily private and not public and for that reason, it is not covered by the special regime of Article 89 GDPR. While this acknowledgment does not make a difference for the appropriate legal ground, it does make a difference as for 'practical implications for companies with commercial interest' (see Section 5.3). In its Guidelines 2/2019, the EDPB explicitly mentions the case where processing takes place for 'service improvement' reasons.[101] In this case, the 'contract' legal basis that we previ-

[95] Article 29WP, Opinion on 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, WP217, 17.

[96] Article 29WP, Opinion on 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, WP217, 17.

[97] European Data Protection Board (EDPB) Guidelines 05/2020 on consent under Regulation 2016/679, version 1.1, Adopted on 4 May 2020, 10.

[98] European Data Protection Board (EDPB), Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects, Version 2.0, 8 October 2019, 7.

[99] Article 29WP, Opinion on 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, WP217, 9.

[100] European Data Protection Board (EDPB), Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects, Version 2.0, 8 October 2019, 9.

[101] European Data Protection Board (EDPB), Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects, Version 2.0, 8 October 2019, 14.

ously discussed is not the appropriate one and for that, the controller should turn towards 'legitimate interest'.

In 2020 the District Court of Midden-Nederland ruled that the Dutch DPA's decision that a purely economic commercial interest can never be a legitimate interest under Article 6(1)(f) GDPR, is wrong. Although the CJEU does not provide a clear description of what constitutes a legitimate interest[102] the Dutch Court rules that in order to assess whether an interest is legitimate, one should apply the 'negative test' asking the question, 'does this interest violate the law?'[103] Coming back to the scenario of this Section, commercial interests do not violate the law.[104] The conditions that the legislature lays down for the legitimate interest ground have already been discussed previously in Section 4 and they are lawfulness of the interest, the necessity of the processing, and a balancing test where the balance tips towards the controller's interest without being overridden by the data subjects' interests. In the case of research by a private company, what should be mentioned is that development of tools more capable of responding to cyber-crime are indeed in the wider interest of our community. Of course, it should not be equated to research performed solely for the public interest given that in the case of private companies there is always the factor of profit as the primary aim. What is crucial to have in mind about the legitimate interest ground is that whether Article 6(1)(f) 'can be relied on will depend on the outcome of the balancing test that follows',[105] i.e. botnet tracking based on legitimate interest is lawful only if the balancing test is in favor of botnet tracking on a case by case basis.

5.3. *Practical implications for companies with commercial interests*

The previous paragraph argued that botnet data collection and further processing by companies with commercial interest is lawful if it is based on the provision of services defined within a contract or under the legitimate interest of improving their service. For the latter, the collection of data needs to withstand the balancing test between the interests of the company and the individuals whose information is collected and processed.

For data collected under the legal basis of legitimate interest, the practical guidelines suggested for the public research in Section 4.3 also apply to companies with commercial interest. However, there are two important distinctions between research in a commercial context and scientific research in the public interest. First, the derogations stated in Section 4.3 apply only to scientific research in the public interest. This is also the case if research performed in a commercial context is published, as the commercial interest prevails. Second, the balancing test between a legitimate interest and the rights of individuals is more likely in favor of research in the public interest than research with a commercial interest. Therefore, publishing the results of commercial research projects could tip the balancing test in favor of the legitimate interest of a company.

As discussed in the previous subsection, companies may also collect and process data based on contractual agreements with their customers. What differentiates this kind of data from tracking data usually collected under the legal basis of legitimate interest is that the identifiable person is directly known to the data controller. Therefore, additional guidelines related to the the data subject's rights stated within the GDPR have to be followed. Most importantly, the following guidelines should be implemented and followed:

- **Information about processing:** Article 13 of the GDPR states that the controller has to inform the data subject about the data and the purpose of the processing when the data is first collected. For example, an anti-virus provider should inform its customers if it collects and processes malware samples obtained from and linked to one of its customers.
- **Purpose limitation:** The collected data may only be used for the purpose specified and agreed to by the data subject. If the data controller wants to use the information for additional processing, the data subject has to be provided with relevant information about the additional purpose. For example, if a data subject agreed to provide their IP address for the configuration of a firewall, this information may not be used to enhance botnet tracking data without prior notification.
- **Access, rectification and erasure:** The data subject should be allowed to obtain access to the information stored about them. Furthermore, the data subject may request that their data be rectified or erased by the data controller. That includes copies of the information provided to third parties for processing. For example, a data subject may request that information about their past IP addresses be removed from the controller's data.

This list should be viewed as a high-level summary containing information relevant for consideration by practitioners. For a complete overview of the rights of data subjects and the obligations of controllers and processors, we ask interested readers to refer directly to the GDPR. We also want to highlight that collecting data is often a crucial part of a company that offers network or system security-related services. Hence, collecting various data to provide better service is critical for them and their business models.

6. Research by internet service providers

While most ISPs are privately owned, there are three important distinctions from the previous scenario of research for commercial interest. First, ISPs' core business is the routing of traffic. This provides them with a unique level of access to all in- and outbound traffic for a large group of customers. Second, ISPs are regulated by the ePrivacy Directive which is

[102] Rechtbank Midden-Nederland, VoetbalTV, ECLI:NL:RBMNE:2020:5111, para 15.

[103] Rechtbank Midden-Nederland, VoetbalTV, ECLI:NL:RBMNE:2020:5111, para 16.

[104] See also the example of 'direct marketing', which is a clear example of commercial practice and interest. Recital 47 GDPR recognises it as a legitimate interest ('[...]The processing of personal data for direct marketing purposes may be regarded as carried out for a legitimate interest.')

[105] Article 29WP, Opinion on 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, WP217, 25.

considered to be lex specialis to the lex generalis which is the GDPR.[106] Third, an ISP can link the IP addresses of their customers to the customers' real names and geographical addresses.

6.1. *Goals and purpose of data collection*

The primary purpose of an ISP is the transmission of Internet data packets and as part of this, a stable daily operation is needed. In order to fulfill this purpose, it is necessary to protect the network infrastructure and customers from DDoS attacks, which are typically initiated by botnets. A secondary purpose of an ISP could be to help customers identify botnet infections and ultimately protect them from such infections by offering customers such a value-added service.

Their access to network traffic of thousands of endpoints, enables ISPs to apply both targeted and non-targeted tracking techniques described in Section 2.2. As one example, Antonakakis et al.[107] describe how they can identify the DGAs of multiple known and unknown botnets by analyzing the DNS traffic of an ISP. An ISP could also employ targeted tracking techniques for botnet detection. Employing targeted techniques such as crawling by an ISP in order to protect its customers was previously discussed by Silva et al.[108] The authors mention that the protection of their customers could constitute a legitimate interest of the ISP and could thus be used as their legal basis, which is similar to our previously discussed scenario for Security Companies (see Section 5). Therefore, the data collection and legal ground described in this section consider only the passive network monitoring technique. These passive techniques are of special interest, as ISPs often are the only entities able to apply them on a large scale.

To achieve the goal discussed above, an ISP would typically choose to collect:

- **NetFlow logs** The deployed network equipment of the ISP can often collect NetFlow logs. These primarily contain information about which IP addresses (hosts) communicate and related metadata like timestamps and amount of packets/bytes.
- **DNS logs** Most subscribers use the ISP's DNS servers to translate from a domain name (like cnn.com) to an IP address. These logs primarily contain information about which domain name was requested at which timestamp by which (subscriber) IP address and what IP address is associated with the domain name.
- **Packet dumps** An ISP could choose to deploy equipment that collects complete Internet data packet dumps of some or all of the traffic. At an ISP scale, this can be very expensive. Therefore it is not typically deployed if any of the above-mentioned options exist. Later in this section it will be shown why this option would not be legal.[109]

While researchers have a clear motivation for sharing the collected data, it is not in the interest of an ISP to store or share any of the above-mentioned data. An exception is sharing data of a detected DDoS attack with the vendor of anti-DDoS equipment or the DNS server software. The shared data would allow the vendor to use this information to improve their detection capabilities. This creates a bridge to the scenario relating to research in commercial interest discussed in Section 5.

6.2. *Appropriate legal ground and data protection principles*

The ePrivacy Directive[110] and its national implementations, regulate among other things, how ISPs are allowed to handle data related to the subscribers' data traffic. The 2009 update of the ePrivacy Directive does not contain any changes relevant to this article.

The following definitions from the ePrivacy Directive are relevant to quote directly:

- **Traffic Data:** (Article 2(b)) ”Traffic data means any data processed for the purpose of the conveyance of a communication on an electronic communications network [..]”. According to Recital 15 ”[..] “Traffic data may include any translation of this information by the network over which the communication is transmitted for the purpose of carrying out the transmission. Traffic data may, inter alia, consist of data referring to the routing, duration, time or volume of a communication, to the protocol used, to the location of the terminal equipment of the sender or recipient, to the network on which the communication originates or terminates, to the beginning, end or duration of a connection. They may also consist of the format in which the communication is conveyed by the network.”
- **Communication:** (Article 2(d)) ”Communication means any information exchanged or conveyed between a finite number of parties by means of a publicly available electronic communications service. [..]”. (Recital 15) ”A communication may include any naming, numbering or addressing information provided by the sender of a communication or the user of a connection to carry out the communication.”

The following paragraphs are relevant:

- **Traffic data:** (Article 6(1)) ”Traffic data [..] must be erased or made anonymous when it is no longer needed for the purpose of the transmission of a communication [..].” (Article 6(3)) ”For [..] the provision of value added services, the

[106] Being lex specialis, the ePrivacy Directive takes precedence over the GDPR. The GDPR will apply only where the lex specialis does not regulate a specific case.

[107] M. Antonakakis, R. Perdisci, Y. Nadji, N. Vasiloglou, S. Abu-Nimeh, W. Lee, D. Dagon, From throw-away traffic to bots: Detecting the rise of DGA-based malware., in: USENIX security symposium, Vol. 12, 2012.

[108] Silva, Karine, and Ruben Roex. ”Zombie alert: Assessing legitimacy of P2P botnet mitigation techniques.” (2014).

[109] For a more comprehensive overview of the types of data sources legally and technically available to a typical ISP, we would like to refer readers to M. Fejrskov, J. M. Pedersen, E. Vasilomanolakis, Cyber-security research by ISPs: A netflow and DNS anonymization policy, in: 2020 International Conference on Cyber Security and Protection of Digital Services, Cyber Security 2020, Dublin, Ireland, June 15-19, 2020, IEEE, 2020, pp. 1-8.

[110] The European Parliament and of the Council, Directive 2002/58/ec (the ePrivacy directive) (2002).

provider [..] may process the [traffic data] to the extent and for the duration necessary for such services [..] if the subscriber or user to whom the data relate has given his/her consent. [..]"
- **Communication:** (Article 5(1))"[..] In particular, they shall prohibit listening, tapping, storage or other kinds of interception or surveillance of communications and the related traffic data by persons other than users, without the consent of the users concerned [..]"

To sum up, the ePrivacy Directive sets the following limitations relevant to this article on processing a subscriber's traffic or location data:

- Data *already* being processed for transmission must be made anonymous before additional processing. This includes NetFlow and DNS data, where the primary PII is the IP address.
- Data *not* being processed for transmission or as part of a value added service cannot be processed. This includes Packet Dumps.
- Data can be processed for a specific value-added service but only if the data subject provides consent. This can, in theory, include all three data sources.

The ePrivacy Directive states that consent for processing all traffic data of a subscriber is valid only in connection with a relevant value-added service, for example a traffic scanning security service. In GDPR terms this would be referred to as a contractual legal basis (ie. Article 6(1)(b)) rather than consent (i.e. Article 6(1)(a)). The ePrivacy Directive therefore does not allow the use of the GDPR legal basis of consent.

Both the EDPB in its Opinion on Anonymization Techniques[111] as well as Recital 26 GDPR, make a clear distinction between pseudonymization and anonymization and make it explicit that a requirement from the ePrivacy Directive to anonymize certain data is not fulfilled by the use of pseudonymization.

6.3. *Practical implications for ISPs*

As it is practically impossible to have *all* subscribers sign up to a value-added service relating to cybersecurity research (and thereby providing a contractual relationship), the use of anonymized NetFlow/DNS data would be the only viable strategy for non-targeted techniques.

An advantage of using anonymized data in research is that the data is no longer personal data and therefore does not fall under the material scope of the GDPR. Consequently, data subject rights such as deletion, correction, access to information, etc., no longer apply. The way of proceeding to the anonymization of data is out of the scope of this article.[112]

Table 2 – Summary of the applicability of legal bases and the Art 89 exemption for scientific research in the three described scenarios. The two remaining legal bases (i.e. legal obligation and vital interest) do not apply in any of the described scenarios. Legend: √= applicable, ✗= not applicable.

	Public	Commercial	ISP
Consent	✗	✗	✗
Contract	✗	√	√
Public Interest	√	✗	✗
Legitimate Interest	√	√	√
Scientific research exemption (Art. 89)	√	✗	✗

The fact that traffic of several computers will be inseparable due to the anonymization could be a significant disadvantage to botnet tracking. The specific implications will depend on the anonymization applied and the method used for botnet tracking.

ISPs are typically obliged by national law to collect data about the IP address assigned to a particular subscriber. Both the ePrivacy Directive and the GDPR provide exemptions to allow this and in a GDPR context, compliance with other laws is a separate legal ground. This information does not in itself, however, tell anything about the prevalence of botnets and is in any case not immediately usable in combination with anonymized NetFlow/DNS data.

7. Conclusion

The research question of the article is: *What are the practical implications for researchers, private companies, and ISPs when engaging in botnet tracking activities so as to comply with the European general legal framework on data protection (GDPR)?*

In Section 3, we established the material applicability of the GDPR in botnet tracking activities, whereby IP addresses are commonly collected and processed. IP addresses qualify as 'personal data' under the GDPR and any activities performed on this data qualify as 'processing operations'. Given the material applicability of the GDPR, we went on to investigate the appropriate legal bases for collecting and further processing botnet tracking data in three main scenarios: research at public institutions (in Section 4), research for commercial interest (in Section 5), and botnet tracking by ISPs (in Section 6). We additionally identified the implications for practitioners and provided practical guidelines for each scenario. An overview of the legal bases applicable in the three scenarios is given in Table 2.

One main conclusion is that in all three scenarios examined, data controllers can, under certain circumstances, rely on the legitimate interest basis to conduct botnet tracking. Researchers for commercial interest may additionally conduct botnet tracking based on contracts with their customers. The main difference between the public and commercial actors lies in the special regime established by Article 89 GDPR, which applies to the processing of personal data for scientific research in the public interest. This special regime does not ap-

[111] Article 29WP, Opinion 05/2014 on Anonymisation Techniques, WP216.

[112] For further information on the topic see M. Fejrskov, J. M. Pedersen, E. Vasilomanolakis, Cyber-security research by ISPs: A netflow and DNS anonymization policy, in: 2020 International Conference on Cyber Security and Protection of Digital Services, Cyber Security 2020, Dublin, Ireland, June 15-19, 2020, IEEE, 2020, pp. 1-8.

ply to research performed for commercial purposes because such research is primarily profit-oriented. Lastly, even though ISPs have exceptional technical capabilities to conduct botnet tracking, they are more strongly regulated by the ePrivacy directive, which is considered lex-specialis to the GDPR and restricts ISPs from taking a major role in the field of botnet tracking.

This article showed that the legal conditions that apply in the three examined scenarios differ as much as the techniques used for conducting botnet research. We hope that the present article will contribute to raising awareness as to what practitioners should do to comply with the GDPR requirements and consequently maximize the outcome of both current and future botnet tracking research projects.

Declaration of Competing Interest

Pieter Wolters (Institutional colleague of one of the authors)

Data Availability

No data was used for the research described in the article.