Using NetFlow to measure the impact of deploying DNS-based blacklists

Andersen, Martin Fejrskov; Pedersen, Jens Myrup; Vasilomanolakis, Emmanouil

Published in:
Security and Privacy in Communication Networks

DOI (link to publication from Publisher):
10.1007/978-3-030-90019-9_24

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

Using NetFlow to measure the impact of deploying DNS-based blacklists *

Martin Fejrskov[1], Jens Myrup Pedersen[2], and Emmanouil Vasilomanolakis[2]

[1] Telenor A/S, Aalborg, Denmark
mfea@telenor.dk
[2] Cyber Security Group, Aalborg University, Copenhagen, Denmark
{jens,emv}@es.aau.dk

Abstract. To prevent user exposure to a wide range of cyber security threats, organizations and companies often resort to deploying blacklists in DNS resolvers or DNS firewalls. The impact of such a deployment is often measured by comparing the coverage of individual blacklists, by counting the number of blocked DNS requests, or by counting the number of flows redirected to a benign web page that contains a warning to the user. This paper suggests an alternative to this by using NetFlow data to measure the effect of a DNS-based blacklist deployment. Our findings suggest that only 38-40% of blacklisted flows are web traffic. Furthermore, the paper analyzes the flows blacklisted by IP address, and it is shown that the majority of these are potentially benign, such as flows towards a web server hosting both benign and malicious sites. Finally, the flows blacklisted by domain name are categorized as either spam or malware, and it is shown that less than 6% are considered malicious.

Keywords: Blacklist · DNS · NetFlow · Ipfix · ISP · RBL · Threat Intelligence.

1 Introduction

Threat Intelligence (TI) in the form of reputation-based blacklists of IP addresses and domain names have been made available by non-profit and commercial organisations for decades [20], and has later been the subject of academic research as well [9]. Improving the accuracy and completeness of the blacklists by the careful selection of entries to maximize the amount of true positives and minimize the amount of false negatives remains a continuous struggle. These metrics describe the blacklist itself, however they do not describe the actual impact of deploying a blacklist in practice. If there is not impact, the time and money

* Funded by Telenor A/S and Innovation Fund Denmark, 2021. This version of the contribution has been accepted for publication, after peer review but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: https://doi.org/10.1007/978-3-030-90019-9_24. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use [18].

spent by the user deploying the blacklist can be considered wasted. Therefore, we argue that the impact is an important metric from a practical perspective.

How to describe and measure the impact will naturally depend on the specific use case in which the blacklist is applied[3]. The most prevalent use cases for blacklists fall in two categories, offering protection to either the originating end of a connection (in antivirus software, in a web browser plugin, in a company firewall, in an Internet Service Providers (ISPs) DNS server, etc.) or the terminating end of a connection (a mail server, at a firewall protecting a web site, etc.). This paper focuses on the impact of deploying blacklists in DNS resolvers at ISPs. Deploying blacklists at ISPs is attractive as it can increase the security posture of all devices that default to use the ISP's DNS resolvers.

Informal conversations with blacklist vendors suggest that a common method for assessing the impact is to let the DNS resolver count the number of performed DNS queries that match an entry on a blacklist. Some ISPs and DNS security vendors even refer to this number directly as the number of blocked threats [3,22]. This is similar to counting the number of emails flagged as phishing by an email server, or counting the number of requests towards a web server originating from an IP address known to be malicious. However a DNS request in itself is only a threat indicator. In order for a user to be at risk, an IP connection towards the malicious host is a minimum precondition, and we therefore consider an IP connection as a stronger threat indicator than a DNS resolution. In this paper, we propose a method based on NetFlow/IPFIX measurements to evaluate the impact of deploying blacklists at an ISP DNS resolver.

Assessing the network-level impact of applying a blacklist at a DNS server will, however, not in itself tell anything about the user-level impact perceived by the user. For instance, blocking a user's connection attempt towards a shared web hosting environment that incidentally also hosts a known spam sender, is likely to be perceived as a nuisance rather than as protection against a threat. On the other hand, connecting to a web server known to solely host malicious payloads represents a high risk to the user. To supplement the measured network-level impact, it is necessary both to identify the cause for the entry to be blacklisted in order to assess the risk level of connecting to the blacklisted entity, and to assess the risk that a connection is in fact made towards the malicious entity.

The contributions of this paper are twofold:

- We show how existing methods for measuring the impact of deploying domain and IP address blacklists in DNS resolvers can be improved by including NetFlow measurements.
- Using the NetFlow method, we quantify the number of malicious and non-malicious flows, and we quantify the number of flows blacklisted by IP address that may be benign.

The paper is organised in 7 sections: Section 2 gives an overview of related work. Section 3 describes the concept of blacklisted flows and the method for merging DNS, NetFlow and blacklist data to identify blacklisted flows. Section

[3] The elaborated definition of impact used in this paper is presented in Section 4.3.

4 describes the 3 data sources used in the paper and the application of the previously described merging method. Section 5 categorizes the blacklisted flows by the type of maliciousness and Section 6 identifies IP addresses that may contain multiple (and possibly both malicious and benign) endpoints. Section 7 combines the results from the previous sections to describe the network-level and user-level impact. Lastly, Section 8 summarizes and concludes the paper.

2 Related work

As outlined in the introduction, the contribution of this paper is to show that existing measurement methods that *measure the impact* of implementing domain and IP address *blacklisting in DNS resolvers* can be improved by including *NetFlow data* based measurements in addition to *DNS data* based measurements. We use the proposed measurement method together with information about the *type of maliciousness* and knowledge about the *type of endpoint* to identify if the endpoint may host both benign and malicious sites simultaneously. The columns of table 1 represent each of the aspects highlighted in the above paragraph, and this section elaborates how related works cover some, but not all, of the aspects.

Many papers such as [24] focus on the creation, quality, accuracy or comparison of blacklists. Bouwman et al. focus on the differences between paid and free lists, and investigate the reasons (price, coverage, false positive rate, etc.) provided by operators/enterprises for choosing specific lists [2]. These topics are considered complementary to this paper, and such efforts will therefore not be the topic of this section. Similarly, papers such as [13, 16] focusing on using blacklists for spam filtering in mail servers are also considered complementary.

		Focus area					
Author	**Year**	Resolver Impact	DNS BL data	NetFlow data	Endpoint		Maliciousness
Sheng et al. [15]	2009		✓	✓			✓
Bermudez et al. [1]	2012			✓	✓		
Connery [4]	2012	✓	✓	✓			
Zhang et al. [23]	2013	✓			✓	✓	✓
Kührer et al. [10]	2014		✓	✓		✓	✓
Foremski et al. [7]	2014			✓	✓		
Satoh et al. [14]	2019	✓					
Spacek et al. [17]	2019	✓	✓	✓			
Wilde et al. [21]	2019	✓	✓	✓			
Li et al. [11]	2019				✓	✓	✓
Telenor Norway [19]	2020	✓	✓	✓			
Griffioen et al. [8]	2020	✓		✓	✓	✓	✓

Table 1. *Related work and their focus areas*

2.1 Network-level impact of blacklisting

Although not focusing on malware and blacklists, the authors of [1] observe that around 50% of DNS responses have an associated flow. This suggests that a flow cannot be assumed to be associated with all *blacklisted* DNS responses either. This forms a motivation for focusing on flows rather than DNS responses.

Zhang et al. measure the impact of applying several (IP address) blacklists on NetFlow records obtained from the routers of a large regional ISP [23]. The paper differentiates between different types of maliciousness and endpoints, and concludes that up to 17% of the traffic measured by volume could be considered tainted. Although this work blacklists NetFlow entries rather than DNS entries, we consider it to be one of the works that are most closely related to our paper.

Sheng et al. evaluate blacklists in browser plugins to protect against phishing websites [15]. This approach represents several advantages to DNS-based filtering, as lists of URLs rather than lists of domain names or IP addresses can be used. The approach is, however, by nature very application and browser specific, thus representing a disadvantage in relation to a DNS-based approach.

Li et al. use telescopes of scanning activities to determine list coverage, thus including some flow level data [11]. Furthermore, the paper uses IP ranges of known CDNs as a source to determine list accuracy. However, the focus is still on assessing the quality of the lists, rather than on the impact of applying them.

Spacek et al. describe many practical considerations in deploying DNS based blacklisting, and elaborates on some of the consequences to the user [17]. These consequences focus on feedback about the blocked site, difficulties in relation to the use of TLS, etc., and does not quantify the impact of the blacklist itself.

Deploying blacklists at an ISP or company DNS server is becoming a common security measure. Public statements such as [19] and [4] with limited descriptions of the impact of such measures exist. Both of these statements measure the impact in terms of visits to a website, to which a user is redirected instead of being blocked. Similarly, DNS firewall/resolver vendors, TI providers, etc. provide use case descriptions focusing on DNS-level measurements only. Furthermore, Wilde et al. examine the blocking behaviour of several publicly available resolvers and conclude that none of them block for security purposes [21]. They also use lists of URLs to quantify to which extent an RPZ enabled DNS resolver would block the list entries. However, no real world traffic is used in the quantification.

Academic papers describe the impact of blacklisting at the router and browser level, and to a certain extent at the DNS level, as outlined above. However, we are not aware of any work that quantifies the impact at the NetFlow level.

2.2 User-level impact of blacklisting

Kuhrer et al. categorize both commercial and public blacklists entries to identify if an endpoint is a sinkhole or a parked domain [10]. The purpose of performing the categorization therefore relates more to the validity of a blacklist entry than to the impact experienced by the user. Furthermore, the paper evaluates the ability of blacklists deployed at a DNS server to detect known botnets.

Using DNS and flow information to determine the used application is the topic of [7]. The application of named flows (flows to which a DNS response can be associated) such as HTTP, Roblox and Skype is identified. This classification, however, focuses solely on the application rather than the type of endpoint.

Determining the type of maliciousness is the main focus of [14]. The authors use Word2Vec to group 388 malicious queries into three clusters, each comprising queries with a common cause. The study focus solely on DNS TXT records, which does not extend well to the majority of queries that do not have TXT records.

Some blacklist vendors and tools such as [12] provide the cause for an entry to be listed. This is in many cases directly related to the type of maliciousness.

Griffioen et al. present several aspects related to our paper [8]. Their main emphasis is to compare open source blacklists, including the impact metric. NetFlow information from a Tier 1 provider is used to assess the timeliness of entries on the lists, but is not used to assess the impact of deploying the lists, which is the main topic of our paper. Instead, information from authoritative DNS servers is used to evaluate the impact of deploying the lists, by analyzing how many domain names were pointing to a particular IP address on the day it was marked as malicious. We will extend this by including other aspects, beyond a high domain name to IP address ratio, and by analysing the domain names and IP addresses to identify different scenarios like shared web hosting.

Both [23] and [8] consider blacklisting on the IP level, for example in firewalls. In our paper, the focus is on invoking the blacklisting in a DNS server, thus considering both domain name and IP address based blacklists. Despite this conceptual difference, we consider these the most closely related to our paper.

2.3 Summary

Although related work exists, the idea of using NetFlow measurements for evaluating a DNS-based blacklist deployment seems to be unexplored, and this will therefore be the topic of Section 3-4. Categorizing existing blacklist entries by type of maliciousness does not seem to be receiving a lot of academic attention, maybe because the categorization can be available as a supplement to the blacklists. Using knowledge about the type of maliciousness and endpoint to provide a risk based view of the blacklisted flows will be the topic of Section 5-6.

3 Method for identifying blacklisted flows

The concept of blacklisted flows is central to the flow based measurement method proposed in this paper. The method to identify blacklisted flows requires three data sources and is comprised of several steps, as illustrated in Figure 1. The three data sources are NetFlow data, DNS data and blacklists containing domain names and IP addresses. The first steps, relating to the practical collection and pre-processing of the three individual data sources are illustrated in blue in Figure 1 and elaborated in Section 4. Combining the three data sources involves two additional processing steps, elaborated in the following subsections. First,

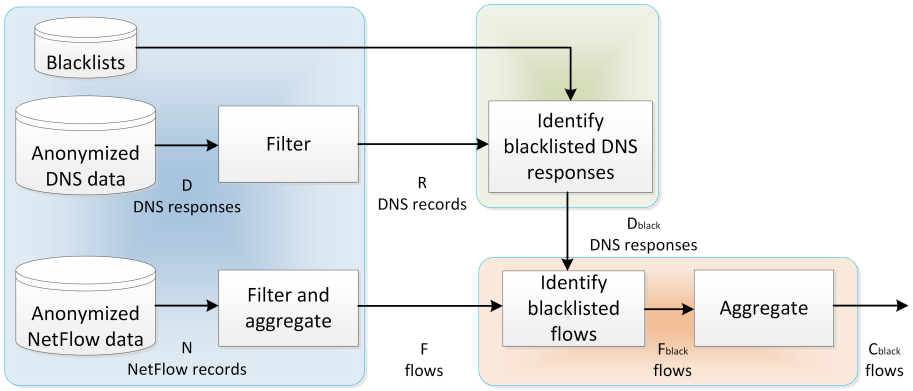

Fig. 1. *The overall dataflow to identify blacklisted flows.*

all blacklisted DNS records are identified (green in Figure 1). Then, all flows relating to the blacklisted DNS records are identified (red in Figure 1).

3.1 Blacklisted DNS data

All DNS records associated with a specific DNS response are considered blacklisted if *any* of these conditions are satisfied for any of the records:

- The Qname or Rname of the DNS record matches a blacklisted domain name
- The Rdata of the DNS record matches a blacklisted IP address or a blacklisted IP prefix

The result of this is that D_{black} blacklisted DNS responses are identified.

3.2 Blacklisted flows

A flow is considered blacklisted by a specific, blacklisted DNS record if *all* of the following conditions are satisfied:

- The DNS record has $rtype = A$
- The DNS record and flow timestamps are less than 30 minutes apart (as elaborated below), $t_{DNS} \leq t_{NetFlow} < t_{DNS} + 30m$
- The blacklisted DNS record is the temporally closest DNS record where the two conditions below are satisfied
- The blacklisted DNS record client IP matches the flow source IP
- The blacklisted DNS record rdata matches the flow destination IP

This yields a number of blacklisted flows, F_{black}.

Both the use of temporal correlation and anonymized IP addresses can cause a number of false positives and false negatives that are not immediately quantifiable as no ground truth exists for verification. The limit of 30 minutes is based

on an analysis of the time difference between the DNS record and the flow. This analysis suggests that the number of matched DNS records and flows converge towards 0 as a function of the time difference between the records, with few matches with a time difference of more than 15 minutes.

In case a flow matches two different DNS records where the only difference is the TTL, the DNS record with the highest TTL is considered a match.

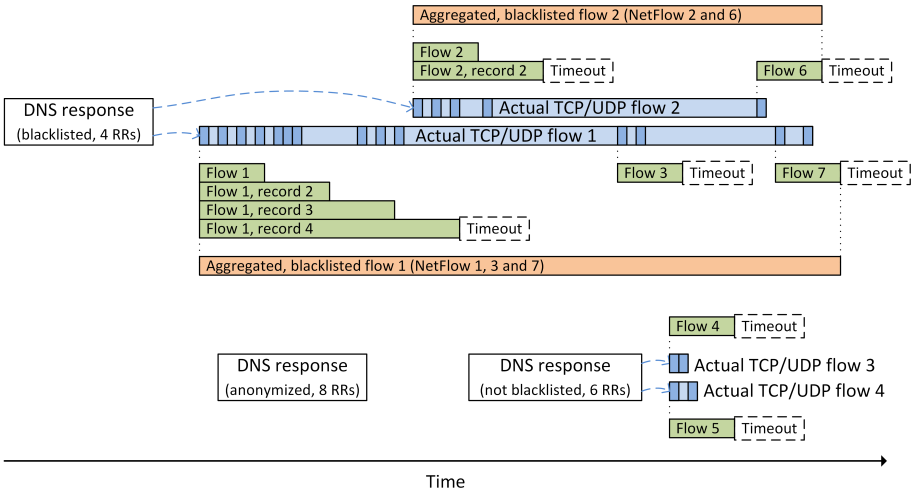

Fig. 2. *Flow aggregation illustration. In this example, D=3, R=10, D_{black}=1, N=11, F=7, F_{black}=5, C_{black}=2 and $C_{black,DNS}$=1*

The merging of NetFlow records into flows is described in Section 4.1. However, a NetFlow emitter may view a single, actual flow as two or more flows due to the use of aggressive timeouts for detection of flow end, especially for UDP traffic. Often this is referred to as flow splitting in related works. The effect is illustrated in Figure 2, where light blue represents the lifespan of actual flows and dark blue represents packets transmitted in the flow. Green represents the lifespan of flows as perceived and reported by the NetFlow emitter in successive flow records. Due to timeouts, the NetFlow emitter perceives the two actual flows as 5 different flows. Therefore, a further aggregation of flows is desirable.

We choose to aggregate all blacklisted flows that are blacklisted by the same DNS record (considered unique by the qname, timestamp and clientip) and that have the same 5-tuple into a single flow, producing C_{black} flows. This aggregated entity is named an aggregated flow to distinguish it from the flow defined by the NetFlow emitter. The aggregated flows are represented in red in Figure 2, where two aggregated, blacklisted flows (red) related to 5 different NetFlows (green), related to 2 actual flows (blue), and related to the same (blacklisted) DNS response (white) are depicted. The aggregated flow record has a cumulative

bytes/packet count and the flow start timestamp that is the earliest timestamp found in the related flows.

4 Data sources and processing

This section will provide details on the selection and pre-processing of the three data sources illustrated in blue in Figure 1 using data from Telenor Denmark's network (Sections 4.1 to 4.3). Furthermore, the section will describe the results of performing the steps described in Section 3 on the data (Sections 4.4 to 4.6).

Metric	**Symbol**	**Week 1**	**Week 2**
Total DNS responses	D	$2,15 \cdot 10^{10}$	$2,25 \cdot 10^{10}$
Total relevant DNS records	R	$1,85 \cdot 10^{10}$	$1,88 \cdot 10^{10}$
Blacklisted DNS responses	D_{black}	$6,81 \cdot 10^{6}$	$4,56 \cdot 10^{6}$
Total NetFlow records	N	$4,63 \cdot 10^{9}$	$4,60 \cdot 10^{9}$
Total relevant flows	F	$3,92 \cdot 10^{8}$	$3,94 \cdot 10^{8}$
Blacklisted flows	F_{black}	185460	191923
Blacklisted, aggregated flows	C_{black}	90796	86854
Unique DNS responses in C_{black}	$C_{black,DNS}$	78312	70134
Blacklisted DNS response ratio	$\frac{D_{black}}{D}$	0,000317	0,000203
Entries in C_{black} matched by IP	C_{ip}	68045	62683
Entries in C_{black} matched by domain	C_{dom}	22842	24486

Table 2. *DNS and NetFlow data metrics*

The three data sources are all collected during two separate weeks for the 1,5M mobile and 100k broadband subscriptions of Telenor Denmark. Notice that multiple users can use the same subscription, such as a household where all members are the users of a single broadband subscription. The data set for week 1 represent 7 full days from 2020-10-29 to 2020-11-04, and the data set for week 2 represent 7 full days of from 2020-11-26 to 2020-12-02. Table 2 lists the key properties for data in these time periods and the following sections will elaborate on these numbers. The following sections will for readability refer to the data from week 1, unless explicitly stated otherwise.

4.1 NetFlow data

NetFlow data is collected at Telenor Denmark's Border Gateway Protocol (BGP) Autonomous System (AS) border routers, representing all Internet traffic entering and exiting Telenor's network, as depicted in Figure 3. As indicated in the figure, two primary types of internal traffic not crossing the border routers exist:

– User-to-user traffic: The amount of user-to-user traffic is considered negligible compared to the amount of traffic crossing the border router and is therefore similarly considered negligible for the purpose of this paper.

– User-to-CDN traffic: A number of Content Delivery Network (CDN) nodes are deployed internally, and these serve a significant volume of traffic. However the types of data hosted on these nodes (Netflix/Youtube videos and similar static content etc.) are considered irrelevant to this paper from a user threat and blacklist perspective.

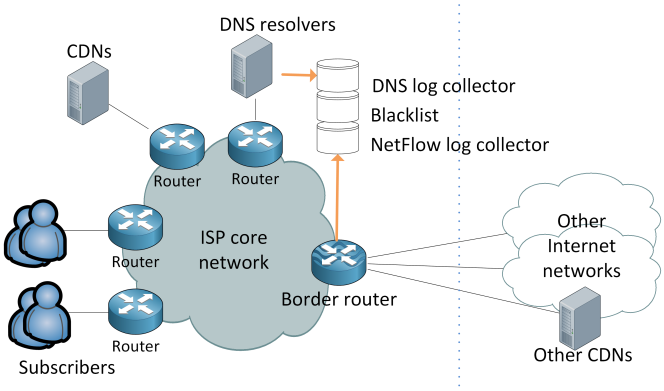

Fig. 3. *A conceptual view of the Telenor network indicating the sources of DNS and NetFlow data.*

A (unidirectional) NetFlow record is created by the border routers at least every 60 seconds for each active 5-tuple flow in each flow direction. A sample rate of Q=512 is used, therefore NetFlow records represent data from $\frac{1}{Q}$ packets. The collected data contains $N = 4,63 \cdot 10^9$ NetFlow records.

For the purpose of this paper, only connections initiated by users as a result of a DNS lookup are relevant. Therefore, only NetFlow records with an internal source address are considered, and for TCP connections only flows in which a SYN packet is seen are considered, as this will make sure that the flow start time actually represents the beginning of the flow. Multiple NetFlow records belonging to the same flow (defined by similar start-time and 5-tuple) are aggregated. As a result of this data reduction, $F = 3,92 \cdot 10^8$ flows are available for comparison with blacklisted DNS records. No application layer proxies are deployed.

NetFlow data is anonymized for legal reasons by truncating the internal (user) IP address to a /24 prefix for non-NAT'ed users (or truncating the port for NAT'ed users), truncating the external IP address to a /16 prefix, reverse truncating the timestamp, as well as a number of other measures less relevant to this paper. The anonymization policy applied follows the guidelines of [6]. Table 3 contains a number of example NetFlow records.

srcip	srcport	dstip	dstport	proto	packets	bytes
129.142.227.0	56065	2.17.0.0	443	TCP	512	32768
83.73.228.0	49906	193.28.147.0	443	TCP	512	32768
85.80.228.0	45820	8.8.0.0	53	TCP	512	30720

Table 3. *Example NetFlow records. Timestamps are omitted for brevity.*

4.2 DNS data

DNS data is collected at Telenor Denmark's DNS resolvers, as depicted in Figure 3. As the queried domain name is also a part of the DNS response packet, and as this study only focuses on syntactically valid DNS requests for which a response is always issued, only the response packets are collected (including for example NXDOMAIN responses). The resolvers are only accessible from Telenor Denmarks network, and are the default choice for all users. The collected data contains $D = 215 \cdot 10^8$ DNS responses. As a response can contain many Resource Records (RRs), the data is stored such that one record represents a unique RR augmented with the information common to all RRs in the same response.

There are no mechanisms preventing the use of 3rd party DNS resolvers residing outside the Telenor network, and therefore it is relevant to assess the prevalence of that type of traffic. NetFlow data contains $N_{DNS} = 5,92 \cdot 10^6$ records for traffic from users towards port 53 (DNS) and 853 (DNS-over-TLS) ($5,78 \cdot 10^6$ and $1,3 \cdot 10^5$ records respectively). It is not legally possible to inspect this traffic further to quantify how many and which queries this traffic represents. Assuming that one NetFlow record represents one DNS query (yielding the worst-case flow sample likelihood of 1:Q), the 3rd party DNS traffic represents only $\frac{QN_{DNS}}{(QN_{DNS}+D)} = 10,8\%$ of all queries. The traffic towards the Telenor DNS resolvers is therefore considered sufficiently representative of the total DNS traffic, and given the lack of legal basis for inspecting the 3rd party DNS resolver traffic, the 3rd party DNS traffic is disregarded for the purpose of this paper. Some anonymity services like TOR use private top level domains like '.onion'. These top level domains are not registered in the public DNS hierarchy. Therefore, such services are not considered relevant to this paper.

Only 0,1% of the DNS records, R, have an *rdata* field referring to a non-CDN IP address within the Telenor Denmark network. This supports the statement made in the NetFlow section that internal network traffic (both user-to-user and user-to-CDN) can be considered negligible to this paper.

DNS data is anonymized for legal reasons by truncating the client (user) IP address to a /24 prefix for non-NAT'ed users (or truncating the port for NAT'ed users), reverse truncating the timestamp, removing the domain name for the 15 most popular domains, and a number of other measures less relevant to this paper. The anonymization policy applied follows the guidelines of [6]. Discounting the anonymized records, $R = 185 \cdot 10^8$ records are therefore available for comparison with blacklists. Table 4 contains a number of example records.

clientip	qname	rtype	rname	rdata	ttl
85.83.74.0	a.config.skype.com.	A	l-0014.l-msedge.net.	13.107.42.23	100-299
85.83.65.0	log.tiktokv.com.	A	a2047.r.akamai.net.	77.214.51.34	1-99
85.83.65.0	log.tiktokv.com.	A	a2047.r.akamai.net.	77.214.51.27	1-99

Table 4. *Example DNS records. The timestamp is omitted for brevity.*

4.3 Blacklists

Blacklists that are available for a fee generally outperform free lists [10]. Therefore, blacklists provided by two well-known, commercial DNS blacklist vendors are used for this paper. After a review of the paper, the vendors opted to stay anonymous. The vendors will therefore be referenced as A and B, and the individual lists provided by each vendor as A_1, A_2, etc. The lists contain both IP addresses, IP prefixes and domains. Some of the lists are updated every minute, and the most realistic result would therefore be produced by doing a real-time correlation of DNS data and blacklists. However, as the DNS data is collected independently of the blacklists for operational and privacy reasons, this has not been possible in practice. Instead, the lists are collected at 23:00 CEST each day and the aggregated list is used for comparison for the whole period. In week 1, the aggregated lists contain 11878657 unique IP addresses, 3389 unique prefixes and 989490 unique domains. In week 2, the aggregated lists contain 16286208 unique IP addresses, 3320 unique prefixes and 1002913 unique domains.

For this paper, the impact of a blacklist describes the effect derived from a specific blacklist deployment. The impact of a blacklist with perfect accuracy and perfect completeness will be zero if a user never visits a malicious website. Conversely, if the completeness of a list is low, but deploying the list in practice would block the majority of traffic anyways, the impact will be high.

4.4 Blacklisted DNS data

The result of the operation described in Section 3.1 is that $D_{black} = 6,81 \cdot 10^6$ blacklisted DNS responses are identified. This represents $\frac{D_{black}}{D}$=0,000317 of the total number of DNS responses. The impact of applying DNS based blacklisting is often measured by the magnitude of this number, with the interpretation that user were protected by $6,81 \cdot 10^6$ threats.

4.5 Blacklisted flows

The result of the operation described in Section 3.2 is that F_{black}=185460 blacklisted flows are identified. After performing aggregation, a total of C_{black}=90796 blacklisted, aggregated flows are identified. Table 5 contains a number of examples of blacklisted, aggregated flows. C_{black} represents the number of flows found in the sampled NetFlows that would have been blocked in the sample week, if DNS based blacklists had been activated for all users.

qname	list	srcip	dstip	dstport	proto	timediff
www-pf-dk.filesusr.com.	A_2	94.145.224.0	34.102.0.0	443	TCP	0
collection.decibelinsight.net.	A_2	94.145.230.0	35.180.0.0	1789	UDP	434
collection.decibelinsight.net.	A_2	94.145.230.0	35.180.0.0	0	ICMP	768
wahoofitness.com.	A_2	2.130.11.0	151.101.0.0	443	TCP	93

Table 5. *Example of the most relevant columns from blacklisted communication records.*

4.6 Discussion

The C_{black}=90796 blacklisted, aggregated flows contain $C_{black,DNS}$=78312 unique DNS responses (defined by DNS timestamp, clientip, qname and ipprotocol). This represents $\frac{C_{black,DNS}}{D_{black}} = 1,1\%$ of all blacklisted DNS responses. However, as packet sampling is employed, this only accounts for the number of observed flows, not the actual number of flows. Techniques exist for estimating the actual number of flows based on the observed number of flows [5]. However, this does not imply that $\frac{C_{black,DNS}}{D_{black}}$ can be scaled by the same techniques, as the non-observed flows could in theory all be related to the DNS responses already found in $C_{black,DNS}$. Therefore, the data available in this study does not allow any further conclusions on the magnitude of $\frac{C_{black,DNS}}{D_{black}}$.

The data sets from week 1 and 2 show that the amount of blacklisted DNS responses in each week differ significantly from $D_{black} = 6,81 \cdot 10^6$ to $4,56 \cdot 10^6$, a drop of 33%. The collected data cannot offer an explanation for this difference, which may simply be attributed to varying activity levels of the malicious actors. As a consequence of this, the fraction $\frac{D_{black}}{D}$ differ proportionately.

It is, however, interesting to note that although D_{black} differ by 33%, the amount of observed flows blocked, C_{black}, only show a drop of 4%, from 90796 to 86854. The estimated ratio of blacklisted DNS requests that result in a TCP flow, $\frac{C_{black,DNS}}{D_{black}}$, does not vary much between the weeks either. This could indicate that the amount of blacklisted flows may be a temporally less variable metric than the amount of blacklisted DNS responses.

For readability, this paper will refer to the set of aggregated flows that are considered blacklisted because of an IP address entry on the blacklist as C_{ip} (68045 entries), the set of aggregated flows that are considered blacklisted because of a domain name entry on the blacklist as C_{dom} (22842 entries), and the set of aggregated flows that are considered blacklisted because of both a domain name and IP address entry on the blacklist as C_{both} (91 entries), where $C_{ip} \cup C_{dom} = C_{black}$ and $C_{ip} \cap C_{dom} = C_{both}$. As C_{both} contains an insignificant number of entries, this category will not be analysed separately in this paper.

5 Type of maliciousness

Two sets of blacklisted flows, C_{ip} and C_{domain}, were identified in the previous section. These are the flows that would have been blocked if DNS based blacklists had been deployed, thus representing a network-level impact of blacklist deployment (subject to scaling due to NetFlow sampling). However, as outlined in the introduction, some blocked flows do not represent a threat to the user

due to different types of maliciousness, and these may be seen as a nuisance instead. To quantify this user-level impact of blacklist deployment, this section will categorize the flows by the type of maliciousness.

Different types of malicious behaviour can cause a domain name or IP address to be blacklisted, but only some of the types should be considered a threat to the user connecting to the blacklist entry. The observations turn out to be different for C_{ip} and C_{domain}, therefore the observations will be described separately.

5.1 Flows blacklisted by domain name

Both the A and B lists provide categories for phishing/malware/botnet related domains, as well as a more general spam category. The latter category includes for example domains in unsolicited mails promoting pills, counterfeits, dating sites etc., and is therefore in terms of badness distinct from malware/phishing domains. Although the sites and goods promoted in the spam category may not be desired to most users, they do not represent a cyber security threat. On the other hand, the phishing and malware related domains in what we will define as the malicious category clearly represent a cyber security threat to the user. In C_{dom}, 3% of the flows are in the malicious category, and the remaining 97% of the flows are in the spam category.

5.2 Flows blacklisted by IP address

Determining the type of maliciousness for entries in C_{ip}, requires different approaches for each list used.

The B lists provide a cause for an IP address to be blacklisted, and 99% of all IP address entries in the B lists are in the malicious category. However, only 5 entries in C_{ip} are blacklisted by B list entries (109 entries in the week 2 data set). As this is an insignificant amount compared to the total amount of entries in C_{ip}, no further analysis of the type of maliciousness of these entries is made.

Two A lists contain IP addresses: The A_1 and A_2 lists. The A_1 registers only spam emitters, and the 18292 flows blacklisted only by the A_1 list (and not also the A_2 list) are therefore considered in the spam category.

The type of maliciousness is not immediately available for the A_2 list. Two distinct groups of A_2 related flows (including flows that relate to both A_2 and A_1) are therefore categorized by other means:

- A subset of A_2, called A_3 is available as a separate list. 3179 entries in C_{ip} are marked by the A_3 (5%) and are therefore in the malicious category.
- A substantial amount of entries (13634, 20% of C_{ip}) relate to a single IP address owned by a laundry company. A manual lookup reveals that this IP address is in the spam category [12].

An informal conversation with list A representatives concluded that the vast majority of A_2 related flows not accounted for above are likely to be in the spam category as well. However, as we cannot quantify this, we will categorize the remaining flows as having unknown type of maliciousness.

5.3 Discussion

The type of maliciousness for the C_{ip} and C_{domain} flow sets are listed in Table 6. An important note is that if Telenor Denmark had only deployed domain name based blacklists, and only blocked the flows that are considered malicious to the user, a total of 1360 observed flows would have been blocked during week 2. The unknown C_{ip} entries are expected to mostly be in the spam category, with an informed guess setting the fraction of malicious flows in C_{ip} to less than 10%.

Type	$\mathbf{C_{ip}}$		$\mathbf{C_{dom}}$	
	Week 1	Week 2	Week 1	Week 2
Spam	31926 (47%)	46918 (75%)	22061 (97%)	23126 (94%)
Malicious	3184 (5%)	1151 (2%)	781 (3%)	1360 (6%)
Unknown	32935 (48%)	14614 (23%)	0	0

Table 6. *Type of maliciousness for blacklisted flow sets.*

Type	$\mathbf{C_{ip}}$		$\mathbf{C_{dom}}$	
	Week 1	Week 2	Week 1	Week 2
Spam	20%	13%	40%	37%
Malicious	11%	47%	72%	61%
Entire data group	39%	18%	40%	38%

Table 7. *Port 80/443 (HTTP/HTTPS) fraction of flows. In the group of flows that are blacklisted by IP address (is in C_{ip}) in the week 2 data set, 13% of the spam-related flows in the group use port 80/443, and 18% of all flows in the group use port 80/443.*

Some DNS based blocking implementations redirect the user to a web page warning the user that he has been blocked for security reasons. Web traffic, defined as traffic towards port 80 and 443, accounts for 40% of the entries in C_{dom}, and 72% of malware/phishing entries in C_{dom}. Further numbers are available in Table 7. Measuring the impact of the DNS based blocking by the number of visits to the warning web site will therefore underestimate the efficiency.

6 Misaligned endpoints

In some scenarios where a user connects to a blacklisted IP address, there is a chance that the user does not in fact connect to the entity that caused the IP address to be blacklisted. A popular example is when a user connects to a web site hosted in a shared web hosting environment. The IP address of the shared hosting environment may be on the blacklist, but it may be included on the blacklist even though only one of the hosted sites serves malicious content. In this case, it is not possible to determine from either NetFlow or DNS data if

the web site actually accessed by the user is benign or malicious. From a user perspective, this will likely be perceived as a nuisance, as the blacklist will then prevent access to benign sites not representing any risk. To assess the user-level impact of deploying DNS based blacklisting, it is therefore relevant to quantify the fraction of flows in C_{ip} where the endpoint of the flow and the endpoint causing the IP address to be blacklisted can differ. We shall refer to these flows as potentially having misaligned endpoints.

An analysis of each individual endpoint IP prefix would be impractical. In order to identify the most prominent groups of C_{ip} flows, we choose to focus the analysis on the groups of flows where:

- Many domain names are associated with a single destination IP prefix (high *qname*/*dstip* ratio)
- Many destination IP prefixes are associated with a single domain name (high *dstip*/*qname* ratio).
- A popular destination IP prefix is used (high *dstip* count)
- A popular domain name is used (high *qname* count)

Based on this analysis, three different scenarios that can cause misaligned endpoints has been identified in the C_{ip} data set, and these three scenarios are elaborated in the following three subsections.

6.1 Shared content providers

The entries in C_{ip} with a high *qname*/*dstip* ratio all have a *dstip* owned by a CDN, shared web hosting or similar cloud content provider like Amazon, Microsoft Azure, Google Cloud, DigitalOcean or Tencent. A total of 29556 entries (43% of C_{ip}) are related to such servers and we consider these flows to potentially having misaligned endpoints.

An number of *dstip*s are owned by Virtual Private Server (VPS) service providers and regular ISP customers. 516 entries are considered ISP customers as well, as they relate to a server with a dynamic IP address, identified by the use of a .duckdns.org domain name, a service used for assigning a permanent domain name to a dynamic IP address. These will not be considered as potentially misaligned endpoints.

It could be argued that all destination IP addresses could easily be enumerated by the use of BGP AS numbers. In practice, however, this turns out not to be viable for a number of reasons. First, only the /16 prefix address is available due to anonymization, and such a prefix may cover several AS numbers. Second, some providers share the AS number between the ISP and hosting part of the company (like OVH). Third, some providers reserve smaller prefixes for specific customers (like Amazon). Fourth, some providers use IP space assigned to other entities (like Tencent using ChinaUnicom owned IP prefixes).

6.2 VPN service providers

VPN service provider (PrivateInternetAccess, Hula, NorthGhost etc.) traffic identified by the *qname* accounts for 12469 (18%) of all entries in C_{ip}. The

specific implementations by the different providers is not known. However, it seems unlikely that a user creating a connection towards such an IP address will be relayed to a host residing behind the VPN service. A connection towards such a server seems more likely to be an attempt to use the service. The VPN provider IP addresses are likely to have been blacklisted because hosts using the service generated traffic that triggered a blacklisting. We shall therefore consider the VPN provider IP addresses as potentially having misaligned endpoints.

6.3 NTP Pool

Traffic towards hosts registered in the ntppool project[4] is identified by the *qname* containing .pool.ntp.org. This traffic accounts for 4006 (6%) of all entries in C_{ip}. A DNS request for a .pool.ntp.org domain will return a number of IP addresses, where each IP address belongs to a pool member. If the IP address of one of the pool members in the DNS response is blacklisted, the entire DNS response is considered blacklisted. Therefore, a connection towards a different pool member will be considered part of C_{ip}. This is not considered a flaw in the data analysis, as it reflects how DNS based blacklisting is implemented in practice. Blacklisting relates to the entire DNS request/response pair, not just to single response resource records. It is therefore likely that these flows have misaligned endpoints.

6.4 Discussion

Table 8 summarizes the amount of flows that may have misaligned endpoints and lists the 3 identified scenarios causing the potential misalignment. As seen in the table, we consider at least 45698 of 68045 C_{ip} entries (67%) as potentially having misaligned endpoints. Blocking these flows involves a risk of blocking benign sites. Although the specific IP address on the blacklist may be correct by reflecting a malicious endpoint using that IP address (a true positive), the user may perceive it as a false positive. When a provider considers deploying DNS based blacklists that includes IP addresses, the willingness of both operators and users to accept this risk should therefore be carefully considered up front.

Cause	**Week 1**	**Week 2**
Shared content providers	29556 (43%)	19370 (31%)
VPN service providers	12469 (18%)	13710 (22%)
NTP pool	4006 (6%)	3505 (6%)
Total	45698 (67%)	36336 (58%)

Table 8. *Amount of entries in C_{ip} with different causes for potential endpoint misalignment. Note that the total is less than the sum, as for example an NTP pool entry may also be a shared content provider entry, and this only counts as 1 in the total.*

Of the 45698 entries, only 5 are tagged by the B lists, while the rest is tagged by A lists. This highlights that the choice of blacklists represent an important

[4] https://www.ntppool.org/

limitation to the results presented in this section. The numbers presented are unlikely to be representative of other blacklists. However, looking outside the scope of this paper, this also suggests that when considering deploying DNS based blacklisting, the concept of IP address blacklists should not necessarily be deselected upfront. The risk of blocking benign sites due to endpoint misalignment can be decreased significantly by the careful selection of IP address blacklists.

As outlined in Section 2, we are only aware of one other paper that evaluates the risk of misaligned endpoints for the individual blacklist entries [8]. However, it is important to notice that the results presented in this section would not be directly comparable, as they relate to the blacklisted flows (C_{black}), not the blacklisted DNS responses (D_{black}) or the individual entries on a blacklist (the latter being the focus of [8]). The primary purpose of our work is to present the method of using NetFlow to measure the impact of deploying blacklists using a specific set of blacklists as examples, and not to compare blacklists. Hence, we consider evaluating a larger number of blacklists as an extension to this paper.

7 Impact

Sections 3-6 identify blacklisted flows, identify the type of maliciousness and assess the risk of endpoint misalignment. This section combines the result of the previous sections to quantify impact of deploying DNS-based blacklists seen from the network perspective and from the user perspective. Furthermore, this section describes interesting future works.

7.1 Network-level impact

The network-level impact of DNS based blocking is usually practically measured by counting the number of blocked DNS requests or by counting the number of visits to a warning page to which a user has been redirected. In this paper, the impact is instead measured using the number of blocked flows instead, and this reveals the fraction of web related flows.

- Approximately 0,02-0,03% of all DNS responses match a blacklist entry, and 1,1-1,5% of these blacklisted DNS responses can be associated with an observed flow, denoted a blacklisted flow. The use of sampled flow data was found to hinder the estimation of the actual fraction of blacklisted DNS responses that can be associated with a flow. Researchers or ISPs with access to non-sampled NetFlow and DNS data should assess the fraction of blacklisted DNS responses that can be associated with a flow. Given a known amount of blacklisted DNS responses, this would make it possible to more accurately assess impact of doing DNS based blocking.
- Some DNS based blocking implementations redirect the user to a website containing a message warning the user that he has been blocked for security reasons. Therefore, such implementations measure only the part of the traffic that is web traffic. Of the flows blacklisted by domain name, 38-40% are

web traffic. Of the flows blacklisted by domain name *and* considered having a high threat level, 61-72% are web traffic. Therefore, this paper shows that measuring the impact of blacklisting by the number of visits to the warning web site underestimates the impact. ISPs and company system administrators should implement measures to also count non-web related connections, in order to get a more correct assessment of the blacklist impact.

These results are specific to a particular week, use particular blacklist vendors, and a particular ISP. Despite the listed limitations, we find the results significant enough to suggest that the method of using NetFlow to measure the impact of applying DNS based blacklists represents an improvement to existing methods.

7.2 User-level impact

Approximately 25% of the blacklisted flows relate to a blacklisted domain name, whereas the remaining 75% of the blacklisted flows relate to a blacklisted IP address.

- The flows blacklisted by domain name are, using the threat type categories provided by the blacklist vendors, divided into two groups. First, a group relating to general spam, considered a nuisance rather than a cyber security threat, accounts for 94-97% of the flows. Second, a group relating to phishing, malware and botnet accounts for the remaining 3-6%. When deploying DNS based blacklisting, it is therefore important to consider if both or only one of these types of traffic should be blocked, as this will have a significant impact on the amount of blocked connections experienced by the user.
- Of the flows blacklisted by IP address, this paper shows that 58-67% may be flows towards benign sites, primarily due to the prevalence of shared web hosting, whereby multiple web sites / domain share the same IP address. From a user and operator perspective, the willingness to risk blocking benign sites must be considered before deploying IP address based blacklists. This study shows that carefully selecting the IP address blacklist vendor can be a significant contribution to minimizing this risk.

These results are also specific to particular blacklist vendors and a particular ISP. Here, however, we show that the specific measurements depend a lot on the particular blacklist used, and therefore it is a clear limitation that these results cannot be generalized to different blacklists.

8 Conclusion

In this paper, we propose a method to measure the impact of deploying blacklists by combining NetFlow and DNS data. We evaluate the method on real data, containing anonymised NetFlow and DNS records collected by Telenor Denmark for two weeks, and combine these with blacklists containing IP addresses and domain names provided by two commercial vendors.

The measurements show that 0,02-0,03% of all DNS responses match a blacklist entry, however only 1,1-1,5% of these blacklisted DNS responses can be associated with an observed flow. Furthermore, only 38-40% of the blacklisted flows are web traffic. These observations suggest that the use of flow data can be used to make a more precise impact assessment than counting the amount of DNS responses matching a blacklist entry or counting the amount of visits to a warning web page.

For flows blacklisted by domain name, 3-6% of the flows related to phishing, malware and botnet domains, while the remaining flows relate to spam domains. For the flows blacklisted by IP address, 58-67% may be flows towards benign sites. These observations show that it the careful consideration of the choice of blacklist type (domain name or IP address) and category (spam, malware etc.) before deployment is essential to avoid undesired impact seen from a user perspective when deploying DNS-based blacklists.

References

1. Bermudez, I.N., Mellia, M., Munafò, M.M., Keralapura, R., Nucci, A.: DNS to the Rescue: Discerning Content and Services in a Tangled Web. IMC: Internet Measurement Conference (2012), https://doi.org/10.1145/2398776.2398819
2. Bouwman, X., Griffioen, H., Egbers, J., Doerr, C., Klievink, B., van Eeten, M.: A different cup of TI? The added value of commercial threat intelligence. USENIX Security Symposium (2020), https://www.usenix.org/system/files/sec20-bouwman.pdf
3. Cisco Umbrella: Better intelligence drives better security (2020), https://umbrella.cisco.com/solutions/reduce-security-infections
4. Connery, H.: DNS: Response Policy Zone (2012), https://dnsrpz.info/spamhaus-rpz-case-study.pdf
5. Duffield, N., Lund, C., Thorup, M.: Properties and prediction of flow statistics from sampled packet streams. IMW: ACM SIGCOMM Internet Measurement Workshop (2002), https://doi.org/10.1145/637201.637225
6. Fejrskov, M., Pedersen, J.M., Vasilomanolakis, E.: Cyber-security research by ISPs: A NetFlow and DNS Anonymization Policy. International Conference on Cyber Security And Protection Of Digital Services (2020), https://doi.org/10.1109/CyberSecurity49315.2020.9138869
7. Foremski, P., Callegari, C., Pagano, M.: DNS-Class: immediate classification of IP flows using DNS. ACM Internation Journal of Network Management (2014), https://doi.org/10.1002/nem.1864
8. Griffioen, H., Booij, T., Doerr, C.: Quality Evaluation of Cyber Threat Intelligence Feeds. ACNS: International Conference on Applied Cryptography and Network Security (2020), https://doi.org/10.1007/978-3-030-57878-7_14
9. Jung, J., Sit, E.: An empirical study of spam traffic and the use of DNS black lists. IMC: Internet Measurement Conference (2004), https://doi.org/10.1145/1028788.1028838
10. Kührer, M., Rossow, C., Holz, T.: Paint It Black: Evaluating the Effectiveness of Malware Blacklists. RAID: Research in Attacks, Intrusions and Defenses (2014), https://doi.org/10.1007/978-3-319-11379-1_1

11. Li, V.G., Dunn, M., Pearce, P., McCoy, D., Voelker, G.M., Savage, S., Levchenko, K.: Reading the Tea leaves: A Comparative Analysis of Threat Intelligence. USENIX Security Symposium (2019), https://www.usenix.org/system/files/sec19-li-vector_guo.pdf
12. MXToolbox: Blacklist check (2021), https://mxtoolbox.com/blacklists.aspx
13. Ramachandran, A., Feamster, N.: Understanding the Network-Level Behavior of Spammers. ACM SIGCOMM Computer Communication Review (2006), https://doi.org/10.1145/1159913.1159947
14. Satoh, A., Nakamura, Y., Fukuda, Y., Sasai, K., Kitagata, G.: A Cause-Based Classification Approach for Malicious DNS Queries Detected Through Blacklists. IEEE Access (2019), https://doi.org/10.1109/ACCESS.2019.2944203
15. Sheng, S., Wardman, B., Warner, G., Cranor, L., Hong, J., Zhang, C.: An empirical analysis of phishing blacklists. CEAS: Conference on Email and Anti-Spam (2009), https://doi.org/10.1184/R1/6469805.V1
16. Sinha, S., Bailey, M., Jahanian, F.: Shades of grey: On the effectiveness of reputation-based "blacklists". MALWARE: International Conference on Malicious and Unwanted Software (2008), https://doi.org/10.1109/MALWARE.2008.4690858
17. Spacek, S., Lastovicka, M., Horak, M., Plesnik, T.: Current Issues of Malicious Domains Blocking. IFIP/IEEE International Symposium on Integrated Network Management (2019), https://ieeexplore.ieee.org/document/8717891
18. SpringerNature: Accepted manuscript terms of use (2021), https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms
19. Telenor Norway: Stanset over 80.000 besøk på falske nettsider på én måned (2020), https://www.mynewsdesk.com/no/telenor/pressreleases/stanset-over-80-dot-000-besoek-paa-falske-nettsider-paa-en-maaned-2986773
20. Wikipedia: Domain Name System-based Blackhole List (2020), https://en.wikipedia.org/wiki/Domain_Name_System-based_Blackhole_List
21. Wilde, N., Jones, L., Lopez, R., Vaughn, T.: A DNS RPZ Firewall and Current American DNS Practice. ICISA: International Conference on Information Science and Applications (2019), https://doi.org/10.1007/978-981-13-1056-0_27
22. Williamson, R.: What do Canada and New Zealand have in common? (2020), https://internetnz.nz/blog/dns-firewall-what-do-canada-and-new-zealand-have-in-common/
23. Zhang, J., Chivukula, A., Bailey, M., Karir, M., Liu, M.: Characterization of Blacklists and Tainted Network Traffic. PAM: International Conference on Passive and Active Network Measurement (2013), https://doi.org/10.1007/978-3-642-36516-4_22
24. Zhauniarovich, Y., Khalil, I., Yu, T., Dacier, M.: A survey on malicious domains detection through DNS data analysis. ACM Computing Surveys (2018), https://doi.org/10.1145/3191329