A Measured Response – On the Nexus of Large-Scale Technical Measurements and Cybercrime Policing

Hugo L.J. Bijmans
TU Delft & TNO
Delft / Den Haag, The Netherlands
h.l.j.bijmans-1@tudelft.nl

Michel J.G. van Eeten
TU Delft
Delft, The Netherlands
m.j.g.vaneeten@tudelft.nl

Rolf S. van Wegberg
TU Delft
Delft, The Netherlands
r.s.vanwegberg@tudelft.nl

Abstract

Over the past two decades, researchers have advanced large-scale technical measurements of cybercrime to analyze techniques, tactics, and procedures (TTP) in cybercrime operations. These quantifications are typically based on analyses of technical artifacts such as domains, binaries, or attack traffic and could potentially inform cybercrime policing – i.e., assisting law enforcement agencies (LEAs) in determining how their scarce resources can be best put to use. Yet, we do not know if this potential is being used, nor how such measurement studies align with LEA needs. This paper investigates the nexus of large-scale technical measurements and cybercrime policing by combining a survey of previous scientific work with a user study involving LEA professionals. We leverage the concept of value chains to structure 38 studies featuring measurements of phishing, booter services, and remote access trojans (RATs). We scrutinize their data sources and characterize their findings to identify common denominators. Then, we let LEA professionals reflect on some of these measurements and jointly identify the unexplored potential for novel measurements that align with current needs in cybercrime policing. We find that many academic studies focus on components in the value chain that are considered less valuable to LEAs and that most measurements lack geographical or attacker differentiation, thereby not allowing for concrete action perspectives.

Keywords

Cybercrime, Measurements, Phishing, DDoS, Booter Services, Remote Access Trojans, Law Enforcement

1 Introduction

Cybercrime – referring to crime facilitated or committed by using a computer [51] – has grown worldwide [26] and even surpassed traditional crime in damages in some countries [8]. Traditionally, LEAs relied heavily on criminological data collection efforts like victimization surveys [54] or police reports [17, 33] for designing and evaluating their interventions. However, it is a well-established fact that the majority of cybercrime incidents are never reported to the police [17, 33], so the actual amount of cybercrime could be much higher than law enforcement agencies (LEAs) estimate. To understand modern-day cybercrime, computer scientists create novel detection methods to perform large-scale technical measurements and capture data on cybercriminal techniques, tactics, and procedures (TTP). Such measurements have the potential to inform LEAs policing cybercrime in a robust and scalable way. Yet, we do not know if this potential is being used, nor how such measurement studies align with LEA needs.

In recent years, a wealth of large-scale technical measurements of cybercrime have been published. Many of these studies revolve around the creation of innovative detection methods, which are afterward deployed to examine a subset of the Internet to assess their workings [15]. For example, within one year Konoth et al. [39], Wang et al. [67], and Kharraz et al. [37] set out to create a robust method able to detect browser-based cryptojacking. Other studies examined phishing through large-scale Web scraping [5, 16, 38, 59, 69, 70], studied DDoS attacks trough network telescopes [30, 34, 42] or deployed honeypots to discover RAT operators [24, 53] Such measurements demonstrate how well a newly designed method functions, but these studies have not been performed with a focus on generating new insights into cybercriminal prevalence or TTP.

Demonstrating the success of a novel detection method can be straightforward. Performing robust measurements of cybercrime and assessing, for example, prevalence is, however, far from trivial. In an overview of cybercrime studies, Clayton et al. [10] found that the existence of concentrations of cybercriminal activity often leads researchers to suggest that such concentrations are potential vantage points for law enforcement interventions. While economic factors cause some of those concentrations, others are the result of measurement biases that mislead researchers into drawing wrong conclusions. Even without such issues, it is unclear what value these studies represent to LEAs. Additionally, in 2013, Anderson et al. [2] reported on the insufficiency of cybercrime statistics and encouraged governments to put more effort into detecting and prosecuting cybercrime. Their follow-up study in 2019 reconfirmed their findings, demanding governments increase detection efforts to improve cybercrime statistics [1]. Recent academic studies involving large-scale technical measurements could assist with these efforts. However, we do not know how such measurements can assist LEAs. To the best of our knowledge, no prior study has investigated how law enforcement professionals assess the value of academic studies featuring large-scale technical measurements of cybercrime for policing efforts.

In this paper, we combine a literature survey and a small user study involving LEA professionals to address this gap and to answer the following two research questions. First, **RQ1:** What are the characteristics of large-scale technical cybercrime measurements? Second, **RQ2:** How do LEAs evaluate these measurement studies regarding their alignment with policing needs?

We answer the first question by surveying the top computer security venues for large-scale technical measurements of cybercrime published in the last fifteen years (2007–2023). For feasibility, we scope our search to three cybercrimes that are extensively empirically studied and received widespread attention by LEAs

worldwide [19, 21, 22], namely phishing, booter services, and remote access trojans (RATs). Measurement studies on these three cybercrimes help us derive relevant characteristics of research that might influence whether a study is more or less aligned with LEA needs. Such characteristics are then relevant for measurement studies beyond these three cybercrimes. To structure our analysis, we leverage the concept of cybercrime value chains [66], which dissects cybercrimes in components and resources. For each study, we identify the components in the value chain it investigates and review its data collection, methods, and findings. Next, we report on a workshop with LEA professionals that elicits their assessment of a sample of studies and answers our second research question. Additionally, we let participants generate ideas on how to improve measurements to assist in cybercrime policing.

By combining our survey of 38 studies with the LEA workshop's results, we find that academic measurements focus on the deployment and execution of cybercrime, whereas LEAs desire to learn more about development and monetization. We observe that the majority of measurement studies lack geographical or attacker differentiation, thereby not allowing for concrete, actionable perspectives for law enforcement.

In short, we make the following contributions:

(1) We survey 38 large-scale technical cybercrime measurements and characterize datasets, methods, and findings, which we combine with the first-ever assessment of scientific measurements by law enforcement.
(2) We find that measurements often focus on the center part of the value chain (deployment and execution), whereas LEAs value insights at the ends of the value chain (development and monetization) to support cybercrime policing.
(3) We observe that current measurement approaches largely overlook geographical and attacker differences, resulting in less actionable measurements for LEAs.

The remainder of this paper is structured as follows: we detail our methodology in Section 2 and present the results of our survey in Sections 3–6. Then, we report on our workshop with law enforcement in Section 7, critically discuss our work in Section 8, and conclude in Section 9.

2 Methodology

This section first introduces the concept of cybercrime value chains, explains our focus on the three selected cybercrimes, and elaborates on our methodology to survey past scientific work.

Cybercrime Value Chains. To structure our synthesis of past scientific work, we leverage the concept of *cybercrime value chains* [66] to map large-scale measurements to value chain components. As cybercrime relies on a delimited mix of resources to turn a profit [66], we can structure components in the value chain with a required set of resources as an input and a resource as an output. For example, as illustrated in Figure 2, to deploy a phishing page, one needs a phishing kit and a domain as input, which results in a phishing page as output. Components in the value chain are not strictly sequential, as some components can be executed in parallel (e.g., a booter service can deploy its storefront website and perform reconnaissance operations simultaneously). Instead of fulfilling all components in the value chain themselves, modern-day cybercriminals rely on specialists to fulfill specific components for them [4, 32, 66]. Specialized third parties provide resources, which can be products (e.g., software) as well as services (e.g., hosting infrastructure). For every outsourced component, a specialist will take a cut. Hence, to avoid these cuts or to expand operations, cybercriminals can decide to self-organize components in the value chain. We refer to this practice as *vertical integration* [61]. Only a few [61] have studied the cybercrime ecosystem with such a holistic view. As we demonstrate in this work, most scholars focus their measurements on specific components of a value chain instead of considering the complete value chain.

Selection of Cybercrimes. We survey a large body of such measurements to study the characteristics of large-scale technical measurements of cybercrime and assess their value for law enforcement. However, as scrutinizing all past measurements of every type of cybercrime is near impossible, we scope our survey to three types of cybercrime that received significant attention from both academics and LEA and argue that identified characteristics are then relevant for measurement studies beyond these three specific cybercrimes. To discover what types of cybercrime measurements are valuable for LEA, we search Europol's 'Internet Organised Crime Threat Assessment' (IOCTA) from 2019 until 2023 [19, 21, 22]. These reports give an overview of the cybercrime landscape, as well as the efforts LEA has made to police it. Cybercrimes mentioned in these reports include ransomware, DDoS attacks, Business Email Compromise (BEC) fraud, dark markets, phishing, bulletproof hosting, botnets, and many more. We constructed value chains for each cybercrime, determined whether measurements were possible at different phases along the chain, and initiated an initial literature search to find large-scale technical measurement studies. It turns out that certain cybercrimes allow for more measurements than others. For example, we could find a plethora of phishing measurement studies in the top venues, but none related to BEC fraud. Consequently, we selected three types of cybercrime that received attention from LEA and were measured frequently by scientists, namely: *phishing*, *booter services*, and *remote access Trojans (RATs)*.

Survey Approach. We employ a systematic approach to search for studies performing large-scale technical measurements of one of the selected cybercrimes. We start by selecting conferences and journals. First, we take the Google Scholar top ten computer security conferences and journals [29], supplemented by their relevant co-located workshops. We extend this list with computer science venues focused on Internet measurements and cybercrime. Appendix A contains the complete list of included conferences and journals. In June 2023, we used the ACM Digital Library and IEEE Xplore to search for papers published in these conferences or journals since 2007 – covering the last 15 years of academic research. Conferences and workshops that are not in these libraries were searched manually. To be included, the paper title or abstract must contain one of the following terms: *phish*∗, *booter*∗, *ddos*, *rat*, *remote access trojan*, *cybercrime*, *cyber crime*. Manual title and abstract screening was performed on the resulting 615 papers. Through this process, we identified 32 papers that contained a large-scale technical measurement of one of the three selected cybercrimes. We added 6 other relevant works discovered during our literature research. In total, we found 38 studies, of which 19 papers measure

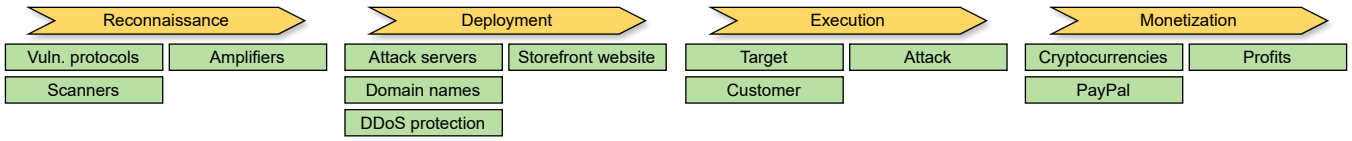

Figure 1: Booter value chain with its components at the top and the required resources at the bottom.

phishing, 14 examine booter services, and 5 investigate RATs. We examine these papers through the lens of cybercrime value chains. We first identify which value chain component(s) the measurement leverages. Next, we classify measurements as either passive (based on an existing dataset, involving no scanning for artifacts or interaction with infrastructure) or active (based on an active collection of artifacts, involving scanning for cybercriminal infrastructure, etc.). Finally, we extract data sources and collect time ranges.

3 Booter Measurements

In a DDoS (Distributed Denial of Service) attack, a server, service, or network is flooded with a massive volume of traffic, rendering it unavailable to legitimate users. Executing a DDoS attack requires significant resources and technical capabilities. To allow low-skilled criminals to perform such attacks, criminal entrepreneurs have set up so-called *booter* or *stresser* services, which offer DDoS attacks as a service. In this section, we examine large-scale technical measurements of booter services in earlier work. The booter service value chain is depicted in Figure 1. Here, we identify four components. First, during *reconnaissance*, the attacker finds the resources (e.g., vulnerable protocols) and scans for infrastructure to abuse. During *deployment*, criminal entrepreneurs set up shop and organize their attack servers, domains, and their protections. *Execution* involves the actual DDoS attack, involving both a client and a target, and *monetization* revolves around all financial aspects of running a booter service. Although DDoS attacks can also originate from botnets or nation-state actors, several studies concur that booter services exert a significant impact on the DDoS landscape [42, 46, 60]. We find 14 large-scale technical measurements of booter services published between 2013 and 2021, list them in Table 1, and detail them in the next paragraphs.

Datasets. Two main data sources are used to study booter services: booter operations data found in databases or website scrapes and DDoS observations through honeypots or darknets. Since 2013, a variety of booter operations databases have been examined. First, an analysis of *TwBooter* in [35], followed by an analysis of 14 booter services in [55]. A year later, [36] scrutinized the databases of *Asylum Stresser* and *Lizard Stresser*, and complemented their data collection with website scrapes from *vDOS*. [60] used website scrapes and API logs from *vDOS* together with the leaked database of *CMDBooter* to validate their honeypot measurements. The *vDOS* database was used again in 2017 by [7]. The other data source used to study booters is honeypots, which capture amplification DDoS attacks. Deploying such honeypots allows researchers to track attacks over time and was initiated in [42], with the design of *AmpPot* – a honeypot network that was deployed at 21 locations worldwide. The same honeypot was later used in [46], which deployed eight of them in Japan, and in [34], which deployed 24 of them worldwide. [60] designed a different honeypot to measure amplification attacks and deployed ±60 of them. Data collection continued, allowing for two more years of data to be analyzed in [11]. The most recent honeypot measurement was presented by [30], in which an entirely new honeypot network was built. This study deployed 549 instances in five public clouds worldwide, demonstrating that the number of honeypots needed to obtain sufficient attack coverage is much higher than shown in the earlier work. Some studies complemented their honeypot measurements with data from darknets – unused IP ranges, also known as network telescopes [30, 34, 42].

Reconnaissance. Scanners are deployed to find vulnerable infrastructure to be used in DDoS attacks, which can be observed through darknets. Deployment of such scanners was limited before 2012, as found in [42]. Since then, scanning for DNS has gained more popularity and increased for more protocols in 2014. However, by attributing scanning IP addresses, [42] found that over 40% of all scanners are operated by universities and security organizations instead of DDoS providers. [60] claimed to have excluded such 'white-hat' scanners from their data and reported on 5,070 IP addresses scanning daily. During the measurement, they noticed an increase in NTP and SSDP scanning and a slight decrease in DNS scanning. The largest honeypot network to study reconnaissance was deployed by [30]. It confirmed [42] by observing prevalent research scans (30% of all scans), finding that responsive IPs make scanners come back twice as fast, and noticing the same packets used for both testing and attacking. Additionally, by periodically switching honeypots between active and passive mode, the existence of a 'memory' of previously exploited servers was discovered, indicating that attackers track vulnerable servers instead of opportunistically selecting them for their attacks. A similar pattern was learned from leaked booter databases in [36] that noticed booters gravitating to using more stable amplifier infrastructure when possible instead of scanning for vulnerable machinery.

Deployment. During deployment, booter operators assemble their attack infrastructure and create a website to serve customers. In 2013, [35] examined *TwBooter* and revealed that only 15 servers were used to perform their attacks, most of them hosted in the Netherlands. Three years later, such servers were purchased by [36], which allowed them to conclude that the required high uplink bandwidth can be obtained with small investments. A different approach was discovered in [55] that found that all but one (*TwBooter*) based their attack infrastructure on Web shells instead. Web shells are scripts that allow backdoor access to compromised machines, making them part of the DDoS infrastructure. Scrutinizing 42 booter websites in [9] learned that websites calling themselves 'stressers' did so to avoid legal problems. Most websites show a verbose page

Table 1: Overview of large-scale booter service measurements. Measurements can be active (*A*), passive (*P*), or both (hybrid, *H*).

Authors	Year	Value chain				Data sources					Misc.			Time frame
		Reconnaissance	Deployment	Execution	Monetization	Databases	Website scrapes	Honeypots	Darknets	ISP/IXP	Measurement	Attack purchased	Intervention	
Karami & McCoy [35]	2013		●	●	●	●					*H*	✓		01/2013 – 03/2013
Chromik et al. [9]	2015		●				●				*A*			09/2014
Krämer et al. [42]	2015	●		●				21	●		*P*			2007 – 2015
Santanna et al. [55]	2015		●	●	●	●					*H*			2011 – 2014
Santanna et al. [56]	2015		●	●			●			●	*H*	✓		2013
Karami, Park & McCoy [36]	2016	●	●	●	●	●	●				*P*	✓	✓	2011 – 2015
Noroozian et al. [46]	2016			●				8			*P*			2014 – 2015
Brunt, Pandey & McCoy [7]	2017			●	●	●	●				*P*		✓	2014 – 2016
Jonker et al. [34]	2017			●				24	●		*P*			03/2015 – 02/2017
Thomas, Clayton & Beresford [60]	2017	●		●		●		60			*P*			2014 – 2017
Collier et al. [11]	2019			●			●	60			*P*		✓	2014 – 2019, 2017 – 2019
Kopp et al. [41]	2019			●			●			●	*P*	✓	✓	2018 – 2019
Griffioen et al. [30]	2021	●		●				549	●		*H*			08/2019 – 11/2019
Kopp, Dietzel & Hohlfeld [40]	2021			●				●		●	*P*			09/2019 – 04/2020

with text and appealing advertisements to sell their services. Additionally, they noticed that nearly all websites have DDoS protection. This was confirmed by [56], who tracked 102 booter websites over time and observed increased use of such protection since 2011. In 2014, none of the analyzed booter services were unprotected. This finding was confirmed in [36], together with the notion that this was also to hamper take-down by law enforcement.

Execution – Customer. Behind every DDoS attack launched by a booter is a customer. The various leaked databases give interesting insights into the customer base of booters. [35] identified three types of customers: gamers (launching attacks of less than 10 minutes), website attackers (launching attacks of one or two hours), and privileged users (performing long attacks for more than two hours). Most of the *TwBooter* customers could be categorized as gamers, attacking roughly three targets per day for a short period. [55] highlighted the importance of differentiating between registered users and paying customers of booters, as the latter is significantly smaller. It turns out that many users are just attracted to take a look at what a service can offer, whereas only a few are interested in performing attacks. [36] and [7] drew similar conclusions, as they found that only roughly 13% and 15-23% of the users in their examined databases ever paid for an attack, respectively. Multiple booter databases showed differences in OPSEC of customers as well [55]. Frequent customers are more likely to take precautions by obfuscating their real IP address – e.g., by using a VPN or Tor.

Execution – Target. Several studies tried to map booter service attack targets. [35] categorized most customers as gamers and, therefore, concluded that most targets were game servers and forums. Subsequent analysis of leaked databases in [36] learned that targets are predominantly residential links and gaming-related servers, with only a handful of higher-profile targets, such as government, media, and law enforcement websites. A finding later confirmed in [7]. Observing ongoing DDoS attacks through honeypots in [42] showed that victimization rates differ much per country. The U.S. stands out (one-third of all victims), followed by China (14%), and France (8.6%). Similar numbers were reported in [11] four years later. Additionally, in [42], it was found that 79% of victims are targeted just once. Analysis of the attacked ports showed a special interest in gaming-related services, such as Xbox Live, Minecraft, and Steam. A similar analysis performed in [34] observed that attacks targeted at HTTP(S) are most prevalent for TCP, whereas the most attacked ports for UDP are associated with various online multiplayer games and Steam. As many websites are hosted on IP addresses operated by large hosting companies, it was evinced that 64% of the `.com`, `.net` and `.org` websites were hosted on IP addresses ever targeted by DDoS attacks. Targeted networks were aggregated by their infrastructure in [40], which showed that content hosting networks were attacked the most (37%), followed by access networks such as ISPs (35%). Using the same honeypot technology as [42], [46] published the most comprehensive work on booter victimization in 2016. It demonstrated that most attacks are directed towards users in access networks and not at hosting or enterprise networks. The number of victims in an ISP network is proportional to the number of ISP customers, just as the victimization rate in a hosting network is proportional to the number of hosted domains. For the identified Web hosting victims, the authors discovered almost no high-profile targets, whereas the largest victim group was again gaming-related websites, mostly related to Minecraft. Noroozian et al. state that "in the Minecraft community specifically, DDoS attacks seem to be part of the culture" [46]. The authors speculate that attackers and victims of booter services are geographically close, and the low entry barrier of booter services allows victims to easily become attackers themselves. In [30], it was concluded that victimization has changed in 2021. The U.S. and China are still popular among attackers. Yet, a disproportional share of attacks on South Africa,

Table 2: Overview of DDoS attack characteristics. Grey-colored rows are based on self-reported numbers; the others are based on network measurements. Protocols in bold are the most popular attack vectors. The average attack duration is converted to seconds, and the number of attacks is per day.

Ref.	Year	UDP	TCP	Attacks	Duration (s)
[35]	2013	DNS	SYN	-	-
[55]	2015	**UDP flood**	SYN	-	260
[42]	2015	**NTP**, DNS	-	10,235	62% ≤ 900
[36]	2016	**SSDP, DNS**	SYN	-	1,620
[46]	2016	**DNS**	-	7,844	272 – 300
[34]	2017	NTP, DNS	**HTTP(S)**	30,000	255 – 454
[60]	2017	**NTP**, DNS	-	5,120	50% ≤ 658
[11]	2019	**LDAP**, NTP	-	30,000	-
[41]	2019	**NTP**, Memcached	-	-	-
[40]	2021	**DNS**, NTP	-	809	±360
[30]	2021	**NTP**, SSDP	-	673	394

Poland, and Kuwait was found – 51% of all targeted IP addresses had an associated domain name. There were also numerous DDoS attacks on residential IP address space.

Execution – Attack. Earlier work has characterized booter DDoS attacks by, for example, attack duration and used protocols. An overview of such studies is presented in Table 2. It shows that most research effort is put into analyzing UDP reflection or amplification attacks. Only four studies [34–36, 55] report on the use of the TCP protocol in DDoS attacks. Looking at the popularity of UDP protocols for abuse, we observe a constant dominance of DNS and NTP, even though their disclosure as DDoS amplification vectors was made long ago [40]. [11] reported on the rise of LDAP as a protocol in DDoS attacks, but this was never confirmed by later research. [40] analyzed several other new attack vectors – such as WS-Discovery, ARMS, and OpenVPN – and confirmed the active abuse of these protocols for DDoS attacks, yet not at the scale of traditional ones. Daily attack numbers are hard to grasp, as each study covers a different dataset and measurement approach. Yet, the numbers in Table 2 show a declining number of daily attacks while attack duration remains relatively stable. Parallel attacks introduce complexity in counting the number of attacks. For example, should a booter customer that launches both an NTP and a DNS reflection attack toward a victim be counted as one or two attacks? [55] noticed that 32% of attacks have been launched in parallel, which means that new attacks against the same target are launched during an ongoing attack. However, they also find that 38% of users do not perform such attacks and perform just one attack per day on average. Attack durations differ per protocol [30] and victim type [46]. Both [60] and [46] noticed spikes in their attack duration measurements. They observed many attacks with a specific duration of either 5, 10, 60, or 120 minutes, likely caused by booter services offering exactly these amounts. Lastly, [30] discovered a new adversary tactic – attack pulses – in which the attacker does not launch its attack as a continuous flow but in powerful, periodic pulses. This maximizes the attack power while minimizing the costs for the attacker.

Monetization. As criminal entrepreneurs run booters, four studies investigated their monetization strategy. Database analysis in [35] learned that a total of 277 active users subscribed to *TwBooter*, totaling a profit of $7,727 a month. A similar conclusion was drawn from analyzing *VDoS* databases in [7]. A median revenue stream of $25,985 was reported, with new customers making up the largest sum of revenue. Although the studied booter supported both PayPal and Bitcoin as payment methods, most profit was generated through clients paying with PayPal – which ease-of-use stands out compared to Bitcoin. The popularity of PayPal was also noted in [36], which monitored the payment infrastructure of 23 booter services. In confirmation with [7], only a small portion of booters accepted Bitcoin payments. [55] learned that most paying customers paid only once to perform their attacks, and over 50% of customers paid $5.00 or less for the booter's services. Although booters offer differently priced services, the cheapest services are the most popular.

Interventions. As booter services impact the DDoS landscape, interventions either by law enforcement [11, 41] or as part of scientific studies [36] have focused on disrupting their business. Both [36] and [7] studied the effects of PayPal interventions on booter operations. While monitoring the payment infrastructure of 23 booter services, [36] reported booter merchant accounts to PayPal. Through this intervention, the average lifespan of such accounts dropped from 8 to 3 days, and PayPal unavailability increased from 20% to above 60% in the days after. Most booters eventually added alternative payment methods, such as Bitcoin. A similar analysis was performed on *VDoS* in [7] and observed decreasing revenue as soon as PayPal was removed as a payment method. This hampered subscriber growth and eventually led to a decreasing user base. Only 11% of the existing customers switched to Bitcoin when PayPal was not available anymore. The smaller user base also resulted in fewer attacks being launched, decreasing by 31% in their analysis period.

The effects of booter takedowns by law enforcement are studied by [41] and [11]. During the analysis in [41], an FBI-led operation seized 15 booter services [65]. Those takedowns caused significant reductions in DDoS traffic to DNS, NTP, and Memcached reflectors, as observed from an IXP perspective, but no significant reduction in traffic from those reflectors to targets. Kopp et al. note that "seizing the domains of booter websites does not improve the situation for DDoS victims, as the underlying infrastructure of reflectors remains online and can be utilized by third parties without disruption" [41]. A longitudinal analysis of UDP amplification attacks in [11] revealed that after each police intervention, the number of observed attacks decreased significantly for a short period but kept increasing in the long run. Search engine adverts (discouraging the use of booters) and the closure of multiple booter websites had a longer-lasting effect on the booter market than arrests.

Booter Takeaways. In contrast to other cybercrimes, much ground-truth data (e.g., leaked databases) is available for research [7, 35, 36, 55], allowing for valuable insights into booter operations, attackers, and targets. As a result, all components in the value chain have been studied within the last 15 years. However, most measurements have focused on the execution component. Reconnaissance has been studied throughout the years, but although booters remain on the radar for LEA worldwide [21], insights into its deployment and

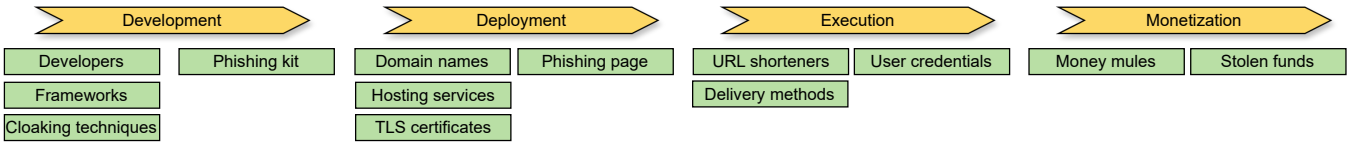

Figure 2: Phishing value chain with its components at the top and the required resources at the bottom.

monetization have not been gathered since 2016 and 2017, respectively. Another unique aspect of this cybercrime is the number of interventions that happened during measurements, which allowed researchers to show that most LEA interventions seem to have a short-term effect [11], whereas other interventions, such as on payment infrastructure [7, 36], seem to have a longer-lasting effect. Customers seem to be diverse, yet a large portion can be related to online gaming. Analysis of groups of booter customers showed great differences, both in terms of victim selection and OPSEC. Attack characteristics across studies show that long-established attack vectors (e.g., DNS or NTP amplification) remain popular while others exist. Research into booter operations also introduces noise, as scanning by security researchers is prevalent [42, 60]. This has to be considered to avoid measurement biases, especially in the reconnaissance phase [30]. Lastly, the popularity of analyzing the same booter databases in [7, 36, 55] suggests that some measurements are driven by data availability.

4 Phishing Measurements

Phishing is the nefarious practice of harvesting user credentials through various means of deception. In our study, we specifically include work on phishing used for gaining direct profits – e.g., obtaining bank credentials to steal funds. We are aware that phishing can also be used to gain initial access to a (company) network, but we do not include such use in our survey. Illustrated in Figure 2, we present the components of the phishing value chain. *Development* is typically served through a *phishing kit* – an off-the-shelf package containing Web pages mimicking a company login page. Occasionally, they feature a back-end *panel* that allows access to the phished credentials. *Deployment* involves registering a domain name and acquiring hosting facilities. During *execution*, victims interact with the phishing website, followed by the *monetization* of the phished credentials. We find 19 large-scale technical measurements of phishing published between 2007 and 2022, which we list in Table 3 and detail in the next paragraphs.

Datasets. A diverse set of data sources can be found in phishing studies, as evident from the overview provided in Table 3. Most studies gather a large body of phishing URLs or domains from a database of known phishing pages to analyze afterward. Until 2019, this source was predominantly PhishTank [52], a community-vetted database of phishing URLs. Until 2019, 7 of the 11 measurements relied on this source [13, 14, 43–45, 50, 63]. This shifted towards other databases such as the APWG eCrime Exchange [3] and OpenPhish [49] in the years that followed. From 2020 onward, combining different data sources gained traction [6, 16, 38, 50]. The introduction and adoption of Certificate Transparency (CT) [28] in 2018 - 2019 opened up new methods to discover phishing domains and was immediately put to use in several studies [5, 6, 38]. *WHOIS* records in phishing research are mostly used to complement other datasets [5, 16, 38, 44, 70]. In terms of data types, we observe a shift from (passive) domain and URL-based measurements [43, 44, 47, 57] towards active measurements using a crawler that inspects live phishing pages [5, 6, 16, 38, 59, 69, 70].

Development – Phishing Kits. The development component of a phishing attack is typically served through a phishing kit. The first study on such kits by [13] showed that more than a third contained obfuscated backdoors exfiltrating phished credentials to a third party. This was quantified in [50] years later, with 5% of phishing pages also disclosing credentials to third-party collectors. Additionally, in [13], it was noted that all kits were written in PHP, likely because of the easy deployment on (shared) hosting servers. This was confirmed twice more, in 2017 by [62], and in 2021 by [5]. Most kits impersonated one specific organization – mostly a U.S bank or PayPal –, and email was found to be the most frequently used method to deliver disclosed information. PayPal was also found to be most mimicked by [31] in 2016 and by [50] in 2019. A similar analysis as in [13] was performed years later in [5] by gathering phishing kits from Telegram and open server directories on live phishing pages. In contrast to [13], high utilization of *multipanels* was observed in [5] – a type of kit that targets multiple organizations simultaneously – and they found many versions of the same *uAdmin* phishing kit, which was the most deployed phishing kit at the time. Like in [62], it was found that only a small number of phishing kits are actively used by many different attackers. Email was no longer used to deliver disclosed credentials, as most kits contained a panel hosted on the same domain to access them instead. Additionally, the use of *decoy* pages was discovered, enabling multi-stage phishing attacks involving multiple brands in one campaign. Phishing websites were grouped through vector clustering on the DOM in [14], which allowed for the observation that attackers only search for a new domain for their attack instead of modifying their phishing pages. However, as pointed out by both [62] and [5], this could result from the same phishing kit being used by many different actors. Through an analysis of live phishing pages [59] report on phishing kit capabilities. A third of pages logged keystrokes and shared them with the attacker as soon as they were typed, and almost half of the websites required users to disclose their credentials in multipage Web forms, hereby confirming the findings of [5]. Although modern phishing sites impersonate a certain brand's Web page, 42% of phishing pages are not direct clones of the corresponding legitimate ones. This was also mentioned in [63] four years earlier, which concluded that such evasions would render visual similarity detection ineffective. Finally, a new type of phishing through man-in-the-middle (MITM) attacks was evinced in [38], circumventing multi-factor protections.

Table 3: Overview of large-scale phishing measurements. Measurements can be active (*A*), passive (*P*), or hybrid (*H*), with a worldwide (*W*) or local (*L*) focus.

Authors	Year	Value chain				Data sources							Data types					Misc.			Time frame
		Development	Deployment	Execution	Monetization	PhishTank	APWG eCX	OpenPhish	Other sources	CT logs	DNS zonefiles	WHOIS	Domains	URLs	Phishing kits	Web pages	E-mails	Measurement	Focus	Crawler	
Moore et al. [45]	2007		●	●		●							●	●				*A*	*W*	✓	02/2007 – 04/2007
Cova et al. [13]	2008	●	●			●			●					●	●			*P*	*W*		04/2008 – 05/2008
McGrath et al. [44]	2008		●	●		●			●		●	●	●	●				*H*	*W*		2007 – 2008
Han et al. [31]	2016	●	●	●											●			*P*	*W*		09/2015 – 01/2016
Cui et al. [14]	2017	●	●			●								●		●		*A*	*W*	✓	01/2016 – 10/2016
Thomas et al. [62]	2017	●		●					●						●			*P*	*W*		03/2016 – 03/2017
Oest et al. [47]	2018	●	●				●						●		●			*P*	*W*		2016 – 2017
Le Page et al. [43]	2018			●		●			●					●				*H*	*W*	✓	01/2016 – 12/2017
Tian et al. [63]	2018		●			●					●		●	●		●		*H*	*W*	✓	04/2018
Peng et al. [50]	2019	●	●	●		●	●	●	●					●	●	●		*A*	*W*	✓	09/2018 – 03/2019
Bitaab et al. [6]	2020		●	●			●	●	●	●			●	●		●	●	*A*	*W*	✓	01/2020 – 05/2020
Oest et al. [48]	2020		●	●	●		●		●					●			●	*H*	*L*		10/2018 – 09/2019
Simpson et al. [57]	2020		●								●	●	●					*P*	*W*		2009 – 2019
Bijmans et al. [5]	2021	●	●	●			●			●		●	●		●	●		*A*	*L*	✓	09/2020 – 01/2021
Kondracki et al. [38]	2021	●	●			●		●		●		●	●	●	●	●		*A*	*W*	✓	03/2020 – 03/2021
Zhang et al. [69]	2021	●					●							●		●		*A*	*W*	✓	06/2018 – 11/2019
Drury et al. [16]	2022		●			●		●	●	●		●		●		●		*A*	*W*	✓	07/2021 – 02/2022
Subramani et al. [59]	2022	●						●						●		●		*A*	*W*	✓	03/2022 – 05/2022
Zhang et al. [70]	2022	●					●					●		●	●	●		*A*	*W*	✓	11/2020 – 07/2021

Development – Cloaking. To prevent phishing websites from being detected, attackers deploy *cloaking* mechanisms to thwart security researchers accessing their phishing websites. Here, the server presents different content to scanners instead of regular visitors. The first analysis of such mechanisms was in [62], which found that many phishing kits deploy a `htaccess` policy or employ a blocklist to frustrate visits from cloud providers, anti-virus brands, and anonymous proxies like Tor. A year later, 2,313 of such `.htaccess` files were examined in [47] to discover that blockades based on IP address, hostname, referrer, and User Agent are most common. Most `.htaccess` files are constantly reused and not kept up to date – as most of them were last modified over a year before their deployment. 23% of all phishing websites were found to implement client-side cloaking techniques in 2018 according to [69], which grew up to 34% in 2019. The most common technique was the use of pop-ups (content remains hidden until a button in a pop-up window is clicked) and click-through interaction (content is shown when a visitor clicks somewhere on the page). Follow-up work in [70] showed that intentionally triggering server-side cloaking behavior could be used as a method to detect phishing websites. Analysis of live phishing kits revealed that 96% employed cloaking techniques in 2022, and [38] showed 85% of MITM phishing kits did too.

Deployment. The deployment of phishing websites in the wild was first studied in [45] by monitoring 1,685 phishing domains from PhishTank [52]. They found an average uptime of 62 hours, with a median of just 20 hours. Additional analysis of domains related to the *RockPhish* gang revealed the use of fast-flux domains, which resulted in a longer average uptime of 95 hours. As listed in Table 4, this research inspired others to perform similar measurements of phishing website life cycles. *WHOIS* records were used in [44] to discover that most domains are used almost immediately after registration. Periodic DNS resolving of the examined domains revealed that, on average, a phishing domain lasts just over three days, but a third of all domains only 55 minutes. An estimated lifespan of eight days for phishing kits installed in a honeypot was calculated in [31]. Honeypot monitoring further revealed that attackers act fast when installing and testing their kits. This was quantified in [48], which found a one-hour window between the first attacker tests and the first victim. Most URLs were hosted on paid domain names, whereas only a very small portion used subdomains offered by free hosting services. Two-thirds of distinct URLs were served over HTTPS, but 86% of the compromised visitors visited over HTTPS, meaning that the use of HTTPS proved more successful than HTTP. A percentage that was much lower one year earlier, when 34% of websites were served over HTTPS in [50]. Subsequent studies leveraged the TLS certificates for HTTPS connections to detect phishing websites in CT logs. In [6], a spike in newly issued certificates for COVID-19-related domains during the pandemic was observed, peaking at more than seven thousand per day. CT logs also allowed for the identification of 1,363 domains targeted at the customers of Dutch financial institutions in [5]. Further analysis confirmed the one-hour testing window found in [48], as most domains had a kit installed one hour after they first responded. A surprising amount of these domains were hosted (73%) and registered (34%) through Namecheap. In the same year, [38] found most MITM domains hosted at DigitalOcean – demonstrating the need

for more demanding requirements (e.g., a VPS instead of shared hosting) for such attacks. TLS certificates were not used for detection in [16], but to timestamp phishing websites. They found TLS certificates to be often requested close to the occurrence of a phishing website on a blocklist – 27.6% was requested within 24 hours of inclusion. Date information included in file headers yielded insights into the resource's creation and indications of resource sharing among phishing websites.

Deployment – URLs & Domains. Much research focused on URLs and domain names used in phishing, which started relatively unstructured. From [13], we learn that 63% of live phishing kits were hosted on trustworthy domains with the targeted brand inserted in the path, and 30% of phishing URLs had no clear relation with the targeted brand. That phishing domains tend to be online for shorter periods than benign ones was discovered in [44], whereas their URLs are typically longer. Additionally, phishing domains typically have fewer unique characters, and more than half of the examined domains contain the targeted brand name. In 2007, the first efforts were made to structure the analysis of phishing URLs. A taxonomy defining four types was proposed in [27], only to be used in [45]. It was later updated in [47] that proposed five types, used in [5] and [48]. Phishing URLs contain either an IP address as hostname and a deceptive path (Type I), a random domain and a deceptive path (Type II), a long and deceptive subdomain (Type III), a deceptive top-level domain (Type IV), or are unintelligible (Type V). 61% Type V domains were found in [47], followed by 21% Type IV domains. Follow-up work in 2020 in [48] showed fewer Type V domains, 29% Type IV domains, an increase in Type II domains (35%), and 28% Type IV domains. Almost solely Type IV domains (95%) were found in [5] by monitoring CT logs for phishing domains. Half of the domains did not contain references to the targeted brands, just deceptive keywords. The gTLDs `.info` and `.com` were among the most popular ones, followed by cheap TLDs such as `.xyz`. Similarly, [38] reported that combo-squatting (Type IV) and target embedding (Type III) were most prevalent but varied significantly per target for the MITM phishing websites. Typosquatting – e.g., replacing one character of a target domain name – was hardly employed. This deceptive method was studied in [63] and [58]. [63] searched proactively for phishing domains by crafting squatting domains and checking for their existence in DNS records. Half of the registered domains were live during the crawling window, and 3% of domains redirected users to domain marketplaces. Verification revealed that just 0.2% were phishing pages. However, 91.5% of these phishing pages remained undetected for at least a month, which suggests that they are more challenging to detect. A similar approach was followed in [58] by combining company registrations with `.com` zone files. 95% of the studied companies had at least one potential visual impersonation domain (VIDN), yet only 7% had at least one registered misspelling during the ten-year analysis period. Historical *WHOIS* records allowed for the clustering of VIDNs and showed that only a handful of companies register VIDNs defensively.

Execution. Half of the examined works have studied the execution of a phishing attack. Analyzing publicly available page logging on phishing websites in [45] estimated that when a phishing website was removed within one day of reporting, the average number of visitors disclosing their credentials was 18, with eight more for each day thereafter. Thirteen years later, an acceleration of this process was reported in [48], that found most visits take place in the nine hours between the first victim visit and detection. During these nine hours, phishing websites lure in 62% of their victims. In the 2007 analysis in [45], half of the responses entered were fake, and many visits to the landing page of a phishing website were observed. This was quantified in [31], that analyzed the visitors of the websites installed in honeypots. Many visits originated from security scanners. Just 9% of all real visitors (security scanners excluded) disclosed any credentials. Crafted credentials were fed into 150 live phishing websites in an experiment in [50] to examine their progression. Only seven leaked accounts received logins quickly after disclosure. Some accounts received multiple attempts from different IP addresses, probably the result of credentials being disclosed to multiple attackers through backdoors in the phishing kit. The longitudinal study of the underground ecosystem fueling credential theft in [62] was in collaboration with Google. 3,785 credential leak dumps were gathered by monitoring various online sources, which were checked against Google's user base. Over two million vulnerable Google users could be tied back to deployed phishing kits, 25% of them with a matching password. Examining the login geolocation of Gmail accounts that were involved revealed that 42% of them were last accessed in Nigeria. In [48], it was shown that the most prevalent geolocations coincide with countries disproportionately associated with cybercrime. This is in contrast to [31], that compared the geolocation of victim IP addresses with the target population of phishing kits and found that many received most of their victims from a single country. This was later confirmed in [5] by examining phishing kit installation times and their manuals.

Table 4: Phishing lifetime measurements.

Ref.	Year	Amount	Type	Uptime
[45]	2007	1,685	URLs	62h average, 20h median
[44]	2008	7,394	Domains	72h average, ±30% 55m
[31]	2016	474	Kits	192h
[63]	2018	1,741	Domains (sq.)	80% ≥ 1 month
[48]	2020	404,628	URLs	21h avarage
[5]	2021	1,288	Domains	45h average, 24h median
[38]	2021	1,220	Websites	40% ≥ 24h, 15% ≥ 480h

The increased use of URL shorteners also allowed for a new method of delivering phishing URLs. In 2008, their abuse was noticed in [44], albeit not very large – only 217 cases. More than ten years later, just 31 short URLs were discovered in the dataset of [50]. Many more were found in [43] that used URL shortening services to compare the life cycle of phishing and malware attacks. Analysis of `bit.ly` short URLs showed that phishing short URLs have a high click-through rate but a short uptime and are most active 4 hours before being reported as malicious.

Monetization. Just one study examined the monetization of phished credentials. [48] found that 7% of real visitors with an active account at the targeted organization suffered a fraudulent transaction, on average, five days after being phished. Additionally, 63% of the

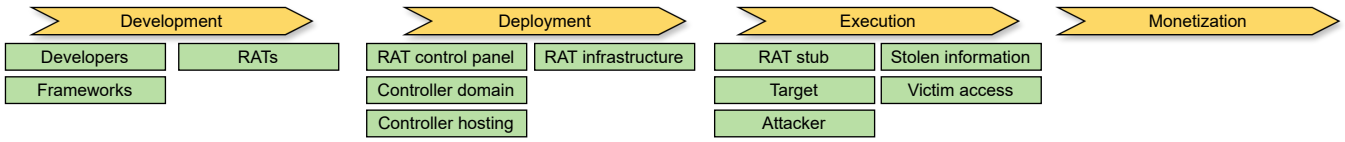

Figure 3: RAT value chain with its components at the top and the required resources at the bottom.

compromised accounts would later appear in a public dump, on average, almost a week later, suggesting that criminals first monetize phished credentials themselves before selling them off.

Phishing Takeaways. From Table 3, we learn that the research focus has shifted towards the left – from execution and deployment towards the development component in the value chain. Yet, only one study [48] yields insights into the monetization of phishing attacks. As a result, insights into how criminals make a profit from phishing attacks remain mostly anecdotal. Over time, more active measurements involving a crawler have been deployed, in contrast to earlier work that focused more on passive (domain) datasets. Consequently, due to changing attacker TTP, analysis of live phishing Web pages is increasingly included in research over time. The increased use of HTTPS by phishers has made their pages more believable [48], at the same time allowing for new possibilities to detect them through Certificate Transparency logs by defenders [5, 38]. We observe a similar pattern for the increased use of cloaking in phishing kits, which researchers have exploited by designing techniques to use it for detection [38, 70]. A closer look at the used data sources reveals the prevalent use of some data sources, such as PhishTank or APWG eCX, which could introduce biases [10]. For example, attacks mimicking PayPal are extensively studied, which raises the question of generalizing these results [31, 50]. Only two studies scoped their measurement to a company [48] or a country [5], while multiple studies suggested the localized nature of phishing [5, 31]. Lastly, just as with booter service measurements, extensive scanning from the security industry [31, 48] hampers robust measurements.

5 RATs Measurements

A Remote Access Trojan (RAT) is a type of malware that allows an attacker to take over a victim's computer [24]. Typically, this includes access to audio and video interfaces, as well as logging of mouse movement and keyboard input. RATs require manual interaction from an attacker and are not designed to execute and exfiltrate automatically, unlike traditional malware [68]. This makes RATs a preferred choice for targeted attacks [23, 24]. Monetary value is created through victim extortion or reselling initial access. A typical value chain for such an attack is depicted in Figure 3, in which we identify four components. In *development*, cybercriminals search for developers and software to achieve their remote access. The required infrastructure to successfully operate the software is set up during *deployment*. Next, in the *execution*, an attacker delivers a *stub*, which can be controlled using a *control panel* or *controller*, hosted at a domain and server under the attacker's control. Lastly, profits are made through the acquired access in the *monetization* component. Although RATs have been around since 1999 [68], they have not been extensively studied. We could identify five large-scale technical measurements published between 2017 and 2022, which we list in Table 5 and further detail in the following paragraphs.

Datasets. VirusTotal is the starting point for most RAT research. Through the years, consecutive studies collected more samples (stubs) for their analyses. First, [24] examined 19k samples in 2017, then [53] with 27k samples in 2018 and [23] used 146k samples in 2020. Only two studies did not use VirusTotal as a source for measurements. Significant effort was spent in [68] to search underground forums for RATs manually. In [25], malware domains from the GT Malware Passive DNS feed were combined with authoritative DNS data. Three studies employed Internet scanning to discover RAT artifacts [23, 24, 53] and two [24, 53] designed honeypots to study interactions. Three studies [23, 24, 53] focused predominantly on two RAT families, namely *DarkComet* or *njRAT*.

Development. An overview of RAT characteristics in [68] comprises static and dynamic analysis on 53 RATs. A stub was generated for each RAT, and it was found that high-level programming languages (such as C# and VB.NET) are the most popular, as they require only a few or no runtime dependencies. 90% of these RAT stubs targeted solely Windows computers. Additionally, analysis of both the stubs and controller panels revealed that over 80% of the RATs were able to log keystrokes, set up a remote shell, download and execute files, and enable the camera. Oftentimes, these functionalities were implemented similarly across different RATs. Both [24] and [23] did not study the development of RATs, but through their analysis of the deployment and execution of DarkComet, they did discover several facts about its development. Analysis in [24] revealed that DarkComet stubs contain a campaign ID to manage infections, a password to encrypt communications, and a list of controller IP addresses. Follow-up work in [23] detailed how to discover DarkComet controllers and download their victim databases.

Deployment. Different approaches have been employed to gain insights into RAT infrastructure. A live overview of DarkComet was obtained in [24] by extracting controller domains and IPs from stubs and through continuous Internet-wide scanning for specific banner responses. 175 online controllers were found at any given time (9,877 in total during an eight-month monitoring period). More controller activity was observed on the weekends compared to the rest of the week. Both [24] and [23] found Turkey and Russia to be hosting many DarkComet controllers, whereas [53] reported on the prevalence of North Africa, the Middle East, Brazil, and Russia as hosting locations. Many controller domains use a dynamic DNS (DDNS) service to rapidly change IPs. A user types mapping of these IPs suggests that roughly 90% of controllers are hosted on residential IP networks, likely with limited OPSEC [24]. Two strategies were deployed in [53] to examine RAT controller operations. Through *RAT-Hole*, a honeypot to mimic a RAT controller,

Table 5: Overview of large-scale RAT measurements. Measurements can be active (*A*), passive (*P*), or both (hybrid, *H*).

Authors	Year	Value chain				Data sources							Misc.				Time frame
		Development	Deployment	Execution	Monetization	VirusTotal	DNS	Internet scans	Hacker forums	RAT databases	Honeypots	Total samples	Measurement	DarkComet	njRAT	Other RATs	
Farinholt et al. [24]	2017		●	●		●	●	●			●	19k	*H*	✓			2016
Rezaeirad et al. [53]	2018		●	●		●		●			●	27k	*H*	✓	✓		2016 – 2017
Farinholt et al. [23]	2020		●	●		●		●		●		146k	*A*	✓			2016 – 2019
Faulkenberry et al. [25]	2022		●				●					245k	*P*	✓	✓	✓	2017 – 2021
Yang et al. [68]	2022	●							●			53	*P*	✓	✓	✓	1999 – 2016

and *RAT-Scan*, to mimic a victim searching for its controller. The latter periodically resolved discovered domains and IPs from sandbox executions and identified 4,584 njRAT and 2,032 DarkComet controller IPs within a seven-month measurement period. The use of DDNS as discovered in [24] was leveraged in [53] to register 6,897 expired DDNS domains previously used by RAT controllers. RAT-Hole revealed that the majority of traffic towards the honeypot originated from scanners and sandbox executions – a limited number of connections arose from victims. Analysis of downloaded DarkComet configurations in [23] discovered 3,518 IP addresses used by 1,162 RAT controllers in a 213-day measurement period. The average uptime of such controllers was 484 days, with some being functional for over three years. A global infrastructure of 399K IPs in 151 countries spread over 202 malware families during a four-year measurement period was observed in [25]. Various RATs were observed in use in 2022, although DarkComet and njRAT were the most prevalent. Additionally, much interest from scanners soon after a domain is listed as malicious was observed, inflating infection population counts if not properly filtered for.

Execution – Attackers. Monitoring attacker behavior in a controlled environment (e.g., a sandbox) reveals common attacker actions on a target machine. Executing 1,165 DarkComet samples in a honeypot in [24] learned that operators commonly access the webcam (61%), steal stored passwords (43%), or explore the victim's file system (40%). An average session lasted four minutes, 45% of the sessions were motivated by access to a human user – e.g., for harassment or extortion – and at least 58% of RAT operators were motivated by access to user credentials. Analysis in [53] showed that 43% of controller domains received only a single victim, 90% received at most 20 victims, and just 5% received over 40 victims. This suggests that some attackers widely distribute their malware, whereas others operate more targeted. A similar disparity was reported in [23], which found a median of two victims per controller, with ten outliers amassing over a thousand victims each. Lastly, [23] found that only 16% of operators use a VPN to hide their tracks.

Execution – Victims. Two studies examined RAT victims. In [53] by letting victims connect to expired DDNS domains and in [23] by scrutinizing downloaded controller databases. Victim connections in [53] showed that most victim IPs are static, and more than half of them have a webcam, making them susceptible to extortion via camera recordings. Infections last long, as 90 days after the controller domain expiration and registration by the researchers, 40% of domains were still receiving victim connections. Almost every country was home to RAT victims, with Brazil being the most prevalent. Correlating the controller and victim locations revealed that they are often located in the same country. This finding was later confirmed in [23], which found that more than 74% of attackers with limited victims are located in the same country as most of their victims. Leveraging downloaded databases in [23] revealed a total of 57,805 victims in a five-year measurement period. Several steps were taken to validate this number and to allow for comparison with [53], yielding an overestimation of the number of victims by 40%. Lastly, DarkComet was found to have collected 79,142 keystrokes and recorded 60 hours of activity over 9.6 days for each victim on average.

RATs Takeaways. Academic measurements focus predominantly on the deployment and execution of RATs, as shown in Table 5. The capabilities of stubs, attacker actions, and infrastructure have been well studied, but how RATs end up at victims or how infections are monetized remains unknown. The prevalent use of VirusTotal as a source for RAT stubs stands out and could introduce a measurement bias, as its database mostly depends on user submissions. This leads to the question of whether the stubs uploaded to VirusTotal are, in fact, a representative sample. Additionally, the discovered RAT characteristics are primarily based on two RATs (DarkComet or njRAT), whereas a great variety of RATs are listed in [25]. To what extent the findings generalize to other types of RATs is unknown. Attacker analysis distinguished two types of attackers: the ones operating on a large scale and targeted attackers operating locally [23, 53], which allows for more in-depth analyses of differences in TTP. Finally, just as with booters and phishing, scanners deployed by the security community make it hard to establish robust observations of victim traffic [25, 53].

6 Measurement Characteristics

Before we report on the results of our workshop with LEAs to elicit their assessment of various studies, we summarize the findings from our survey. When scholars measure cybercrime, they measure mostly its deployment and execution, not the monetization component in the value chain, as illustrated by Tables 1, 3, and 5. Many measurements leverage the same datasets or methods, like

the use of AmpPot in booter measurements [34, 42, 46], PhishTank in phishing measurements [13, 14, 16, 38, 43–45, 50, 63], and VirusTotal in RAT measurements [23, 24, 53]. The prevalent use of such datasets could introduce measurement biases. For example, a community-vetted database like PhishTank does not contain every phishing URL but does include all known PayPal phishing URLs due to their collaboration [10]. Which, consequently, results in an overestimation of PayPal's popularity as a target among phishers. And, although differentiation in temporal and geographical characteristics is mentioned in measurements of every type of cybercrime in our research [5, 23, 46], just two of the 38 studies had their focus adjusted to a specific geographical area or company. Internet-scale measurements provide a nice panoramic view of the global cybercrime landscape. Yet, studies have found phishers operating in one geographical or language area – often the same as theirs [5], gamers launching attacks at their neighboring rivals through booter services [46], and RAT operators targeting victims locally [53]. Such localities can easily get lost in an analysis on an Internet scale, where a geographical focus is absent. Therefore, the takeaways of such analyses represent an average across countries, not within a country, making them less actionable for LEAs operating within local jurisdictions. From the attacker analyses across the examined cybercrimes, we can differentiate two types of actors: cybercriminal groups performing large-scale untargeted attacks – e.g., originating from countries known for cybercrime [48, 62] – and individuals performing small-scale attacks locally (e.g., many RAT operators with few victims [53]). A clear differentiation between these attackers is essential in measurements. Just as geographical diversity, including both groups in one measurement, can result in senseless averages – as these groups rely on very distinct TTP. Lastly, multiple studies across the three cybercrimes mention the amount of scanning by other researchers that hamper robust measurements [42, 48, 53], which should be accounted for properly.

7 Measurements for Law Enforcement

As detailed in previous sections, scholars have spent significant effort measuring cybercrime. However, little is known about what kind of measurements (i.e., data, methods, and analyses) cater best to LEA needs. Some papers specifically mention how their measurements could assist LEAs in policing cybercrime [23, 45], but does law enforcement agree? To answer our second research question, we engage with LEA professionals in the Netherlands during a workshop with two objectives. First, we want to elicit their assessment of the added value of specific measurements to see which characteristics align with LEA needs. Second, we wanted their insights into what innovative measurements can cater to LEA needs. Dutch LEAs have been involved in several high-profile cybercrime interventions, such as the takedown of QakBot [64], Webstresser [18], and Bestmixer [20]. Such experiences make them a good partner for this study.

Workshop approach. We invited a diverse group of seven LEA professionals consisting of four participants from the Dutch National Police (ranging from analysts to project managers – all in dedicated cybercrime units), two from the Dutch Public Prosecution Service tasked with cybercrime cases, and one from the cybercrime unit of the Dutch Fiscal Information and Investigation Service (FIOD).

Table 6: Scores assigned by LEA participants, ranging from 1 (totally not agreed) to 7 (totally agreed) ($N = 7$).

	Reference	*Understanding*	*Connection*	*Action*
Booters	Santanna et al. [56]	4.0	2.2	2.8
	Brunt et al. [7]	4.3	3.8	3.9
	Kopp et al. [41]	4.8	4.0	3.8
Phish.	Peng et al. [50]	4.0	2.5	3.1
	Oest et al. [48]	5.0	3.3	3.1
	Bijmans et al. [5]	5.0	4.3	4.1
RATs	Farinholt et al. [24]	4.9	3.3	3.3
	Rezaeirad et al. [53]	4.1	3.3	3.0
	Farinholt et al. [23]	4.7	4.3	4.9

Their experience was evenly spread and ranged from one to 15+ years. All participants gave verbal informed consent to participate in our study under the condition that their names or any other PII would not be used in any publication. They also agreed to record and transcribe the workshop. To collect their perspectives on what makes a study more or less aligned with their needs, we present them with specific studies rather than abstract overviews of a large set of studies. Exposing them to all 38 studies would not be feasible; thus, we selected three studies with different characteristics in terms of data, methods, and findings for each of the three crime types – nine studies in total. In the workshop, we presented 100-word summaries of each study. Each summary detailed the data source(s), methods, and main findings. Participants were asked to rate each study on three criteria, using seven-point Likert scales. First, the participants were asked about *understanding the phenomenon*: To what extent does this paper add to your understanding of the measured cybercrime? Second, about the *connection to police work*: To what extent does this study connect to your daily work as a law enforcement professional? And third, about the *actionable perspectives*: To what extent do elements of this study offer clear and actionable perspectives? Additionally, we asked participants to reflect on the study in a five-minute discussion. This open-ended approach allowed for a reflexive process of inductive reasoning to discover the characteristics of studies that align with LEA needs. One researcher distilled such characteristics from the transcriptions and audio recordings, and the other researcher who attended the workshop agreed with its findings. In the second part of the workshop, we asked our participants how to advance cybercrime measurements by bringing new measurement ideas forward in a brainstorming session. To allow for fresh ideas, we structured this around three different cybercrimes than the ones already discussed, namely *ransomware*, *online stolen data markets* (e.g., data from phishing or data breaches), and *bulletproof hosting* (hosting providers that do not comply with LEA inquiries when hosting malicious content). For each of these crimes, the participants were asked to write down ideas for new measurements on Post-its. These Post-its were collected per cybercrime on a large sheet of paper. After reading all the ideas, every participant was given 10 points per cybercrime, which they were asked to divide. The highest-scoring ideas were discussed to capture the rationale behind them. The results of both workshop parts are detailed in the next sections.

Booters. As shown in Table 6, all studies were rated moderately well on contributing to a better understanding of the phenomenon. The work in [56] was met with skepticism, and questions were raised regarding the relevancy of this work, as it was published eight years ago. Additionally, the finding that a third of all purchased attacks at booters are not executed was deemed not interesting as "criminals scamming criminals is not our priority", according to Participant 2. The responses to the measurements presented in [41] were more positive. Ironically, multiple participants agreed with Participant 3, who stated: "It is great to see how little effect these police interventions have. This should be taken into account when designing other interventions". Participant 1 added that "taking down booters is just a game of whack-a-mole, it's better to aim higher". However, this person also added that "police actions are also normative, especially with types of crime that are difficult to control", indicating that such actions send a message to other criminals even though direct results are limited. This study inspired Participant 5 to state that "it would be nice to collaborate with academics whenever we act. They could measure before and after our actions to analyze the effects." Insights into monetizing booters in [7] gave our participants concrete and actionable perspectives. The popularity of a certain payment service provider can spark future cooperation, especially when such a provider is willing to work with law enforcement. According to Participant 4, this shows that public-private partnerships to police cybercrime are effective, as "profits diminished despite customers changing payment methods" in [41].

Phishing. All presented studies were rated moderately well on contributing to a better understanding of the phenomenon but differed in being well-connected to police work and whether or not they offered an action perspective, as shown in Table 6. This discrepancy is well illustrated by the discussion that followed upon the life cycle measurements in [48]. This paper illustrates the short life cycle of phishing URLs, being online for only 21 hours on average. Participant 1 from the Dutch police said: "It is shocking to see these short time frames, as we can never act in this time frame, it's simply too short." The measurements in [48] and [5] strengthened some participants in their beliefs that phishing can not be stopped through criminal investigations alone, but also by taking preventive actions or through public-private partnerships. Various participants appreciated the local focus in [5] because its findings directly apply to their work. Additionally, the "insights into what phishing kits are popular could steer our investigations towards the developers of the most popular kits.", according to Participant 2. The methods and results in [50] did not align well with police practice, according to our participants. Purposely leaking credentials and observing their evolution did not contribute to anything LEAs can act upon. Despite an idea for a public-private partnership ("if PayPal is really that prevalent, we should cooperate with them", participant 5), it offered no useful insights for criminal investigations.

RATs. All presented studies featured analyses of either one or two different RATs, which made Participant 2 raise the question: "there are hundreds of different RATs active at the moment, did the researchers know how prevalent this RAT was?". Besides this question, all papers were rated moderately well regarding both phenomenon understanding, as shown in Table 6, yet [23] scored significantly better on actionable perspective. The geographic distributions of both attackers and defenders reported in [23] were well-received: "It is always interesting to see the numbers. Such an overview shows that victim notification is possible" (Participant 5), referring to the victim counts per country in [23]. Additionally, the long RAT controller uptime finding showed that "there is input for criminal investigations, because of the long uptimes. In the 400 days that some RATs remain active, we can easily do a full investigation!" (Participant 3). Participant 4 from the FIOD noted that "we often think about cybercrime on a global scale, this paper shows that victim and perpetrator are much closer to each other, which makes it more worthwhile to investigate", referring to the fact that prosecuting a Dutch perpetrator is much easier than a foreign one. Additionally, the analysis of perpetrator OPSEC gave relevant hints for LEA action. The low use of VPNs by RAT controllers could be exploited for identification. The findings of [53] showed that dependencies of malicious actors on legitimate services could be exploited for law enforcement investigations or in public-private partnerships. Regarding the popularity of one VPN provider, Participant 1 stated that "the use of commercial products, like these VPN services, makes them ideal targets for public-private partnerships."

Advancing Measurements. For ransomware, participants wanted to learn about money laundering within the ecosystem. Since the victim payments are transacted in cryptocurrencies, what do criminals do next to convert their assets to fiat currencies? The second most-voted measurement was related to victims, emphasizing both geographical and sectoral analysis. Does ransomware tend to target specific organizations, or is it mostly opportunistic, capitalizing on initial access? Such insights could assist in designing preventive measures. Exact profit calculations or the means of initial access were less popular. As data breaches and stolen credentials enable many different forms of cybercrime [61], our participants expressed interest in learning more about the data types. Is every piece of data equally valuable, or are some more interesting – and more expensive – than others? And, to what extent does it influence the price of stolen data? The total volume of leaked credentials offered in the underground economy was less interesting to the participants. To learn which hosting providers can be considered bulletproof, our participants question the ratio of malicious and benign content that would classify a provider as bulletproof. Additionally, they are curious to know how such providers advertise in the underground economy. Customers are interesting as well, and our participants expressed the desire to understand both their background and payment methods.

To summarize, the LEA participants valued scientific measurements, as Participant 5 stated: "we fail in keeping a close eye on what science does, we should hire more researchers that could introduce these scientific insights into our daily work." The high scores related to the *"understanding the phenomenon"* column in Table 6 illustrate this. However, the lower scores in the other two columns emphasize that although scientific measurements can help to better understand

cybercrime, they can not directly assist in combating it, as they do not offer concrete action perspectives. The few higher-scoring studies generated actionable perspectives related to geographical differentiation [5, 23], development [5], monetization [7] or the effects of interventions [7, 41]. Measurements of solely the deployment or execution components were less valued. Accurate numbers related to their jurisdiction allow LEAs to act, monetization insights allow for public-private partnerships with, for instance, payment service providers, whereas development insights could steer criminal investigations toward the developers of malicious software instead of their many clients downstream. Findings originating from these scientific measurements could be leveraged to set up public-private partnerships with companies (ab)used by criminals. Brainstorming on innovative measurements highlighted the wish to gain predominantly insights into the monetization of cybercrime, illustrated by the desire to know more about money laundering in ransomware, stolen data price mechanics, and bulletproof hosting advertising.

8 Discussion

In this section, we discuss the inherent limitations of our work and mark interesting avenues for future work.

Limitations. We identify three limitations in our research that could have influenced our findings: our selection of cybercrimes, paper collection, and the recruitment of workshop participants. First, we selected only three cybercrimes to survey the academic field. As mentioned in § 2, we selected booter services, phishing, and RATs based on three years of Europol reporting [19, 21, 22]. We generalize our conclusions for all cybercrime research based on our survey of studies covering only these three cybercrimes, which could introduce a bias. Given the large differences in value chains across these cybercrimes, we, however, don't think our conclusions would be vastly different if we had selected other cybercrimes. Second, although we took a systematic approach to survey past cybercrime measurements, as elaborated upon in § 2, it is possible that we have missed or incorrectly discarded studies as part of our selection process. Lastly, the workshop with law enforcement professionals included only a limited number of participants. Although these participants came from a variety of agencies and had different roles, the group could have been biased.

Future work. We envision three avenues for future work from the results of this combined survey and user study. First, we encourage researchers to perform more large-scale measurement studies with a sole focus on cybercrime measurements instead of creating detection methodologies. Since one of the pillars of science is to build on the shoulders of previous researchers, we promote research that takes already published detection methodology from earlier work to perform robust measurements of cybercrime. Second, in performing such measurements, we argue that well-demarcated studies in terms of jurisdiction, geography, or language could aid in better connecting scientific research to police operations. We acknowledge that such research requires good relations with local law enforcement – which can be difficult to achieve – and that such research is not appropriate at every university, nor approved by every institutional research board. Finally, comparative studies between the aforementioned demarcations have the potential to demonstrate differences in actor TTP and in the law enforcement efforts to police them. Such future work would help improve our understanding of cybercrime across the globe and discover what approaches work to combat cybercrime.

9 Conclusions

Combining our survey of large-scale technical measurements and the workshop with LEA professionals allows us to answer our research questions and find the nexus between cybercrime science and policing. In short, we find a mismatch between LEA needs and academic measurements. Most measurements focus on the deployment and execution of cybercrime, whereas LEAs desire to learn more about development and monetization. Although the deployment and execution components are paramount to measure, analyses on the source or how profits are made would also be beneficial, especially for LEAs to build an intervention repertoire. They provide clues that could bring investigators closer to the people facilitating these types of cybercrime, allowing for proactive and disruptive infrastructural policing with longer-lasting effects [12]. That is, monetization insights allow for public-private partnerships with, for instance, payment service providers, whereas development insights could steer criminal investigations toward the developers of malicious software instead of their many clients downstream. As noted in [14], taking down individual phishing websites is far less efficient than policing phishing kit developers, which is what LEAs – given their scarce resources – should focus on. Law enforcement doesn't prioritize investigations into hundreds of individual phishing pages, but this changes as soon as all these websites can be associated with one attacker [45]. To arrive at this conclusion, one needs large-scale technical measurements – which academics could design. Such measurements need a geographical focus, which is essential to LEAs as they operate in conjunction with local jurisdictions. Also, LEA interventions lean on attacker differentiation, as policing individuals requires a different enforcement repertoire than organized cybercriminal groups. Aligning academic and LEA opportunities, considering all value chain components, actor differentiation, and geographical diversity, augments the nexus between large-scale technical measurements and cybercrime policing.

References

[1] Ross Anderson, Chris Barton, Rainer Böhme, Richard Clayton, Carlos Gañan, Tom Grasso, Michael Levi, Tyler Moore, Stefan Savage, and Marie Vasek. 2019. Measuring the Changing Cost of Cybercrime. In *The 18th Annual Workshop on the Economics of Information Security (WEIS 2019)*. Boston, Massachusetts.

[2] Ross Anderson, Chris Barton, Rainer Böhme, Richard Clayton, Michel J. G. van Eeten, Michael Levi, Tyler Moore, and Stefan Savage. 2013. Measuring the Cost of Cybercrime. In *The Economics of Information Security and Privacy*, Rainer Böhme (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 265–300. https://link.springer.com/10.1007/978-3-642-39498-0_12

[3] APWG. 2023. The APWG eCrime Exchange (ECX). https://apwg.org/ecx/

[4] Rasika Bhalerao, Maxwell Aliapoulios, Ilia Shumailov, Sadia Afroz, and Damon McCoy. 2019. Mapping the Underground: Supervised Discovery of Cybercrime Supply Chains. In *2019 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, Pittsburgh, PA, USA, 1–16. https://doi.org/10.1109/eCrime47957.2019.9037582

[5] Hugo Bijmans, Tim Booij, Anneke Schwedersky, Aria Nedgabat, and Rolf van Wegberg. 2021. Catching Phishers By Their Bait: Investigating the Dutch Phishing Landscape through Phishing Kit Detection. In *30th USENIX Security Symposium*. USENIX Association, Virtual Event, 3757–3774. https://www.usenix.org/conference/usenixsecurity21/presentation/bijmans

[6] Marzieh Bitaab, Haehyun Cho, Adam Oest, Penghui Zhang, Zhibo Sun, Rana Pourmohamad, Doowon Kim, Tiffany Bao, Ruoyu Wang, Yan Shoshitaishvili, Adam Doupe, and Gail-Joon Ahn. 2020. Scam Pandemic: How Attackers Exploit Public Fear through Phishing. In *2020 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, Boston, MA, USA, 1–10. https://doi.org/10.1109/eCrime51433.2020.9493260

[7] Ryan Brunt, Prakhar Pandey, and Damon McCoy. 2017. Booted: An Analysis of a Payment Intervention on a DDoS-for-Hire Service. In *Workshop on the Economics of Information Security (WEIS)*. San Diego, 6–26.

[8] Centraal Bureau voor de Statistiek (CBS). 2022. Minder Traditionele Criminaliteit, Meer Online Criminaliteit. https://www.cbs.nl/nl-nl/nieuws/2022/09/minder-traditionele-criminaliteit-meer-online-criminaliteit

[9] Justyna Joanna Chromik, Jose Jair Santanna, Anna Sperotto, and Aiko Pras. 2015. Booter Websites Characterization: Towards a List of Threats. In *33rd Brazilian Symposium on Computer Networks and Distributed Systems (SBRC)*. IEEE, Vitória, Brazil, 445–458.

[10] Richard Clayton, Tyler Moore, and Nicolas Christin. 2015. Concentrating Correctly on Cybercrime Concentration. In *14th Annual Workshop on the Economics of Information Security (WEIS)*. Delft, The Netherlands, 16.

[11] Ben Collier, Daniel R. Thomas, Richard Clayton, and Alice Hutchings. 2019. Booting the Booters: Evaluating the Effects of Police Interventions in the Market for Denial-of-Service Attacks. In *Proceedings of the Internet Measurement Conference*. ACM, Amsterdam Netherlands, 50–64. https://doi.org/10.1145/3355369.3355592

[12] Ben Collier, Daniel R. Thomas, Richard Clayton, Alice Hutchings, and Yi Ting Chua. 2022. Influence, Infrastructure, and Recentering Cybercrime Policing: Evaluating Emerging Approaches to Online Law Enforcement through a Market for Cybercrime Services. *Policing and Society* 32, 1 (Jan. 2022), 103–124. https://doi.org/10.1080/10439463.2021.1883608

[13] Marco Cova, Christopher Kruegel, and Giovanni Vigna. 2008. There Is No Free Phish: An Analysis of "Free" and Live Phishing Kits. In *2nd USENIX Workshop on Offensive Technologies (WOOT)*. USENIX Association, San Jose, CA, USA.

[14] Qian Cui, Guy-Vincent Jourdan, Gregor V. Bochmann, Russell Couturier, and Iosif-Viorel Onut. 2017. Tracking Phishing Attacks Over Time. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Perth Australia, 667–676. https://doi.org/10.1145/3038912.3052654

[15] Avisha Das, Shahryar Baki, Ayman El Aassal, Rakesh Verma, and Arthur Dunbar. 2020. SoK: A Comprehensive Reexamination of Phishing Research From the Security Perspective. *IEEE Communications Surveys & Tutorials* 22, 1 (2020), 671–708. https://doi.org/10.1109/COMST.2019.2957750

[16] Vincent Drury, Luisa Lux, and Ulrike Meyer. 2022. Dating Phish: An Analysis of the Life Cycles of Phishing Attacks and Campaigns. In *Proceedings of the 17th International Conference on Availability, Reliability and Security*. ACM, Vienna Austria, 1–11. https://doi.org/10.1145/3538969.3538997

[17] Benoit Dupont. 2017. Bots, Cops, and Corporations: On the Limits of Enforcement and the Promise of Polycentric Regulation as a Way to Control Large-Scale Cybercrime. *Crime, Law and Social Change* 67, 1 (Feb. 2017), 97–116. https://doi.org/10.1007/s10611-016-9649-z

[18] Europol. 2018. World's Biggest Marketplace Selling Internet Paralysing DDoS Attacks Taken down. https://www.europol.europa.eu/media-press/newsroom/news/world%E2%80%99s-biggest-marketplace-selling-internet-paralysing-ddos-attacks-taken-down

[19] Europol. 2019. *Internet Organised Crime Threat Assessment (IOCTA) 2019*. Technical Report. Publications Office of the European Union, Luxembourg. https://www.europol.europa.eu/publications-events/main-reports/internet-organised-crime-threat-assessment-iocta-2019

[20] Europol. 2019. Multi-Million Euro Cryptocurrency Laundering Service Bestmixer.Io Taken Down. https://www.europol.europa.eu/media-press/newsroom/news/multi-million-euro-cryptocurrency-laundering-service-bestmixerio-taken-down

[21] Europol. 2021. *Internet Organised Crime Threat Assessment (IOCTA) 2021*. Technical Report. Publications Office of the European Union, Luxembourg. https://www.europol.europa.eu/publications-events/main-reports/internet-organised-crime-threat-assessment-iocta-2021

[22] Europol. 2023. *Internet Organised Crime Threat Assessment (IOCTA) 2023*. Technical Report. Publications Office of the European Union, Luxembourg. https://www.europol.europa.eu/publication-events/main-reports/internet-organised-crime-assessment-iocta-2023

[23] Brown Farinholt, Mohammad Rezaeirad, Damon McCoy, and Kirill Levchenko. 2020. Dark Matter: Uncovering the DarkComet RAT Ecosystem. In *Proceedings of The Web Conference 2020*. ACM, Taipei Taiwan, 2109–2120. https://doi.org/10.1145/3366423.3380277

[24] Brown Farinholt, Mohammad Rezaeirad, Paul Pearce, Hitesh Dharmdasani, Haikuo Yin, Stevens Le Blond, Damon McCoy, and Kirill Levchenko. 2017. To Catch a Ratter: Monitoring the Behavior of Amateur DarkComet RAT Operators in the Wild. In *2017 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Jose, CA, 770–787. https://doi.org/10.1109/SP.2017.48

[25] Aaron Faulkenberry, Athanasios Avgetidis, Zane Ma, Omar Alrawi, Charles Lever, Panagiotis Kintis, Fabian Monrose, Angelos D. Keromytis, and Manos Antonakakis. 2022. View from Above: Exploring the Malware Ecosystem from the Upper DNS Hierarchy. In *Proceedings of the 38th Annual Computer Security Applications Conference*. ACM, Austin, TX, USA, 240–250. https://doi.org/10.1145/3564625.3564646

[26] FBI. 2022. Internet Crime Report 2022. https://www.ic3.gov/Media/PDF/AnnualReport/2022_IC3Report.pdf

[27] Sujata Garera, Niels Provos, Monica Chew, and Aviel D. Rubin. 2007. A Framework for Detection and Measurement of Phishing Attacks. In *Proceedings of the 2007 ACM Workshop on Recurring Malcode*. ACM, Alexandria Virginia USA, 1–8. https://doi.org/10.1145/1314389.1314391

[28] Google. 2023. Certificate Transparency - Working Together to Detect Maliciously or Mistakenly Issued Certificates. https://certificate.transparency.dev/

[29] Google. 2023. Computer Security & Cryptography - Top Publications. https://scholar.google.com/citations?view_op=top_venues&vq=eng_computersecuritycryptography

[30] Harm Griffioen, Kris Oosthoek, Paul Van Der Knaap, and Christian Doerr. 2021. Scan, Test, Execute: Adversarial Tactics in Amplification DDoS Attacks. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Virtual Event Republic of Korea, 940–954. https://doi.org/10.1145/3460120.3484747

[31] Xiao Han, Nizar Kheir, and Davide Balzarotti. 2016. PhishEye: Live Monitoring of Sandboxed Phishing Kits. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Vienna Austria, 1402–1413. https://doi.org/10.1145/2976749.2978330

[32] Keman Huang, Michael Siegel, and Stuart Madnick. 2019. Systematically Understanding the Cyber Attack Business: A Survey. *Comput. Surveys* 51, 4 (July 2019), 1–36. https://doi.org/10.1145/3199674

[33] ISACA. 2019. *State of Cybersecurity 2019 Part 2: Current Trends in Attacks, Awareness and Governance*. Technical Report. ISACA, Schaumburg, IL, USA. https://www.isaca.org/-/media/files/isacadp/project/isaca/why-isaca/surveys-and-reports/state-of-cybersecurity-2019-part-2_res_eng_0619

[34] Mattijs Jonker, Alistair King, Johannes Krupp, Christian Rossow, Anna Sperotto, and Alberto Dainotti. 2017. Millions of Targets under Attack: A Macroscopic Characterization of the DoS Ecosystem. In *Proceedings of the 2017 Internet Measurement Conference*. ACM, London United Kingdom, 100–113. https://doi.org/10.1145/3131365.3131383

[35] Mohammad Karami and Dammon McCoy. 2013. Rent to Pwn: Analyzing Commodity Booter DDoS Services. *login Usenix Magazine* 38, 6 (2013). https://www.usenix.org/publications/login/december-2013-volume-38-number-6/rent-pwn-analyzing-commodity-booter-ddos

[36] Mohammad Karami, Youngsam Park, and Damon McCoy. 2016. Stress Testing the Booters: Understanding and Undermining the Business of DDoS Services. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Montréal Québec Canada, 1033–1043. https://doi.org/10.1145/2872427.2883004

[37] Amin Kharraz, Zane Ma, Paul Murley, Charles Lever, Joshua Mason, Andrew Miller, Nikita Borisov, Manos Antonakakis, and Michael Bailey. 2019. Outguard: Detecting In-Browser Covert Cryptocurrency Mining in the Wild. In *The World Wide Web Conference on - WWW '19*. ACM Press, San Francisco, CA, USA, 840–852. https://doi.org/10.1145/3308558.3313665

[38] Brian Kondracki, Babak Amin Azad, Oleksii Starov, and Nick Nikiforakis. 2021. Catching Transparent Phish: Analyzing and Detecting MITM Phishing Toolkits. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Virtual Event Republic of Korea, 36–50. https://doi.org/10.1145/3460120.3484765

[39] Radhesh Krishnan Konoth, Emanuele Vineti, Veelasha Moonsamy, Martina Lindorfer, Christopher Kruegel, Herbert Bos, and Giovanni Vigna. 2018. MineSweeper: An In-depth Look into Drive-by Cryptocurrency Mining and Its Defense. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Toronto Canada, 1714–1730. https://doi.org/10.1145/3243734.3243858

[40] Daniel Kopp, Christoph Dietzel, and Oliver Hohlfeld. 2021. DDoS Never Dies? An IXP Perspective on DDoS Amplification Attacks. In *Passive and Active Measurement - 22nd International Conference (Lecture Notes in Computer Science, Vol. 12671)*. IEEE, Virtual Event, 284–301. https://doi.org/10.1007/978-3-030-72582-2_17

[41] Daniel Kopp, Matthias Wichtlhuber, Ingmar Poese, Jair Santanna, Oliver Hohlfeld, and Christoph Dietzel. 2019. DDoS Hide & Seek: On the Effectiveness of a Booter Services Takedown. In *Proceedings of the Internet Measurement Conference*. ACM, Amsterdam Netherlands, 65–72. https://doi.org/10.1145/3355369.3355590

[42] Lukas Krämer, Johannes Krupp, Daisuke Makita, Tomomi Nishizoe, Takashi Koide, Katsunari Yoshioka, and Christian Rossow. 2015. AmpPot: Monitoring and Defending Against Amplification DDoS Attacks. In *Research in Attacks, Intrusions, and Defenses (RAID) (Lecture Notes in Computer Science, Vol. 9404)*. Springer, Kyoto Japan, 615–636.

[43] Sophie Le Page, Guy-Vincent Jourdan, Gregor V. Bochmann, Jason Flood, and Iosif-Viorel Onut. 2018. Using URL Shorteners to Compare Phishing and Malware Attacks. In *2018 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, San Diego, CA, 1–13. https://doi.org/10.1109/ECRIME.2018.8376215

[44] D Kevin McGrath and Minaxi Gupta. 2008. Behind Phishing: An Examination of Phisher Modi Operandi. In *First USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*. USENIX Association, San Francisco, CA, USA.

[45] Tyler Moore and Richard Clayton. 2007. Examining the Impact of Website Take-down on Phishing. In *Proceedings of the Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit*, Vol. 269. ACM, Pittsburgh Pennsylvania USA, 1–13. https://doi.org/10.1145/1299015.1299016

[46] Arman Noroozian, Maciej Korczyński, Carlos Hernandez Gañan, Daisuke Makita, Katsunari Yoshioka, and Michel Van Eeten. 2016. Who Gets the Boot? Analyzing Victimization by DDoS-as-a-Service. In *Research in Attacks, Intrusions, and Defenses - 19th International Symposium, RAID 2016 (Lecture Notes in Computer Science, Vol. 9854)*. Springer, Paris, France, 368–389.

[47] Adam Oest, Yeganeh Safei, Adam Doupe, Gail-Joon Ahn, Brad Wardman, and Gary Warner. 2018. Inside a Phisher's Mind: Understanding the Anti-Phishing Ecosystem through Phishing Kit Analysis. In *2018 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, San Diego, CA, 1–12. https://doi.org/10.1109/ECRIME.2018.8376206

[48] Adam Oest, Penghui Zhang, Brad Wardman, Eric Nunes, Jakub Burgis, Ali Zand, Kurt Thomas, Adam Doupé, and Gail-Joon Ahn. 2020-08-12/2020-08-14. Sunrise to Sunset: Analyzing the End-to-end Life Cycle and Effectiveness of Phishing Attacks at Scale. In *29th USENIX Security Symposium (USENIX Security 2020)*. USENIX Association, Virtual Event, 361–377. https://www.usenix.org/conference/usenixsecurity20/presentation/oest-sunrise

[49] OpenPhish. 2023. OpenPhish - Phishing Intelligence. https://openphish.com/

[50] Peng Peng, Chao Xu, Luke Quinn, Hang Hu, Bimal Viswanath, and Gang Wang. 2019. What Happens After You Leak Your Password: Understanding Credential Sharing on Phishing Sites. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security*. ACM, Auckland New Zealand, 181–192. https://doi.org/10.1145/3321705.3329818

[51] Kirsty Phillips, Julia C. Davidson, Ruby R. Farr, Christine Burkhardt, Stefano Caneppele, and Mary P. Aiken. 2022. Conceptualizing Cybercrime: Definitions, Typologies and Taxonomies. *Forensic Sciences* 2, 2 (April 2022), 379–398. https://doi.org/10.3390/forensicsci2020028

[52] PhishTank. 2023. PhishTank - Join the Fight against Phishing. https://phishtank.org/

[53] Mohammad Rezaeirad, Brown Farinholt, Hitesh Dharmdasani, Paul Pearce, Kirill Levchenko, and Damon McCoy. 2018. Schrödinger's RAT: Profiling the Stakeholders in the Remote Access Trojan Ecosystem. In *27th USENIX Security Symposium*. USENIX Association, Baltimore, MD, USA, 1043–1060.

[54] Markus Riek and Rainer Böhme. 2018. The Costs of Consumer-Facing Cybercrime: An Empirical Exploration of Measurement Issues and Estimates. *Journal of Cybersecurity* 4, 1 (Jan. 2018). https://doi.org/10.1093/cybsec/tyy004

[55] Jose Jair Santanna, Romain Durban, Anna Sperotto, and Aiko Pras. 2015. Inside Booters: An Analysis on Operational Databases. In *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE, Ottawa, ON, Canada, 432–440. https://doi.org/10.1109/INM.2015.7140320

[56] Jose Jair Santanna, Roland van Rijswijk-Deij, Rick Hofstede, Anna Sperotto, Mark Wierbosch, Lisandro Zambenedetti Granville, and Aiko Pras. 2015. Booters – An Analysis of DDoS-as-a-service Attacks. In *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE, Ottawa, ON, Canada, 243–251. https://doi.org/10.1109/INM.2015.7140298

[57] Geoffrey Simpson and Tyler Moore. 2020. Empirical Analysis of Losses from Business-Email Compromise. In *2020 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, Boston, MA, USA, 1–7. https://doi.org/10.1109/eCrime51433.2020.9493250

[58] Geoffrey Simpson, Tyler Moore, and Richard Clayton. 2020. Ten Years of Attacks on Companies Using Visual Impersonation of Domain Names. In *2020 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, Boston, MA, USA, 1–12. https://doi.org/10.1109/eCrime51433.2020.9493251

[59] Karthika Subramani, William Melicher, Oleksii Starov, Phani Vadrevu, and Roberto Perdisci. 2022. PhishInPatterns: Measuring Elicited User Interactions at Scale on Phishing Websites. In *Proceedings of the 22nd ACM Internet Measurement Conference*. ACM, Nice France, 589–604. https://doi.org/10.1145/3517745.3561467

[60] Daniel R. Thomas, Richard Clayton, and Alastair R. Beresford. 2017. 1000 Days of UDP Amplification DDoS Attacks. In *2017 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, Scottsdale, AZ, 79–84. https://doi.org/10.1109/ECRIME.2017.7945057

[61] Kurt Thomas, Danny Yuxing Huang, David Wang, Elie Bursztein, Chris Grier, Thomas J Holt, Christopher Kruegel, Damon McCoy, Stefan Savage, and Giovanni Vigna. 2015-06-22/2015-06-23. Framing Dependencies Introduced by Underground Commoditization. In *14th Annual Workshop on the Economics of Information Security (WEIS)*. Delft, The Netherlands.

[62] Kurt Thomas, Frank Li, Ali Zand, Jacob Barrett, Juri Ranieri, Luca Invernizzi, Yarik Markov, Oxana Comanescu, Vijay Eranti, Angelika Moscicki, Daniel Margolis, Vern Paxson, and Elie Bursztein. 2017. Data Breaches, Phishing, or Malware?: Understanding the Risks of Stolen Credentials. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Dallas Texas USA, 1421–1434. https://doi.org/10.1145/3133956.3134067

[63] Ke Tian, Steve T. K. Jan, Hang Hu, Danfeng Yao, and Gang Wang. 2018. Needle in a Haystack: Tracking Down Elite Phishing Domains in the Wild. In *Proceedings of the Internet Measurement Conference 2018*. ACM, Boston MA USA, 429–442. https://doi.org/10.1145/3278532.3278569

[64] U.S. Attorney's Office. 2023. Qakbot Malware Disrupted in International Cyber Takedown. https://www.justice.gov/usao-cdca/pr/qakbot-malware-disrupted-international-cyber-takedown

[65] U.S. Department of Justice. 2018. Criminal Charges Filed in Los Angeles and Alaska in Conjunction with Seizures Of 15 Websites Offering DDoS-For-Hire Services. https://www.justice.gov/opa/pr/criminal-charges-filed-los-angeles-and-alaska-conjunction-seizures-15-websites-offering-ddos

[66] Rolf van Wegberg, Samaneh Tajalizadehkhoob, Kyle Soska, Ugur Akyazi, Carlos Ganan, Bram Klievink, Nicolas Christin, and Michel van Eeten. 2018. Plug and Prey? Measuring the Commoditization of Cybercrime via Online Anonymous Markets. In *27th USENIX Security Symposium*. USENIX Association, Baltimore, USA, 1009–1026. https://www.usenix.org/conference/usenixsecurity18/presentation/van-wegberg

[67] Wenhao Wang, Benjamin Ferrell, Xiaoyang Xu, Kevin W. Hamlen, and Shuang Hao. 2018. SEISMIC: SEcure In-lined Script Monitors for Interrupting Cryptojacks. In *Computer Security*, Javier Lopez, Jianying Zhou, and Miguel Soriano (Eds.). Vol. 11099. Springer International Publishing, Cham, 122–142. http://link.springer.com/10.1007/978-3-319-98989-1_7

[68] Runqing Yang, Xutong Chen, Haitao Xu, Yueqiang Cheng, Chunlin Xiong, Linqi Ruan, Mohammad Kavousi, Zhenyuan Li, Liheng Xu, and Yan Chen. 2022. RATScope: Recording and Reconstructing Missing RAT Semantic Behaviors for Forensic Analysis on Windows. *IEEE Transactions on Dependable and Secure Computing* 19, 3 (May 2022), 1621–1638. https://doi.org/10.1109/TDSC.2020.3032570

[69] Penghui Zhang, Adam Oest, Haehyun Cho, Zhibo Sun, Rc Johnson, Brad Wardman, Shaown Sarker, Alexandros Kapravelos, Tiffany Bao, Ruoyu Wang, Yan Shoshitaishvili, Adam Doupe, and Gail-Joon Ahn. 2021. CrawlPhish: Large-scale Analysis of Client-side Cloaking Techniques in Phishing. In *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Francisco, CA, USA, 1109–1124. https://doi.org/10.1109/SP40001.2021.00021

[70] Penghui Zhang, Zhibo Sun, Sukwha Kyung, Hans Walter Behrens, Zion Leonahenahe Basque, Haehyun Cho, Adam Oest, Ruoyu Wang, Tiffany Bao, Yan Shoshitaishvili, Gail-Joon Ahn, and Adam Doupé. 2022. I'm SPARTACUS, No, I'm SPARTACUS: Proactively Protecting Users from Phishing by Intentionally Triggering Cloaking Behavior. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Los Angeles CA USA, 3165–3179. https://doi.org/10.1145/3548606.3559334

A Included Conferences & Journals

Conferences and journals included in our literature survey.

Type	Conference or journal
Top 10 Computer Security	IEEE Symposium on Security and Privacy (S&P)
	IEEE Transactions on Information Forensics and Security
	ACM Symposium on Computer and Communications Security (CCS)
	USENIX Security Symposium
	Computers & Security
	Network and Distributed System Security Symposium (NDSS)
	IEEE Transactions on Dependable and Secure Computing
	International Conference on Theory and Applications of Cryptographic Techniques (EUROCRYPT)
	International Cryptology Conference (CRYPTO)
	Journal of Information Security and Applications
Co-located workshops	USENIX Workshop On Offensive Technologies (WOOT)
	USENIX Large-Scale Exploits and Emergent Threats (LEET)
Internet measurements	ACM Internet Measurement Conference (IMC)
	Passive and Active Measurements (PAM)
Cybercrime	APWG Symposium on Electronic Crime Research (eCrime)
	International Symposium on Research in Attacks, Intrusions, and Defenses (RAID)
	Workshop on the Economics of Information Security (WEIS)
	ACM The Web Conference (WWW)