Assessing the Effectiveness of Domain Blacklisting Against Malicious DNS Registrations

Thomas Vissers*, Peter Janssen†, Wouter Joosen*, Lieven Desmet*
*imec-DistriNet, KU Leuven
†EURid VZW

***Abstract*—Blacklists are widely-used in security research. However, there is little insight into how they operate, what their main focus is, and how effective they are. In this paper, we combine DNS traffic measurements with domain registration and blacklisting data. This allows us to assess and support to what extent researchers can extrapolate on existing blacklist sources. We focus on large-scale malicious campaigns that register thousands of domain names used in orchestrated attacks to evaluate this situation. We show that blacklist operators use both reactive, and to a lesser extent, proactive detection methods. Furthermore, by examining behavioral aspects of these malicious domains, we can pinpoint when blacklists fail to detect campaign domains.**

I. Introduction

DNS continues to serve as a major facilitator of internet-based crime. From phishing and spam to botnet communication and malware distribution: most cyber attacks require domain names to be operational. While some malicious actors compromise existing domain names, many register new ones to provision their attacks. The amount of domain names that are newly registered for malicious purposes is substantial [6], [18].

In our previous study, we extensively analyzed the ecosystem of malicious registrations within `.eu` [18]. We found that the vast majority of blacklisted registrations could be attributed to a small set of cybercriminal registrants. We found that these cybercriminals continously set up lage-scale campaigns, producing thousands of domain names used in cyber attacks.

An important finding of this study is that a substantial amount of campaign registrations[1], while clearly affiliated to cybercrime, never ends up on a blacklist. One possible explanation is that some campaign registrations are never actively used in attacks. Alternatively, blacklist operators might simply fail to detect some malicious behavior. At this time, there is no clear understanding of this discrepancy, in part because blacklist methods are somewhat opaque, as they typically combine multiple tactics to achieve detection. However, the security community heavily depends on blacklists and often treats them as oracles. For example, many detection and prevention systems are modelled using blacklists as their ground truth for maliciousness (e.g. [1], [4], [6]). Furthermore, the understanding of cybercrimal ecosystems relies on analyses using blacklists as a main indicator of malice (e.g. [7], [15], [18]). A lack of understanding and transparancy limits these initiatives.

In this paper, we set out to further understand how malicious campaigns operate and interact with blacklisting. We combine behavioral traffic data with registration and blacklisting information to analyze the different strategies of malicious campaigns and blacklist curators, and how they affect each other. More specifically, by looking at incoming DNS requests for malicious domains, we can infer their involvment in attack operations. This enables us to observe campaign specific attack patterns. Following these insights, we can further assess the effectiveness of domain blacklisting of campaign registrations.

The main findings of this paper are:

- We demonstrate that domains registered as part of campaign are deployed in a coordinated fashion. Furthermore, we discern the presence of campaign-specific behavioral patterns.
- We report on the usage of reactive and proactive blacklisting strategies to detect the attacks that these campaign exhibit.
- We provide insights into missed detections in relation to active and dormant registrations.
- We further develop the understanding of how campaigns approach the large-scale registration and deployment of their domains.

The remainder of this paper is structured as follows. In Section II, we introduce the data and subjects of this study. Next, we give a few examples of attack activity in malicious campaigns in Section III. In Section IV, we design a measure for domain activity in order to assess and understand blacklisting effectiveness. Afterwards, in Section V, we study the lifespan of campaigns in terms of registration, attack deployment and blacklisting. We discuss our analysis and related work in Section VI and VII. We state our concluding remarks in Section VIII.

II. Dataset and campaign identification

In this section, we first describe the data used in this paper. Next, we establish the starting point of our research by identifying the five most active campaigns present in our dataset.

[1] A *campaign* encompasses the entire set of domain registrations made by the same malicious registrant

A. Dataset

Registration data We analyze the data of 6 months of new incoming registrations within `.eu`, starting from January 1, 2018. Overall, this is encompasses 304K registrations. This data includes the registered name, the time of registration, along with the contact information given by the registrant. This lists the company name, email address, phone, as well as postal address information.

Blacklist data For each of these new registrations, we want to assess if and when they are placed on a blacklist. To that extent, we query a set of public blacklists twice per day. Each new domain is continuously monitored for at least three months once it has been registered. We consult the following widely-used blacklists: Spamhaus DBL [17], SURBL [16] and Google's Safe Browsing list [5]. Overall, we detect blacklisting events for 15K domains, or 5% of the total amount of registrations in the examined period.

DNS request data We process the passively-logged DNS requests of two `.eu` TLD name servers during the same six month period. One of the servers is located in the UK, the other one in Slovenia. Both name servers receive DNS requests for all `.eu` domains, although they each only see a part of the traffic (requests are distributed among 7 redundant name server of `.eu`). It should be noted that TLD name servers have a unique vantage point. They receive DNS requests for the entire set of domain names present in the TLD zone file. However, as they are not the final authoritative name server for the second-level domains, they normally only see the initial and cache-expired DNS requests from resolvers. A resolver does not query the TLD nameservers for follow-up requests for that domain.

Previous work concludes that the vast majority of malicious behavior and domain blacklisting, occurs within 30 days after registration [7], [18]. Following this insight, we limit the data processing to requests of domains that are less than 35 days old at the time the name servers received them.

We extract the name and record type (e.g. A or TXT) of each request. Furthermore, we make note of the origin country of the client that sent us the request (i.e. the DNS resolver or forwarder) using data from MaxMindDB [11].

B. Campaign identification

We use the insights of the previous ecosystem study [18] to identify campaigns in our current dataset. Specifically, we find the largest malicious campaigns in our dataset based on the distinct use of registrant contact details within our blacklisted set. As can be seen in Figure 1, the top five most active malicious registrants are responsible for 79% of all blacklisted registrations. These five registrants will serve as the starting point for our further campaign-centric analysis. The five campaigns are shown in Table I.

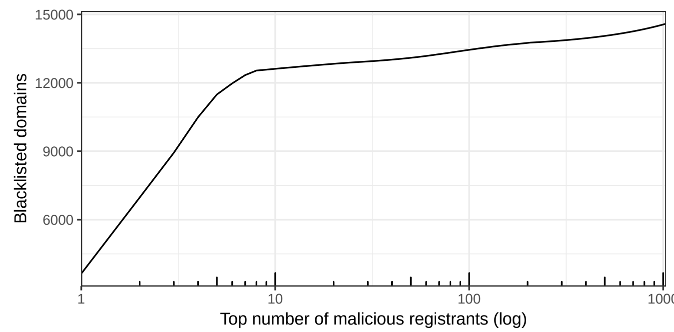

Fig. 1. The cumulative amount of blacklisted registrations that are made by the top malicious registrants.

Campaign	Registrations	Blacklisted	Non-blacklisted	Distinct registration days
A	3,661	3,634	27	22
B	4,351	3,337	1,014	4
C	2,045	1,962	83	24
D	2,086	1,558	528	105
E	1,730	995	735	1

TABLE I
THE NUMBER OF REGISTRATIONS IN THE FIVE LARGEST MALICIOUS CAMPAIGNS, AND THE NUMBER OF DISTINCT DAYS WE RECEIVED REGISTRATIONS FOR THEM.

III. OBSERVING ATTACKS THROUGH DNS REQUESTS

In this section, we demonstrate how we can observe campaign-orchestrated attacks by looking at the DNS requests for domains.

An incoming DNS request implies that some client on the web wants to request information for that domain. Many malicious operations trigger requests to attacker-controlled domains. For instance, sending out spam emails triggers the receiving entity to query SPF [9], DMARC [10] and DKIM [2] records to validate the sender's domain. Similarily, when an email cannot be delivered, the MX record of the sender's domain will be requested in order to respond with a bounce message. Other malicious activity, such as phishing websites and C&C servers, will trigger DNS requests from their victims as well. Given this characteristic, we can use incoming DNS requests as an indicator of registrant-induced activity.

Thus by looking at this activity indicator, we are able to map out coordinated attacks across multiple domains within the same campaign. For instance, Figure 2 shows the incoming A record requests over time of four blacklisted domains. All four domains were registered around the same time (Jan 6-8) However, according to our registrant-based campaign identification, the first three domains are part of campaign B, while the last one is associated with campaign A.

The different campaigns are clearly reflected in the measured activity. Domains B.1 and B.2 are undoubtedly operating in a coordinated fashion: they both exhibit a very similar burst of activity at the exact same time, 20 days after their registration. Domain B.3 exhibits similar activity in the first week after registration: several short burst of requests. We hypothesize that requests are artefacts of an attack preparation stage. However, B.3 does not exhibit the same timed burst as B.2 and B.3. A possible cause is the early blacklisting of B.3 (Jan 13), while B.1 and B.2 were only blacklisted at the time they exhibit the burst behavior (Jan 27).

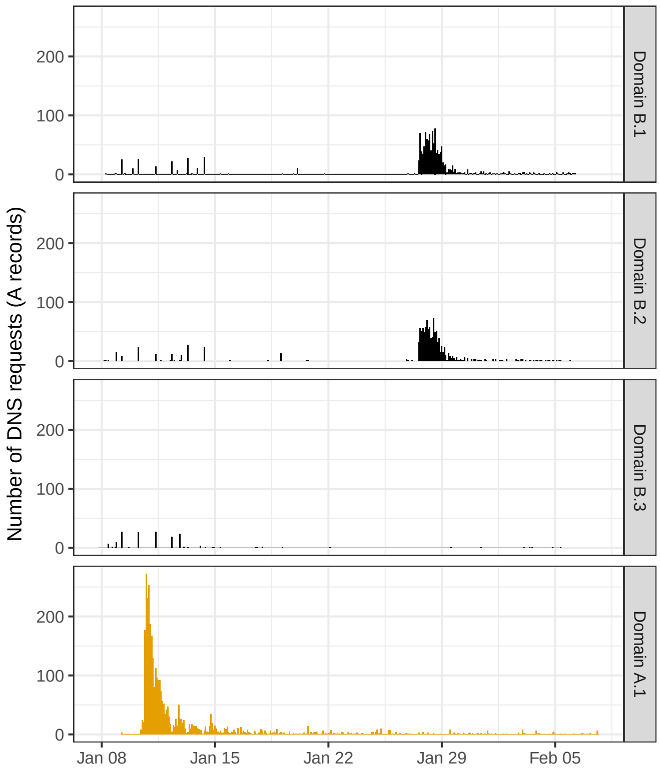

Fig. 2. The amount of DNS A record requests received in 2-hour windows after registration for three different domains in Campaign B and one domain of campaign A.

Domain A.1, a blacklisted domain registered around the same time, exhibits an entirely different activity pattern. The activity burst is much stronger and takes place soon after registration. This further illustrates how each campaign is managed by a single entity that controls specifc and distinct attacks across its domains.

IV. Quantifying effectiveness of blacklists

Previous work has found that (1) there is a substantial amount of domain names registered as part of cybercriminal campaigns that never ends up on a blacklist [18], and that (2) in some cases, blacklists flag domains before they exhibit malicious behavior [7]. In other words, researchers have reported that blacklists both flag and miss registrations that are linked to other malicious domains. In this section, we want to quantitatively assess the effectiveness of blacklisting by taking into account the domain's behavior. We consider four distinct cases in which we can place the malicious campaign registrations:

1) **Blocked.** The campaign domain was *active* and placed on a blacklist.
2) **Missed.** The campaign domain was *active*, but blacklists failed to detect it.
3) **Proactive.** The campaign domain was *not active*, but was proactively blacklisted. Presumably through other signs of maliciousness (e.g. linked to existing malicious campaign)
4) **Unused.** The campaign domain was *not active* and was not placed on a blacklist. Even though our data indicates that this registration was made by a malicious actor.

In order to place the campaign registrations in these categories, we have to determine which ones actively took part in an attack. To that extent, we design a behavioral measure based on DNS traffic that concentrates on representing burst activity and DNS requests that are not seen by all registrations.

A. *Designing a behavioral measure*

To enable meaningful comparisons between behavioral patterns, we make use of Dynamic Time Warping (DTW) [13], a similarity measure between time series that allows non-linear stretching and compressing to map two series together before calculating the distance. Intuitively, it allows us to find similarities between time series even if they are time-shifted. Thus, enabling the behavioral measure to align anomalous activity patterns, such as the bursts shown in Figure 2.

To compare the behavioral patterns of two domains, we pre-process the DNS request data of each domain. We shift the requests' timestamps to a time relative to the domain's registration time. For each domain, we establish differently weighted and standardized timeseries for each distinct DNS record type and request origin observed. The reasoning here is that, for instance, A record type requests are vastly more prevalent than TXT requests. However, the relative significance of one extra TXT request is much greater than an extra A record request.

After this pre-processing phase, the DTW distance between different domains can be used to assess behavioral similarity. For the purpose of determining an activity level for each domain, we compare the pre-processed time series with a dummy timeseries with no activity, i.e. zero DNS requests.

Using this measure for intra- and inter-campaign comparisons is left for future work.

1) Determining a threshold for dormant domains: To determine a threshold to differentiate between **dormant** and **active** domains, we take the 13,873 campaign registrations from our dataset and include another 13,873 randomly sampled benign registrations. Next, we calculate the activity level as described above. To find an appropriate threshold, we inspect the distribution of the activity level across the blacklisted domains in each campaign, as shown in Figure 3 To give a visual impression, the inactive domain B.3 that was shown earlier in Figure 2, falls into the first curve of campaign B. In comparison, the clearly active B.1 and B.2 domains lie in the second curve. This suggests that an accurate threshold falls in between B's two curves. We further manually verify several samples and confirm that, for instance, domains in campaign E are dormant. Interestingly, campaign D has domains across a large part of the activity level spectrum. In this case, manual inspection across the activity spectrum simply reveals gradual increasing activity with no clear threshold between B's two curves. To prevent drawing inaccurate conclusions from the threshold, we make conservative decisions by establishing a

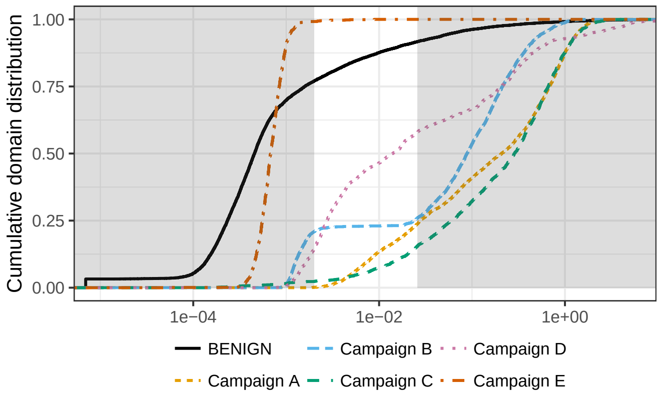

Fig. 3. The cumulative distribution of the activity level across all blacklisted campaign registration and the benign sample set. The left shaded area marks the dormant zone, the right marks the active zone.

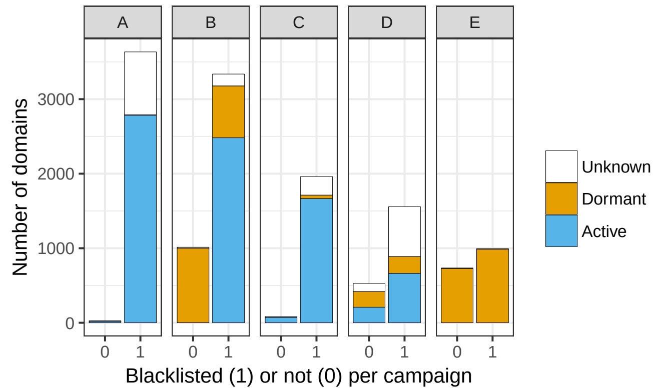

Fig. 4. The distribition of active and dormant domains amongst the blacklisted and missed campaign registrations.

	Active	Dormant
Blacklisted	**Blocked** 54.8%	**Proactive** 2.9%
Non-blacklisted	**Missed** 14.1%	**Unused** 14.0%

TABLE II
DISTRIBUTION OF CAMPAIGN REGISTRATIONS IN DIFFERENT CATEGORIES BASED ON THEIR OCCURENCE ON A BLACKLIST AND ACTIVITY LEVEL. EXCLUDES 15% OF REGISTRATIONS IN AN UNKNOWN ACTIVITY STATE.

broad margin from 0.0020 to 0.0250 (as shown in Figure 3). We consider any domain below this margin as dormant, and any domain above as active. The 15% of campaign domains present in the margin are excluded from the results.

B. Results

We show the resulting distribution of the domains amongst the four different categories in Table II. The majority of domains registered as part of campaign are blocked, i.e. they exhibited malicious activity and were blacklisted. Proactive blacklisting happens in the wild, but is found to still be rather rare (2.9% of campaign registrations). A large portion of campaign domains are missed by blacklists. While reactive blacklisting is a well-adopted practice, we still witness 14.1% of campaign domains flying under the radar even though they exhibited active behavior. Another 14.0% of unused campaign registrations could arguably be flagged on top of this by linking them to their malicious campaign proactively.

Figure 4 gives the breakdown of the different cases in each campaign. Campaign A and C have the most straightforward results. Nearly all of their registrations were active and picked up by a blacklist.

Interestingly, campaign E is fully dormant and thus, from our data's perspective, flagged in an entirely proactive fashion. However, proactive blacklisting requires historical knowledge of malicious domains. To further investigate this situation, we analyze the `.eu` registration data from 2017 and find 6,090 additional domains registered by the registrant of campaign E on 38 different days. As a matter of fact, the single batch of registrations made by campaign E in our current dataset, was the last one. Presumably, the malicious actor abandoned the tainted registrant credentials as soon as new registrations were aggresively and proactively blacklisted.

We note a similar scenario in the case of Campaign B. Although there are many active domains, there is also a substantial amount of dormant domains, both blacklisted and missed. Notably, 75% of those dormant domains were all registered on the last day this campaign was active. Once more suggesting that registrant credentials were abandoned once they became tainted.

The results of campaign D are not straightforward, as was to be expected from the activity level distribution. Unfortunately, at this point, we cannot draw clear conclusions for this campaign.

One important caveat for all these results is that blacklisting may influence the activity level of a domain name. For instance, once blacklisted, clients might be blocked from connecting to the domain. Contrarily, security intelligence services and researchers might start requesting information for that domain name. As such, the activity level is just an indicator for potential malicious behavior.

V. CAMPAIGN STRATEGIES AND LIFECYCLES

In this section, we explore the lifecycles of malicious campaigns and the domain names that are deployed as a part of it. In particular, we analyze different strategies that cybercriminals use when registering and deploying domain names for their attacks. Furthermore, we observe the different blacklisting strategies changing across the lifespan of a malicious campaign.

A. Data analysis

We analyze the cumulative amount of domains that have been registered and blacklisted for every campaign over time[2], as shown in Figure 5. Additionally, we keep track of the

[2] We only take into account the blacklisted domain names of the registrant here. The non-blacklisted campaign registrations are excluded.

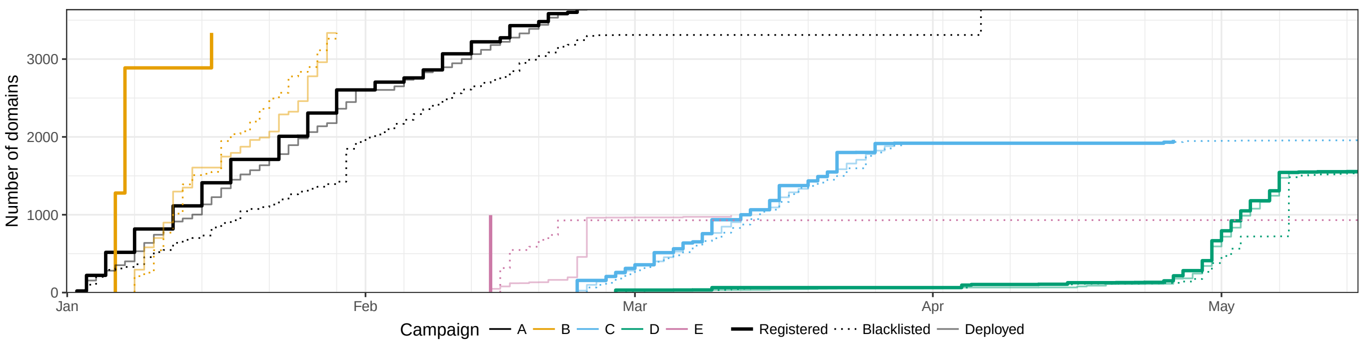

Fig. 5. For each campaign, the cumulative amount of domains that are registered (thick line), blacklisted (dotted line) and deployed (thin line) over time.

amount of domains that have been potentially deployed in an attack by using their *most active day* as a proxy. Specifically, for each domain, we note the day the TLD nameserver received the most DNS requests for it. The sequences of these events allows us to witness the different registration and blacklisting strategies.

B. Campaign registration strategies

The campaign registration data, as shown by the thick lines in Figure 5, confirms the existince of two distinct strategies. Campaigns B and E are typical examples of **bulk registration**. They are active on a limited number of days on which they register a very large amount of domain names in bulk. The other strategy, **continuous registration**, is seen in campaigns A, C and D. In these cases, the malicious actors continuously register smaller batches of domains.

C. Campaign deployment strategies

The continuous registration strategy (as exhibited by A, C and D) is clearly related to a specific deployment strategy. These campaigns deploy (thin lines) their domain names **in tandem** with their batches of registrations over time. For instance, in campaign A, we can clearly see how a new batch of registrations is made as soon as the domains in the previous batch have been actively deployed. This further validates the hit-and-run hypothesis, which suggests that new registrations are made as soon as previous domains are tainted.

Campaign B, one of the bulk-registering campaigns, does not adhere to this tandem situation. While the registration of domains happens in bulk, here, they are **gradually deployed** over time. This suggests that some attackers proactively stock up on domain names some time before the actual attacks is executed.

We will not discuss the deployment of campaign E, as it was shown in Section IV-B that this campaign was completely dormant.

D. Campaign blacklist timing

When comparing the time of deployment and time of blacklisting on Figure 5 (dotted lines), we are able to observe the **reactive mechanism**. In these campaigns, the domain is blacklisted after it was active (i.e. the cumulative blacklisted line runs behind the cumulative deployment line). This scenario is cleary illustrated by campaign C, where we note a very tight repetitive process of registering, subsequently deploying and thereafter becoming blacklisted. A similar situation is again observed in campaign A, however here blacklisting generally happens much later than the deployment step. This suggests that this campaign is more effective at avoiding detection by blacklists and potentially was able to sustain his attack for a longer period of time for each deployed domain.

Interestingly, we find that this granular reactive mechanism is not the sole blacklisting method. There are cases where exceptionally large sets of campaign domains are blacklisted at once. For instance, on January 30, 422 domains of campaign A were suddenly blacklisted. Similarly, on May 8, campaign D had 759 of its registrations blacklisted. Both of these larger takedowns suggests that blacklists operators are not only flagging reactively on domain-per-domain basis. They are **flagging batches** of related domains.

As mentioned in Section IV-B, campaign B was was likely discontinued due to being affected by **proactive blacklisting** at the end of its lifespan. Figure 5 demonstrates this process clearly. Starting from the last day registrations are made for campaign B (Jan 16), domains are getting blacklisted before they are they become active. Moreover, we determined that in these cases, those domain names are simply dormant and actualy never really deployed. The similar situation for Campaign E is also reflected in this graph.

VI. Discussion

A. Limitations

As mentioned earlier, there are certain limitations of using TLD name server DNS request to analyze behavioral acitivity. Most prominently, the impact of caching. Furthermore, blacklisting itself might impact the amount of DNS requests arriving at the TLD name server, potentially influencing our measured acitivty level.

We did not have access to a real-time feed from blacklist operators. This prevents us from accurately determining the exact time a domain name was detected.

B. Ethical considerations

The analysis is performed as part of the registry's security and trust program [3], to detect and prevent abusive behavior within its TLD. Ethical considerations have been made to conduct the analysis in a privacy-preserving manner, in line with the terms and conditions. The main purpose of this study is to understand the malicious domain name ecosystem, and study the effectiveness of current blacklisting approaches.

As part of the analysis, registrant information of a domain name, as well as DNS queries, to this domain name have been studied. Registrant information, as requested by the registry as part of the domain registration process, has only been used to identify the biggest abusers of the TLD. Based on data from external domain blacklists, the analysis has been scoped to the biggest abusers of the research corpus; as well as a randomly selected set of non-blacklisted domain names. For the query analysis, traffic arriving at the TLD name servers (managed by the TLD registry) has been passively monitored. Hereby, only the query type, the requested name and the originating country of the DNS query are used for the analysis. The resolver's IP address nor the response of the query are part of the analysis.

This research required us to periodically request information from blacklists [5], [16], [17]. This entailed public data that we consulted in compliance with the respective terms of use.

Only aggregated and pseudomized results have been disclosed in the context of this research.

VII. Related work

Kidmose et al. [8] recently reported on the difficulties of assessing the value of using blacklists. They opine that researchers introduce errors when using highly imperfect blacklists as their main ground truth source. This vision strengthens the motivation to shine a brighter light on blacklisting effectiveness.

Using blacklists as the starting point for studying cybercrime and designing detection systems is very common. Several noteworthy examples have been given in the introduction ([1], [6], [7], [15], [18]). Zhauniarovich et al. gives an overview of how DNS data has been used to detect malicious domains [19]. They specifically report on domain blacklist as common source of ground truth.

Other related academic work focuses on improving proactive detection through campaign-like associations [4], or by early discovery of domains generated by DGAs (e.g. [12], [14]).

VIII. Conclusion

In this study, we combined DNS traffic measurements with domain registration and blacklisting information to strengthen our understanding of blacklist effectiveness. We bring forward important insights in the ambiguity and incompleteness of blacklists for the security community. Researchers namely rely on blacklists as a starting point of studies, and as ground truth for modelling and evaluating detection systems.

We also confirm, in line with earlier findings, that a substantial amount of campaign registrations are still being missed by blacklist operators. Our data suggests there is potential in more effective proactive approaches.

References

[1] L. Bilge, S. Sen, D. Balzarotti, E. Kirda, and C. Kruegel, "Exposure: a passive dns analysis service to detect and report malicious domains," *ACM Transactions on Information and System Security (TISSEC)*, vol. 16, no. 4, p. 14, 2014.

[2] D. Crocker, T. Hansen, and M. Kucherawy, "Domainkeys identified mail (dkim) signatures," Tech. Rep., 2011.

[3] EURid. (2019) Trust .eu. [Online]. Available: https://trust.eurid.eu

[4] M. Felegyhazi, C. Kreibich, and V. Paxson, "On the potential of proactive domain blacklisting," in *Proceedings of the 3rd USENIX Conference on Large-scale Exploits and Emergent Threats: Botnets, Spyware, Worms, and More*, 2010, pp. 6–6.

[5] Google. (2016) Google Safe Browsing. [Online]. Available: https://developers.google.com/safe-browsing/

[6] S. Hao, A. Kantchelian, B. Miller, V. Paxson, and N. Feamster, "Predator: Proactive recognition and elimination of domain abuse at time-of-registration."

[7] S. Hao, M. Thomas, V. Paxson, N. Feamster, C. Kreibich, C. Grier, and S. Hollenbeck, "Understanding the domain registration behavior of spammers," in *Proceedings of the 2013 Conference on Internet Measurement Conference*, 2013, pp. 63–76.

[8] E. Kidmose, K. Gausel, S. Brandbyge, and J. M. Pedersen, "Assessing usefulness of blacklists without the ground truth," in *International Conference on Image Processing and Communications*. Springer, 2018, pp. 216–223.

[9] S. Kitterman, "Sender policy framework (spf) for authorizing use of domains in email, version 1," Tech. Rep., 2014.

[10] M. Kucherawy and E. Zwicky, "Domain-based message authentication, reporting, and conformance (dmarc)," Tech. Rep., 2015.

[11] MaxMind, Inc. (2016) GeoLite2 Free Downloadable Databases. [Online]. Available: https://dev.maxmind.com/geoip/geoip2/geolite2/

[12] D. Plohmann, K. Yakdan, M. Klatt, J. Bader, and E. Gerhards-Padilla, "A comprehensive measurement study of domain generating malware," in *25th USENIX Security Symposium*, 2016, pp. 263–278.

[13] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE transactions on acoustics, speech, and signal processing*, vol. 26, no. 1, pp. 43–49, 1978.

[14] J. Spooren, D. Preuveneers, L. Desmet, P. Janssen, and W. Joosen, "Detection of algorithmically generated domain names used by botnets: a dual arms race." Association for Computing Machinery, 2019, pp. 1902–1910. [Online]. Available: $$Uhttps://lirias.kuleuven.be/retrieve/529700$$Ddga.pdf[freelyavailable]

[15] B. Srinivasan, A. Kountouras, N. Miramirkhani, M. Alam, N. Nikiforakis, M. Antonakakis, and M. Ahamad, "Exposing search and advertisement abuse tactics and infrastructure of technical support scammers," in *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2018, pp. 319–328.

[16] SURBL. (2016) SURBL - URI Reputation Data. [Online]. Available: http://www.surbl.org

[17] The Spamhaus Project Ltd. (2016) The Domain Block List. [Online]. Available: https://www.spamhaus.org/dbl/

[18] T. Vissers, J. Spooren, P. Agten, D. Jumpertz, P. Janssen, M. Van Wesemael, F. Piessens, W. Joosen, and L. Desmet, "Exploring the ecosystem of malicious domain registrations in the. eu tld," in *International Symposium on Research in Attacks, Intrusions, and Defenses*. Springer, 2017, pp. 472–493.

[19] Y. Zhauniarovich, I. Khalil, T. Yu, and M. Dacier, "A survey on malicious domains detection through dns data analysis," *arXiv preprint arXiv:1805.08426*, 2018.