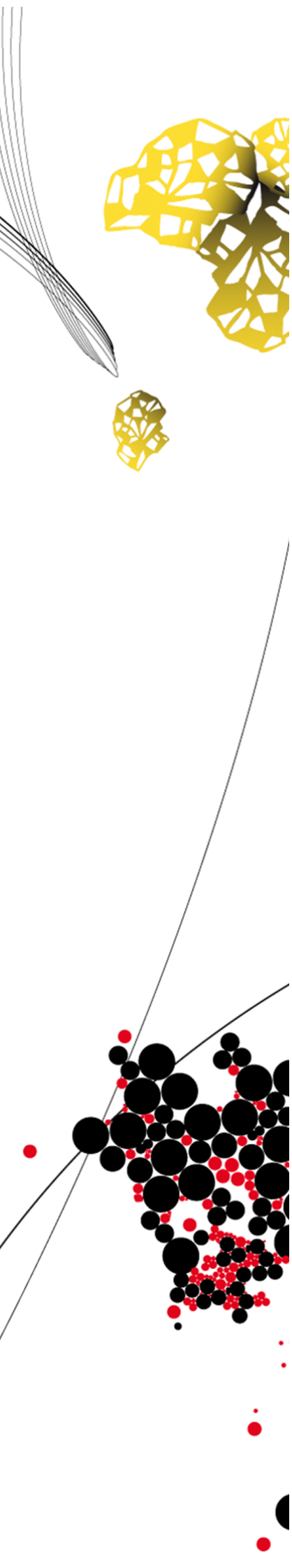

UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering,
Mathematics & Computer Science

Complete Characterization of Publicly Available Domain Blacklists

Ivan Lukman

M.Sc. Thesis
August 2019

Graduation committee:
dr. A. Sperotto (Anna)
dr.ing. E. Tews BSc (Erik)

Design and Analysis of Communication Systems Group
Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Preface

I would first like to express my sincere gratitude to dr. A. Sperotto (Anna) for giving me the opportunity to conduct this research under her supervision. I had a good time working with her in one of my favorite topics in cyber security. Her continuous support and proficiency in this field have allowed me to finish this research on time.

I am also very grateful for my family and friends for accompanying my journey pursuing my Master's degree in the Netherlands. Their presence not only encouraged my academic life, but also supported me to enjoy my days in Enschede.

Abstract

Domain names are not only used for *benign* purposes, like sharing information or buying/selling items. Numerous categories of cyber incidents, such as phishing, mail spamming, or distributing malicious software, also involve domain names. Domain blacklists (DBLs) aim to collect these *malicious* domains and store them in a list to lower the number of victims of cyber-crime.

However, currently, there are many different sources that publish blacklisted domain names, also with different blacklisting methodologies. In this study, the DBLs used were accessible for free in the Internet, meaning that everybody can access the blacklisted domain names without any charge. This research was aimed to provide a complete characterization of thirteen different publicly available DBLs, in terms of how well they document and maintain their database.

This study is one of the first project that completely characterize multiple public DBLs. Similar previous studies have been conducted under different scenarios, one of them was related with only mail-spamming activities. Nevertheless, some of the approaches introduced could still be applied to achieve the main goal of this research, which is to understand the maintenance and the documentation of public DBLs.

This research shows that there is no *perfect* DBL. One of the metrics defined later in this report indicates that all public DBLs used in this research have false positives (blacklisted benign domains). In addition, not all of the blacklisted domain names were active during the blacklist time. The reported malicious domain names might have been removed already. Another interesting result is that, DBL that publish a large number of domain names per day might not explain how the domain names got blacklisted or publish the details of the blacklisted domain names.

One additional metric to investigate how well public DBLs were maintained is *liveliness*. This estimates the ratio of active machines from the published blacklists from each DBLs. Unfortunately, this metric needs special considerations and attentions to be implemented. Firstly, the application is required to be efficient because of the massive number of blacklisted domain names per day. In addition, *touching at* lots of malicious machines could raise some problems, such as ethical and security concerns.

Contents

List of Acronyms

AS Autonomous System

C2dom OSINT Feeds from Bambenek Consulting

CCTracker CyberCrimeTracker

DBLs Domain Blacklists

DNS Domain Name System

MDL MalwareDomainList

MS Mail Spam

MW Malware Distribution

P Phishing

RWTracker RansomwareTracker

SLDs Second-level Domain Names

ZTracker ZeusTracker

Chapter 1

Introduction

The bonding between human and the Internet is growing stronger and stronger each day. People are taking more advantages from more web-based applications, Internet of Things (IoT) services, or social media. These technologies are relying on the Internet. In principle, one of the basic mechanisms for the functioning of the Internet hinges on the interaction between two resources, namely the Internet Protocol (IP) addresses and domain names. The first resource, the IP address, is a numerical label assigned as an identification for every device connected to IP-based network, including, the Internet. The second one, the domain names, are unique human-readable labels that are understood by both the Internet users and the system. A service called Domain Name System (DNS) does the translation of domain names into IP addresses. In addition, one IP address can also host several domain names. Instead of memorizing sequence of numbers of the IP address of a website, the Internet users can just simply memorize its domain name. For example, memorizing `google.com` is comparatively simpler than `172.217.20.78`.

On the other hand, in the last few years, the number of cyber security incidents, such as malware distributions, phishing, email spamming, botnet infections, is showing an increasing trend. Taking the advantage of both the easy-to-access Internet services and the stronger bond between human activities with the Internet, attackers can spread their malicious software, broadcast spams, distribute fake websites much easier and faster.

One of the effective methods to reduce the number of victimized Internet users is by creating a blacklist, where these “malicious” systems are contained into a list. In general, there are two types of blacklists, namely the domain-based blacklists and IP-based blacklists. As the name suggests, Domain Blacklists (DBLs) contain “malicious” domain names, whereas IP blacklists contain the IP addresses of “malicious” systems.

Based on the availability of the data, blacklists can be categorized into two main groups, which are the publicly available, and premium blacklists. Publicly available

blacklists, like MalwareDomainList [1] or Joewein [2], publish their data for free in the Internet. Meanwhile, premium blacklists, such as ESET Anti-Phishing database [3] or blacklists provided by antivirus programs, only give their database access to subscribed users. There are also combinations of both, such as blacklist provided by SpamHaus [4]. SpamHaus release part of their blacklisted domain names for free, but users need to pay for their complete database.

By investigating the websites and forums of different DBLs, different blacklist sources show different characteristics of blacklisting, such as:

- Detail of a blacklist.
  Different blacklist sources provide different level of detail of blacklisted machines. Some blacklists contain only the domain names or the IP addresses, whilst other blacklists include WHOIS information or the machine's Autonomous System (AS) Number.

- Categories of malicious behaviors.
  Different blacklists focus on different category of malicious activities, such as Mail Spam (MS), Malware Distribution (MW), or Phishing (P). Some sources only blacklist systems that are related with spam campaign, while other DBLs contain machines that are used for multiple cyber incident categories.

- Blacklist update frequency.
  Most of the blacklists are maintained based on reports submitted by their members or through their own *sinkhole*. Hence, the delay between the first appearance of a malicious incident until the system gets blacklisted could be very diverse. Some sources update their blacklist entries once per day, while the others could update their list once every 5 minutes, or even real-time.

- Blacklisting/de-listing methodology and verification procedures.
  Based on the described procedures in their web pages or fora, there are multiple ways of registering a domain name or IP address of a malicious system into blacklists. For instance, members could submit the malicious domain names through online form, forum messages, or email, to the DBL's administrators. Furthermore, different sources could carry out different verification strategies to analyze the submitted domain names, whether they are indeed malicious or just some benign domains reported by mistakes.

- Volume.
  Since not all blacklists are maintained by their members, the number of active members and their frequency of submitting malicious domain names could vary considerably. As a result, the number of new domain names appearing in the released blacklist could vary a lot.

1.1 Motivation

Until this research was conducted, different DBLs have been used in different studies. For instance, Kührer *et. al.* [5] analyzed systems that were used just for distributing malicious applications. On the other hand, Sheng *et. al.* [6] used only data from phishing blacklists in their study. However, the number of studies that comprehensively investigates how the publicly available DBLs are maintained and documented is still minimum. It is also important to make sure that the released data of DBLs are large and unbiased enough to cover the overall situation of malicious activities, so that the studies could give essential knowledge.

As can be seen from the short summarized characteristics of publicly available DBLs in the previous section, different DBL has different characteristics, notably the number of domain names captured by each DBLs. For instance, the number of domain names that got blacklisted by HostFile is much more superior than ThreatExpert. However, this difference does not necessarily mean that HostFile's data are more useful or suitable than ThreatExpert. More detailed information about this will be discussed later in this report at Section 4.1.

This is not the only difference that can be spotted in different sources of DBLs. Therefore, it is important to understand how each DBL is maintained, to provide information about which blacklist publishes data that are more suitable in which condition, than the others.

1.2 Research Goal

The main goal of this research is to understand how different sources update their DBL of different malicious categories, as well as, how detailed they are in giving information about their blacklisted domain names.

Therefore, firstly, getting insights on the state-of-the-art metrics and understanding the applicability into the data set used in this research are critical in determining how to achieve the goal of this study. Then, by applying the suitable metrics into the data from domain blacklists, information about how each DBL is maintained and its suitability can be determined.

1.3 Research Question and Approach

To meet the aforementioned research goal, the following Research Question (RQ), with several sub-questions, is defined.

"How well are publicly available domain blacklists from different categories documented and maintained?"

With the following sub-questions:

1. In which proportion does a DBL source contribute to the overall new blacklisted domains intake?
   (a) Do DBL sources also include benign domain names?
2. What is the level of details each DBL source provide?
3. How quick does a DBL source blacklist and remove domain names?
   (a) How long do blacklisted domain names stay in each DBL source?
   (b) Do blacklisted domains re-appear at a later point of time?
   (c) Are blacklisted domains also found in other DBLs?
4. Do DBL sources contain domain names that are currently active?

To answer the four sub-questions, the following approaches are defined.

1. Sub-question 1 can be answered by performing pairwise comparison and finding exclusive domains for each DBL source, which will be described in more detail at Section 3. Then, pairwise comparison with Alexa Top Global Sites [7] is performed to determine how many blacklisted domain names are also found in Alexa's list of popular websites.
   Based on preliminary mini-research, taking Alexa's Top 100k popular website is considered to be a large enough data set of domain names that is also almost completely benign. This mini-research showed that less than 0.5% of Alexa's top 100k website list could be associated to malicious activities. This result was generated by cross-checking Alexa's Top 100k website against VirusTotal URL scanner [8]. Of course, using a larger Alexa's list of popular websites will cover more benign domain names, but it is also important to note that the number, also possibly the ratio, of malicious domains in the list will also increase.

2. To find the answer of sub-question 2, finding and performing analysis of the information provided by each DBL source can be done. In most of the publicly available DBLs, the descriptions were not just posted in their website, but also their fora or posted in their related services.

3. Temporal analysis can be performed to determine the answer for the third sub-question. Finding the duration a domain name “stays” in and disappear from a DBL, and the existence of the same domain name from different blacklists, or at a later time, are essential to answer this sub-question. Then, the quickness of a DBL source in blacklisting and de-listing domain names can be estimated by comparing the first and last appearance date of malicious campaigns found at multiple DBL sources.

4. To estimate the number of blacklisted domain names that are still active at the blacklisted date, live testing can be conducted. The *liveliness application* introduced in this research will be executed to check whether specific ports of a domain name are accessible, as well as to try retrieving its HTTP and HTTPS response codes.

1.4 Report organization

The remainder of this report is structured as follows. In Chapter 2, the existing studies and state-of-the-art metrics are explained. Then, in Chapter 3, the data sets, considerations and selected metrics are discussed. The results of this research and the analysis are shown and elaborated at Chapter 4. One of the selected metrics, the *Liveliness* is completely described and analyzed at Chapter 5. This report concludes at Chapter 6 as the conclusions and the discussions of the limitations and future works.

Chapter 2

Related Works and Existing Metrics

This chapter elaborates the related existing studies and state-of-the-art metrics that can be extracted from the previous studies.

2.1 Related Works

As far as this study is carried out, this is one of the first study that comprehensively investigate how publicly available DBLs collect, update, and archive malicious domain names from multiple categories. However, similar prior studies have been done and some of their approaches and analysis are relevant and useful to guide this research to answering the RQ defined in Chapter 1.3. In this chapter, some of these studies are discussed and summarized to highlight the keynotes and their related approaches in comparing different blacklists. Then, the usability of the existing metrics into the data set and the relevance with this study are also explained.

2.2 Existing Studies

1. Taster's Choice: A Comparative Analysis of Spam Feeds [9].
   This paper investigated Second-level Domain Names (SLDs) related with email spam campaigns. The authors attempted to understand the suitability of spam-related domain blacklists (feeds) to be used for further research analysis. This was done by comparing the contents of ten different sources of spam-related domain names. As stated in the paper, the blacklists used should not be "too small or too biased to be used for all purposes".
   In this paper, five distinct sources of spam-advertised domain names used were botnets, MX honeypot, seeded honey accounts, human identified, and domain blacklists. These sources had different levels of "purity" and "volume"

quality in capturing spam emails due to the different approaches in each methods. The ground truth in verifying their results was collected using the "Click Trajectories" project, which was a collection of spam value chain.
Several interesting points can be taken from this study. Firstly, the existence of the ground truth when verifying blacklisted domains is difficult to achieve. Capturing all spam campaigns occurring at the same time is almost impossible to do. Secondly, there is no perfect feed that are usable for all purposes. Even the best domain blacklist for spam campaign, if they exist, as also mentioned in this research, may still include benign domain names. This information is essential because further analyses might want to only use "bad" domain names and filtering out the benign ones. Therefore, it is important to not just take spam domains from a single feed without validating with other sources.
Four metrics used in this research to compare the quality of spam feeds are:

(a) Purity.
This metric measures how much of a given feed is actually spam-advertised domain names. To calculate the final indicator, this metric is determined using five approaches. The first one is by determining whether the domains are real or not by cross-checking the DNS zone files based on several major top level domains. The second approach is to test whether the domain names respond to an HTTP request. Then, the third point is validating their existence with Click Trajectories project. The fourth and fifth approach determines the percentage of benign domain names in the blacklists by cross-checking with Open Directory Project and Alexa top 1 million websites.

(b) Coverage.
This calculates what fraction of spam is captured by a particular feed. To determine the coverage, there are two approaches, which are by comparing the domain names that only appear in one feed and not in the others, and the domain names that appear in multiple blacklists. The first approach is referred as "exclusive domains", while the second one is determined by performing "pairwise comparison" for each feeds.

(c) Proportionality.
This evaluates the accuracy of a feed including the relative frequency. In the paper, not all feeds could be used to determine this metric because only two of them contained the volume information. This metric is determined by computing the Kendall rank correlation coefficient and comparing these values for the two feeds.

(d) Timing.

The last metric estimates the accuracy of a spam feed in representing the spam period. This metric measures how well each spam feed captures the timing of spam campaigns. This is determined by approximating the first and last appearance time of spam-advertised domain names in each blacklists.

2. Paint It Black: Evaluating the Effectiveness of Malware Blacklists [10].
This paper used both SLDs and IP addresses related with malware distribution and tried to evaluate the completeness and accuracy of malware blacklists. In this paper, 15 public malware blacklists and 4 blacklists maintained by antivirus vendors were used. To categorize the blacklist contents and understand the nature of the blacklisted domain names and IP addresses, first, the data sets were split into two categories, the *current* and *historical* domain names. Then, several mechanisms were introduced to identify parked domains (domains that are registered to display web advertisements) and *sinkholed* entries, such as by extracting unique features that were only found at *sinkholed* and parked domains. Using these mechanisms, this paper investigated how much of real-world malware domain names were actually blacklisted by these sources.
Parked domains have seven distinguishable features that were identified using Support Vector Machine (SVM) classifier and evaluated using 10-fold validation. Using similar approach to identify sinkholes, graph exploration was then used to capture actual sinkholes.
In this study, metrics that were used to evaluate the effectiveness of malware blacklists are:

   (a) Coverage (parked domains and sinkholes ratio).
   These ratios can be calculated by identifying parked domains and sinkholes from each blacklist using similar approaches as mentioned in the previous paper.

   (b) Completeness.
   This metric measures how much malicious domains are blacklisted and how much are not. To calculate the ratio, the *ground truth* is captured from dynamic malware analysis platform called `Sandnet`. The completeness of the blacklists is evaluated by computing the ratio of malicious domain names that appear in both `Sandnet` and the blacklists.

   (c) Reaction Time.
   This metric estimates how long it took for a malicious domain name to appear in the blacklists once they are seen in `Sandnet`.

   (d) Accuracy.
   This metric ensures that blacklists provide accurate information, since

blacklists may become outdated and the maliciousness of a domain name or IP address may change at a different point of time.

(e) Agility.
This metric keeps track of the number of active, new, and de-listed domain names from a blacklist on a daily basis. This shows which blacklist is more active than the others, also which one removes outdated entries more constantly.

3. Empirically Characterizing Domain Abuse and the Revenue Impact of Blacklisting [11].
This study analyzed IP addresses and SLDs related with multiple categories of malicious activities. The authors investigated the nature of the abused domains and the economic impact on the revenue from domain blacklisting. This paper used data from URIBL blacklist and spam-advertised domain names from Pitsillidis *et. al.*'s paper [9]. They studied the possible revenue from advertising via email spam, web searches, and usage of internet infrastructure, like free web hosting services. This paper took the performance of domain blacklists into account when measuring the revenue, since the speed and coverage could impact financially.
Metrics that were mentioned considering domain blacklists are:

(a) Speed.
This considers the delay for a spam domain to appear on a domain blacklist.

(b) Coverage.
This metric measures the overlaps and disjoints of multiple blacklists.

4. Phoneypot: Data-driven Understanding of Telephony Threats [12].
In general, this paper used telephone numbers to analyze telephony abuse and did not aim to characterize domain blacklists. However, several techniques that were explained to measure the quality of telephony abuse intelligence could be applied to DBLs. This paper aimed to understand telephony threats and introduce Phoneypot, the first large-scale telephony honeypot. The ground truth for this research was taken from Federal Trade Commission (FTC), a US government instance for people to submit complaints of abusive calls, and 800notes, a crowd sourced data set. The result of this research was the findings of misuse of telephone numbers, like used by debt collectors and telemarketers, and could lead to telephony denial-of-service attack.
Metrics that were explained to evaluate the quality of telephony abuse intelligence are:

(a) Completeness.
This metric evaluates how much telephony abuse are captured to have a complete picture of a certain threat. Completeness of telephony abuse intelligence can be estimated by finding the overlap of abuse report submitted to Phoneypot and FTC, two major source of telephony abuse reports.

(b) Accuracy.
This metric is defined as how detail a telephony abuse report should be described. More accurate description of an abuse report means that the report is submitted correctly. The extra information also provides reasons why the reported number is abusive.

(c) Timeliness.
This refers to how quickly a telephony abuse is reported. The duration ranges from one day to several weeks after the call is received.

5. Developing Security Reputation Metrics for Hosting Providers [13].
This paper did not aim to characterize domain blacklists. By analyzing SLD-IP Address pairs, this paper tried to investigate the security performance of hosting providers against cyber abuses. Comparison and analysis of data feeds were, however, explained and applicable to analyze domain blacklists. Metrics that were explained to determine the quality of data feeds are:

(a) Coverage.
This metric measures how much overlap is found between the different data feeds. The coverage is calculated by performing pairwise comparison on each data feeds and intersection analysis of these blacklists.

(b) Purity.
This quantifies how much of the blacklisted domains actually host malicious contents. All abuse feeds contain some domain names that are legitimate (false positives). To measure the purity of data feeds, domain names from each blacklist are checked and *a posteriori* analyzed whether they appear in Alexa top 25k list or not.

6. Blacklists Assemble: Aggregating Blacklists for Accuracy [14].
This paper aimed at aggregating multiple IP-based blacklists from various types of malicious activities into one master blacklist. The final product of this paper was a sophisticated approach to filter, merge, and selectively expand only the relevant information from various blacklists. This product was called BLAG. Three (not 100% accurate) ground truths were used in this paper, consisting of combination of Mailinator, Mirai, Darknet, Alexa top 500K websites,

and Ham. To validate benign domain names, every entry of Alexa and Ham lists was checked with Google Safe Browsing API. When combining domain blacklists into a master blacklist, *score matrix* containing reputation scores of each blacklists was used.
Several limitations were found in this study are first, blacklist sources often depend on specific attack type and thus, will miss out domain names that also used for different malicious activity. Secondly, the accuracy of blacklists may vary a lot, as it is quite difficult to capture all malicious activities from all over the world. Then, blacklists may also contain false positives, where legitimate traffics are falsely filtered because of dynamic addressing of IP addresses. Metrics that were implemented to determine the performance of BLAG are:

(a) Recall.
This metric measures the percentage of malicious behaviors that are blacklisted, based on several ground truth sets.

(b) Specificity.
This metric is the opposite of recall and estimates the percentage of benign hosts that are not blacklisted.

7. A Long Way to the Top: Significance, Structure, and Stability of Internet Top Lists [15].
This paper investigated nothing about malicious activities but the nature and evolution of top lists that were used in studies, such as Alexa Global, Cisco Umbrella, and Majestic Million lists. The characteristics analyzed for the three lists were significance, structure, stability, ranking mechanisms, and research result impact. Significance investigated the existence of rank manipulation and measures how important the Internet top lists are to scientific papers. Structure aimed to understand the properties of domains in top lists. Stability investigated how much changes occurred in each lists. In addition, this paper also studied the ranking mechanisms of each lists.
Some of the metrics that were explained in this paper are:

(a) Intersection between lists.
This metric is used to measure the level of inconsistency between each lists. This metric could be determined by checking whether domain names appear in similar rank for each lists.

(b) Stability of top lists.
This metric investigates the daily changes and weekly patterns of each lists. By comparing each data sets, daily fluctuations of domain ranks could be learned. In addition, weekly patterns of domain rank are ob-

served, for instance, some domains that turned out to be more popular in the weekends. Not only the periodic fluctuations, but this metric also keeps track of new or in-and-out domains.

2.3 Existing Metrics

Not all state-of-the-art techniques and metrics discussed from the previous section are usable and relevant in this research. This section sums up all metrics and their approaches, based on the keynotes of existing studies from the previous section. The usability of each metrics on the data sets used in this research is also discussed.

2.3.1 Purity

- Definition: The percentage of actual spam or malicious contents in a blacklist [9], [13].
- Approach:
  1. Finding the proportion of unique domain names in a feed that were registered based on several major top-level domains.
  2. Finding the ratio of unique domain names that responded to an (*a posteriori*) HTTP request.
  3. Finding the ratio of unique domain names that lead to storefronts or are *tagged* in Click Trajectories Project.
  4. Calculating the fraction of unique domain names appearing in Open Directory Project listings.
  5. Calculating the ratio of unique domain names appearing in Alexa Top 1 million websites.
- Usability: Yes.
  This metric can be used for analyzing domain blacklists as this does not depend on any ground truths. However, not all of the approaches can be applied, such as approach 3 and 4, because the additional projects used are exclusively created for the corresponding studies.

2.3.2 Coverage

- Definition: The percentage of actual spam or malicious contents that are blacklisted [9]–[11], [13].

- Approach:

    1. Finding the ratio of unique domain names that appear only on a single feed and not in other feeds (exclusive domains).
    2. Finding the ratio of unique domain names that appear on multiple feeds (pairwise comparison).

- Usability: Yes.
  This metric can be used for comparing different domain blacklists alongside with the approaches.

2.3.3 Proportionality

- Definition: How well a blacklist accurately represents the relative volume of different campaigns [9].

- Approach:

    1. Finding the distribution of domains, relative to the number of times a domain is seen in spam.

- Usability: No.
  This metric is not usable since the data does not contain any volume information about the frequency a domain was seen in a malicious campaign.

2.3.4 Timing

- Definition: How accurate a blacklist estimates the start and end time of a spam or malicious campaign [9].

- Approach:

    1. Finding the first appearance time of a domain name in domain blacklists.
    2. Finding the last appearance time of a domain name in domain blacklists.
    3. Estimating the relative duration of the campaign by subtracting the last appearance time by the first appearance time.

- Usability: Yes.
  This metric can be used to determine the distribution of the duration of a domain name appearing in different blacklists.

2.3.5 Speed / Timeliness

- Definition: How quick a spam or malicious domains to appear on a blacklist [10]–[12].

- Approach:

    1. Comparing the first appearance time of a domain name in domain blacklists with the defined ground truth.

- Usability: No.
  This metric can not be used since the existence of the ground truth in analyzing domain names is difficult.

2.3.6 Recall

- Definition: How many offenders are blacklisted [14].

- Approach:

    1. Finding the ratio of unique blacklisted domains that are also observed in the ground truth.

- Usability: Not completely.
  This metric is difficult to perform due to the lack of ground truth, as also experienced by Pitsillidis [9]. However, aggregation of all malicious domains could be considered as the collection of all malicious activities.

2.3.7 Specificity

- Definition: How many legitimate domains are not blacklisted [14].

- Approach:

    1. Creating list of benign domains: cross-checking domain names from Alexa list that appears in Google Safe Browsing API.
    2. Finding the ratio of unique benign domains that are not blacklisted.

- Usability: Yes.
  This metric can determine the performance of a blacklist.

2.3.8 Historical and Current

- Definition: How many blacklisted domains are new or became de-listed [10].
- Approach:
  1. Current: Finding the number of new, unique blacklisted domains.
  2. Historical: Finding the number of unique blacklisted domains that are de-listed during the research period.
- Usability: Not completely.
  This metric can provide insights on characteristics of a blacklist feed. However, based on the observations result, many blacklists used in this study do not publish the history of the de-listed domain names. Therefore, the domain names that once appear in a blacklist and then disappear could be categorized as historical data.

2.3.9 Completeness

- Definition: How well blacklists perform in covering all domains for popular malware families [10], [12].
- Approach:
  1. Finding the percentage of unique blacklisted domain names with reference to the ground truth (`Sandnet`).
- Usability: No.
  This metric is difficult to perform due to the lack of ground truth.

2.3.10 Accuracy

- Definition: The details or consistency of each abuse report [10], [12].
- Approach:
  1. Finding the details, such as accurate date-time or reporting party, of abuse reports.
- Usability: Yes.
  This metric can be used since different feeds provide different depth of details of the blacklists.

2.3.11 Agility / Stability

- Definition: The consistency of domain names / ranking in lists [10], [15].
- Approach:
  1. Finding the daily fluctuations of domain rankings.
  2. Finding weekly patterns of domain rankings.
  3. Keeping track of new or in-and-out domain names.
- Usability: Not completely.
  Ranking, like in Scheitle's study [15], in domain blacklists is not that relevant. Once a domain is marked as malicious, the only way to change its rank is by de-listing. However, measure of counting the domain names that enter and exit a domain blacklist is useful in determining how DBLs are maintained each day.

Chapter 3

Settings and Methodologies

This chapter elaborates the data sets used in this study and discusses the considerations of the metrics defined in the previous section and lists the selected metrics to be used in this research.

3.1 Data Sets

In this research, data captured from thirteen distinct publicly available DBLs within different time stamps are used. This means that this research only covers DBL sources that distribute their database for free through the Internet. The published domain names were crawled on a daily basis, since the least frequent update is once per day. In this research, the term "domain" and "domain name" are referred to the SLDs, such as `google.com`. These thirteen DBL sources are described in this subsection and the statistics are computed only when the DBL published their daily updates.

The following description of each DBL contains the general information and statistics of the crawled blacklisted domain names. `Minimum` and `maximum` shows the minimum and maximum number of unique domain names that are found in the daily updates during the crawling period. `Q1`, `median`, and `Q3` indicates the 25^{th}, median, and 75^{th} percentile of the number of unique domain names per day during measurement period respectively. Similarly, `average`, `variance`, and `standard deviation` contain the average, variance, and the Standard Deviation (SD) of the number of unique domain names during the observation period.

1. MalwareDomainList (MDL) [1].
   This source covers multiple categories of malicious activities and provides information about the blacklisting and removal procedures in its forum. Based on several posts and replies, removing a domain name from the blacklist took about one hour. The data used from this blacklist were captured since July 8,

2016 until February 12, 2019. On average, MDL blacklisted around 900 unique domain names each day. The statistics of the data captured from MDL are as follows.

- Minimum: 72, Maximum: 994.
- Q1: 881, Median: 900, Q3: 909.
- Average: 908.61, Variance: 2,354.30, Standard Deviation: 48.52.

Based on the statistics above, on average, MDL published almost 1,000 unique blacklisted domain names each days. In addition, based on the variance and the standard deviation, the number of domain names blacklisted by MDL day-to-day was quite constant, although the number of blacklisted domain names reached its minimum at just 72 domains a day.

2. Joewein [2].
Joewein is the only source used in this research that specifically contains domain names related with mail spamming. The domain blacklisting service provided by Joewein is also used by SURBL [16] and PhishTank [17], as mentioned in their website. The data from Joewein were taken from July 8, 2016 until February 12, 2019. Each day, Joewein approximately released 1,200 unique domain names, which can be seen through the statistics below.

   - Minimum: 396, Maximum: 5,666.
   - Q1: 770, Median: 1,040, Q3: 1,532.
   - Average: 1,289.13, Variance: 505,999.95, Standard Deviation: 711.34.

   The stats shows that Joewein published more than 1,000 unique domain names each day. However, during the measurement period, the number of unique domain names fluctuated quite frequently, as indicated by the variance and SD values. The stats shows that Joewein could publish just 396 unique domain names, or more than 5,000 unique domain names, on a single day.

3. Malc0de [18].
Malc0de is one of the popular DBL among researchers, where more than 60 papers have been published using Malc0de's data [19], although there has been no clear explanation about their blacklist and removal procedures in their website. For instance, one of the papers using Malc0de's blacklist is Paint It Black [10], which is also used in this study. Around 100 unique domain names were blacklisted each day, since July 8, 2016 until February 12, 2019. The statistics of Malc0de can be seen as follows.

- Minimum: 5, Maximum: 333.
- Q1: 30, Median: 63, Q3: 148.
- Average: 91.52, Variance: 6,709.08, Standard Deviation: 81.91.

Malc0de is shown to be one of the smaller DBLs, when taking the number of unique domain names as consideration. On average, less than 100 unique domain names were published daily from their blacklist. Considering this average value, the variance and SD computation results show that Malc0de is also one of the DBLs that fluctuate frequently. Malc0de has been spotted publishing only 5, up to 333, unique domain names on a single day.

4. ZeusTracker (ZTracker) [20].
   ZeusTracker is one of the sub-projects conducted by Abuse.ch [21], which focuses on domain names related with malware spreading of Zeus family, although the database also contains Ice IX, Citadel, and KINS malware family. Submitted domain names for blacklisting and removal are taken for verification before published into the daily updates. The statistics of this source, taken between July 8, 2016 and February 12, 2019, are as follows.

   - Minimum: 335, Maximum: 430.
   - Q1: 339, Median: 355, Q3: 382.
   - Average: 363.50, Variance: 786.36, Standard Deviation: 28.04.

   It is visible from the stats above that ZeusTracker is one of the smaller DBLs that blacklisted quite constantly. On average, the number of unique domain names blacklisted is approximately 363, and the maximum is 430 and the minimum is 335. The variance and SD values indicates that the number of blacklisted domain names did not change a lot during the measurement period.

5. RansomwareTracker (RWTracker) [22].
   RansomwareTracker is also one of the sub-projects under Abuse.ch [21]. This service focuses on domain names that are used for distributing ransomware or used as botnets' command and control servers. RansomwareTracker updates their database every five minutes, which makes them one of the services that publish their database more frequent than the others. On average, RansomwareTracker published more than 1,000 unique domain names each day, based on data captured from July 8, 2016 until February 12, 2019. The statistics of the data published by RansomwareTracker can be seen as follows.

   - Minimum: 1, Maximum: 1,668.
   - Q1: 1,298, Median: 1,640, Q3: 1,664.

- Average: 1,441.51, Variance: 125,440.43, Standard Deviation: 354.18.

This sub-project of Abuse.ch is shown to contain more domain names each day than ZeusTracker. On average, more than 1,000 unique domain names were blacklisted, and the variance and SD hint that RansomwareTracker has a wide spread of the number of blacklisted domain names each day during the observation period.

6. URLHaus [23].
   URLHaus is another Abuse.ch's sub-projects, focusing on general malware distribution. Their service is also used to feed Google Safe Browsing [24], SpamHaus DBL [4], and SURBL [16]. There is a verification mechanism for putting a domain name into their blacklist. However, the removal procedures are not mentioned in their website. Based on the captured data between December 31, 2018 until February 12, 2019, the stats of URLHaus are as follows.

   - Minimum: 29,627, Maximum: 36,733.
   - Q1: 29,897, Median: 31,589, Q3: 34,656.
   - Average: 32,344.80, Variance: 5,889,007.80,
     Standard Deviation: 2,426.73.

   URLHaus is one of the DBLs used in this research that publish relatively large number of unique domain names. On average, URLHaus blacklisted more than 32 thousands of unique domain names. In addition, the maximum and minimum number of unique domain names spotted during the measurement period lies approximately 10% around the average value.

7. HostFile [25], both "partial" and "full" file update.
   Initially, HostFile split their database into "partial", smaller in size but more frequent updates, and "full", containing relatively larger number of domain names but less frequent updates, file update. However, the full file update was deprecated at 2018 and all updates of malicious domain names are contained into the partial file update. However, the data were still crawled until the end of this study. The blacklisting verification process is based on "hpHosts Inclusion Policy" as mentioned in their website. The statistics of both HostFile data are as follows.

   (a) Full `hphosts` file update: October 1, 2016 to February 12, 2019.

   - Minimum: 6,715, Maximum: 396,310.
   - Q1: 239,202, Median: 248,873, Q3: 277,365.

- Average: 254,013.52, Variance: 2,830,539,702.24, Standard Deviation: 53,202.82.

(b) Partial `hostfile` file update: July 8, 2016 until February 12, 2019.

- Minimum: 100, Maximum: 166,566.
- Q1: 10,254, Median: 20,636, Q3: 60,540.
- Average: 44,009.38, Variance: 2,238,362,446.85, Standard Deviation: 47,311.34.

Based variance and SD from the stats above, it is visible that the number of unique domain names for both file updates of HostFile fluctuate a lot during the measurement period. In general, the full file update contains more unique domain names compared to the partial file update. This seems logical, as the domain names found at partial file update *feeds* the full file update.

8. ThreatExpert [26].
ThreatExpert, also known as the Internet Storm Center (ISC), crawled data from many sources, as explained in their website, such as MalwareDomainList, DNSBH (MalwareDomains), RansomwareTracker, and ZeusTracker. The published data are categorized into low (more false positives), medium, and high (least false positives) sensitivity level. Since October 1, 2016 to December 30, 2018, the data from ThreatExpert were crawled from one of its services, Network Security [27], and the stats are as follows.

- Minimum: 230, Maximum: 282.
- Q1: 243, Median: 247, Q3: 254.
- Average: 250.56, Variance: 107.79, Standard Deviation: 10.38.

During the measurement period, ThreatExpert is shown to be one of the smaller DBLs. The maximum number of unique domain names published by ThreatExpert did not reach 300, and the average number of blacklisted domain names was around 250, relatively lower than some other sources used in this research.

9. OpenPhish [28].
OpenPhish is one of the widely used blacklist for domain names related with phishing. OpenPhish offer different subscription plans with different blacklist update frequency. The free 'Community' version is updated once per hour, while 'Premium' plan is updated once every five minutes. Based on the captured data from October 28, 2017 to February 12, 2019, the preliminary statistics are as follows.

- Minimum: 503, Maximum: 5,181.
- Q1: 792, Median: 962, Q3: 1,853.
- Average: 1,456.95, Variance: 972,460.49, Standard Deviation: 986.13.

The stats above shows that OpenPhish could be categorized as the medium-sized DBLs used in this study. In addition, the variance and SD values hint that OpenPhish tends to publish different number of unique domain names each day.

10. CyberCrimeTracker (CCTracker) [29].
CyberCrimeTracker contains domain names that are related with malware distribution and its command and control servers, such as Zeus family, Pony, Lokibot, `etc`. This source has maintained its database since August 2012. In this research, the data from CyberCrimeTracker were collected since December 31, 2018 until February 12, 2019 and the stats are:

    - Minimum: 9,922, Maximum: 9,974.
    - Q1: 9,939, Median: 9,945, Q3: 9,962.
    - Average: 9,946.66, Variance: 210.91, Standard Deviation: 14.52.

    On average, CyberCrimeTracker blacklisted less than 10 thousands unique domain names per day, as shown by the statistics above. However, although CCTracker blacklisted a lot of domain names per day, the owner seldom publish more, or less, than the average value computed above. This is visible from the small values of the variance and SD.

11. DNSBH [30].
DNSBH, also referred as MalwareDomains, lists domain names that are used for propagating malwares and spywares, and some of these domain names are also found in other sources, such as VirusTotal [8], OpenPhish [28] or PhishTank [17]. The stats of this source based on data taken from December 31, 2018 to February 12, 2019 are as follows.

    - Minimum: 23,032, Maximum: 23,054.
    - Q1: 23,038, Median: 23,042, Q3: 23,052.
    - Average: 23,042.91, Variance: 61.76, Standard Deviation: 7.86.

    DNSBH is quite similar to CCTracker, in terms of the number of blacklisted domain names and its maintenance. In fact, the average number of unique domain names blacklisted each day was more than the double of CCTracker. The variance and SD of DNSBH also shows that during the measurement period, DNSBH released quite a constant number of blacklisted domain names.

12. VXVault [31].
    VXVault contains domain names that are used for distributing malicious applications since 2006. Furthermore, this source does also publish a list of URLs containing downloadable malwares. In this research, the daily updates of VXVault have been captured since December 31, 2018 to February 12, 2019, and the stats are as follows.

    - Minimum: 42, Maximum: 95.
    - Q1: 61, Median: 78, Q3: 86.
    - Average: 74.07, Variance: 244.97, Standard Deviation: 15.65.

    VXVault is one of the smallest DBLs used in this research. The number of blacklisted domain names has never reached 100 during the measurement period. However, the number of blacklisted domain names could vary moderately.

13. OSINT Feeds from Bambenek Consulting (C2dom) [32].
    The OSINT Feeds contains only domain names that are used as command and control servers of numerous malware families, for instance, Mirai, CryptoLocker, Kraken, *etc*. Statistics of the data captured from December 31, 2018 until February 12, 2019 are as follows.

    - Minimum: 677, Maximum: 1,851.
    - Q1: 704, Median: 715, Q3: 729.
    - Average: 787.91, Variance: 61,612.26, Standard Deviation: 248.22.

    The statistics above display that C2dom is one of the actively updated DBL, as shown by the variance and the SD results above. In addition, based on the observation results, it is shown that the fluctuations could range from 677 up to 1,851 unique domain names per day.

The thirteen unique DBLs used in this research can be grouped into three major maliciousness categories, namely Phishing, Malware (that also includes ransomware, botnets and command-and-control servers), and Mail Spam. This can be seen at Figure 3.1, which is a Venn Diagram of the categories of DBLs. As can be seen from the figure, only Hostfile and MalwareDomainList intersects with all three categories. All other sources tend to focus on a single malicious category.

Among all domain names captured from 13 sources, there were 108 unique domain names in 4,344 occurrences that need to be discarded, since these domain names were either invalid or not parse-able into ASCII characters. This makes

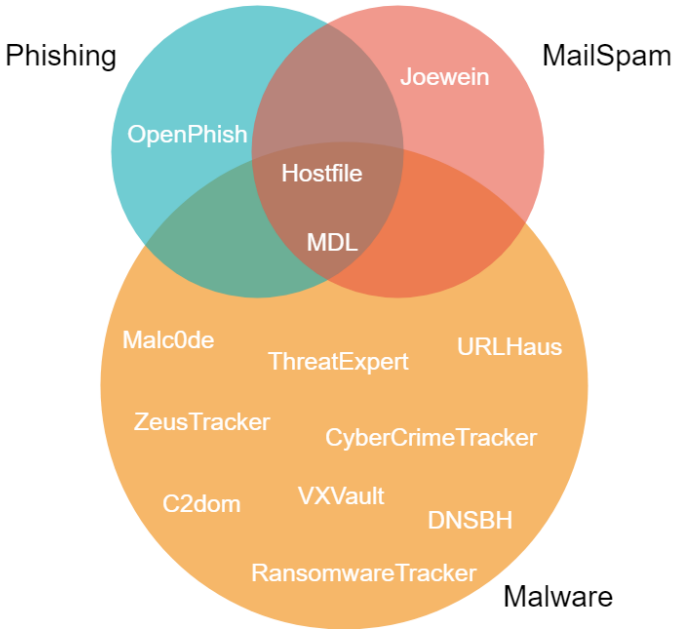

Figure 3.1: Malicious Categories of DBLs

these domains can not be used for further analysis. Considering that these "special" domain names contribute to close to zero percent of the total domain names, discarding these domain names will not make significant difference to the results of this research.

Besides the main data from the 13 DBLs, Alexa's top 1M websites were also crawled daily. This measurement period started from July 8, 2016 until February 12, 2019. This list of popular websites is used to measure some of the metrics, like the "Purity", which will be described later in the next chapter.

In addition, another data set that is essential in this research is the WHOIS database. This database contains the blacklisted domain names, their registrars, registrant, registration date, expiration date, and record update date. WHOIS information of domain names were captured using `pywhois` Python library [33] since July 22, 2018 until February 12, 2019. However, the original data set also needs to be filtered because there are several domain names with no registrars or ambiguous registrar content, such as "No registrar" or "root SA". Out of 4,031 distinct registrars contained in this database, 4,026 valid registrars could be used further in this research, while 5 others need to be discarded from this research.

Besides the registrars, the WHOIS database contains more invalid registrant. Out of 42,016 unique registrants, around 20 unique registrants can not be used because they contain "no registrant", "***", "-", ".", or some other invalid combinations.

3.2 Considerations on Existing Studies

Based on the aforementioned studies, some methodologies for comparing lists could be applied for multiple sectors. For instance, finding the intersections and disjoints between different lists could be applied to analyze domain and IP-based blacklists, telephony abuse data sets, or ranked lists.

Unfortunately, the definition of the ground truths itself remains an open issue. Different studies have different interpretation of ground truths, which makes one useful approach in one paper unusable for the other studies. One of the examples can be observed at a research conducted by Kührer *et al.* [10], where the delay between the first appearance of malicious domain names and the appearance of these domain names in the blacklists could be calculated. On the other hand, this measurement was difficult to be conducted for Gupta *et al.* [12], since the delay could vary indefinitely.

As also mentioned in Pitsillidis *et al.*'s paper [9], obtaining the ground truths for domain blacklists is not a simple task. In addition, based on prior knowledge on behaviors of domain blacklists, most of them are maintained manually by their members and administrators. This makes the creation of accurate *all-in-one* collection of malicious activities and reports almost impossible. Therefore, what can be done is to compare different lists using existing, combined, or modified metrics to provide insights on each blacklists' characteristics.

In this research, firstly, to analyze the characteristics of DBLs, comparisons are conducted only using the second-level fully qualified domain names. This excludes domain names that consists of special characters, as mentioned in the previous chapter.

Secondly, the ground truth naïvely can be created by combining all blacklists and remove the duplicates. One of the better approach has been introduced in "Blacklists Assemble" [14], by creating a better aggregation result. Unfortunately, not all of the techniques introduced in this research could be fully used since not every domain blacklists include their de-listing history. This makes this measure difficult to be used in this research. For example, one of these unusable metrics is the calculation of the addresses' history of offense.

To check the non-maliciousness of domain names, usage of only Google Safe Browsing API is also mentioned as one of the limitations in [14]. Therefore, usage of some other services, such as VirusTotal web checker [8] or Comodo Web Inspector [34], can be used. However, scanning websites only using VirusTotal API is considered to be enough, because they also do cross-check the domain names with Google Safe Browsing, Comodo Web Inspector, and many other online scanning services.

In this study, VirusTotal public API is used, although the private API key could cut quite a lot of time, since their private API key is used for specific purposes only. The public API key is used to verify the ratio of malicious domain names found in Alexa top websites.

Using different *URL scanners* have also been considered to scan the maliciousness of blacklisted domain names. However, each scanner has different blacklist categories and procedures. For instance, the type of threats that Google Safe Browsing API publish are `UNSPECIFIED`, `MALWARE`, `SOCIAL_ENGINEERING`, `UNWANTED_SOFTWARE`, or `HARMFUL_APPLICATION`. This categorization might not be the case for other scanner services, for instance, BitDefender URL scanner. In addition, scanning hundreds of thousands of domain names per day might instead require more resources and take longer time than using VirusTotal's service.

Another consideration made in this research is, the list of benign domain names can be created by scanning each of the domain names with the VirusTotal API mentioned in the previous point. While Ramanthan *et al.* created their ground truth from Alexa Top 500K websites, only around 60% of them were actually benign. This was because some domain names that were used for any kind of malicious activities were also accessed quite frequently. It is also expected that using Alexa top 1M websites list, the number of malicious domain will be increased. Therefore, in this research, the list of benign domain names will contain only Alexa top 100K "cleaner" websites.

3.3 Selected Metrics

Based on the existing studies and aforementioned considerations, the final metrics and approaches that are usable and will be implemented in this research are discussed in this section.

3.3.1 Purity

This metric estimates how much of a domain blacklist is actually malicious. This metric is combined with Specificity in Subsection 2.3.7, when dealing with list of benign domain names. The approaches that can be done to measure Purity are:

1. Finding the proportion of unique domain names in a feed that were registered based on several major top-level domains.

2. Calculating the ratio of unique domain names appearing in Alexa top website list.

3.3.2 Coverage

This metric measures the ratio of actual malicious contents that are blacklisted. This metric can be combined with Recall in Subsection 2.3.6 and has similar purpose with Completeness in Subsection 2.3.9. The approaches are:

1. Finding the ratio of unique domain names that appear only on a single feed and not in other feeds (exclusive domains).

2. Finding the ratio of unique domain names that appear on multiple feeds (pair-wise comparison).

3. Finding the ratio of unique domain names that also appear in the aggregated blacklist.

3.3.3 Timing

This metric estimates the duration of a malicious campaign. This metric provides insight about the distribution of the duration of a domain name appearing in different blacklists. To measure the timing, the following approaches are defined.

1. Finding the relative first appearance time of a domain name in domain blacklists.

2. Finding the relative last appearance time of a domain name in domain blacklists.

3. Computing the duration of the campaign by subtracting the last appearance time by the first appearance time.

4. Comparing the relative first and last appearance time of a domain name from multiple blacklists.

3.3.4 Responsiveness

Based on the relative campaign duration from the previous metric, the responsiveness of a DBL can be estimated. This metric could indicate which DBL is more responsive, or have tendency to be late, than other DBLs. The methods to measure the Responsiveness are:

1. Determining the relative campaigns' disappearance duration to be set as the campaign threshold.

2. Computing the difference of the campaign start date (from multiple DBLs) with the first appearance date of a domain name in a DBL.

3. Computing the difference of the campaign end date (from multiple DBLs) with the last appearance date of a domain name in a DBL.

4. Estimating the DBLs' tendency, whether they are likely to blacklist a domain name earlier, or later, than other blacklists.

5. Estimating the DBLs' tendency, whether they are likely to remove a domain name earlier, or later, than other blacklists.

3.3.5 Specificity

This metric calculates the ratio of benign domain names that are not blacklisted. Specificity can be estimated by:

1. Finding the ratio of unique benign domains from Alexa top 100k website list that are not blacklisted.

3.3.6 Accuracy

This metric determines how detailed the information of a domain is in a blacklist. Based on the observation of several sources, the contents of the blacklists are quite different. A more detail and complete information about a blacklisted domain could indicate that the malicious behaviors did actually happen. The approach to determine the accuracy is:

1. Finding the details of the blacklisted domain names, such as domain name, IP Address, report date, country, registered name servers, type of malicious behavior, *etc* from each DBL's web or forum pages.

3.3.7 Agility

This metric is similar to Subsection 2.3.11. The stability of a domain blacklist is measured by counting how many new malicious domains are captured as well as how many disappears from the blacklist each day. Agility can be measured by:

1. Finding the number of new malicious domain names that appear in each blacklist.

2. Measuring the number of domain names that disappear from domain blacklists.

3. Visualizing the fluctuations in a graph.

3.3.8 Liveliness

This metric measures how much of blacklisted domain names do actually exist and *active* when they appear in a blacklist. This metric is a new branch from approach 2 of Purity at Subsection 2.3.1, with deeper investigation into the blacklisted domain names. Liveliness of a DBL can be estimated by the following approaches.

1. Finding the existence of a domain name by checking it with DNS resolver using `nslookup` or `dig` command, or Python's `pywhois` library.
2. Finding the liveliness of a domain name by:
   - Pinging the server.
   - Checking HTTP response code from port 80 and HTTPS response code from port 443.
   - Checking the status of selected ports, in which in this case the ports selected are port 20, 21, 22, 23, 25, and 53.
3. Visualizing the liveliness of a DBL based on the number of blacklisted domain names and *live* machines each day during the observation period.

Chapter 4

Blacklists Analysis

This section contains the processes and the results of the aforementioned approaches to answer the Research Question defined at Section 1.3.

4.1 Current Situation of DBLs

The summarized descriptions of publicly available domain blacklists that are used in this research can be seen at Table 4.1 to 4.3. The columns' title of the three tables are described as follows.

- `Domain Blacklist` (`#` and `Name`). This contains the number and name of the DBL.

- `Category`. This column indicates under which category the domain blacklist belongs to, whether the malicious activities can be categorized as phishing (`P`), malware distribution (`MW`), or mail spamming (`MS`).

- `Date`. Sub-column `From` of this header shows the very first appearance of malicious behavior in a blacklist, while sub-column `To` contains the latest addition to a blacklist. `Last Checked` indicates the last inspection time to the respective domain blacklist.
  In some sources, such as VXVault, their database also contains ambiguous date, such as "`0000-00-00`". To some extent, this hints that the database might contain inaccurate contents.

- `Update Frequency`. This field contains how often domain blacklist sources update their lists. For instance, some sources directly update their blacklists as soon as new reports are received or validated, while others update their database once per day.

- `Blacklisting` procedures. This column shows how different sources provide ways for members to submit malicious domains. Some sources also provide information about how the submitted domains are checked and validated before being published to their blacklist.

- `Removal` (de-listing) procedures. This column is similar to blacklisting procedures, but for de-listing process.

- `Notes`. This column shows additional information about each blacklist sources.

As can be seen from Table 4.1 to 4.3, different domain blacklist has different characteristics. For example, `MalwareDomainList` includes multiple types of malicious behaviors, such as phishing/fraud, trojan distribution, fake antivirus, backdoor, *etc*. Their blacklist is maintained based on members' reports that can be delivered using online form, forum post and messages, or personal messages to the administrator. Then, the list is updated when a new submission is manually verified. On the other hand, `OpenPhish` only contains domain names that are associated with phishing activities. Their blacklist is updated once every hour, or 5 minutes for premium users, and the submitted domain names are verified automatically.

4.2 DBLs Start and End Date

In this section, the start and the end date of capturing data from each DBL are documented, summarized at Figure 4.1. The earliest date of capturing several DBLs data is August 7, 2016 and the latest collection date is February 12, 2019. To provide a fair and complete understanding of how publicly available DBLs used are maintained, when comparing DBLs against each other, the start date used is the maximum of the compared start date, while the selected end date is the minimum of the compared end date. This ensures that the comparison is conducted with the existing data from both DBLs.

4.3 Statistics and Analysis

This section shows the approaches and the results of executing the metrics defined in the previous chapter.

In general, the coloring schemes used in the analysis can be seen at Table 4.4. The percentages are split into five categories, and the coloring scheme is just meant to aid visualizing the tables and to distinguish the groupings.

Table 4.1: Complete Domain Blacklists Analysis.

Domain Blacklist		Category	Date			Update Frequency	Procedures				Notes
							Blacklisting		Removal		
#	Name		From	To	Last Checked		Method	Verification?	Method	Verification?	
1	MDL [1]	P, MS, MW	22-Mar-2009	21-Feb-2019	31-Jul-2019	Unknown, Submission-based.	Online form, forum messages and posts, personal messages, email.	Yes. Manually by JohnC.	Online form, forum messages and posts, personal messages, email.	Yes. Manually.	De-listed domain names are archived [35]. Blacklisting or de-listing processes take around 1 hour. Related servces: MalZilla [36].
2	Joewein [2]	MS	1-Jan-2015	29-Jul-2019	31-Jul-2019	Unknown. Hourly is recommended.	Unknown.	Yes. Structured, manually.	Email.	Yes. Manually.	Updated archive is stored. False positive rate is less than one per month. Both verification are done automatically and manually. Related services: SURBL [16], PhishTank [17].
3	Malc0de [18]	MW	30-Nov-2018	23-May-2019	31-Jul-2019	Daily.	Unknown.	Unknown.	Unknown.	Unknown.	Blog has been inactive since 2010. Public list only contain malicious domains from the last 30 days.
4	ZTracker [20]	MW (Zeus family)	12-May-2011	08-Jul-2019	31-Jul-2019	Unknown.	Online form.	Yes. Unknown.	Email.	Yes. Unknown.	Sub-project of Abuse.ch [21]. More complete report is also available [37]. Discontinued at 08-07-2019

Table 4.2: Complete Domain Blacklists Analysis (2).

Domain Blacklist		Category	Date			Update Frequency	Procedures				Notes
							Blacklisting		Removal		
#	Name		From	To	Last Checked		Method	Verification?	Method	Verification?	
5	RWTracker [22]	MW (Ransomware)	2-Mar-2015	31-Jul-2019	31-Jul-2019	Every 5 minutes.	Email.	Unknown.	Email.	Unknown.	Sub-project of Abuse.ch [21]. Related services: CryptoWall Tracker [38].
6	URLHaus [23]	MW	5-Mar-2018	31-Jul-2019	31-Jul-2019	Submission-based.	API / web interface.	Yes. Unknown.	Unknown.	Unknown.	Malicious links can be submitted through API or web form. Related services: Google Safe Browsing [24], Spamhaus DBL [4], SURBL [16].
7	HostFile [25]	P, MW	17-Jun-2018	31-Jul-2019	31-Jul-2019	Submission-based.	Email.	Yes. Manually.	Email.	Yes. Manually.	Full update has been deprecated. De-listed domains are stored [39]. Related services: MalwareBytes [40].
8	ThreatExpert [26]	MW	1-Jan-2000	30-Dec-2018	31-Jul-2019	Daily.	Online Form.	Yes. Manually.	Online Form.	Yes. Manually.	Taken from Network Security Suspicious Domain List [27]. Weighted lists based on tracking and malware lists from different sources (MalwareDomainList [1], Abuse.ch [21], VirusTotal [8]).

Table 4.3: Complete Domain Blacklists Analysis (3).

Domain Blacklist		Category	Date			Update Frequency	Procedures				Notes
							Blacklisting		Removal		
#	Name		From	To	Last Checked		Method	Verification?	Method	Verification?	
9	OpenPhish [28]	P (and data breaches).	Unknown	Unknown	31-Jul-2019	Community: Once per hour. Premium: Once in 5 minutes.	Email and automatic engine.	Yes. Automatically.	No.		Fully automated (and via email) self-contained platform for phishing intelligence. Free version only contains limited number of URLs.
10	CCTracker [29]	MW	19-Jul-2012	31-Jul-2019	31-Jul-2019	Submission-based.	Online form.	Yes. Unknown.	Unknown.	Unknown.	This blacklist also contains real malware samples.
11	VXVault [31]	MW	24-Jun-2006	31-Jul-2019	31-Jul-2019	Submission-based	Unknown.	Unknown.	Unknown.	Unknown.	Also contains active and blacklisted malware database.
12	Malware Domains (DNSBH) [30]	MW	24-Jun-2014	22-Jul-2019	31-Jul-2019	Daily	Unknown	Unknown	Email.	Yes.	Blacklisting method most likely by scraping other blacklists (PhishTank [17], OpenPhish [28], ThreatExpert [26], *etc.*). Blacklist removal will be responded within 24 hours. Complete domain files are archived [41].
13	C2dom [32]	MW	Unknown	31-Jul-2019	31-Jul-2019	Daily	Unknown	Unknown	Unknown	Unknown	Master list is also available [42].

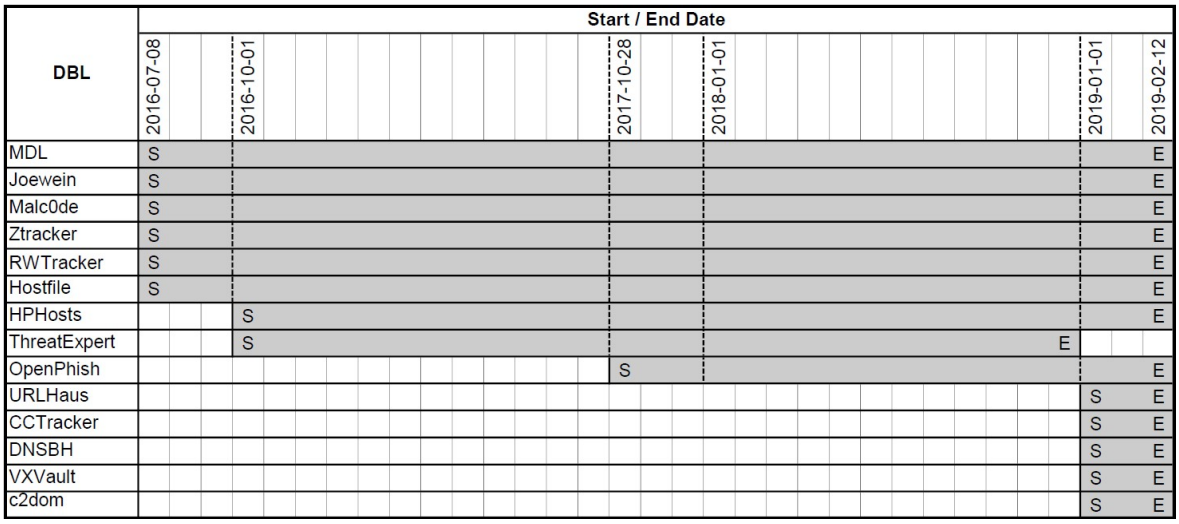

Figure 4.1: Start and End Date of DBLs Observations

Table 4.4: Coloring Categories

Quality	Ratio
Very High	80% - 100%
High	60% - 80%
Medium	40% - 60%
Low	20% - 40%
Very Low	0% - 20%

4.3.1 Purity

As defined in Chapter 3.3.1, there are two approaches to measure the purity of a domain blacklist. The first approach is by checking the existence of the blacklisted domain names by looking for their respective registrars and registrants using the WHOIS database. The second way is by finding the number of blacklisted domain names that also appear in Alexa top website list. The results of conducting these two approaches are discussed in this section.

Registered Domains

In this research, investigating the existence of a blacklisted domain name was conducted by considering the registrar and registrant of that respective domain names. In a valid domain name registration, a registrar is an entity that receives the request, from a registrant, for a domain name. After receiving all of the required information and validating them, the registrar will then process the request and if the request is processed smoothly, the registry zone files will eventually be updated and the domain name is ready to be used. Therefore, an existing domain name should also

have a registrar or a registrant associated with it.

After filtering out the WHOIS database, the result of finding the number of registered domains can be found at Table 4.5. The measurement period started from July 22, 2018 until February 12, 2019. From Table 4.5, category `entries` indicates the total number of domains captured from each DBL, whilst `unique` calculates the number of unique domains, by removing duplicates, from each DBL. In both of these categories, column `valid` indicates the number of domain names that have valid registrants and/or registrars in the database. Then, the `Ratio` is computed by dividing the number of valid domains by the total number of domain names. The `Rank` is sorted by the ratio of unique domain names in a descending order, since a higher number of registered domain names might indicate that the DBL captures real malicious campaigns.

A domain name is considered to be valid if:

1. Its registrar or registrant are not `null` or empty string,
2. Its registrar does not contain “no registrar” or ambiguous strings, like “root sa”,
3. Its registrant does not contain “no registrant”, or ambiguous strings, like sequences of “***”, “-”, or “.”,
4. The domain name is created before, or at least on the same day of, the appearance in a DBL, and
5. The domain name’s expiry date is after the date of appearance in a DBL.

Based on the results provided at Table 4.5, only Malcode and Joewein contain a relatively high ratio of valid domain names. In general, the ratio of valid blacklisted domain names lies around 40-60%. On the other hand, it is difficult to estimate how much blacklisted domain names were registered for ThreatExpert, since the data captured during this period contained no domain names from ThreatExpert. There is an indication that ThreatExpert changed their system or output structure at some point before July 22, 2018.

Intersection with Alexa Top 100k Websites

Another approach that can be used to determine the purity of a DBL is by cross-checking the blacklisted domain names with list of benign websites. In this research, this list of benign websites used is Alexa’s top 100k websites. This list is considered to be “benign” enough that almost all of domain names contained in this list are benign domains. Therefore, a higher number of domain names found in both DBLs data and Alexa’s list indicates that the DBL might also blacklist more benign domain

Table 4.5: Ratio of Registered Domains

DBL	Entries			Unique			Rank
	Total	Valid	%	Total	Valid	%	
VXVault	3,259	1,780	54.62	570	325	57.02	3
Hostfile	24,678,662	12,417,777	50.32	177,628	88,086	49.59	4
URL-Haus	1,423,190	607,408	42.68	37,007	15,809	42.72	7
MDL	169,552	70,775	41.74	913	375	41.07	9
Threat-Expert	112	0	0.0	1	0	0.0	14
RW-Tracker	331,888	49,294	14.85	1,712	267	15.60	13
CC-Tracker	437,697	164,956	37.69	10,108	3,812	37.71	10
HPHosts	50,272,548	24,830,092	49.39	261,652	127,783	48.84	5
DNSBH	1,013,932	419,585	41.38	23,342	9,663	41.40	8
C2dom	34,668	13,886	40.05	3,336	533	15.98	12
Open-Phish	166,602	80,037	48.04	27,049	12,988	48.02	6
Malc0de	10,147	6,666	65.70	368	238	64.67	1
Joewein	201,323	125,025	62.10	19,946	12,417	62.25	2
ZTracker	69,057	23,636	34.23	362	125	34.53	11

names.

The results of finding the intersection with Alexa top 100k websites can be seen at Table 4.6. In this table, column `Domain Names` shows the number of unique domain names found in each DBL's list. `Alexa found` indicates the number of domain names that are also found at Alexa list. `Ratio` shows the ratio of domain names that intersect with Alexa list, and the `Rank` indicates the order of the DBLs based on the ratio, in an increasing order. In this measurement, the higher ratio hints that the DBL has quite a large number of domain names intersected with benign domains. Therefore, it is desired that a DBL should have this *false positives* as low as possible.

As can be seen in Table 4.6, all DBLs have relatively low number of domain names intersected with Alexa Top 100k websites. This indicates that in general, public DBLs performed quite well, in terms of the false positives, based on this metric applied to the data set. Joewein, RansomwareTracker, and C2dom had the lowest ratio of intersection with Alexa top 100k websites. Only Malcode contained the highest ratio, almost 10%, of their blacklisted domain names that were also found in

Table 4.6: Intersection of Blacklisted Domain Names with Alexa Top 100k Websites

DBL	Domain Names	Alexa Found	Ratio	Rank
VXVault	3,259	59	1.81%	10
Hostfile	38,091,901	230,558	0.61%	6
URLHaus	1,423,171	13,174	0.93%	7
MDL	752,559	27,351	3.63%	13
ThreatExpert	88,327	1,751	1.98%	11
RWTracker	1,200,821	825	0.07%	2
CCTracker	437,653	2,099	0.48%	5
HPHosts	189,020,588	2,385,888	1.26%	8
DNSBH	1,013,888	4,267	0.42%	4
C2dom	34,668	131	0.38%	3
OpenPhish	558,160	7,332	1.31%	9
Malc0de	76,637	7,508	9.80%	14
Joewein	1,075,949	309	0.03%	1
ZTracker	304,051	9,899	3.25%	12

Alexa Top 100k websites.

4.3.2 Coverage

Coverage measures the ratio of domain names that are used for malicious activities that are blacklisted. This ratio can be determined by finding the exclusive domains and conducting pairwise comparison with every DBLs and the aggregated blacklist.

Exclusive Domains

The results of extracting exclusive domains from all DBLs used in this research can be seen at Table 4.7. Finding the domain exclusiveness of a DBL was compared against all data from all DBLs from July 8, 2016 until February 12, 2019. The domain exclusiveness of a DBL indicates how much contribution a DBL give to the overall malicious activities.

Table 4.7 shows the exclusiveness of each DBL used in this research. Column `Exclusive` shows the number of exclusive domain names, out of the `Total` number of blacklisted domain names. The `Ratio` is computed by taking the ratio of the exclusive domains compared to the total number of unique domain names. Then, the `Rank` is ordered by the ratio of exclusive domains in a decreasing order.

Intuitively, it is desired for a DBL to have a high number of exclusive domains.

This indicates that the DBL might contain many domain names that are not captured by other DBLs. Thus, analyzing DBLs with a high number of exclusive domains might uncover more useful information that cannot be predicted from other sources. However, this number alone could also indicate that the source might have a bias towards some directions. Therefore, the exclusiveness needs other metrics to provide more useful information about the maintenance of each DBL.

In general, as can be seen from Table 4.7, Joewein and RansomwareTracker contributed a considerable ratio of exclusive domain names. When taking the number of exclusive domains, RansomwareTracker was not one of the largest since the start of the measurement for this DBL was a bit later than other sources. On the other hand, VXVault had the least percentage of exclusive domains since most of its blacklisted domains were also found at other DBLs.

Taking deeper investigation into each malicious categories, the result of finding the exclusive domains of MDLs related with phishing activities can be found at Table 4.8. It is more visible that MalwareDomainList had the highest ratio of exclusive domains in the phishing category. However, by considering the total number of exclusive domain names, MalwareDomainList was the lowest, compared to other sources containing phishing activities.

Table 4.7: DBL Exclusiveness (Overall)

Source	Exclusive	Total	Ratio	Rank
VXVault	14	570	2.46%	14
Hostfile	142,795	522,481	27.33%	11
URLHaus	13,068	36,746	35.56%	7
MDL	629	1,028	61.19%	5
Joewein	113,864	117,289	97.08%	1
RWTracker	1,377	1,684	81.77%	2
CCTracker	6,403	9,974	64.20%	4
HPHosts	116,482	452,947	25,72%	13
ZTracker	223	640	34.84%	8
c2dom	2,509	3,332	75.30%	3
OpenPhish	31,504	71,691	43.94%	6
DNSBH	6,901	23,083	29.90%	10
ThreatExpert	79	292	27.05%	12
Malc0de	706	2,249	31.39%	9

Table 4.9 shows the DBL exclusiveness of domain names that are used for spamming campaigns. Based on this table, it is shown that Joewein had the highest ratio of exclusive domains. One of the reason for this is, since the other DBLs do not only

contain domain names related with spamming, the ratio of domain names used for multiple malicious campaigns in different DBL is also higher.

Exclusiveness of DBLs containing domains that are related with distributing malwares can be seen at Table 4.10. This table shows a similar trend as the overall investigation, as can be seen at Table 4.7. RansomwareTracker produced the highest ratio of exclusive domain names, and Hostfile showed the largest number of exclusive domain names.

Table 4.8: DBL Exclusiveness (Phishing)

Source	Exclusive	Total	Ratio	Rank
Hostfile	169,488	522,481	32.44%	3
MDL	639	1,028	62.16%	1
HPHosts	117,512	452,947	25,94%	4
OpenPhish	35,244	71,691	49.16%	2

Table 4.9: DBL Exclusiveness (Mail Spam)

Source	Total	Exclusive	Ratio	Rank
MDL	639	1,028	62.16%	2
Joewein	113,966	117,289	97.17%	1
Hostfile	186,697	522,481	35.73%	3
HPHosts	117,874	452,947	26.02%	4

Pairwise Comparison

Conducting pairwise comparison required relatively more details than measuring other metrics, since the start and end date of the compared DBL might be different. Investigating the number of domain names that exist on two DBLs must be conducted in the same time span to provide a complete information about the pairwise comparison. The results of comparing the content of DBLs can be seen at Table 4.11 and Table 4.12.

In these tables, the pairwise comparisons were conducted by taking the DBLs on the Y-axis as the pivot against the DBLs on the X-axis. Each cell in these tables contains the ratio of domain names that exist on both compared DBL, start date and end date of each comparison. In addition, the coloring scheme is the same as defined in Table 4.4.

For instance, notation $403/570\ (70.70\%)$ in the first row second column of Table 4.11 indicates that 403 out of 570 entries from VXVault were also found at Hostfile. 70.70% is computed by dividing 403 by 570. '2019-01-01' indicates the start date

Table 4.10: DBL Exclusiveness (Malware Distribution)

Source	Total	Exclusive	Ratio	Rank
VXVault	14	570	2.46%	12
Hostfile	159,137	522,481	30.46%	9
URLHaus	13,113	36,746	35.69%	6
MDL	629	1,028	61.19%	4
RWTracker	1,377	1,684	81.77%	1
CCTracker	6,414	9,974	64.31%	3
HPHosts	116,842	452,947	25.80%	11
C2dom	2,509	3,332	75.30%	2
ZTracker	224	640	35.00%	7
DNSBH	10,562	23,083	45.76%	5
ThreatExpert	79	292	27.05%	10
Malc0de	706	2,249	31.39%	8

(January 1, 2019) and '2019-02-12' indicates the end date (February 12, 2019) of the pairwise comparison measurement. This applies to all table contents.

As can be seen at Table 4.11 and Table 4.12, the majority of DBLs had relatively low intersection against each other. However, it is also visible that domain names blacklisted by Hostfile (both Hostfile and HPHosts) were more frequent to be intersected with other DBLs. This can be inferred by seeing that more greens and yellows are found at column Hostfile. One of the factors that cause this was the number of unique domain names published by Hostfile that were relatively much more significant than other sources. In addition, Hostfile includes blacklisted domain names from multiple categories. These factors could also increase the probability of domain names published by "smaller" DBLs to be also found at Hostfile's daily updates.

Pairwise with Aggregated List

Finally, the coverage of a DBL can be estimated by finding how much contribution a DBL contributes to the overall malicious domain name list. Similar to the pairwise comparison for each DBL, the start and end date of observation period is critical when measuring this approach.

Executing pairwise with aggregated list can be split into two major groups, the ones before (*historical*) and after (*newer*) December 31, 2018. This distinction is considered to be necessary, because of the following reasons.

- Using July 08, 2016 as the start date and February 12, 2019 as the end

Table 4.11: DBLs Pairwise Comparison

DBL	DBL													
	VXVault	Hostfile	URLHaus	MDL	Joewein	RWTracker	CCTracker	HPHosts	ZTracker	c2dom	OpenPhish	DNSBH	ThreatExpert	Malc0de
VXVault	-	403/570 (70.70%)	510/570 (89.47%)	0/570 (0.00%)	0/570 (0.00%)	0/570 (0.00%)	7/570 (1.23%)	4/570 (0.70%)	1/570 (0.18%)	0/570 (0.00%	1/570 (0.18%)	3/570 (0.53%)		186/570 (32.63%)
		2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	-	2018-12-31
		2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12		2019-02-12
Hostfile	403/166,665 (0.24%)	-	22,047/ 166,665 (13.23%)	108/ 522,481 (0.02%)	3,307/ 522,481 (0.63%)	283/ 522,481 (0.05%)	1,081/ 166,665 (0.65%)	300,173/ 487,836 (61.53%)	245/ 522,481 (0.05%)	1/ 166,665 (0.00%)	33,239/ 304,766 (10.91%)	4,757/ 166,665 (2.85%)	49/473,464 (0.01%)	1,405/ 522,481 (0.27%)
	2018-12-31		2018-12-31	2016-07-08	2016-07-08	2016-07-08	2018-12-31	2016-10-01	2016-07-08	2018-12-31	2017-10-01	2018-12-31	2016-10-01	2016-07-08
	2019-02-12		2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-01-01	2019-02-12
URLHaus	510/36,746 (1.39%)	22,047/ 36,746 (60.00%)	-	11/36,746 (0.03%)	14/36,746 (0.04%)	0/36,746 (0.00%)	332/ 36,746 (0.91%)	1,083/ 36,746 (2.95%)	18/36,746 (0.05%)	3/36,746 (0.01%)	98/36,746 (0.27%)	767/ 36,746 (2.09%)		206/ 36,746 (0.56%)
	2018-12-31	2018-12-31		2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	-	2018-12-31
	2019-02-12	2019-02-12		2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12		2019-02-12
MDL	0/860 (0.00%)	108/1,028 (10.51%)	11/860 (1.28%)	-	3/1,028 (0.29%)	0/1,028 (0.00%)	17/860 (1.98%)	375/1,016 (36.91%)	3/1,028 (0.29%)	0/860 (0.00%)	21/904 (2.32%)	28/860 (3.26%)	2/1,016 (0.20%)	7/1,028 (0.68%)
	2018-12-31	2016-07-08	2018-12-31		2016-07-08	2016-07-08	2018-12-31	2016-10-01	2016-07-08	2018-12-31	2017-10-01	2018-12-31	2016-10-01	2016-07-08
	2019-02-12	2019-02-12	2019-02-12		2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-01-01	2019-02-12
Joewein	0/3,996 (0.00%)	3,307/ 117,289 (2.82%)	14/3,996 (0.35%)	3/117.289 (0.00%)	-	0/117,289 (0.00%)	2/3,996 (0.05%)	2,241/ 112,162 (2.00%)	7/117,289 (0.01%)	0/3,996 (0.00%)	116/51,455 (0.23%)	1/3,996 (0.03%)	1/109,105 (0.00%)	2/117,289 (0.00%)
	2018-12-31	2016-07-08	2018-12-31	2016-07-08		2016-07-08	2018-12-31	2016-10-01	2016-07-08	2018-12-31	2017-10-01	2018-12-31	2016-10-01	2016-07-08
	2019-02-12	2019-02-12	2019-02-12	2019-02-12		2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-01-01	2019-02-12
RWTracker	0/1,668 (0.00%)	283/1,684 (16.81%)	0/1,668 (0.00%)	0/1,684 (0.00%)	0/1,684 (0.00%)	-	0/1,668 (0.00%)	290/1,679 (17.27%)	3/1,028 (0.29%)	0/1,668 (0.00%)	0/1,670 (0.00%)	18/1,668 (1.08%)	1/1,679 (0.06%)	0/1,028 (0.00%)
	2018-12-31	2016-07-08	2018-12-31	2016-07-08	2016-07-08		2018-12-31	2016-10-01	2016-07-08	2018-12-31	2017-10-01	2018-12-31	2016-10-01	2016-07-08
	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12		2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-01-01	2019-02-12
CCTracker	7/9,974 (0.07%)	1,080/9,974 (10.83%)	332/9,974 (3.33%)	17/9,974 (0.17%)	2/9,974 (0.02%)	0/9,974 (0.00%)	-	1,440/ 9,974 (14.44%)	113/9,974 (1.13%)	1/9,974 (0.01%)	38/9,974 (0.38%)	253/9,974 (2.54%)		4/9,974 (0.04%)
	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31		2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	-	2018-12-31
	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12		2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12		2019-02-12

Table 4.12: DBLs Pairwise Comparison (continued).

DBL	DBL													
	VXVault	Hostfile	URLHaus	MDL	Joewein	RWTracker	CCTracker	HPHosts	ZTracker	c2dom	OpenPhish	DNSBH	ThreatExpert	Malc0de
HPHosts	4/248,873 (0.00%)	300,173/ 452,947 (66.27%)	1,083/ 248,873 (0.44%)	375/ 452,947 (0.08%)	2,241/ 452,947 (0.49%)	290/ 452,947 (0.06%)	1,440/ 248,873 (0.58%)	-	271/ 452,947 (0.06%)	271/ 248,873 (0.11%)	18,233/ 488,726 (0.04%)	5,078/ 248,873 (0.02%)	31/452,947 (0.01%)	856/ 452,947 (0.19%)
	2018-12-31	2016-10-01	2018-12-31	2016-10-01	2016-10-01	2016-10-01	2018-12-31		2016-10-01	2018-12-31	2017-10-01	2018-12-31	2016-10-01	2016-10-01
	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12		2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-01-01	2019-02-12
ZTracker	1/344 (0.29%)	245/640 (38.28%)	18/344 (5.23%)	3/640 (0.47%)	7/640 (1.09%)	1/640 (0.16%)	113/344 (32.85%)	271/587 (46.17%)	-	0/344 (0.00%)	21/390 (5.38%)	27/344 (7.85%)	31/585 (5.30%)	8/640 (1.25%)
	2018-12-31	2016-07-08	2018-12-31	2016-07-08	2016-07-08	2016-07-08	2018-12-31	2016-10-01		2018-12-31	2017-10-01	2018-12-31	2016-10-01	2016-07-08
	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12		2019-02-12	2019-02-12	2019-02-12	2019-01-01	2019-02-12
C2dom	0/3,332 (0.00%)	554/3,332 (16.62%)	3/3,332 (0.09%)	0/3,332 (0.00%)	0/3,332 (0.00%)	0/3,332 (0.00%)	1/3,332 (0.03%)	271/3,332 (8.13%)	0/3,332 (0.00%)	-	0/3,332 (0.00%)	31/3,332 (0.93%)	-	0/3,332 (0.00%)
	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31		2018-12-31	2018-12-31		2018-12-31
	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12		2019-02-12	2019-02-12		2019-02-12
OpenPhish	1/8,104 (0.01%)	33,239/ 71,691 (46.36%)	98/ 8,104 (1.21%)	21/ 71,691 (0.03%)	116/ 71,691 (0.16%)	0/71,691 (0.00%)	38/8,104 (0.47%)	18,233/ 71,691 (25.43%)	21/71,691 (0.03%)	0/8,104 (0.00%)	-	220/ 8,104 (2.71%)	2/65,880 (0.00%)	35/ 71,691 (0.05%)
	2018-12-31	2017-10-01	2018-12-31	2017-10-01	2017-10-01	2017-10-01	2018-12-31	2017-10-01	2017-10-01	2018-12-31		2018-12-31	2017-10-01	2017-10-01
	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12		2019-02-12	2019-01-01	2019-02-12
DNSBH	3/23,083 (0.01%)	4,756/ 23,083 (20.60%)	767/ 23,083 (3.32%)	28/ 23,083 (0.12%)	1/23,083 (0.00%)	18/23,083 (0.08%)	254/ 23,084 (1.10%)	5,078/ 23,083 (21.34%)	27/23,083 (0.12%)	31/23,083 (0.13%)	220/23,083 (0.95%)	-	-	2/23,083 (0.01%)
	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31	2018-12-31			2018-12-31
	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12			2019-02-12
ThreatExpert	-	49/292 (16.78%)	-	2/292 (0.68%)	1/292 (0.34%)	1/292 (0.34%)	-	31/292 (10.62%)	31/292 (10.62%)	-	2/246 (0.81%)	-	-	3/292 (1.03%)
		2016-10-01		2016-10-01	2016-10-01	2016-10-01		2016-10-01	2016-10-01		2017-10-01			2016-10-01
		2019-01-01		2019-01-01	2019-01-01	2019-01-01		2019-01-01	2019-01-01		2019-01-01			2019-01-01
Malc0de	186/226 (82.30%)	1,405/2,249 (62.47%)	206/226 (91.15%)	7/2,249 (0.31%)	2/2,249 (0.09%)	0/2,249 (0.00%)	4/226 (1.77%)	856/2,122 (40.34%)	8/2,249 (0.36%)	0/226 (0.00%)	35/1,219 (2.87%)	2/226 (0.88%)	3/2,050 (0.15%)	-
	2018-12-31	2016-07-08	2018-12-31	2016-07-08	2016-07-08	2016-07-08	2018-12-31	2016-01-01	2016-07-08	2018-12-31	2017-10-01	2018-12-31	2016-10-01	
	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-02-12	2019-01-01	

date might produce misleading results because the majority of the aggregated blacklist will be populated with obsolete (or even expired) domain names, especially with Hostfile's domains. These domains could have turned into NXDOMAINs when the observation using newer DBLs took place.

- Using only the data from December 31, 2018 might not be large enough, as the data captured in this duration might only cover approximately 10% of the overall observation period.

Hence, the separation of the data set into two groups could provide pairwise information about the *historical* and the *newer* data sets. The results of conducting pairwise comparison with the aggregated list can be seen at Table 4.13 and Table 4.14. Table 4.13 shows the pairwise comparison with *historical* aggregated data and Table 4.14 displays the results of executing pairwise comparison with aggregated *newer* data. In these tables, `Domains` indicates the total number of unique domain name a DBL published for each timestamp. `Total` shows the total number of unique domain names from all DBLs for each category. `Ratio` is computed by dividing the `Domains` by `Total`, and the rank is sorted in a decreasing order, since a higher contribution of a DBL to the overall malicious blacklist is considered to be better.

As can be seen from Table 4.13, using the historical data set, both Hostfile file

Table 4.13: Historical DBLs Pairwise Comparison with Aggregated List.

DBL	Domains	Total	Ratio	Rank
MDL	1,027	793,536	0.13%	7
Joewein	113,985		14.36%	3
Malc0de	2,099		0.26%	5
ZTracker	636		0.08%	8
RWTracker	1,682		0.21%	6
Hostfile	507,864		64.00%	1
HPHosts	452,940		57.08%	2
ThreatExpert	290		0.04%	9
OpenPhish	65,795		8.29%	4

updates contributed to more than 50% of the aggregated blacklisted domain names. ThreatExpert was the smallest contributor, since the total number of unique domain names published was less than 1% of the total aggregated list. Joewein could also be recognized as one of the major contributors to the overall blacklisted domain names, because although they just included domain names related with mail spamming, they registered more than 14% of the aggregated blacklisted domain names.

From Table 4.14, it is visible that Hostfile still dominated the aggregated blacklisted domain names using the newer data set. URLHaus and DNSBH followed the

Table 4.14: Newer DBLs Pairwise Comparison with Aggregated List.

DBL	Domains	Total	Ratio	Rank
MDL	860		0.19%	10
Joewein	3,996		0.87%	7
Malc0de	226		0.05%	13
ZTracker	344		0.07%	12
RWTracker	1,668		0.36%	9
Hostfile	166,664		36.15%	2
HPHosts	248,873	461,043	53.98%	1
OpenPhish	8,104		1.76%	6
URLHaus	36,746		7.97%	3
CCTracker	9,974		2.16%	5
DNSBH	23,083		5.01%	4
VXVault	570		0.12%	11
C2dom	3,332		0.72%	8

ranking below the Hostfile feeds. Considering the results from both historical and newer pairwise comparisons, the contributions of MalwareDomainList, ZeusTracker, and RansomwareTracker remained below 1%. It is likely that in the future, the ratios might not change a lot, considering the number of domain names published was considerably less significant compared to Hostfile or URLHaus.

4.3.3 Timing

The timing measures the relative start and end date of a domain name to appear in a domain blacklist. Using these dates, the overall malicious campaign duration can then be estimated.

Blacklisted Duration

Firstly, to determine the start and end date of a blacklisted domain name in a specific DBL, all contents in the data set were merged to find continuous appearance of a domain name in a DBL. Based on this, the start and end date of a malicious event using this particular domain could be determined. Please note that this is not the complete estimation of the start and end date of a malicious campaign. This process was conducted to group the data set by domain name and each DBL. This means that, cases where a domain disappeared from a DBL and re-appeared the day after, are considered to be two different cases. Finally, the estimation of the duration of a domain name to stay in a DBL can be calculated by subtracting the end date by the

start date. In general, the statistics of the duration of a domain name to stay in a DBL are as follows.

- Total: 17,770,425.
- Minimum: 1, Maximum: 363.
- Q1: 1.0, Median: 3.0, Q3: 7.0
- Average: 14.69, Variance: 946.71, SD: 30.77.

In the statistics provided above, `Total` shows the total number of independent malicious events from July 8, 2016 until February 12, 2019 regardless of their appearance, end date, and the respective DBL. This means that this number also includes cases where a domain name appeared in multiple DBLs. `Minimum` shows the minimum difference between the start and end date of a malicious event, where `Maximum` indicates the maximum duration of a domain name to stay in a DBL. `Q1`, `Median`, and `Q3` show the 25th, 50th, and 75th percentile respectively, in regard to how long a domain stays in a DBL. Finally, information about the average, variance, and the standard deviation of this duration are also provided.

Based on the statistics, in general, blacklisted domain names stayed in a DBL for around two weeks, although in some cases, one domain name could stay in a blacklist for almost one year. This indicates that in general, malicious domain names had relatively short lifespan, before the machines got blacklisted or shut down. Taking the public DBLs as the focus point, short-lived domain names could mean two-fold, either these DBLs actually blacklisted real malicious domains (that eventually they got shut down), or the domain they blacklisted might not exist in the first place, that they needed to remove from their database. To find out which one is more likely to happen, metric "Liveliness", which will be discussed later, could provide hints about how many domain names are actually *live*.

The blacklisted duration for every single DBL can be seen at Table 4.15. In this table, column `Total` shows the total number of "stays" inside a DBL. `Min`, `Q1`, `Med`, `Q3`, and `Max` indicates the minimum, 25th percentile, median, 75th percentile, and maximum number of days blacklisted domain names stay in the DBL respectively. `Avg`, `Var`, and `SD` shows the average, variance, and the standard deviation of the blacklist duration of each DBL.

As can be seen in Table 4.15, on average, Joewein had the lowest average blacklist duration. The blacklist duration of a domain name in Joewein was also relatively constant, since the variance is the smallest compared to other DBLs. This indicates that although the total number of blacklisted domain names was relatively high, Joewein was one of the DBL that update their database frequently.

Table 4.15: DBL Blacklisted Duration

DBL	Total	Min	Q1	Med	Q3	Max	Avg	Var	SD
VXVault	592	1	20	21	28	76	26.28	245.73	15.68
Hostfile	1,763,553	1	2	7	42	363	25.89	1,337.21	36.57
URL-Haus	36,945	1	44	76	76	76	56.24	554.38	23.55
MDL	39,991	1	1	3	11	363	19.97	2,282.48	47.78
Threat-Expert	5,485	1	1	2	5	363	18.54	2,963.57	54.44
RW-Tracker	74,710	1	1	3	11	363	16.94	1,933.59	43.97
CC-Tracker	9,975	1	44	44	44	76	51.46	188.45	13.73
HP-Hosts	15,339,287	1	1	3	7	363	13.40	888.06	29.80
ZTracker	15,230	1	1	3	11	363	21.82	3,103.04	55.70
C2dom	869	44	76	76	76	76	75.74	8.18	2.86
Open-Phish	166,685	1	2	4	7	126	11.78	538.26	23.00
DNSBH	23,084	1	44	44	76	76	57.06	250.60	15.83
Joewein	287,520	1	1	2	5	363	4.14	50.17	7.08
Malc0de	6,499	1	2	6	29	363	18.58	1,007.54	31.74

On the other hand, the average of blacklist duration was topped by c2dom. Blacklisted domain names in c2dom tend to stay longer than two months in the database by taking the average blacklist duration as the reference. The variance and standard deviation results also show that the number of domain names blacklisted during the observation period fluctuated less frequently.

Disappearance Duration

Based on the observed behaviors of blacklisted domain names from the data set, there were some domain names that got de-listed at some point of time and re-appeared later, whether in the same DBL, or different DBL. For example, domain `img001.com` was blacklisted by MalwareDomainList from July 8, 2016 until August 20, 2016. Then, on August 22, 2016, this domain re-appeared in the same blacklist. Therefore, it is also interesting to investigate the statistics of the disappearance duration of a domain from a DBL, besides the blacklisted duration. This information is shown below.

- Total: 16,528,138.
- Minimum: 1, Maximum: 912.
- Q1: 3.0, Median: 4.0, Q3: 8.0
- Average: 12.65, Variance: 807.25, SD: 28.41.

As can be seen from the statistics above, on average, de-listed domains might reappear within two weeks, or even several years, after their de-list-ed date. Short disappearance duration hints that the DBL is updated more frequently than other DBLs. They re-check the maliciousness of submitted domain names and directly put these domains into their blacklist database. On the other hand, it could also mean that the verifying procedure for de-listing might be less strict than other DBLs. This means that, the domains were de-listed before they were completely removed from the Internet.

The disappearance duration for each DBL can be seen at Table 4.16. Similar to Table 4.15, the `Total` column shows the total number of domain names that got de-listed and re-appear in a specific blacklist. Column `Min`, `Q1`, `Med`, `Q3`, and `Max` indicates the minimum, 25th percentile, median, 75th percentile, and maximum number of days one domain name disappear from a DBL. `Avg`, `Var`, and `SD` shows the average, variance, and the standard deviation of the disappearance duration of each DBL.

As can be seen at Table 4.16, in most of DBLs, de-listed domain names were likely to re-appear after around two weeks. Joewein, again, shows the lowest average disappearance duration. One of the possible reasons is that, Joewein contains domain names related with spamming activities. As broadcasting emails could be stopped and re-executed one click away, blacklisted domain names related with spamming might disappear and re-appear in a blacklist more frequently than other malicious activities. For instance, MalwareDomainList, ZeusTracker, and ThreatExpert are related with malware distribution and their average disappearance duration were more than two weeks.

4.3.4 Responsiveness

This metric is an extension of the previous metric, the Timing. After noticing how blacklisted domain names “behave” in the DBLs, how fast these DBLs update their database could be investigated.

Firstly, as mentioned in Sub-section 3.3.4, the first step is to determine the relative campaigns’ disappearance duration. This is to distinguish malicious domain names of different campaigns by considering the disappearance duration. Reappearance of a domain in a DBL after one year of disappearance might indicate

Table 4.16: DBL Disappearance Duration

DBL	Total	Min	Q1	Med	Q3	Max	Avg	Var	SD
VXVault									0
Hostfile	1,240,785	2	2	4	8	912	13.67	1,251.88	35.38
URL-Haus									0
MDL	38,964	2	3	6	12	364	19.16	2,240.39	47.33
Threat-Expert	5,195	2	3	4	6	261	17.75	2,413.14	49.12
RW-Tracker	73,028	2	2	4	9	364	15.53	1,697.94	41.21
CC-Tracker									0
HP-Hosts	14,886,343	1	3	4	8	422	12.61	768.25	27.72
ZTracker	14,592	2	3	4	9	364	18.81	2,709.57	52.05
C2dom									0
Open-Phish	94,768	2	2	6	14	402	13.42	613.58	24.77
DNSBH									0
Joewein	170,223	2	2	3	5	507	4.53	64.03	8.00
Malc0de	4,240	2	2	3	6	661	8.88	1,069.80	32.71

that the domain name was used for different campaign. Therefore, this campaign duration is set to be the threshold in separating malicious campaigns.

As shown in the previous results, setting the threshold to be 30 days seems logical. Since the average blacklisted duration was around two months and disappearance duration was around two weeks, using 30 days as the boundary to distinguish different campaign is considered to be long enough to separate two distinct malicious campaigns.

Based on this threshold, if a domain d first appeared at DBL DBL_1 at date $start_1$ and disappeared at end_1, and it also appeared at different DBL DBL_2 (with $DBL_1 \neq DBL_2$) at date $start_2$ and disappeared at end_2, these two events will be considered to be different campaigns using threshold θ, under the following conditions:

1. $(start_2 - end_1 > \theta)$ or $(start_1 - end_2 > \theta)$. This represents cases where the first appearance of domain d at DBL_2 was considerably later than the last appearance date at DBL_1, and the last appearance of domain d at DBL_2 was

much earlier than the first appearance date at DBL_1.

2. $(start_2 - start_1 > \theta)$ and $(end_1 - end_2 > \theta)$. In this scenario, the first appearance at DBL_2 was later than DBL_1, but greater than the threshold. Then, the disappearance at DBL_2 was earlier than DBL_1, but greater than the threshold. This means that, the appearance of domain d at DBL_2 was somewhere in between $start_1$ and end_1, but the start and end date differences are greater than the threshold. This rule is applicable *vice versa*.

Other than the defined rules, the intersections of the appearance of a domain name on multiple blacklists was considered as one same campaign, just that one DBL blacklisted quicker than the others.

After defining the boundary to distinguish a same and different malicious campaign, finding head-to-head responsiveness of a DBL against other DBLs could be performed. This was done by first finding domain names that are categorized as the same campaign with all other DBLs. This means that the domain names must exist in at least one other DBLs within the time threshold.

The first appearance date of each domain name in a DBL were compared against each other to see which DBL blacklist quicker or slower than the others. Accumulating the total count of intersections and the total time difference of appearance for each domain names in each DBL allowed the average *early* and *late* appearance of domain names to be computed. Average *early* estimates how early, in days, a DBL puts a domain name into their blacklist, when other DBLs are also blacklisting it a bit later. Average *late* estimates how late, in days, a DBL puts a domain name into their blacklist, when other DBLs have already blacklisted it. This computation also goes the same way for the last appearance date of a domain name in a DBL. Therefore, each DBL would have their own average *early* and *late* score for both start and end date of a malicious campaign.

The final computation was conducted by taking the average of all *early* and *late* score for each DBLs. The final result of this computation can be seen at Table 4.17. In this table, column `Early` shows how many days earlier a DBL blacklist a domain name, compared to other DBLs. Similarly, column `Late` shows how late a DBL blacklist a domain name, compared to other DBLs. These two columns are using days as the unit. Then, the early and late score are compared to determine the DBL's tendency, whether they are likely to blacklist quicker or slower than other DBLs. The higher score of `Early` or `Late` indicates the higher tendency of a DBL to be quick or slow respectively in updating their database.

As can be seen at Table 4.17, Hostfile, MalwareDomainList, and Joewein had a tendency to blacklist and de-list earlier than other DBLs. On the other hand, database of URLHaus and ZeusTracker seemed to be updated a bit slower than

Table 4.17: DBL Responsiveness

DBL	Responsiveness Score					
	Start			End		
	Early	Late	Tendency	Early	Late	Tendency
VXVault	4.00	10.25	Late	19.29	0.00	Early
Hostfile	14.92	7.78	Early	22.33	17.23	Early
URLHaus	5.50	16.41	Late	17.50	19.56	Late
MDL	14.33	0.00	Early	14.00	0.00	Early
ThreatExpert	0.00	0.00	-	2.00	0.00	Early
RWTracker	8.75	17.40	Late	0.00	0.00	-
CCTracker	0.00	18.39	Late	0.00	0.00	-
HPHosts	17.67	14.95	Early	17.25	25.70	Late
ZTracker	17.86	20.86	Late	16.00	16.53	Late
C2dom	0.00	0.00	-	0.00	0.00	-
OpenPhish	11.13	16.23	Late	17.34	13.21	Early
DNSBH	0.00	16.81	Late	0.00	0.00	-
Joewein	16.70	13.18	Early	18.62	10.13	Early
Malc0de	15.35	12.26	Early	14.34	16.12	Late

the other DBLs. During the observation period, the data was not enough to estimate the tendency of C2dom, as it had no intersecting campaigns with other DBLs.

4.3.5 Specificity

Specificity measures how much of benign domain names are not blacklisted. This is quite the opposite of one of the approach in measuring the purity. In this research, the specificity was conducted by finding the intersection of domain names that were found in both daily updates of each DBL and Alexa top 100k websites on the day a DBL publish their data. The result of this can be seen at Table 4.18.

In this table, column `Total` shows the total number of domains, based on the number of days the data from a DBL was captured, and multiplied by $100,000$, since the data used are only Alexa's top 100k website. For instance, the observation of VXVault started from January 1, 2019 until February 12, 2019, which lasted 44 days. Therefore, the value of the `Total` column is $4,400,000$. Column `Blacklisted` shows the total number of blacklisted domain names of each DBL, and `Not Blacklisted` was computed by subtracting the `Blacklisted` values from `Total`. Then, `Ratio` shows the percentage of the values from `Not Blacklisted` against the `Total`.

As can be seen at Table 4.18, in general, the specificity of all public DBLs used in

Table 4.18: DBL Specificity

DBL	Total	Blacklisted	Not Blacklisted	Ratio
VXVault	4,400,000	59	4,399,941	100.00%
Hostfile	83,200,000	230,558	82,969,442	99.72%
URLHaus	4,400,000	13,174	4,386,826	99.70%
MDL	82,900,000	27,351	82,872,649	99.96%
ThreatExpert	46,300,000	1,751	46,298,249	100.00%
RWTracker	83,300,000	825	83,299,175	100.00%
CCTracker	4,400,000	2,099	4,397,901	99.95%
HPHosts	74,700,000	2,385,888	72,314,112	96.81%
DNSBH	4,400,000	4,267	4,395,733	99.90%
C2dom	4,400,000	131	4,399,869	100.00%
OpenPhish	39,200,000	7,332	39,192,668	99.98%
Malc0de	83,800,000	7,508	83,792,492	99.99%
Joewein	82,700,000	309	82,699,691	100.00%
ZTracker	83,600,000	9,899	83,590,111	99.99%

this research were quite high. This indicates that these public DBLs did not blacklist benign domains. The lowest-scoring DBL was HPHosts. One of the possible reason for this is because the number of blacklisted domain names, as shown in Table 4.6, was much more significant compared to the number of domain names used in the specificity. Therefore, the increase of the number of domain names found in Alexa top 100k websites seems logical.

4.3.6 Accuracy

Each DBL has their own policy regarding how detail they want to publish the data about the blacklisted and de-listed domain names. Table 4.19 summarizes what information are published by each DBL regarding the blacklisted domain names. This table considers some information that exist in at least one of the DBLs used. The accuracy considered are: domain names, complete URL to the malicious activities, IP Address of the machine, WHOIS information, reverse IP lookup, system status and uptime, domain registration date, submitted report date, date of appearance of a domain in a blacklist, country of the machines, registered name servers, Autonomous System (AS) numbers and names, and the category of malicious activities each DBL contain. Finally, the blacklist score indicates how many information is provided by each DBL. In addition, the row at the bottom shows the percentage of appearance of each detailed information in the DBLs.

In general, most of publicly available DBLs publish the domain names, IP Address, and the date of blacklisting into their daily updates. However, it is rare to find extra information about the blacklisted domains, such as WHOIS information or domain registration date, from most of public DBLs used in this study. Among 13 different public DBLs used in this research, DBLs that are under Abuse.ch project, such as ZeusTracker, RansomwareTracker, and CybercrimeTracker, showed more information for the blacklisted domain names than other DBLs. Their websites also look more modern and well-maintained compared to other DBLs, like the website of MDL or Joewein. On the other hand, DNSBH and ThreatExpert did not put any extra information other than the domain names on the published blacklists.

Besides the detail of blacklisted domain names, different DBLs also show different details of de-listed domain names. Table 4.20 shows how DBLs provide information about the de-listed domain names. As can be seen in Table 4.20, in general, most of the publicly available DBLs published little to no information about the de-listed domain names. Only ZeusTracker and HostFile documented the de-listed domain names in detail. MalwareDomainList, ThreatExpert, or OpenPhish just contained the de-listed domain names without any further information.

Based on the level of detail of black-and-de-listed domain names of each DBL, the ranking can be seen at Table 4.21. This table summarizes how much information was published by each DBL, categorized by blacklisting, de-listing, and overall ranking. Column `Sum` indicates the total number of information released by each DBL. Column `Total` shows the maximum information that can be gathered for each category. Then, the `Rank` is sorted based on the `Sum` in decreasing order. The more information the DBL contain, the higher the rank.

Based on Table 4.21, ZeusTracker provides the highest amount of information about the black-and-de-listed domain names. On the other hand, DNSBH publishes the least amount of additional information regarding the black-and-de-listed domain names.

4.3.7 Agility

The agility of a DBL estimates how responsive a DBL is based on the number of domain names that are newly found in the database and domain names that disappear from the database. The results of measuring the agility of each DBL are discussed below.

In the figures visualizing the number of domain names entering and disappearing from a DBL below, there are four lines, namely green straight, red straight, blue dotted, and yellow dotted line. Green straight line represents the *plus* (the number of domain names that were newly found in the daily updates of a DBL). Red straight

Table 4.19: DBLs Blacklisting Accuracy

DBL	Domain Names	Complete URL	IP Address	WHOIS DB	Reverse IP Lookup	System Status	System Uptime	Registered Date	Report Date	Blacklisted Date	Country	Registered NS	AS Number	AS Name	Malicious Type	Blacklist Score
MDL	Yes	Yes	Yes	No	Yes	No	No	No	No	Yes	Yes	No	Yes	No	Yes	8
Joewein	Yes	No	No	Yes	No	No	No	Yes	No	Yes	No	Yes	No	No	No	5
Malc0de	Yes	Yes	Yes	No	No	No	No	No	No	Yes	Yes	No	Yes	Yes	No	7
ZTracker	Yes	No	Yes	No	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	9
RWTracker	Yes	No	Yes	No	No	Yes	No	No	No	Yes	Yes	Yes	Yes	No	Yes	8
URLHaus	Yes	Yes	No	No	No	Yes	No	No	Yes	Yes	No	No	No	No	Yes	6
Hostfile	Yes	No	Yes	No	No	No	No	No	No	Yes	Yes	No	Yes	No	Yes	6
ThreatExpert	Yes	No	No	No	No	No	No	No	No	No	No	No	No	No	No	1
OpenPhish	Yes	Yes	No	No	No	No	No	No	No	Yes	No	No	No	No	No	3
VXVault	Yes	Yes	Yes	No	No	Yes	No	No	No	Yes	No	No	No	No	No	5
c2dom	Yes	No	Yes	No	No	No	No	No	No	Yes	No	Yes	No	No	No	4
CCTracker	Yes	Yes	Yes	No	No	No	No	No	No	Yes	Yes	No	Yes	Yes	Yes	8
DNSBH	Yes	No	No	No	No	No	No	No	No	No	No	No	No	No	No	1
	100.00%	46.15%	61.54%	7.69%	7.69%	30.77%	7.69%	7.69%	7.69%	84.62%	46.15%	30.77%	46.15%	23.08%	38.46%	

Table 4.20: DBLs De-listing Accuracy

DBL	Domain Names	IP Address	De-list Date	Country	AS Number	AS Name	Removal Reason	De-list Score
MDL	Yes	No	No	No	No	No	No	1
Joewein	No	No	No	No	No	No	No	0
Malc0de	No	No	No	No	No	No	No	0
ZTracker	Yes	Yes	Yes	Yes	Yes	Yes	Yes	7
RWTracker	No	No	No	No	No	No	No	0
URLHaus	No	No	No	No	No	No	No	0
Hostfile	Yes	Yes	Yes	Yes	Yes	No	Yes	6
ThreatExpert	Yes	No	No	No	No	No	No	1
OpenPhish	Yes	No	No	No	No	No	No	1
VXVault	No	No	No	No	No	No	No	0
c2dom	No	No	No	No	No	No	No	0
CCTracker	No	No	No	No	No	No	No	0
DNSBH	No	No	No	No	No	No	No	0
	38.46%	15.38%	15.38%	15.38%	15.38%	7.69%	15.38%	

Table 4.21: DBLs Overall Accuracy Ranking

DBL	Blacklisting			De-listing			Overall		
	Sum	Total	Rank	Sum	Total	Rank	Sum	Total	Rank
MDL	8	15	2	1	7	3	9	22	3
Joewein	5		8	0		4	5		8
Malc0de	7		5	0		4	7		6
ZTracker	9		1	7		1	16		1
RWTracker	8		2	0		4	8		4
URLHaus	6		6	0		4	6		7
Hostfile	6		6	6		2	12		2
ThreatExpert	1		12	1		3	2		12
OpenPhish	3		11	1		3	4		10
VXVault	5		8	0		4	5		8
C2dom	4		10	0		4	4		10
CCTracker	8		2	0		4	8		4
DNSBH	1		12	0		4	1		13

line visualizes the opposite of the green straight line, which is the *minus* (the number of domain names that disappeared from daily updates of a DBL). Blue dotted line shows the number of domain names that were contained in each DBL's database. This value is calculated by comparing each daily updates with the previous (could be one to several days before) update published. Finally, the yellow dotted line represents the number of domain names that each DBL published each day.

The idea behind introducing the blue and yellow dotted lines was because for some DBLs, they might publish nothing, but several days later, the service were fixed and they started publishing blacklisted domain names again. This might be caused by technical problems from the DBL, or from the scraping scripts. In this scenario, the blue dotted line will stay at the position of the previous update, but the yellow dotted line would plunge to zero and return to the same place as the blue dotted line. The *plus* and *minus* are referring to the blue dotted line, not the yellow dotted line.

The statistics were computed from the second day of observation onward. This is to discard the "initialization" first day, where the whole domain names captured from the first measurement date were considered as the *plus*.

MalwareDomainList

Complete visualization of the agility of MalwareDomainList can be seen at Figure 4.2. The statistics of MDL's agility are also provided below. From Figure 4.2, since

July 8, 2016 until February 12, 2019, the number of domain names published by MDL was around 800 domain names each day. During the measurement period, several days where MDL published nothing are visible from the fluctuating yellow dotted line from the figure.

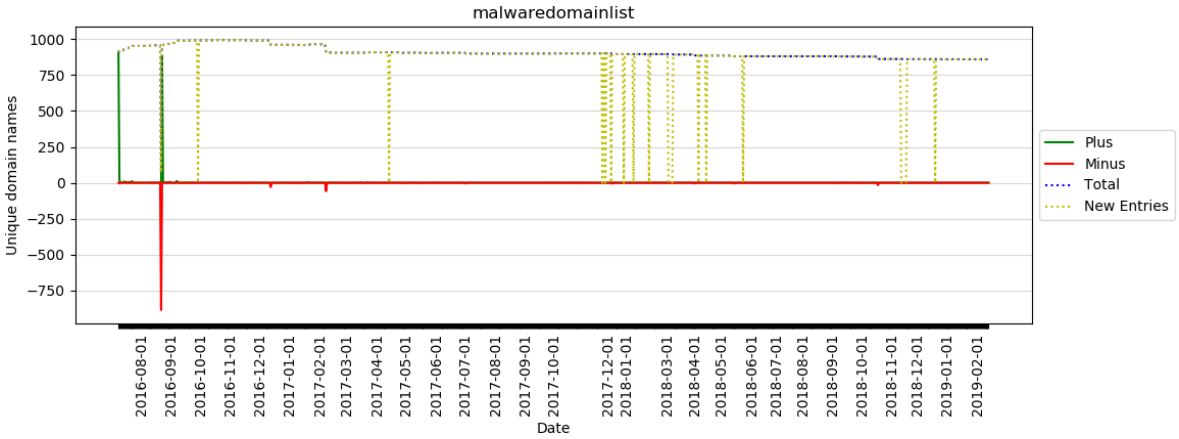

Figure 4.2: Agility of MalwareDomainList

- Statistics:

  1. Duration: 949 days (2016-07-09 - 2019-02-12).
  2. Total number of unique domain names: 1,027.
  3. Average database content: 908.12, average new entries: 885.32.
  4. Plus:
     - Minimum: 0, Maximum: 886.
     - Q1: 0.0, Median: 0.0, Q3: 0.0.
     - Average: 1.12, Variance: 874.94, Standard Deviation: 29.58.
  5. Minus:
     - Minimum: 0, Maximum: 886.
     - Q1: 0.0, Median: 0.0, Q3: 0.0.
     - Average: 1.18, Variance: 879.18, Standard Deviation: 29.65.

In addition, the maintenance of MDL was relatively small, only less than 1% of its total published domain names. This can be inferred from the statistics, where the average *plus* and *minus* were much smaller than the average number of domain names contained in MDL.

Joewein

Figure 4.3 shows the complete visualization of the agility of Joewein. From July 8, 2016 to February 12, 2019, more than 1,000 domain names were blacklisted each day. From the visualization, it can be seen that Joewein was one of the well-maintained service, since the *plus* and *minus* fluctuated a lot during the measurement period.

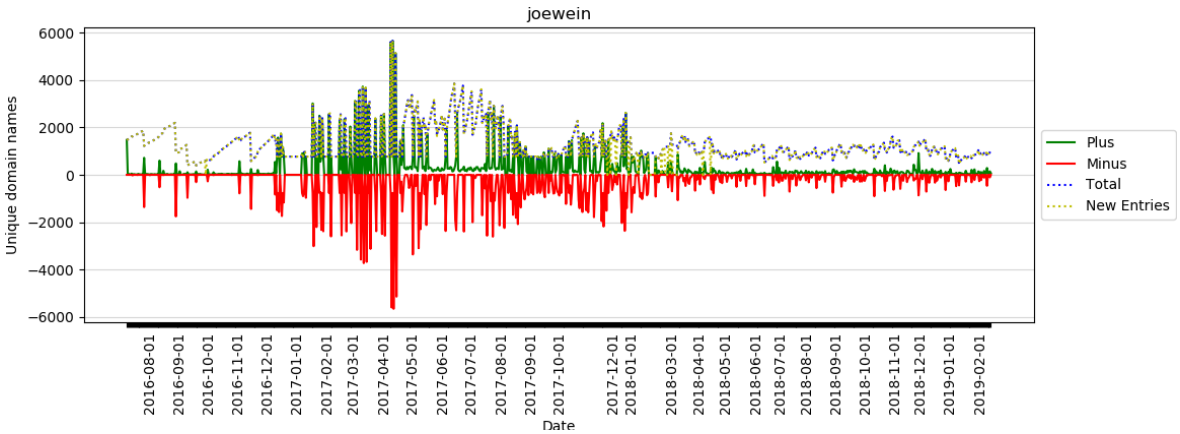

Figure 4.3: Agility of Joewein

- Statistics:
  1. Duration: 949 days (2016-07-09 - 2019-02-12).
  2. Total number of unique domain names: 117,288.
  3. Average database content: 1,284.05, average new entries: 1,253.04.
  4. Plus:
     - Minimum: 0, Maximum: 5,664.
     - Q1: 28.0, Median: 79.0, Q3: 237.0.
     - Average: 290.25, Variance: 359,844.94, Standard Deviation: 599.87.
  5. Minus:
     - Minimum: 0, Maximum: 5,664.
     - Q1: 0.0, Median: 2.0, Q3: 268.0.
     - Average: 290.74, Variance: 421,704.43, Standard Deviation: 649.39.

Based on Joewein's statistics, it is visible that Joewein was one of the frequently-updated blacklist. Taking the average as reference, the number of new domain names found in the database and the number of domain names disappeared from the blacklist were around 20% of their total number of domain names in their database.

The variance and SD of the statistics hint that during the observation period, the number of domain names entering and leaving the database fluctuated quite frequently.

Malc0de

Figure 4.4 displays the complete visualization of the agility of Malc0de. From July 8, 2016 to February 12, 2019, the maximum number of domain names blacklisted on their database was less than 350 items. The maintenance of Malc0de was relatively good, by seeing small number of new blacklisted and de-listed domain names each day.

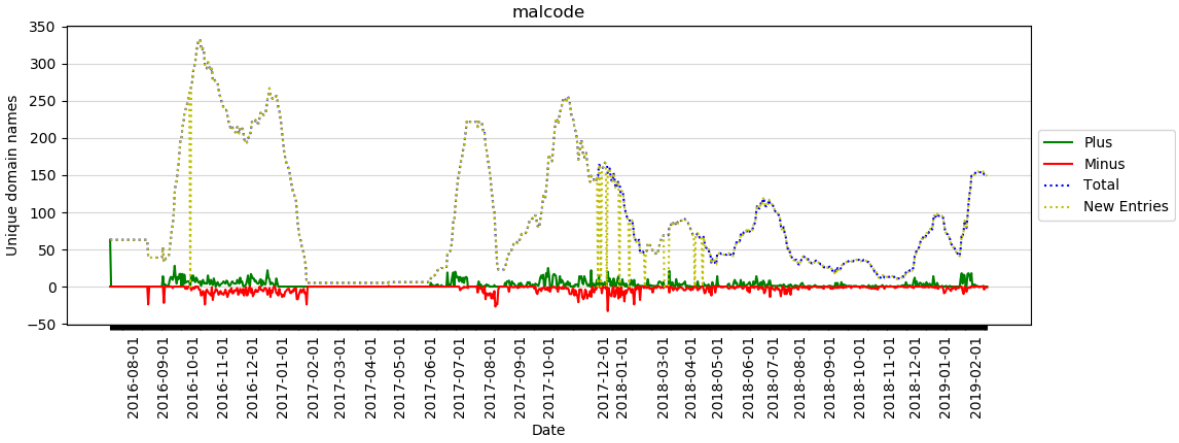

Figure 4.4: Agility of Malc0de

- Statistics:
  1. Duration: 949 days (2016-07-09 - 2019-02-12).
  2. Total number of unique domain names: 2,249.
  3. Average database content: 91.78, average new entries: 90.12.
  4. Plus:
     - Minimum: 0, Maximum: 28.
     - Q1: 0.0, Median: 0.0, Q3: 4.0.
     - Average: 2.69, Variance: 20.03, Standard Deviation: 4.48.
  5. Minus:
     - Minimum: 0, Maximum: 33.
     - Q1: 0.0, Median: 1.0, Q3: 3.0.
     - Average: 2.60, Variance: 19.15, Standard Deviation: 4.38.

The statistics above hints that Malc0de was one of the DBL with frequent updates, but the updates were relatively small. On the daily basis, almost three distinct domain names were newly found in, and disappeared from, the database. This information is supported by Figure 4.4 above, where the green and red lines were swinging around the zero line.

ZeusTracker

Complete visualization of the agility of ZeusTracker is shown by Figure 4.5. As can be seen in the figure, the number of blacklisted domain names kept decreasing since the start of the measurement period, which is July 8, 2016. Upon deeper investigation, one of the factors leading to this decline was the decreasing popularity of malware of Zeus family. Yearly report on top 10 malware list at January 2018 [43] and August 2017 [44] showed that the ratio of virus infections caused by Zeus malware family decreased by around 3%.

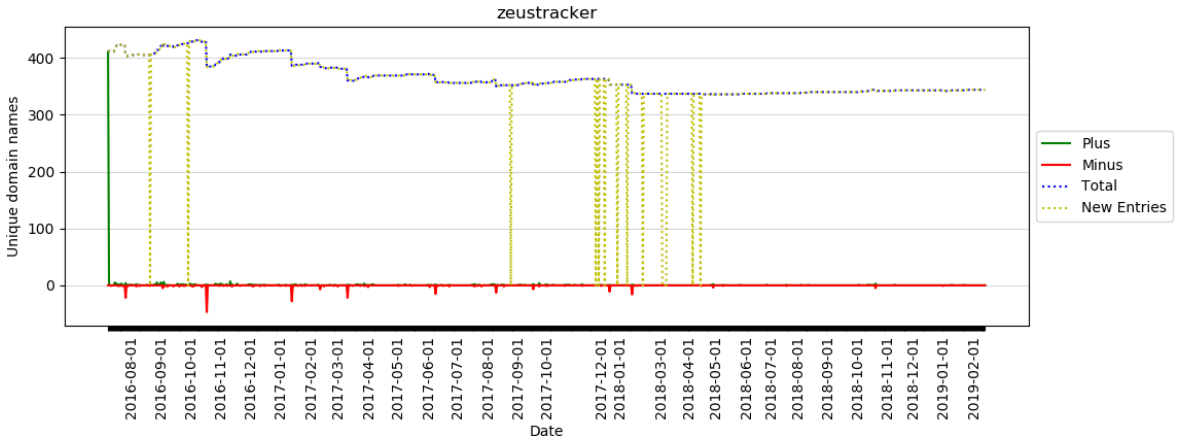

Figure 4.5: Agility of ZeusTracker

- Statistics:
  1. Duration: 949 days (2016-07-09 - 2019-02-12).
  2. Total number of unique domain names: 638.
  3. Average database content: 364.23, average new entries: 357.90.
  4. Plus:
     - Minimum: 0, Maximum: 7.
     - Q1: 0.0, Median: 0.0, Q3: 0.0.
     - Average: 0.26, Variance: 0.49, Standard Deviation: 0.70.
  5. Minus:

  - Minimum: 0, Maximum: 47.
  - Q1: 0.0, Median: 0.0, Q3: 0.0.
  - Average: 0.33, Variance: 5.51, Standard Deviation: 2.35.

Zeustracker's statistics indicates that the number of domain names in their database was relatively stable, with not much changes during the measurement period. Furthermore, the changes trend shows that the number of unique domain names blacklisted in Zeustracker kept decreasing since August 2016 to February 2019.

RansomwareTracker

Agility of RansomwareTracker is displayed in Figure 4.6. As can be seen in the figure, the number of blacklisted domain names showed an increasing trend since the start, July 8, 2016, to the end, February 12, 2016, of the measurement period. By the end of the measurement period, daily updates of RansomwareTracker contained more than 1,500 unique domains.

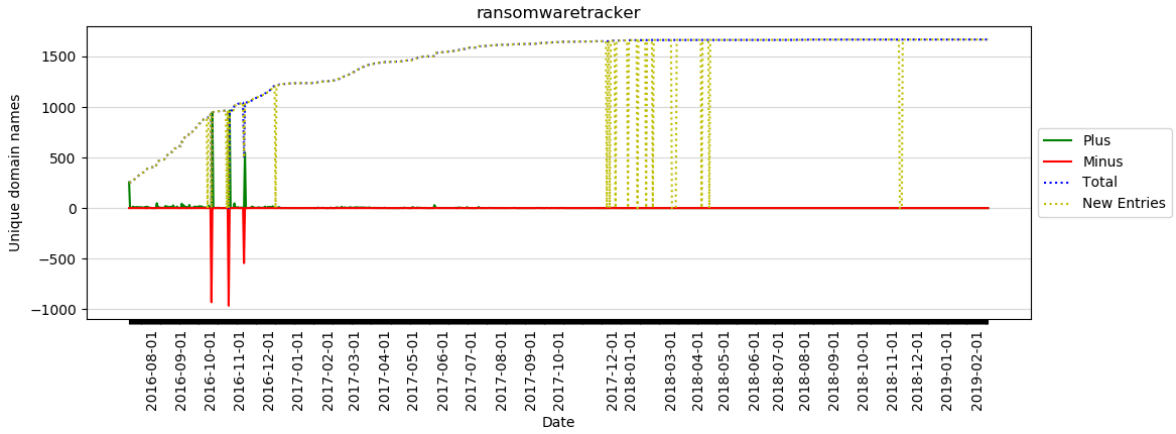

Figure 4.6: Agility of RansomwareTracker

- Statistics:

  1. Duration: 949 days (2016-07-09 - 2019-02-12).
  2. Total number of unique domain names: 1,682.
  3. Average database content: 1,444.0, average new entries: 1,409.09.
  4. Plus:
     - Minimum: 0, Maximum: 966.
     - Q1: 0.0, Median: 0.0, Q3: 1.0.
     - Average: 4.32, Variance: 2,379.85, Standard Deviation: 48.78.
  5. Minus:

- Minimum: 0, Maximum: 966.
- Q1: 0.0, Median: 0.0, Q3: 0.0.
- Average: 2.74, Variance: 2,332.33, Standard Deviation: 48.29.

Taking the average of plus and minuses of RansomwareTracker as reference point, the number of unique domain names newly entering the blacklist was almost doubled of the number of domain names disappearing from the database. However, generally these changes were relatively small, as shown by the Q1, median, and Q3 of both plus and minus of Ransomware's statistics.

Hostfile

Figure 4.7 displays the complete visualization of the agility of the "partial" file update of Hostfile. From July 8, 2016 to February 12, 2019, it is visible that over some period of time, from the beginning of the observation period until April 2018, Hostfile normally added new domain names gradually, and after some time, around 90 days, they flushed out these domain names from their database. However, this behavior stopped at around April 2018 and the number of blacklisted domain names kept increasing. One of the possible reason for this is because, at some point of time in 2018, their other "full" file update was deprecated. Since then, this "partial" file update contained all malicious domains. Nonetheless, from Figure 4.7, it can be inferred that Hostfile maintained their database actively.

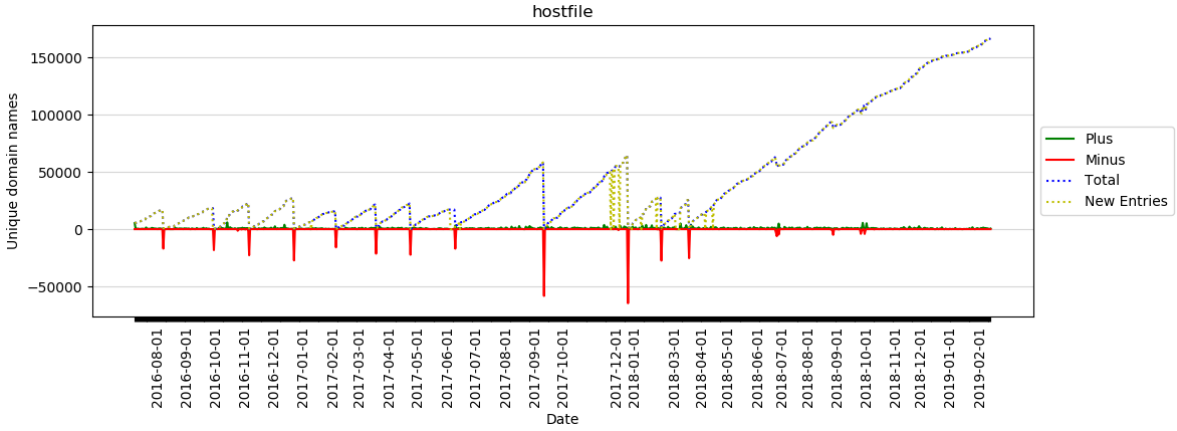

Figure 4.7: Agility of Hostfile

- Statistics:

  1. Duration: 949 days (2016-07-09 - 2019-02-12).
  2. Total number of unique domain names: 522,477.
  3. Average database content: 43,481.01, average new entries: 43,022.82.

4. Plus:
   - Minimum: 0, Maximum: 5,887.
   - Q1: 153.0, Median: 430.0, Q3: 833.0.
   - Average: 597.02, Variance: 441,221.36, Standard Deviation: 664.24.
5. Minus:
   - Minimum: 0, Maximum: 64,686.
   - Q1: 0.0, Median: 0.0, Q3: 2.0.
   - Average: 417.16, Variance: 13,771,393.58, Standard Deviation: 3,710.98.

The "partial" file update of Hostfile is one of the active DBL, as can be seen from the statistics. On average, there were almost 600 new domain names entering the DBL and around 400 leaving from the database during the measurement period. In general, the number of new domain names found in the database on daily basis were much larger in quantity compared to the number of de-listed domain names.

HPHosts

The agility of the "full" file update of Hostfile can be seen at Figure 4.8. As can be seen in the figure, in their "full" file update, periodic increase in the number of domain names found was related with Hostfile's "partial" file update. By also looking at Figure 4.7, the plunge of the number of blacklisted domain names in the "partial" full update occurred at the same day as the increase in blacklisted domain names in the "full" file update. This indicates that periodically, domain names from the "partial" file update were migrated into their "full" full update. However, since they no longer update this HPHosts file, the number of domain names remained static from April 2018 onward.

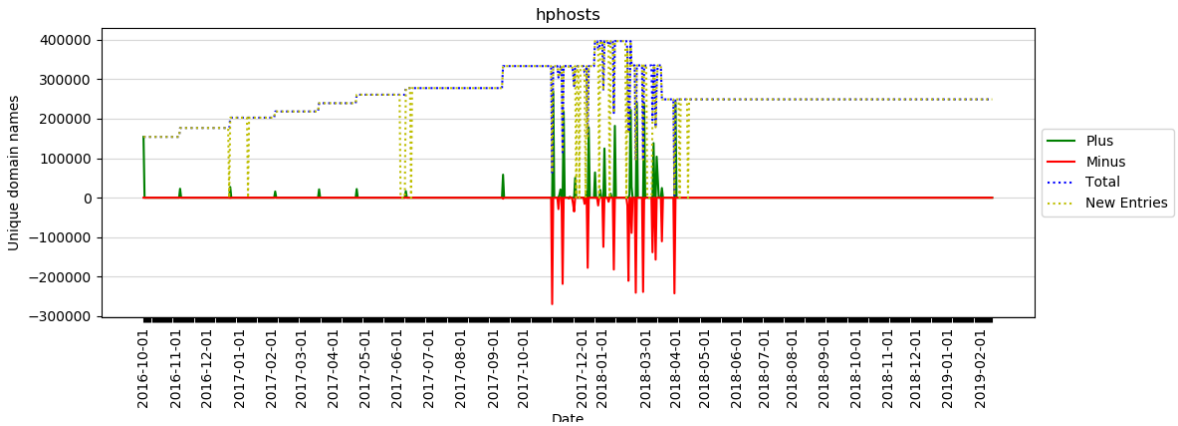

Figure 4.8: Agility of HPHosts

- Statistics:
    1. Duration: 864 days (2016-10-02 - 2019-02-12).
    2. Total number of unique domain names: 452,947.
    3. Average database content: 255,295.37, average new entries: 247,568.60.
    4. Plus:
        - Minimum: 0, Maximum: 269,301.
        - Q1: 0.0, Median: 0.0, Q3: 0.0.
        - Average: 3,257.54, Variance: 569,764,923.25, Standard Deviation: 23,869.75.
    5. Minus:
        - Minimum: 0, Maximum: 269,301.
        - Q1: 0.0, Median: 0.0, Q3: 0.0.
        - Average: 3,140.61, Variance: 587,392,798.36, Standard Deviation: 24,236.19.

It can be inferred that, based on the statistics above, the "full" file update of Hostfile contained quite a large number of unique domain names, while the changes could vary hugely as indicated by the large value of the variance and the standard deviation of both plus and minus of HPHosts. These changes could vary between 0 to more than 250,000 unique domain names on a single day.

ThreatExpert

The agility of ThreatExpert is shown at Figure 4.9. During the 820 days of observation, the number of blacklisted domain names was quite static. In terms of newly found and disappearing domain names, ThreatExpert were also less active, compared to other sources. The observation then stopped at 2019 because the source no longer update their database.

- Statistics:
    1. Duration: 820 days (2016-10-02 - 2018-12-30).
    2. Total number of unique domain names: 290.
    3. Average database content: 208.92, average new entries: 120.65.
    4. Plus:
        - Minimum: 0, Maximum: 2.
        - Q1: 0.0, Median: 0.0, Q3: 0.0.

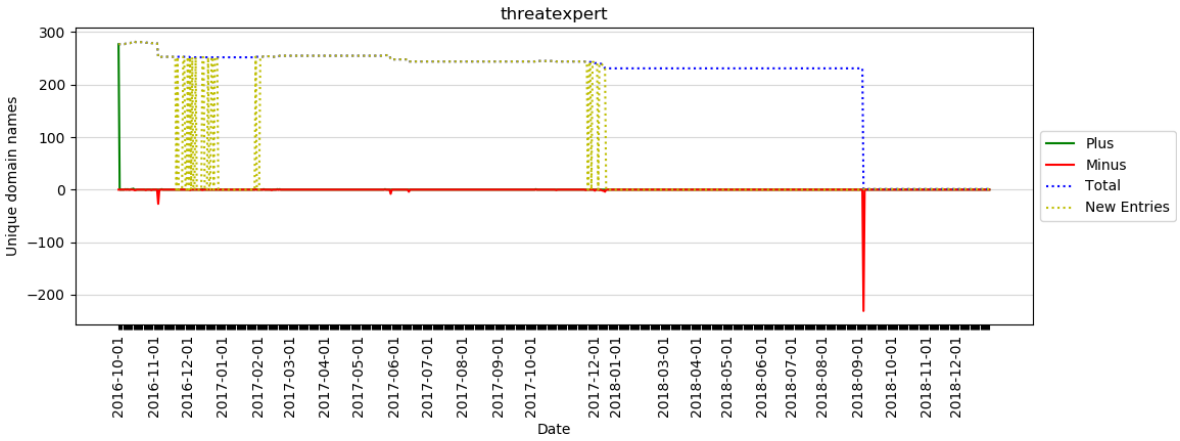

Figure 4.9: Agility of ThreatExpert

  - Average: 0.02, Variance: 0.02, Standard Deviation: 0.15.

5. Minus:
  - Minimum: 0, Maximum: 231.
  - Q1: 0.0, Median: 0.0, Q3: 0.0.
  - Average: 0.38, Variance: 70.44, Standard Deviation: 8.39.

The statistics shows that ThreatExpert was also one of the relatively stable DBL. Out of more than 200 domain names, ThreatExpert tended to remove one domain name every three days, while the number of new domain names found was quite hard to spot.

OpenPhish

The agility of OpenPhish is illustrated at Figure 4.10. As can be seen in the figure, OpenPhish was also one of the well-maintained publicly available DBLs based on the number of new and de-listed malicious domain names. The number of domain names they published showed a decreasing trend, from around 4,000 domain names in 2017 and early 2018 to around 1,000 domain names in 2019.

- Statistics:

1. Duration: 473 days (2017-10-28 - 2019-02-12).
2. Total number of unique domain names: 71,691.
3. Average database content: 1,477.67, average new entries: 1,410.09.
4. Plus:
  - Minimum: 0, Maximum: 4,041.
  - Q1: 165.0, Median: 220.0, Q3: 317.0.

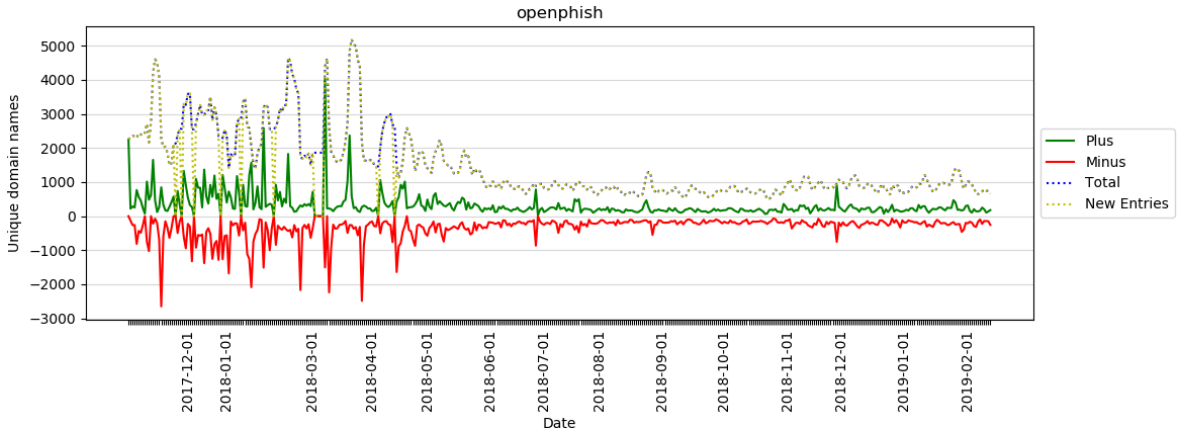

Figure 4.10: Agility of OpenPhish

   - Average: 314.78, Variance: 119,083.99, Standard Deviation: 345.09.

5. Minus:
   - Minimum: 0, Maximum: 2,648.
   - Q1: 162.0, Median: 220.0, Q3: 338.0.
   - Average: 318.56, Variance: 113,317.70, Standard Deviation: 336.36.

From the statistics, it is also visible that OpenPhish was one of the actively-maintained DBL. In total, more than 600 distinct domain names were newly found and disappeared from the database. Both plus and minus statistics showed a similar trend, except the maximum number of changes per day, where the maximum number of new domain names found was around 4,000 domain names.

URLHaus

Figure 4.11 shows the responsiveness of URLHaus. As can be seen in the figure, URLHaus was one of the public DBL that contributes a lot to the global blacklisted domain names based on the number of unique domain names published. This measurement started on December 31, 2018 until February 12, 2019. On average, URLHaus published more than 30,000 domain names each day, and as the time goes by, the number was increasing.

- Statistics:

  1. Duration: 43 days (2018-12-31 - 2019-02-12).
  2. Total number of unique domain names: 36,746.
  3. Average database content: 32,408.00, average new entries: 32,408.00.
  4. Plus:

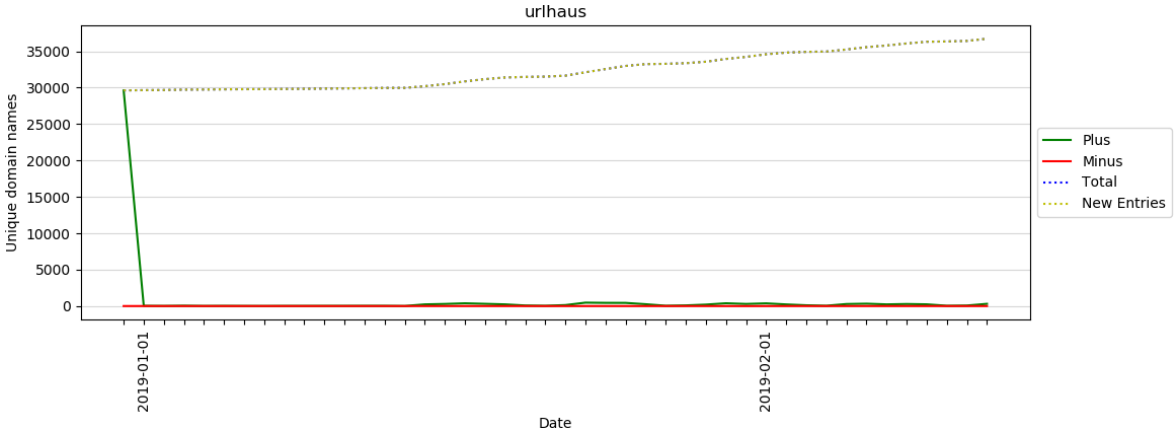

Figure 4.11: Agility of URLHaus

- Minimum: 10, Maximum: 467.
- Q1: 33.0, Median: 106.0, Q3: 282.5.
- Average: 165.56, Variance: 20,509.18, Standard Deviation: 143.21.

5. Minus:
   - Minimum: 0, Maximum: 6.
   - Q1: 0.0, Median: 0.0, Q3: 0.0.
   - Average: 0.30, Variance: 1.00, Standard Deviation: 1.00.

The statistics above shows that URLHaus was one of the DBL that tend to put new domain names into their database more than de-listing them. On average, more than 150 new domain names were found in URLHaus database, while the average number of de-listed domain names was less than one per day.

CyberCrimeTracker

Figure 4.12 shows the agility of CyberCrimeTracker. As can be seen in the figure, the number of blacklisted domain names remained quite constant, at around 10,000 entries, during the shorter observation period.

- Statistics:

1. Duration: 43 days (2019-01-01 - 2019-02-12).
2. Total number of unique domain names: 9,974.
3. Average database content: 9,947.23, average new entries: 9,947.23.
4. Plus:
   - Minimum: 0, Maximum: 6.
   - Q1: 0.0, Median: 0.0, Q3: 2.0.

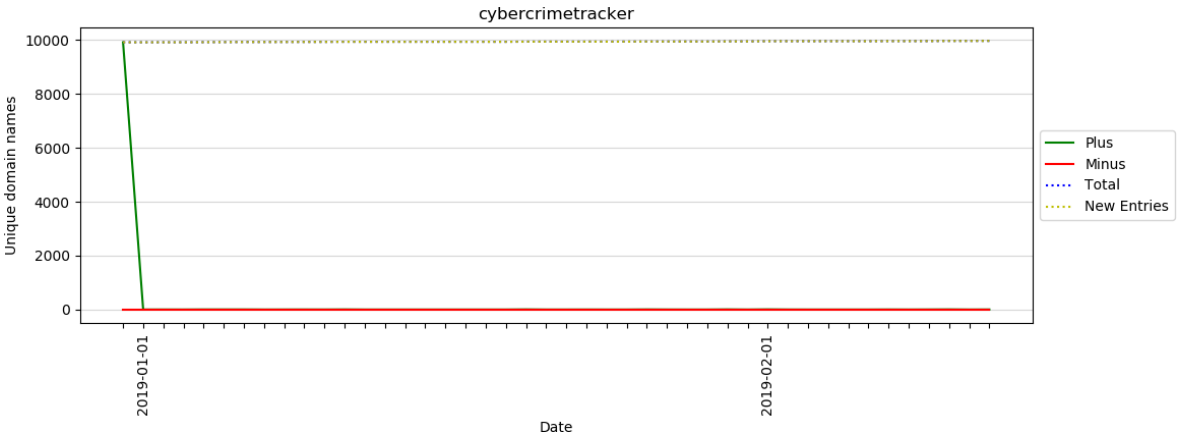

Figure 4.12: Agility of CyberCrimeTracker

- Average: 1.21, Variance: 3.00, Standard Deviation: 1.73.

5. Minus:
   - Minimum: 0, Maximum: 0.
   - Q1: 0.0, Median: 0.0, Q3: 0.0.
   - Average: 0.0, Variance: 0.0, Standard Deviation: 0.0.

As can be seen from the statistics, there were not much changes made to CyberCrimeTracker's database during the measurement period. The number of domain names published by this DBL remained constant around 10,000 unique domain names each day. During the 43-day measurement period, on average, one domain name was added into the blacklist, while none was removed, each day.

DNSBH

Figure 4.13 visualizes the agility of DNSBH. As can be seen in the figure, the number of blacklisted domain names also remained relatively constant, at around 23,000 entries, during the observation period that started on 2019. DNSBH was one of the DBL that less frequently update their database.

- Statistics:

  1. Duration: 43 days (2019-01-01 - 2019-02-12).
  2. Total number of unique domain names: 23,083.
  3. Average database content: 23,043.16, average new entries: 23,043.16.
  4. Plus:
     - Minimum: 0, Maximum: 9.
     - Q1: 0.0, Median: 0.0, Q3: 2.0.

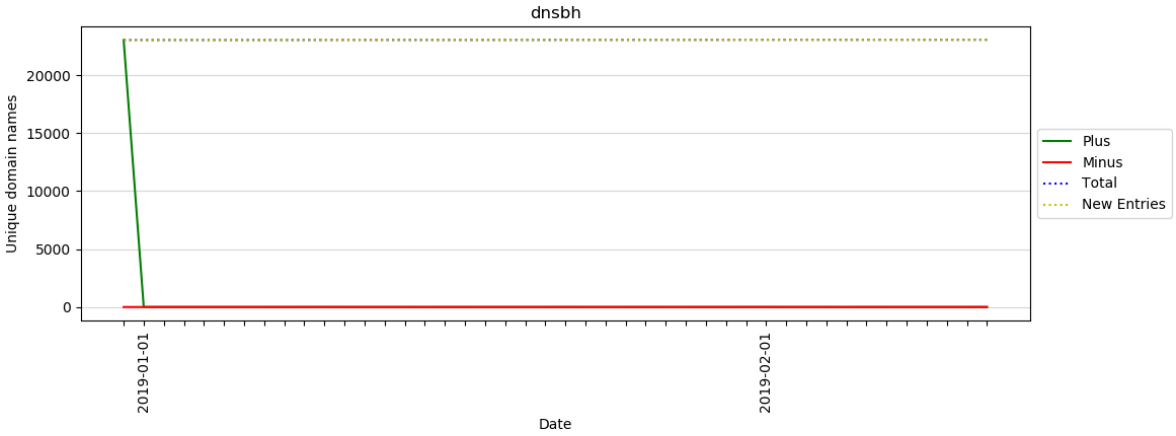

Figure 4.13: Agility of DNSBH

   - Average: 1.19, Variance: 3.73, Standard Deviation: 1.93.

5. Minus:
   - Minimum: 0, Maximum: 9.
   - Q1: 0.0, Median: 0.0, Q3: 1.0.
   - Average: 0.70, Variance: 2.77, Standard Deviation: 1.66.

DNSBH showed a steady number of unique domain names during the measurement period. There were relatively very small changes in the database, although it was more likely to see new domain names appearing into the database, instead of spotting de-listed domain names.

VXVault

Complete visualization of the agility of VXVault can be seen at Figure 4.14. From Figure 4.14, it can be inferred that VXVault was also one of the well-maintained DBL. Although the number of domain names contained in their database was relatively smaller compared to other blacklists, they had frequent changes of domain names.

- Statistics:

  1. Duration: 43 days (2019-01-01 - 2019-02-12).
  2. Total number of unique domain names: 570.
  3. Average database content: 73.88, average new entries: 73.88.
  4. Plus:
     - Minimum: 0, Maximum: 92.
     - Q1: 0.0, Median: 1.0, Q3: 9.50.
     - Average: 11.40, Variance: 518.47, Standard Deviation: 22.77.

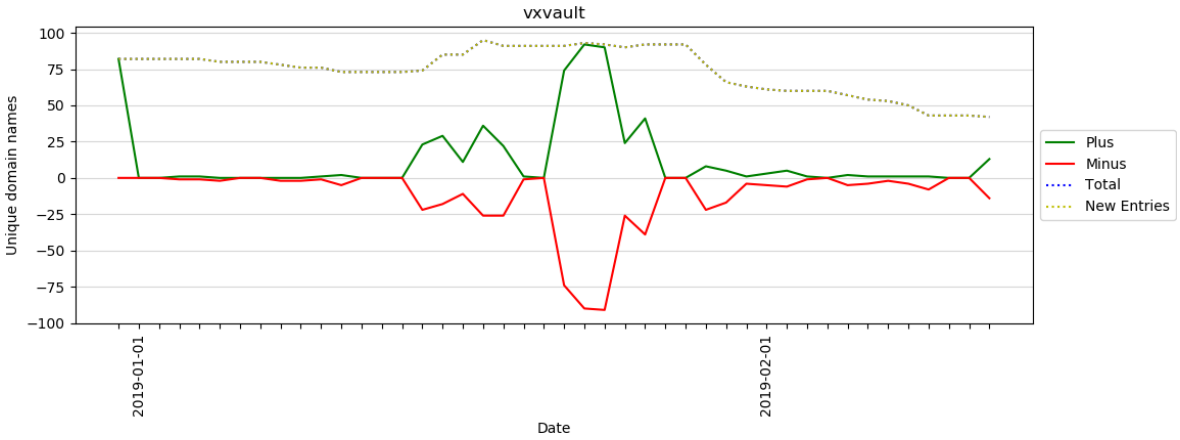

Figure 4.14: Agility of VXVault

5. Minus:
   - Minimum: 0, Maximum: 91.
   - Q1: 0.0, Median: 2.0, Q3: 15.50.
   - Average: 12.33, Variance: 488.73, Standard Deviation: 22.11.

This DBL showed a quite active behavior during the observation period. On average, the number of new domain names entered the database was quite similar with the number of domain names de-listed from the database. In addition, the changes could vary considerably, by seeing at the standard deviation and the variance values.

C2dom

Complete visualization of the responsiveness of the OSINT Feeds from Bambenek Consulting can be seen at Figure 4.15. From Figure 4.15, the published domain names did not have many significant changes over the observation period. The number of unique domain names was also relatively small, which could be caused by the fact that investigating command and control servers from a malware is relatively not as simple as tracing phishing websites or other malicious activities. One interesting information that can be achieved from this visualization is that, at late February 2019, the number of blacklisted domain names used for command and control servers doubled.

- Statistics:

  1. Duration: 43 days (2019-01-01 - 2019-02-12).
  2. Total number of unique domain names: 3,332.
  3. Average database content: 789.72, average new entries: 789.72.

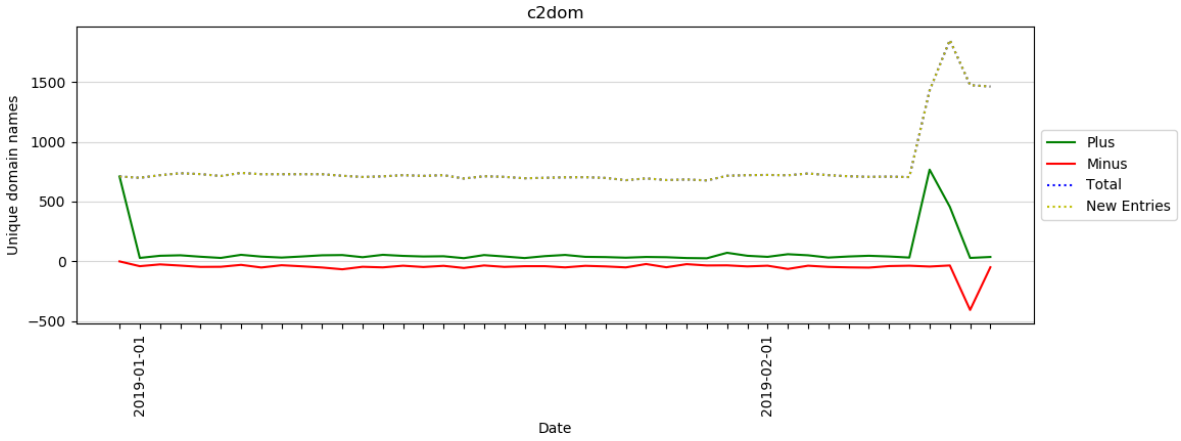

Figure 4.15: Agility of C2dom

4. Plus:
   - Minimum: 26, Maximum: 767.
   - Q1: 33.5, Median: 41, Q3: 51.
   - Average: 67.79, Variance: 15,645.14, Standard Deviation: 125.08.
5. Minus:
   - Minimum: 22, Maximum: 406.
   - Q1: 36, Median: 42, Q3: 49.50.
   - Average: 50.30, Variance: 3,099.61, Standard Deviation: 55.67.

The statistics of C2dom suggests that this DBL was also one of the frequently updated DBL. In 43 days of observation, more than 50 domain names were changed from the database.

Analysis on Agility of DBLs

Based on the results, the agility of public DBLs can be summarized in Table 4.22. In this table, `Update Days` shows the total number of days a DBL showed some changes in their released database, either a new domain name was found, or a domain name disappeared from the blacklist. The ratio was computed by taking the number of changes divided by the total duration of the observation. Column `Daily Changes` contains information about the average number of changes during the measurement period. Then, the rank, `#`, is computed by taking the average of *plus* and *minus* from each DBL, then divided by the average database content of each DBL. Both ranks for `Update Days` and `Daily Changes` are ordered by the ratio in descending order, since more frequent updates indicates better maintenance of a DBL.

As can be seen in this table, URLHaus and C2dom always updated their blacklist. However, for URLHaus, the number of updated domain names was relatively low

Table 4.22: Summary of DBLs Agility

DBL	Criteria					
	Update Days			Daily Changes		
	Total	Ratio	#	Average	Ratio	#
MDL	60 / 949	6.32%	12	1.15 / 908.12	0.13%	10
Joewein	777 / 949	81.88%	5	290.50 / 1,284.05	22.62%	1
Malc0de	607 / 949	63.96%	7	2.65 / 91.78	2.89%	5
ZTracker	195 / 949	20.55%	11	0.30 / 364.23	0.08%	12
RWTracker	296 / 949	31.19%	10	3.53 / 1,444.00	0.24%	9
Hostfile	846 / 949	89.15%	3	507.09 / 43,481.01	1.17%	7
HPHosts	53 / 864	6.13%	13	3,199.08 / 255,295.37	1.25%	6
Threat-Expert	31 / 820	3.78%	14	0.20 / 208.92	0.10%	11
OpenPhish	408 / 473	86.26%	4	316.67 / 1,477.67	21.43%	2
URLHaus	43 / 43	100.00%	1	82.94 / 32,408.00	0.26%	8
CCTracker	19 / 43	44.19%	8	0.62 / 9,947.23	0.01%	13
DNSBH	16 / 43	37.21%	9	0.95 / 23,043.16	0.00%	14
VXVault	30 / 43	69.77%	6	11.87 / 73.69	16.10%	3
C2dom	43 / 43	100.00%	1	59.05 / 789.72	7.48%	4

compared to the average database size. Overall, the ranks of C2dom show that they were one of the well-maintained blacklist, since they frequently change the database with considerable average number of changes. On the other hand, ThreatExpert's database was one of the least-frequently updated.

Chapter 5

Blacklists Liveliness

The complete explanation of measuring the liveliness of a DBL and its results are discussed in this chapter.

5.1 Description

One of the measures to determine how well public DBLs are maintained is to check whether their published domain names are actually active. It is important to verify whether the blacklisted domain names are actually active at the same time when they appear in a blacklist. When a domain name is put into a blacklist, it is expected that the domain name is still active and legitimate Internet users might be endangered by its existence. This is also one of the reasons why DBLs exist in the first place.

Liveliness of a DBL can be estimated by first checking the existence of a domain name. This can be done by using `dns.resolver` library from `dnspython` for Python. The next step is to *ping* the domain name and do port-scanning the blacklisted domain names using several commonly used ports. In this study, some of the selected Transmission Control Protocol (TCP) / User Datagram Protocol (UDP) ports are:

- Port 21. This port is commonly used as File Transfer Protocol (FTP) to transfer files from and to the server. This port is selected because it is expected that the administrators of the blacklisted domain names might need this port to transfer their source code into the server.

- Port 22. This port is commonly used for Secured Shell (SSH) to interact with the server. This port is selected because it is expected that most active servers will have this port open, since this port is one of the preferred secured communication method for the system administrators to interact with the system.

- Port 23. This port is commonly used as Telnet to interact with the server. This

port is selected because besides SSH, Telnet is one of the alternatives for a system administrator to communicate with the server. Telnet provides similar functionality as SSH, but in a less secure communication method.

- Port 25. This port is commonly used for Simple Mail Transfer Protocol (SMTP). This port is selected because some domain names relate with phishing might have their own mail service running in their system.

- Port 53. This port is commonly used for Domain Name System (DNS). This port is selected because some domain names might be mapped to several internal IP addresses. This is where the DNS service comes into play.

- Port 80. This port is commonly used for Hypertext Transfer Protocol (HTTP). This port is selected because it is expected that most of phishing, and some other malicious activities, websites will leave this port open so that their victims can access their website. When this port is open, generally, HTTP server is running in the server. Then, it is possible to retrieve HTTP response code of the blacklisted domain names.

- Port 443. This port is commonly used for Hypertext Transfer Protocol Secure (HTTPS). This port is selected because it serves the same purposes as port 80, but in a more secured way. One of the indicators to identify a benign website is by seeing the validity of its HTTPS certificate. As malicious domain names try to imitate legitimate domain names, serving HTTPS services could attract their victims to fall to their trap. When this port is open, generally, HTTP server with HTTPS support is running in the server. Then, it is possible to retrieve HTTPS response code of the blacklisted domain names.

However, one of the limitations on port scanning the aforementioned ports is the system's firewall, since it can be configured to only reply communication requests from specific IP addresses only. For instance, a server's firewall can be configured such that SSH-ing this machine could only be done using the system administrators' office network.

In addition, a machine could also be configured to run specific services using different ports. For instance, SSH service could be configured to use some random port instead of port 22. One of the reasons of doing this is that the malicious domains' administrators might want to hide their communication port, and one of the simple ways is to change the recommended communication port into some specific port.

To prevent any misunderstandings and simplify the naming, the Python application for investigating the liveliness of blacklisted domain names is called the *liveli-*

ness application and the machine used to run the liveliness test is referred as the *liveliness server*.

5.2 Requirements of The System

There are several challenges in checking whether blacklisted domain names are *live* or not. These challenges are described as follows.

1. System placement.
   Since the nature of blacklisted domain names might contain sophisticated active dangerous websites, it is recommended for the system to be placed in a secured isolated environment. Although the main goal of the application is to check the status of the server by looking for replies from several ports, the attackers could pinpoint the location of the liveliness server and attack this machine instead. Therefore, the system should be put in a special environment to minimize any harms caused by these sophisticated attackers and also to counter-attack this measurement.

2. System efficiency.
   Based on the data set used in this research, many different domain names could be mapped into one single IP address. This occurs quite often because one domain could have sub-domains and other names. However, accessing these domains could be redirected into one single machine. After checking with the WHOIS database, one of the examples is, one IP address was mapped into thousands of domain names. One of the possible reason for this is the web hosting service might use some DDoS Protection Service (DPS) to prevent Distributed Denial of Service (DDoS) attacks [45]. As stated in this paper written by Jonker *et. al.* [45], the web hosting providers can map several domain names into a DPS-assigned IP address.
   Therefore, it is crucial to keep track of the IP addresses when checking the liveliness of a domain name. When a lot of domain names are mapped into one single IP address, pinging and port-scanning the machine will return the same results. Therefore, it is important to prevent the application from inspecting the same server multiple times. Not only improving the efficiency of the application, port-scanning lots of entries at the same time could raise the respective system administrators' attention as they might think that this liveliness machine is used for some kind of cyber threats. In this case, it is for the good of both parties, since the blacklisted domain names' administrators might think that the liveliness check is some kind of random access, and for the liveliness server, this avoid real attackers into hacking into the machine.

3. Ethical considerations.
   "Poking" thousands of malicious domain names might bring several ethical consequences. Firstly, based on the data used in this research, more than 500,000 malicious domains were blacklisted every day. Sending liveliness requests and receiving the replies could flood the network infrastructure also the traffic to the web hosting providers.
   Secondly, as also mentioned in the previous point, a real attacker might administering blacklisted domain names. Poking some of their malicious domains might caught their attentions and attack this system instead. This could endanger any machines connected with the liveliness server.

Therefore, optimizing the source code to be able to check the liveliness of blacklisted domain names as efficient as possible is crucial. Not only to reduce the time taken to conduct the daily measurements, but also to prevent any unintended accidents to happen. Based on these requirements, the liveliness application is required to:

1. Finish the daily liveliness tests within 24 hours.
   Since the data from every DBL come daily, the liveliness test must be finished before the new data come. Therefore, the complexity of the liveliness test must be minimum to reduce the execution time and to be able to inspect more than 500,000 domain names each day.

2. Prevent inspecting the same IP address more than once.
   Based on the aforementioned considerations, it is important for the liveliness application to not "poking" at the same server more than once. Therefore, the list of checked IP addresses should be stored, so that the application will just take the result from the first inspection.

3. Also check associated domain names, such as with `www` label.
   In some cases, `domain.name` and `www.domain.name` could be mapped to different IP addresses. Therefore, it is important to also check the associated domain names, to also investigate their "side-services".

4. Be flexible.
   The desired application needs to be flexible in terms of which port to use and how it performs. The application should work as the users instruct. For instance, the users should be able to manually select which port to inspect, or set each port-scan timeout duration.

5.3 Application Flow

The general sequence of the processes in the liveliness application can be seen at Figure 5.1. Firstly, blacklisted domain names from each DBL is loaded into the application. Then, for each one of them, liveliness test is performed and the results are stored into `IP & Domain Database`. This database is required to prevent the liveliness application to inspect the same IP address or domain multiple times. Finally, the results of scanning domain names and checked IP addresses are exported from the database and stored into the output and log file. Besides these two output, general information, such as the application configuration and total time taken to conduct the liveliness test, are also stored in separate files.

The detailed instructions inside the liveliness test can be seen at Figure 5.2. This

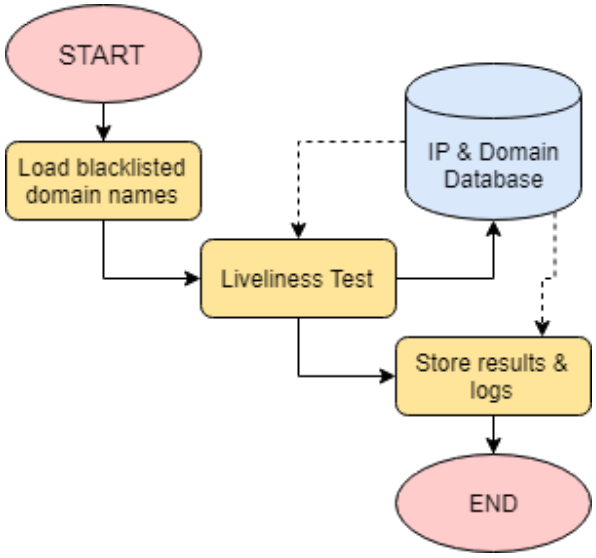

Figure 5.1: General Flowchart of Liveliness Application

flow is performed for each blacklisted domain name and the `IP & Domain Database` keeps track of the checked domain names and IP addresses. Note that, for each of the domain name, `dig` is performed on both `domain.name` and `www.domain.name`. Firstly, for each domain name checked, its existence is checked, whether querying the domain name into DNS servers will return the corresponding IP address or not. If the domain name does not exist (NXDomain), it will directly be stored into the database and the liveliness test for this domain stops.

The process continues if the domain name can be resolved to an IP address. Before performing the port-scanning, the IP address is first checked from the database. If the IP address has already been scanned, the information is retrieved from the DB and the process jumps to checking port 80. If the IP address does not exist in the database, the port-scan is performed first before continuing

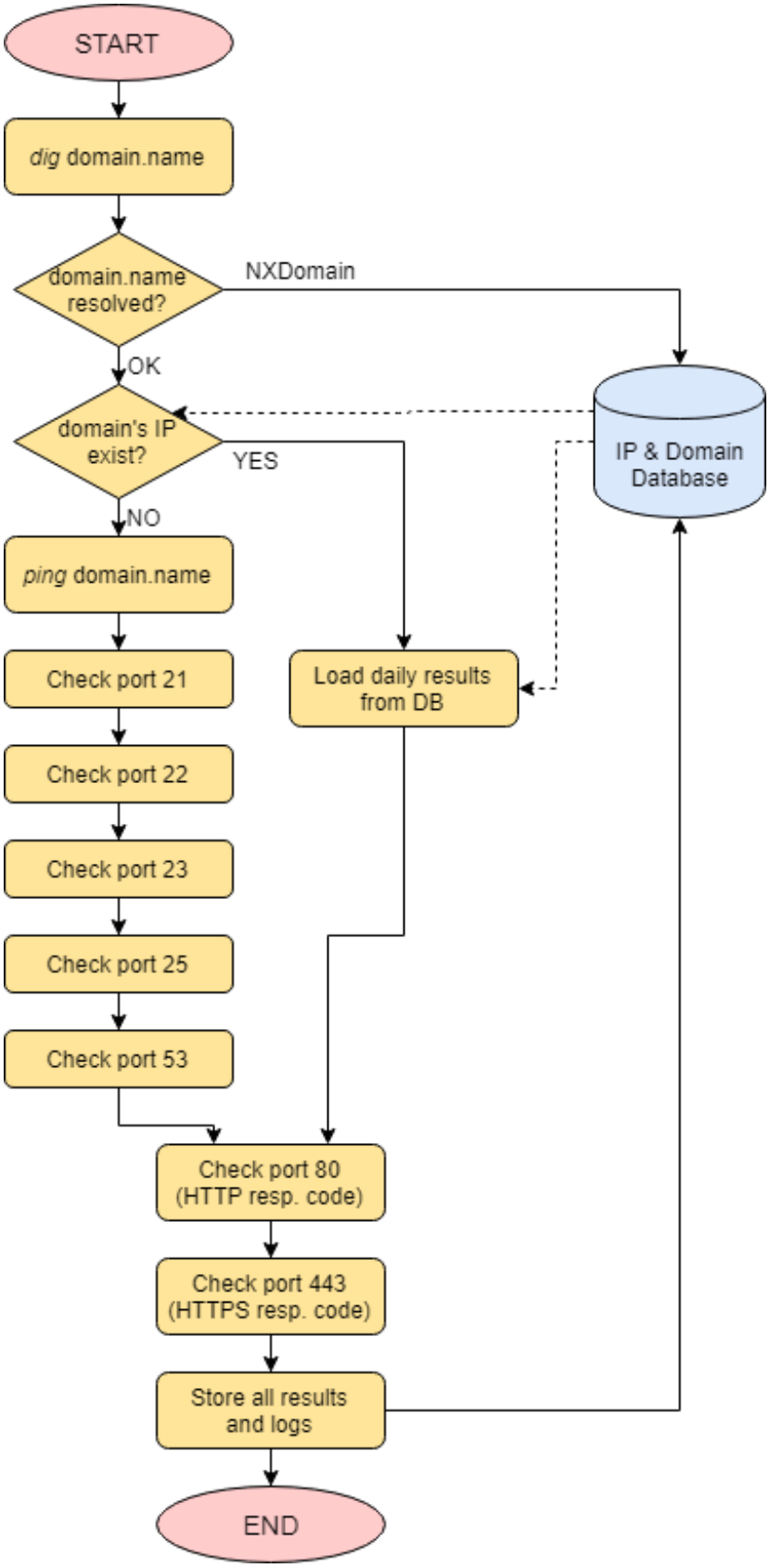

Figure 5.2: Detailed Flowchart of Liveliness Application

to checking HTTP response code. Finally, the results of port-scanning (if it is the first time scanning the IP address) and checking the liveliness of a domain name is stored into the database. The process continues with the next domain names.

5.4 Performance Measurements

After running the liveliness application for 22 days with two different scenarios, normally the script needed around 10 hours to execute one complete run. On average, more than 800,000 unique domain names were found each day. These domain names were mapped to around 140,000 unique IP addresses. This means that at least, one IP address served more than 5 different domain names.

The liveliness tests can be split into two phases. Firstly, from `2019-06-24` until `2019-07-05` the application just checked *ping* answers, SSH port status, HTTP and HTTPS response codes. Over the two-week period, the liveliness application was proved to be stable and quick enough to handle a huge number of unique domain names. Based on the 11 results, the system took around 10 hours to completely execute the application.

Then, more ports were added to the script since `2019-07-05` onward. The tests were executed to also check port 21, 23, 25, and 53, on top of SSH, port 80 and 443 from the previous tests. These measurements took approximately 10 hours and 30 minutes to complete.

CPU-consumption-wise, both scenarios used around 110% of CPU usage. This happened because of the optimization using Python's multiprocessing thread library and the timeout for each port-scans. The job assignments were split into the workers to check the liveliness of each domain name. Then, the workers were joined together after testing all domain names and the process continued with preparing and storing the output files.

5.5 Preliminary Results

The liveliness application has been executed to completely test the liveliness for 22 days (June 24, 2019 until July 15, 2019) and the results are shown in this section.

5.5.1 General Information

In general, the statistics of input files are as follows.

- Total days: 22

- Minimum: 500,711, Maximum: 531,990.
- Q1: 530,615.75, Median: 530,983.50, Q3: 531,257.50.
- Average: 528,641.23, Variance: 45,270,041.72, Standard Deviation: 6,728.30.

When running the liveliness application, the number of domain names to be checked was increased by around 50% because some domain names had different IP address when `www.` was added at the front of the blacklisted domain names. The stats of the number of domain names to be checked daily are as follows.

- Minimum: 748,968, Maximum: 812,182.
- Q1: 806,556.25, Median: 807,779.00, Q3: 810,361.75.
- Average: 804,607.64, Variance: 164,550,252.50, Standard Deviation: 12,827.71.

On the other hand, the number of IP addresses shrunk for about 70% when compared with the initial number of unique domain names from the input files. The stats of the number of unique IP address "poked" each day are as follows.

- Minimum: 143,282, Maximum: 151,948.
- Q1: 146,016.25, Median: 146,101.50, Q3: 146,187.75.
- Average: 146,131.59, Variance: 2,255,765.88, Standard Deviation: 1,501.92.

5.5.2 Phases Description

As mentioned in the previous section, currently, the results can be split into two categories, the simple and the complete version. Each of these categories are described in this section.

Simple Liveliness Test

This test was meant to measure the capabilities of the liveliness application. In this phase, the application just *ping*-ed the domains' IP address, checked the status of port 22, and retrieving HTTP and HTTPS response code. Nevertheless, the liveliness of a domain can still be tested, based on the status of port 22 and response code of HTTP and HTTPS requests. Based on the results of executing this test for 12 days, the visualization of the test can be seen at Figure 5.3.

Figure 5.3 shows the daily changes of the number of *live* domain names during this simple liveliness test. In this graph, there are nine different lines representing

different results of performing the liveliness test. Green straight line indicates the total number of unique domain names per day. Red straight line visualizes the number of domain names that actually exist and can be mapped into a specific IP address. Blue dotted line shows the number of NXDomains, which are the non-existence domain names. Yellow straight line represents the number of domain names that can be *ping*-ed during the observation period. Red dotted line indicates the result of sending SSH request to the machines. Cyan straight line represents the number of machines responded 200 as the HTTP requests, while magenta dotted line shows other replies from sending HTTP requests. Finally, cyan dotted line represents the number of machines responded 200 as the HTTPS requests, while green dotted line shows other replies from sending HTTPS requests.

As can be seen from Figure 5.3, there were more NXDomains than the domain

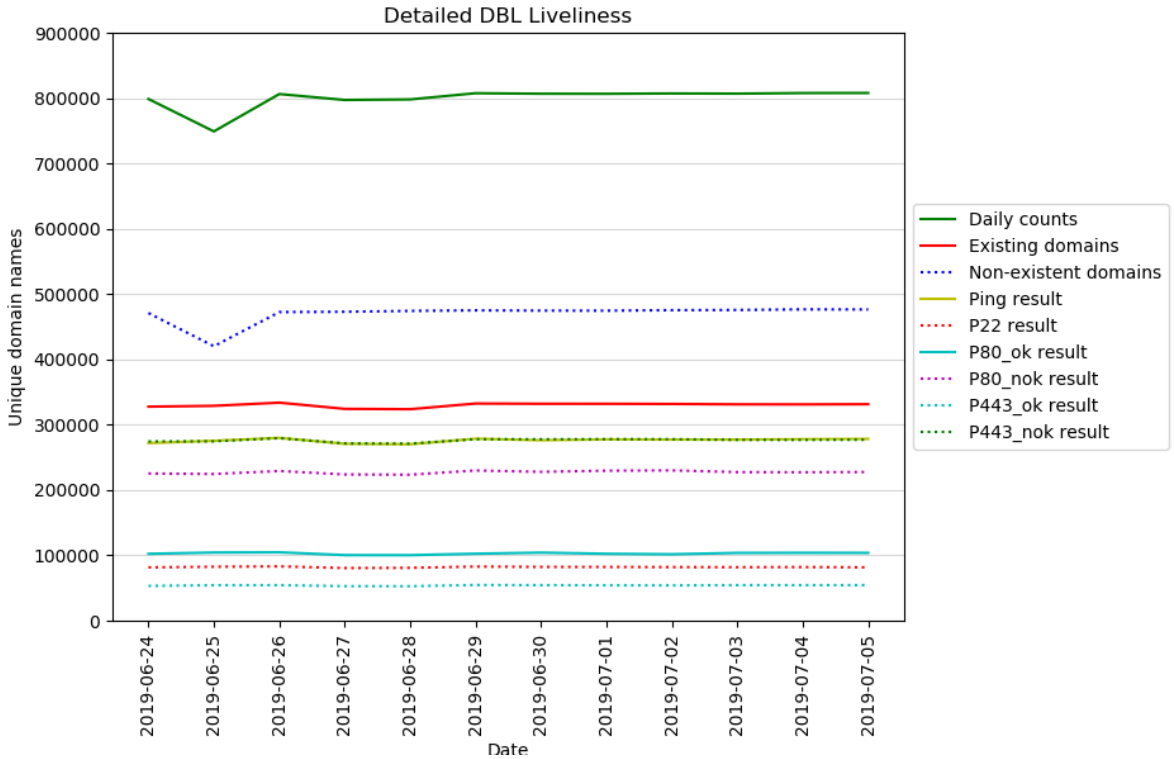

Figure 5.3: Detailed Graph of DBL liveliness from 2019-06-24 to 2019-07-05

names that actually existed. On average, almost 60% of the blacklisted domain names actually did not exist, while the existing domains contributed to only around 40% of the total number of unique domain names per day. 80.60% of these existing domains were actually *live* and could be *ping*-ed. In addition, both HTTP and HTTPS responses showed that more not-OK responses, such as 403 (*Forbidden*), 301 (*Moved Permanently*), or 503 (*Service Unavailable*), were retrieved, compared to HTTP 200 OK responses. The detailed statistics of the checked domain names in 12-day observation are shown at Table 5.1.

Similar to the previous statistics published in this research, the columns represent the stats category, minimum value, 25th, 50th, and 75th percentile, maximum, average, variance and the standard deviation of the number of unique domain names captured during the observation period. Additionally, row *Live* `machines` indicates the machines that were likely to be active during the blacklisted date.

In this simple liveliness test, a machine is considered to be active if:

1. The domain exists, and

   (a) It replies to ping requests, or

   (b) It replies to SSH requests (open port 22), or

   (c) It sends `200 OK` as either HTTP (port 80) or HTTPS (port 443) response codes.

To summarize this simple liveliness test, Figure 5.4 displays the general visualization of the number of *live* machines during this 12 day of measurement. On average, out of around 330 thousands domain names that existed, around 90% of them were active. This is indicated by yellow straight line in Figure 5.4

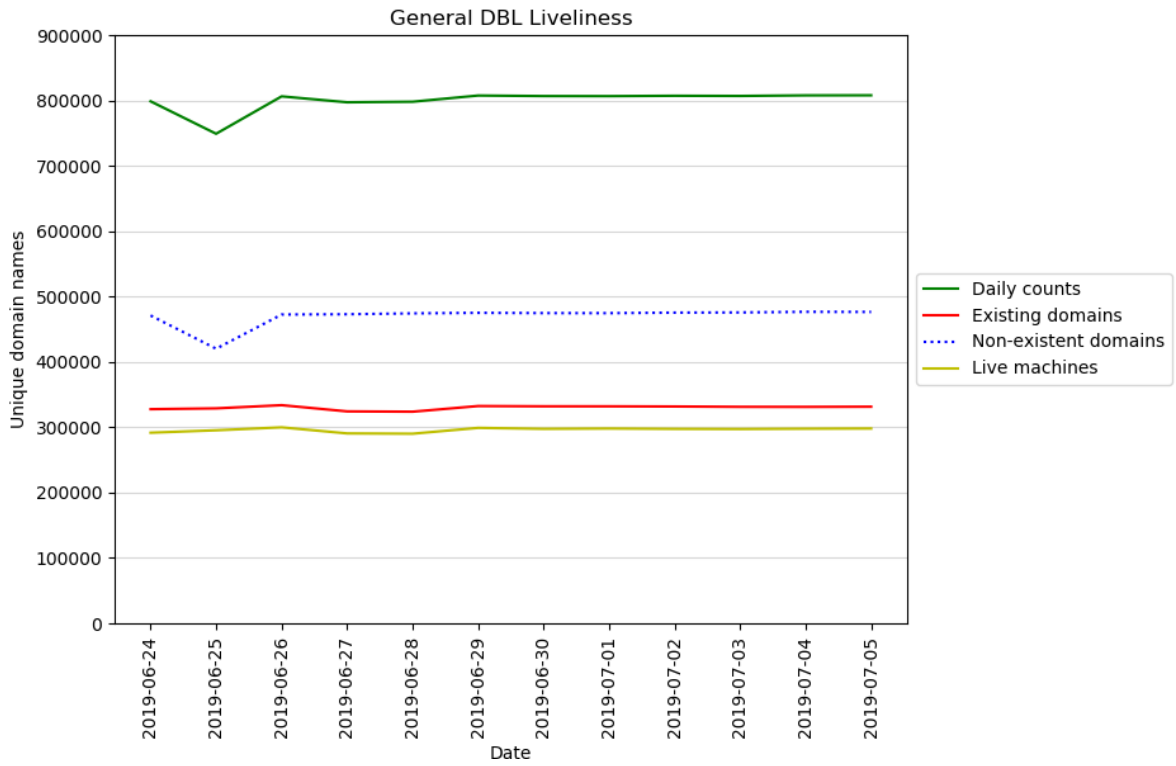

Figure 5.4: General Graph of DBL liveliness from 2019-06-24 to 2019-07-05

Table 5.1: Simple Liveliness Test Result

Category	Min	Q1	Med	Q3	Max	Avg	Var	SD
All domains	748,869	798,418.25	806,584.50	807,167.75	807,804	799,897.92	251,281,124.74	15,851.85
Existing domains	323,758	328,564.25	331,294.00	332,007.25	333,750	330,014.00	9,555,622.17	3,091.22
NXDomains	420,091	472,754.00	474,570.00	475,401.50	476,573	469,883.92	227,876,873.58	15,095.59
Ping OK	270,094	274,458.25	277,211.00	277,937.75	279,827	275,894.08	9,499,790.24	3,082.17
Open p22	80,591	81,664.75	82,067.50	82,369.00	83,105	81,988.08	483,104.74	695.06
HTTP OK	100,318	102,185.00	103,145.00	103,967.75	104,667	102,865.50	2,048,416.25	1,431.23
HTTP NOK	223,440	225,088.00	227,492.50	229,207.25	230,056	227,148.50	5,119,702.25	2,262.68
HTTPS OK	52,782	53,786.25	54,306.50	54,347.50	54,627	53,969.33	378,629.06	615.33
HTTPS NOK	270,976	274,560.50	276,938.00	277,769.50	279,404.00	276,044.67	6,445,099.22	2,538.72
Live machines	290,090	294,407.25	297,526.00	298,038.00	299,846.00	296,052.67	10,481,510.22	3,237.52

Complete Liveliness Test

The second testing scenario was an extension of the Simple Liveliness Test, but with more ports to be scanned. This test is considered to be the real liveliness tests of blacklisted domain names.

Based on the 11-day measurement period, the visualization graph can be seen at Figure 5.5 below. In this graph, in general, there was not much changes in the number of domain names during the measurement period, as the lines are relatively stable. Green straight line shows the total number of unique domain names daily. Blue and red straight lines indicate the non-existent and existing domain names respectively. Just below the 300,000 mark, there are two lines, the yellow straight line representing the *ping* result and green dotted line indicating the other responses from HTTPS requests. Then, the magenta dotted line shows the other responses from HTTP requests. Another green dotted line, around 100,000 unique domain names, represents the opened port 21. Yellow dotted, cyan straight, purple straight, and red dotted lines indicate the results of scanning port 25, HTTP OK response, port 53, and port 22 respectively. Finally, the cyan dotted line and blue dotted line shows the OK response of HTTPS requests and opened port 23.

As can be seen at Figure 5.5, again, the number of non-existent domains

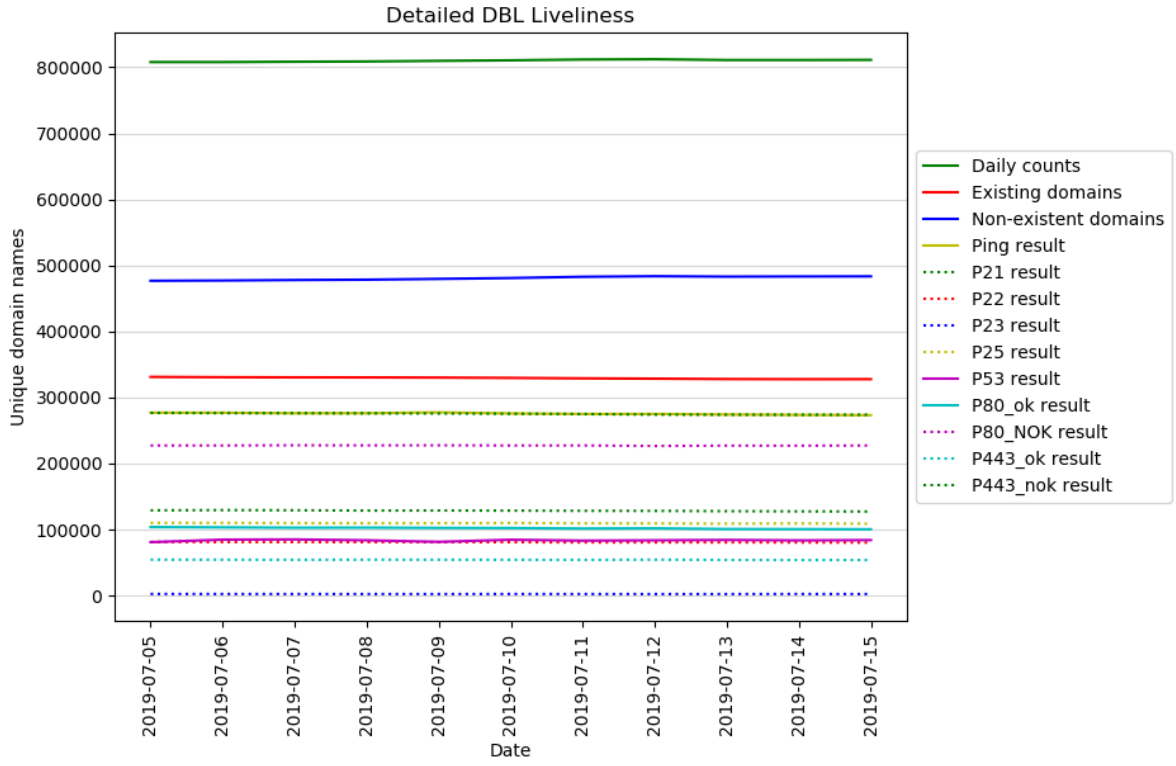

Figure 5.5: Detailed Graph of DBL liveliness from 2019-07-05 to 2019-07-09

was higher than the existing ones. In addition, taking both HTTP and HTTPS re-

sponses as the consideration, the number of OK responses was visibly lower than other responses, like timeout or not found. The number of existing domain names that opened common ports were quite spread. Out of around 330k existing domain names, the most common open port was port 21, which stayed around 39%. The least common port to be used by domain names was port 23, which was predictable since most of the usage of Telnet has mostly been replaced by SSH. The detailed statistics of the observed domain names can be seen at Table 5.2.

In this complete liveliness test, a machine is considered to be active if:

1. The domain exists, and
   (a) It replies to ping requests, or
   (b) It replies to any of FTP, SSH, Telnet, SMTP, or DNS requests (open port 21, 22, 23, 25, or 53), or
   (c) It sends `200 OK` as either HTTP (port 80) or HTTPS (port 443) response codes.

The summary of the liveliness test of blacklisted domain names can be seen at Figure 5.6. Similar to the result of simple liveliness test, the number of active machines lied around 91% of the number of domain names that exist.

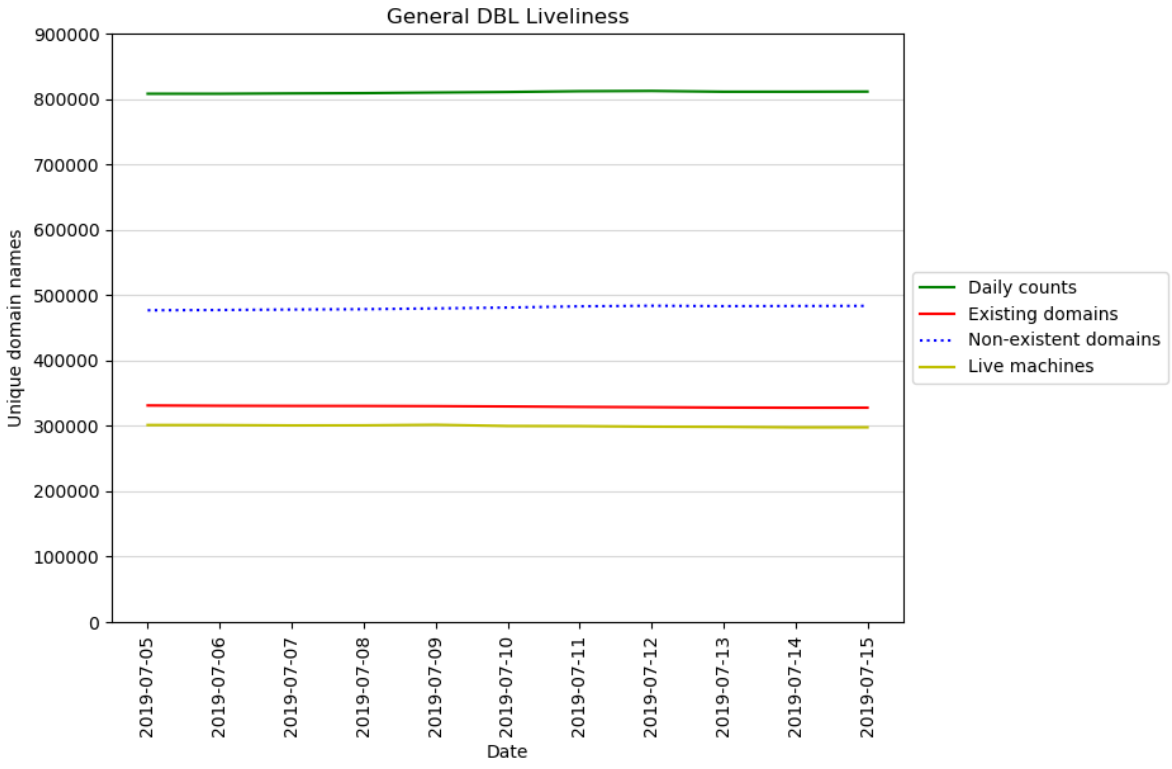

Figure 5.6: General Graph of DBL liveliness from 2019-07-05 to 2019-07-09

Table 5.2: Complete Liveliness Test Result

Category	Min	Q1	Med	Q3	Max	Avg	Var	SD
All domains	807,865	606,601.50	810,568.00	811,105.50	812,182	810,036.09	2,238,361.54	1,496.12
Existing domains	327,719	328,225.00	329,667.00	330,462.00	331,165	329,413.00	1,485,809.45	1,218.94
NXDomains	476,729	478,139.50	480,901.00	483,156.50	483,670	480,622.91	6,878,229.72	2,622.64
Ping OK	273,226	274,591.00	275,824	276,601.50	277,445	275,516.09	1,816,936.81	1,347.94
Open p21	127,363	128,128.00	128,662.00	129,021.00	129,487	128,528.91	427,533.72	653.86
Open p22	80,350	80,615.50	80,885.00	81,009.00	81,212	80,815.64	65,777.69	256.47
Open p23	2,369	2,393.00	2,413.00	2,462.50	2,520	2,427.27	2,030.38	45.06
Open p25	108,856	109,286.00	109,616.00	109,887.50	110,135	109,550.45	172,152.79	414,91
Open p53	80,948	83,439.00	83,794.00	84,455.00	85,094	83,600.36	1,555,369.32	1,247.14
HTTP OK	100,433	101,189.50	102,333.00	102,184.00	103,818	102,101.09	1,156.665.90	1,075.48
HTTP NOK	226,617	227,231.50	227,347.00	227,456.00	227,685	227,311.91	81,485.54	285.46
HTTPS OK	53,874	54,048.50	54,248.00	54,373.00	54,421	54,202.91	38,232.26	195.53
HTTPS NOK	273,845	274,011.50	275,419.00	276,158.00	276,744	275,210.09	1,146,217.17	1,070.62
Live machines	297,638	298,567.00	299,758.00	300,987.50	301,678	299,763.55	1,965,020.07	1,401.79

5.5.3 Liveliness of DBLs' Blacklisted Domain Names

After knowing the general liveliness of blacklisted domain names, it is interesting to find the ratio of active blacklisted domain names in each DBL. To do this, results from complete liveliness test were used.

Liveliness Visualization

One of the quickest method to see how much live domains were contained in a DBL is by using line chart. In the charts below, three distinct lines are used. Firstly, the green straight line represents the total number of domain names that were contained by each DBL per day. Secondly, the yellow straight line indicates the existence of the domain names. This means that the domain names could be translated into some IP addresses. The third line, which is the purple dotted line, shows the number of active machines. The results of performing liveliness test on 13 domain blacklists can be seen at Figure 5.7 to 5.19.

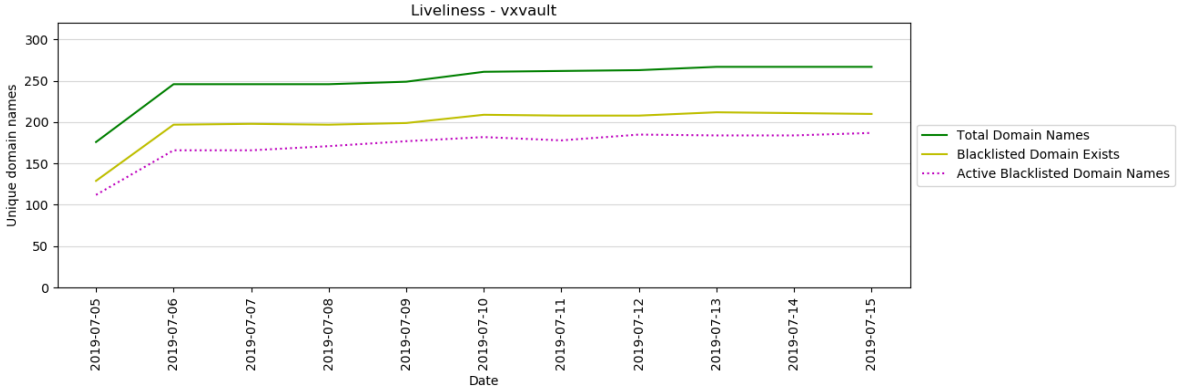

Figure 5.7: DBL liveliness of VXVault

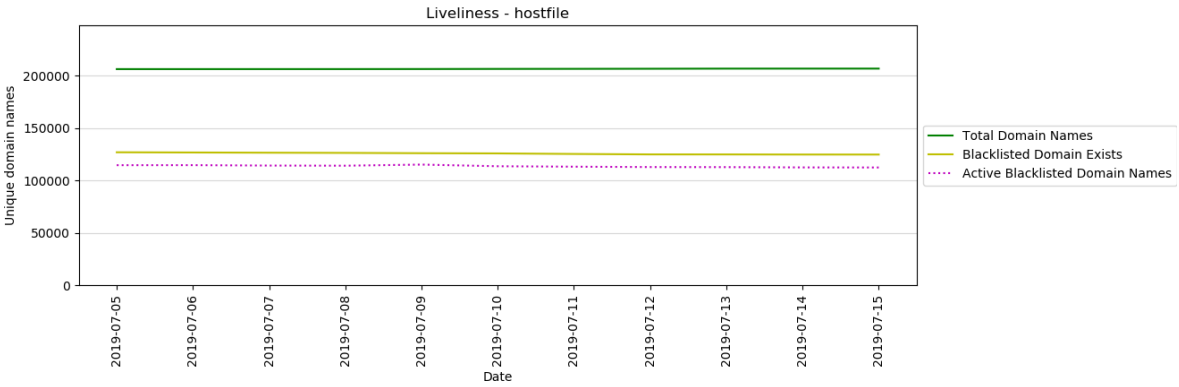

Figure 5.8: DBL liveliness of "partial" file update of Hostfile

From these figures, three liveliness categories could be extracted. Firstly, VX-Vault, OpenPhish, Joewein, and C2dom showed a relatively high ratio of active domain names. Secondly, Hostfile, URLHaus, Malc0de, MalwareDomainList, and HPHosts blacklisted active machines at around half of their published domain names. Thirdly, RansomwareTracker, DNSBH, ZeusTracker, and CyberCrimeTracker published a relatively low amount of active servers out of their published blacklisted domain names.

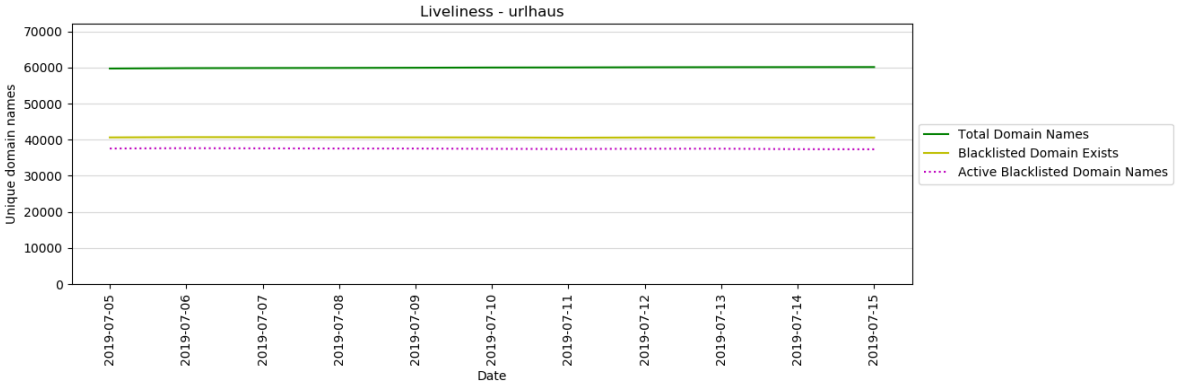

Figure 5.9: DBL liveliness of URLHaus

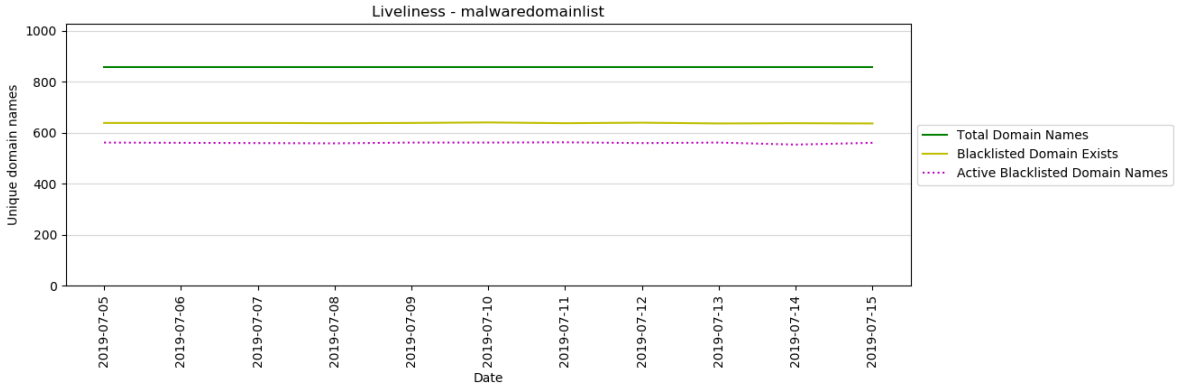

Figure 5.10: DBL liveliness of MalwareDomainList

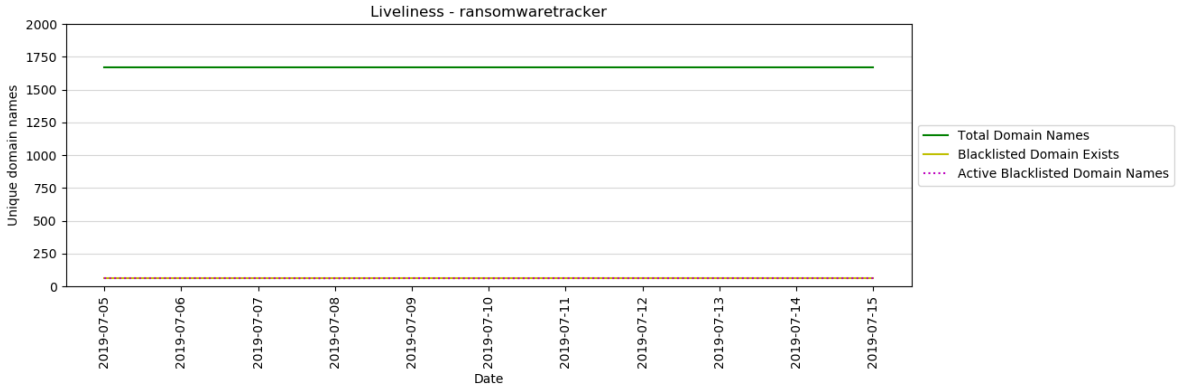

Figure 5.11: DBL liveliness of RansomwareTracker

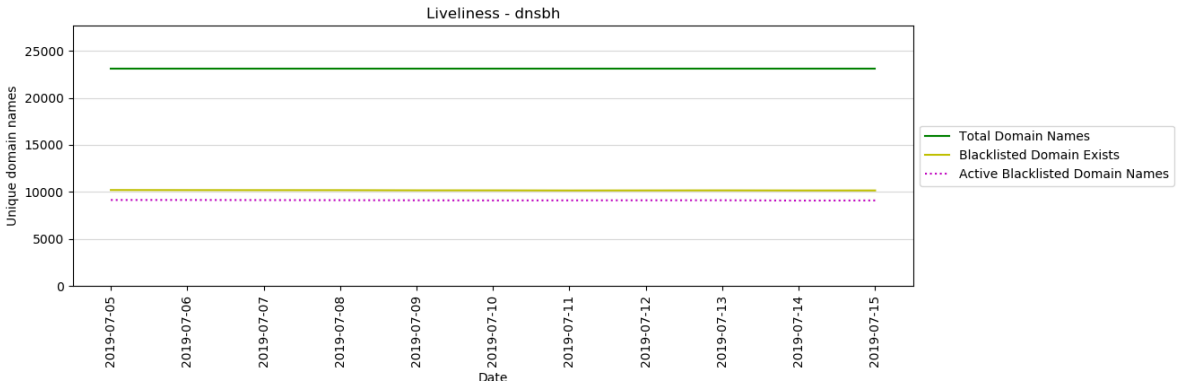

Figure 5.12: DBL liveliness of DNSBH

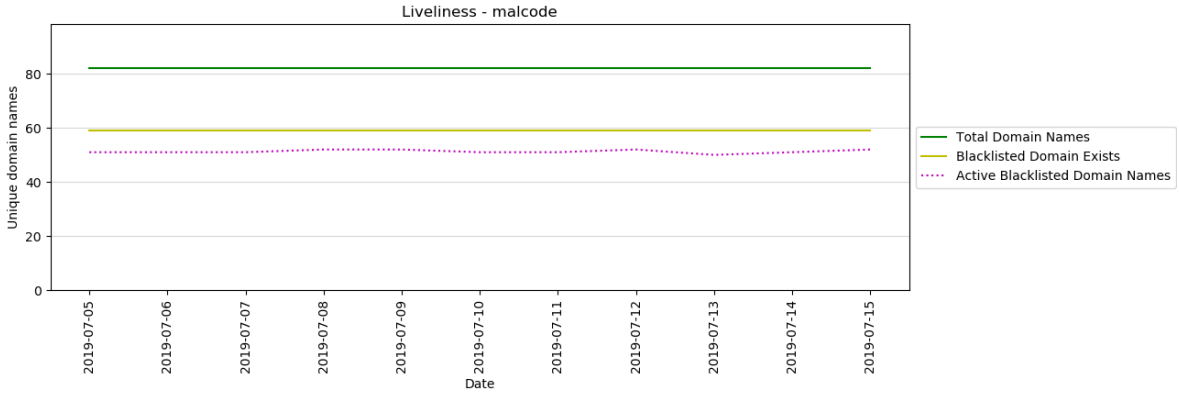

Figure 5.13: DBL liveliness of Malc0de

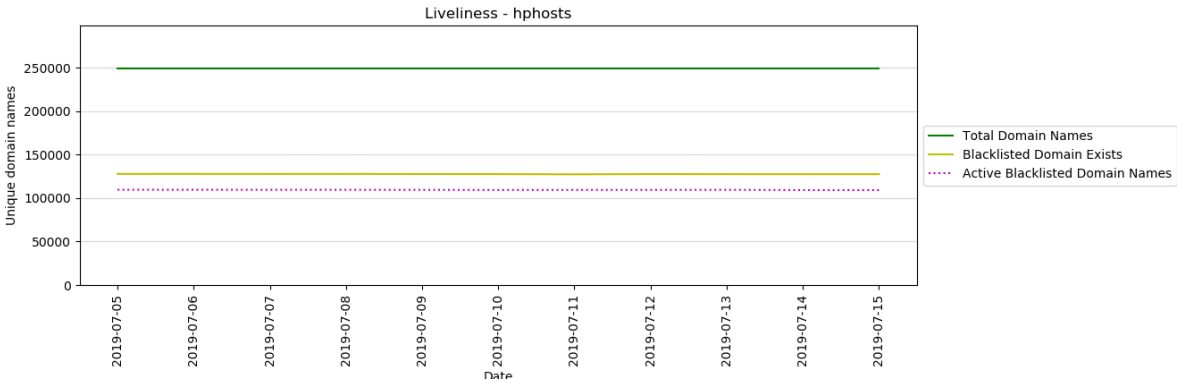

Figure 5.14: DBL liveliness of HPHosts (“full” file update of Hostfile)

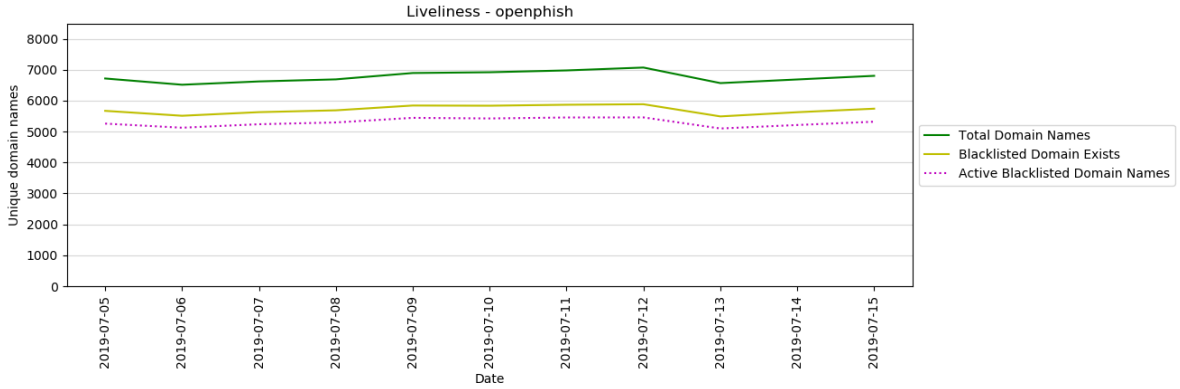

Figure 5.15: DBL liveliness of OpenPhish

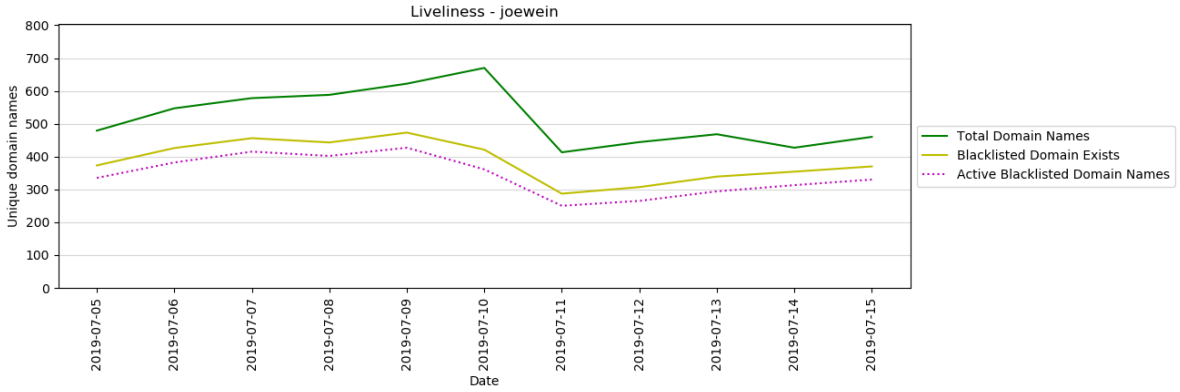

Figure 5.16: DBL liveliness of Joewein

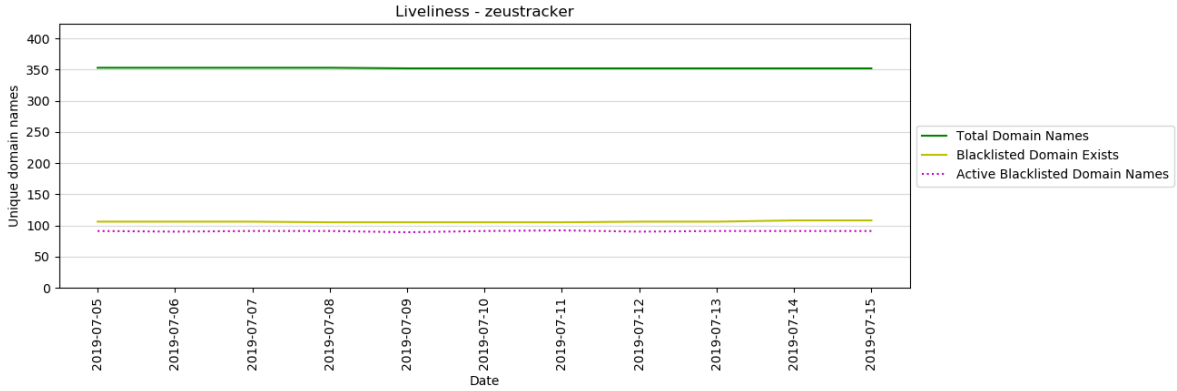

Figure 5.17: DBL liveliness of ZeusTracker

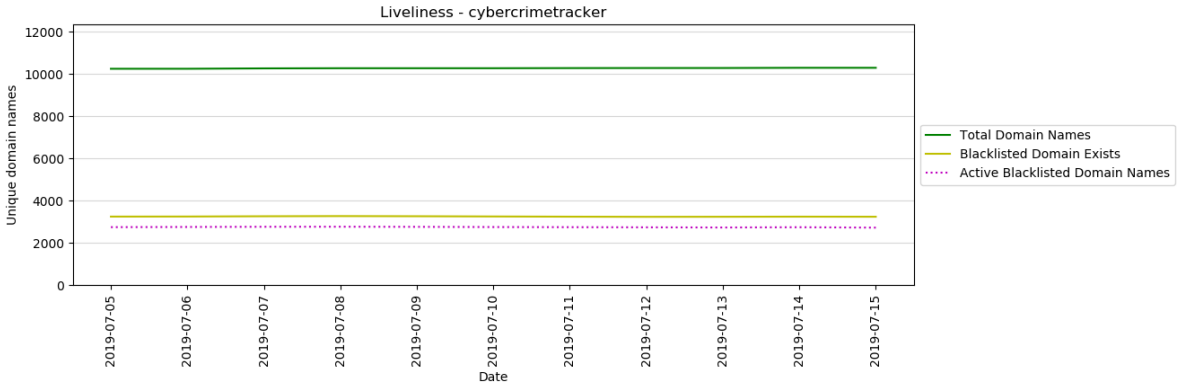

Figure 5.18: DBL liveliness of CyberCrimeTracker

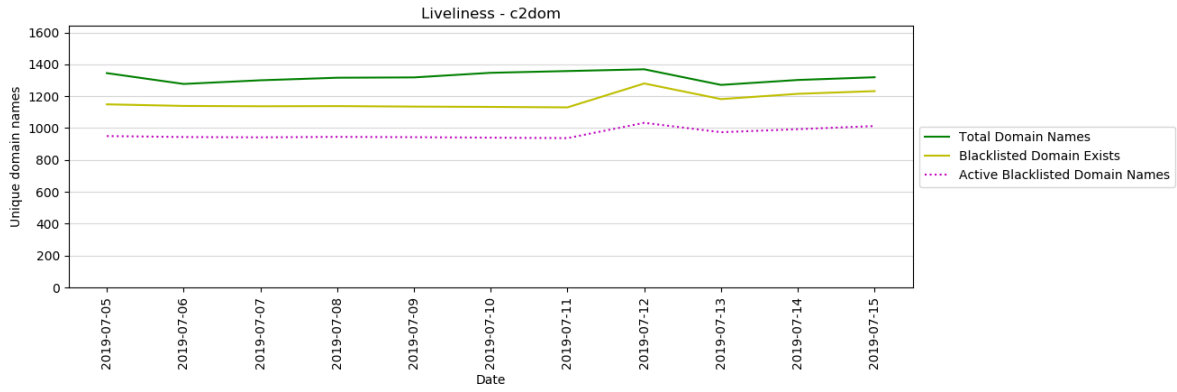

Figure 5.19: DBL liveliness of C2dom

DBL Liveliness Statistics

Another way to determine the liveliness of a DBL is by comparing the statistics of the active servers associated with blacklisted domain names. The overall comparison can be seen at Table 5.3.

Table 5.3: DBL Liveliness Stats

DBL	Avg. Total	Avg. Exist	Avg. Active	Liveliness Ratio	Rank
VXVault	257.40	204.90	178.00	69.15%	3
Hostfile	206,672.90	125,646.00	113,561.20	54.95%	8
URLHaus	60,002.00	40,634.90	37,498.70	62.50%	7
MDL	856.00	637.60	559.40	65.35%	5
RWTracker	1,667.00	63.0	61.80	3.71%	13
DNSBH	23,081.00	10,163.90	9,104.20	39.44%	10
Malc0de	82.00	59.00	51.30	62.56%	6
HPHosts	248,873.00	127,662.00	109,447.50	43.98%	9
OpenPhish	6,772.50	5,710.10	5,306.00	78.35%	1
Joewein	521.70	387.60	343.90	65.92%	4
ZTracker	352.30	106.00	90.70	25.75%	12
CCTracker	10,281.10	3,245.30	2,742.60	26.68%	11
C2dom	1,317.70	1,172.10	966.40	73.34%	2

In this table, column `DBL` lists the blacklists used in this study. `Avg. Total` indicates the average total number of blacklisted domain names by each DBL. Then, `Avg. Exist` shows the average number of domain names that existed. Column `Avg. Active` contains the values of the average number of domain names that were actually active during the blacklist date. The *activeness* of a DBL is considered based on the responses of *ping*, FTP (port 21), SSH (port 22), Telnet (port 23), SMTP (port 25), DNS (port 53) requests, as well as HTTP and HTTPS (port 80 and 443) response codes. Then, column `Liveliness Ratio` is computed by finding the percentage of `Avg. Active` out of `Avg. Total` for each DBL. The last column, `Rank`, indicates which DBL had the highest and lowest ratio of active machines ordered by the ratio in decreasing order.

Based on the stats, surprisingly, smaller DBLs were more likely to have more active servers, compared to the DBL that published a higher total number of blacklisted domain names. As can be seen in the table, the top five of the liveliness rank is mostly filled with DBLs that published less than 1,000 domain names per day. DBLs that published larger number of blacklisted domain names, such as Hostfile, only contained around 50% of active machines. On the other hand, RansomwareTracker had the least ratio of active machines at just less than 4% of their released blacklisted domain names.

Chapter 6

Discussions, Future Work, and Conclusions

This chapter explains the limitations of this study, future works to be done, and the summary of what have been done in this study.

6.1 Discussions

In this study, some points to be considered are as follows.

Existence of a *Ground Truth* Until this research is conducted, the existence of a ground truth that contains all malicious activities happening in the Internet still remains an open issue. Firstly, capturing all cyber incidents from multiple categories requires a lot of manpower and high costs. Collecting all malicious activities intuitively requires help from the victims themselves to report the real incidents to the correct parties, such as DBLs or the police. This means that the people also need to be educated about the cyber incidents awareness to do necessary responses that hopefully could reduce the damage and future threats.

Secondly, even if the ground truth does exist, processing them could require relatively high costs (storage, computing power, and network infrastructures). In this research, the daily files that contain blacklisted domain names from multiple blacklists consumed around 25MB storage capacity. Collecting these files for a year would consumed a considerably large amount of disk space. This also means that, transferring and processing these data would consume a lot of bandwidth and computing power.

Some of the previous studies have attempted to introduce some methods to capture a wider coverage of malicious activities, such as through sinkholes. However, these approaches are not enough to capture the whole malicious activities in the

world. For instance, some phishing attempts that were made in a specific country, might not be captured by sinkholes focusing on another countries.

Liveliness Test on Real Malicious Domains Another important point to consider based on this research is related with the impacts of *poking* malicious domain names. In this research, the liveliness tests were conducted only by using the commonly used ports for seven basic services of a server. As mentioned beforehand, system administrators of malicious domains might want to hide their services from the *cyber cops*, since what they are doing is illegal in some countries. This could be done by re-configuring the servers to use some random ports other than the default ports. This means that, domain names that are considered to be *inactive* from the liveliness test in this research might actually be active, using different ports.

However, it is unethical to perform complete port-scans to the servers, because not all of the blacklisted domain names are malicious. With the latest security technologies, it is quite easy to spot port-scans against a server. This means that, the liveliness server might be mistakenly considered as malicious instead. Even with only the seven specific ports, this has actually happened in this study. Reports saying that the IP address of the liveliness server appearing in a blacklist have been received and created some problems for the server providers. This was mainly caused by sending and receiving an abnormal amount of network traffic each day.

In addition, a DBL might not fully contain real malicious domain names. Some of the blacklisted domain names were actually sinkholes. They were designed and monitored by cyber security organizations for some non-malicious purposes, such as educating the citizens. However, it is not simple to distinguish real malicious domain names and sinkholed domains. Performing port-scan against these servers might also be one of the factors causing the liveliness server's IP address getting blacklisted instead.

6.2 Future Works

As the extensions of this study, several ideas are interesting to investigate, namely:

Characterization of *Premium* DBLs Firstly, it is interesting to compare the maintenance and documentation of premium DBLs, that could also be used by antivirus applications. This could provide a more complete understandings of the *behavior* of more domain blacklists.

Purity Improvements The purity of a DBL (how much of a domain blacklist is actually malicious) can be improved in further studies. In this research, the maliciousness of a domain name could not be fully determined since the private API key for VirusTotal could not be acquired. Using just the public key, it is impossible to verify the maliciousness of domain names published on daily updates from many DBLs due to the requests limits set by VirusTotal. However, there are many ways to check the history of cyber incidents involving blacklisted domain names. One of the recommended approach is by using private key for VirusTotal's web scanner. This allows blacklisted domain names to be scanned and their maliciousness could be verified by using multiple scanning services supported by VirusTotal.

Detection and Distinction of Sinkholes One of the limitations in this research is that, some of the blacklisted domain names analyzed were blacklisted for non-malicious purposes. Therefore, it is interesting to identify these sinkholes and evaluate the behavior differences of the real malicious domain names and the sinkholed domains.

Liveliness Test Improvements Another essential work to be done in the future is to make the liveliness application run more efficiently and create less problems. For instance, this can be done by distributing the liveliness test to several servers, so that each servers will not need to make a considerably high amount of traffic.

6.3 Conclusions

To conclude this report, in this study, the complete characterization of DBLs has been conducted by taking eight metrics. The answers for the main Research Question and its sub-questions are summarized in this section.

The answer for the first sub-research question, "*In which proportion does a DBL source contribute to the overall new blacklisted domains intake?*", is answered at Section 4.3.2 and 4.3.1. In short, Joewein, RansomwareTracker, and C2dom have the highest ratio of exclusive domains, while VXVault, HPHosts, and ThreatExpert are the bottom-three in terms of domain exclusiveness. With regard to the pairwise comparison, domain names blacklisted by Hostfile have a higher tendency to appear at other DBLs, as Hostfile is one of the DBLs with the largest number of unique domain names. This is also supported by the ratio of domain names that intersect with the aggregated list that can be seen at Table 4.14.

The second sub-question, "*What is the level of details each DBL source pro-*

vide?", is answered at Section 4.3.6. By considering both the blacklisting and de-listing information provided by each DBL, ZeusTracker, Hostfile, and MalwareDomainList provide more information about the blacklisted domain names compared to other sources used in this study. On the other hand, DNSBH and ThreatExpert are two of the DBLs that publish the least amount of information about their blacklisted domain names.

Then, "*How quick do a DL source blacklist and remove domain names?*" has been answered at Section 4.3.3 and 4.3.4. The results show that Hostfile, MalwareDomainList, and Joewein have the tendency to blacklist and de-list domain names earlier than other DBLs. ZeusTracker is shown to be one of the DBLs that update their database a bit later than other DBLs.

Finally, "*Do DBL sources contain domain names that are currently active?*" has been answered at Chapter 5. The results at Table 5.3 show that OpenPhish, C2dom, and VXVault contain the highest ratio of active domain names from their total blacklisted domain names. On the other hand, RansomwareTracker, ZeusTracker, and CyberCrimeTracker show the highest ratio of inactive machines.

Answering the main Research Question, "*How well are publicly available domain blacklists from different categories documented and maintained?*", Table 6.1 could provide a summary of the results of this study.

As can be seen at Table 6.1, each DBL has its own strengths and weaknesses. For instance, ZeusTracker is one of the DBL that publish more information about the blacklisted domain names, however, in terms of the responsiveness, it is one of the slowest to update their database. Taking Hostfile as another example, this is one of the DBL with more *very good* results compared to other sources in most of the compared metrics. However, they have relatively low number of exclusive domains.

Table 6.2 to 6.5 summarize the strengths and weaknesses exposed from each DBLs used in this research. As can be seen from these strengths and weaknesses, it is difficult to pick one single DBL as the best, based on the used metrics. Each DBL, like Hostfile, could have several good points. However, they also showed some limitations.

To conclude, this research has shown the characteristics of 13 publicly available domain blacklists. Metrics used in this research were selected to measure the maintenance and documentation of public DBLs although the ground truth is still relatively difficult to achieve. Furthermore, the specified metrics could still be applied to determine the characteristics of many other public DBLs in the future. Even though three of the DBLs used in this research were deprecated during the observation period, the methodologies and approaches could still be applied to provide useful information about public DBLs.

Table 6.1: DBLs Blacklisting Accuracy

DBL	Purity		Coverage			Responsiveness		Specificity	Accuracy	Agility		Liveliness
			Exclusive Domains	Pairwise with Aggregated List								
	Registered Domains	Intersection with Alexa		Historical	Newer	Start	End			Update Days	Daily Changes	
C2dom	12	3	3	-	8	-	-	4	10	1	4	2
CCTracker	10	5	4	-	5	Late	-	10	4	8	13	11
DNSBH	8	4	10	-	4	Late	-	11	13	9	14	10
Hostfile	4	6	11	1	2	Early	Early	12	2	3	7	8
HPHosts	5	8	13	2	1	Early	Late	14		13	6	9
Joewein	2	1	1	3	7	Early	Early	1	8	5	1	4
Malc0de	1	14	9	5	13	Early	Late	7	6	7	5	6
MDL	9	13	5	7	10	Early	Early	9	3	12	10	5
OpenPhish	6	9	6	4	6	Late	Early	8	10	4	2	1
RWTracker	13	2	2	6	9	Late	-	2	4	10	9	13
ThreatExpert	14	11	12	9	-	-	Early	5	12	14	11	-
URLHaus	7	7	7	-	3	Late	Late	13	7	1	8	7
VXVault	3	10	14	-	11	Late	Early	3	8	6	3	3
ZTracker	11	12	8	8	12	Late	Late	6	1	11	12	12

Table 6.2: DBLs Strengths and Weaknesses Summary

DBL	Strengths	Weaknesses
C2dom	• Frequently updated. • More than 70% of the blacklisted domains were active. • Low intersection with Alexa (few benign domains blacklisted).	• Low number of registered domains. • Few information about blacklisted domain names provided.
CCTracker	• More details of the blacklisted domains were described.	• Low number of registered and active domains. • Blacklisted many benign domains.
DNSBH	• One of the large contributors to the overall malicious activities. • One of the fewest intersection with benign domain name list.	• Low number of exclusive domains. • Low number of active blacklisted machines.

Table 6.3: DBLs Strengths and Weaknesses Summary (continued).

DBL	Strengths	Weaknesses
Hostfile	• Provide details of the blacklisted domains. • One of the major contributors to the overall malicious activities. • Frequently updated. • Tend to blacklist and delist earlier than other DBLs.	• Around 50% of blacklisted domain names were active.
Joewein	• High number of exclusive domains for mail spamming DBLs. • High number of registered domains. • Frequently updated.	• Less information provided about the blacklisted domain names.
Malc0de	• High number of registered domains.	• High intersection with list of benign domains. • Low number of published domain names.
MDL	• More details given about blacklisted domain names. • High responsiveness.	• High intersection with Alexa top 100k websites. • Database update frequency was low.

Table 6.4: DBLs Strengths and Weaknesses Summary (continued).

DBL	Strengths	Weaknesses
OpenPhish	• Frequently updated. • High number of active domain names.	• Low accuracy of blacklisted domain
RWTracker	• High number of exclusive domains. • Low intersection with list of benign websites.	• Low number of registered domains. • Less frequently updated.
ThreatExpert	• Medium specificity.	• This service has been deprecated. • Low number of registered domains. • High intersection with list of benign websites. • Low maintenance of the database.
URLHaus	• One of the major contributors to the overall malicious activities. • Database was updated quite frequently.	• Low reponsiveness. • Medium number of active blacklisted domain names.

Table 6.5: DBLs Strengths and Weaknesses Summary (continued).

DBL	Strengths	Weaknesses
VXVault	• High number of registered blacklisted domains. • Quite frequently updated. • High number of live domain names.	• Relatively high intersection with Alexa top 100k websites. • Low number of exclusive domains. • Less information given about blacklisted domain names.
ZTracker	• One of the DBLs with detailed information about blacklisted domain names.	• This DBL was discontinued since July 8, 2019. • Low number of registered and active domains. • Database was less frequently updated. • Tended to be low responsiveness. • High intersection with list of benign websites.

Bibliography

[1] "Mdl," http://www.malwaredomainlist.com/, accessed: August 5, 2019.

[2] "joewein.de llc - fighting spam and scams on the internet," http://www.joewein.net, accessed: August 5, 2019.

[3] "Anti-phishing — eset," https://www.eset.com/au/phishing/, accessed: August 5, 2019.

[4] "Dbl - the spamhaus project," https://www.spamhaus.org/dbl/, accessed: August 5, 2019.

[5] M. Kührer and T. Holz, "An empirical analysis of malware blacklists," *Praxis der Informationsverarbeitung und Kommunikation*, vol. 35, no. 1, pp. 11–16, 2012.

[6] S. Sheng, B. Wardman, G. Warner, L. Cranor, J. Hong, and C. Zhang, "An empirical analysis of phishing blacklists," 2009.

[7] "Alexa top 500 global sites," https://www.alexa.com/topsites, accessed: August 5, 2019.

[8] "Virustotal," https://www.virustotal.com/#/home/url, accessed: August 5, 2019.

[9] A. Pitsillidis, C. Kanich, G. M. Voelker, K. Levchenko, and S. Savage, "Taster's choice: a comparative analysis of spam feeds," in *Proceedings of the 2012 Internet Measurement Conference*. ACM, 2012, pp. 427–440.

[10] M. Kührer, C. Rossow, and T. Holz, "Paint it black: Evaluating the effectiveness of malware blacklists," in *International Workshop on Recent Advances in Intrusion Detection*. Springer, 2014, pp. 1–21.

[11] N. Chachra, D. McCoy, S. Savage, and G. M. Voelker, "Empirically characterizing domain abuse and the revenue impact of blacklisting," in *Proceedings of the Workshop on the Economics of Information Security (WEIS)*, vol. 4, 2014.

[12] P. Gupta, B. Srinivasan, V. Balasubramaniyan, and M. Ahamad, "Phoneypot: Data-driven understanding of telephony threats." in *NDSS*, 2015.

[13] A. Noroozian, M. Korczynski, S. Tajalizadehkhoob, and M. van Eeten, "Developing security reputation metrics for hosting providers," in *8th Workshop on Cyber Security Experimentation and Test (CSET 15)*, 2015.

[14] S. Ramanthan, J. Mirkovic, and M. Yu, "Blacklists assemble: Aggregating blacklists for accuracy."

[15] Q. Scheitle, O. Hohlfeld, J. Gamba, J. Jelten, T. Zimmermann, S. D. Strowes, and N. Vallina-Rodriguez, "A long way to the top: Significance, structure, and stability of internet top lists," *arXiv preprint arXiv:1805.11506*, 2018.

[16] "Surbl," http://www.surbl.org/, accessed: August 5, 2019.

[17] "Phishtank — join the fight against phishing," http://phishtank.org/, accessed: August 5, 2019.

[18] "Malc0de," http://malc0de.com/, accessed: August 5, 2019.

[19] ""malc0de.com" - google scholar," https://scholar.google.nl/scholar?hl=en&as_sdt=0%2C5&q=%22malc0de.com%22&btnG=, accessed: August 5, 2019.

[20] "Zeus tracker :: Home," https://zeustracker.abuse.ch, accessed: August 5, 2019.

[21] "abuse.ch — it-security blog," https://abuse.ch, accessed: August 5, 2019.

[22] "Ransomware tracker," https://ransomwaretracker.abuse.ch, accessed: August 5, 2019.

[23] "Urlhaus — malware url exchange," https://urlhaus.abuse.ch/, accessed: August 5, 2019.

[24] "Safe browsing – google safe browsing," https://safebrowsing.google.com/, accessed: August 5, 2019.

[25] "hphosts online - simple, searchable & free!" https://hosts-file.net, accessed: August 5, 2019.

[26] "Internet storm center - sans internet storm center," https://isc.sans.edu/, accessed: August 5, 2019.

[27] "Network security suspicious domain list," https://www.networksec.org/grabbho/block.txt, accessed: August 5, 2019.

[28] "Openphish - phishing intelligence," https://openphish.com, accessed: August 5, 2019.

[29] “Cybercrime,” http://cybercrime-tracker.net/, accessed: August 5, 2019.

[30] “Dns-bh — malware domain blocklist by riskanalytics,” http://www.malwaredomains.com/, accessed: August 5, 2019.

[31] “Vx vault last 100 links,” http://vxvault.net/URL_List.php, accessed: August 5, 2019.

[32] “Master feed of known, active and non-sinkholed c&c domain,” http://osint.bambenekconsulting.com/feeds/c2-dommasterlist.txt, accessed: August 5, 2019.

[33] “pywhois - pypi,” https://pypi.org/project/pywhois/, accessed: August 5, 2019.

[34] “Website malware scanner — free online web scan for malware infections,” https://app.webinspector.com/, accessed: August 5, 2019.

[35] “Mdl-delisted,” http://www.malwaredomainlist.com/hostslist/delisted.txt, accessed: August 5, 2019.

[36] “Malzilla - malware hunting tool,” http://malzilla.sourceforge.net/, accessed: August 5, 2019.

[37] “Zeus tracker :: Monitor,” https://zeustracker.abuse.ch/monitor.php?filter=all, accessed: August 5, 2019.

[38] “Cryptowall tracker:: Overview,” https://www.cryptowalltracker.org/, accessed: August 5, 2019.

[39] “hphosts online - simple, searchable & free!” https://hosts-file.net/?s=History, accessed: August 5, 2019.

[40] “Malwarebytes cybersecurity for windows, mac, android & ios — malwarebytes,” https://www.malwarebytes.com/, accessed: August 5, 2019.

[41] “Dns-bh — malware domain blocklist by riskanalytics — bh dns files,” http://www.malwaredomains.com/?page_id=66, accessed: August 5, 2019.

[42] “Osint master feed,” osint.bambenekconsulting.com/feeds/c2-masterlist.txt, accessed: August 5, 2019.

[43] “Top 10 malware january 2018,” https://www.cisecurity.org/blog/top-10-malware-january-2018/, accessed: August 5, 2019.

[44] “Top 10 malware of august 2017,” https://www.cisecurity.org/blog/top-10-malware-of-august-2017/, accessed: August 5, 2019.

[45] M. Jonker, A. Sperotto, R. van Rijswijk-Deij, R. Sadre, and A. Pras, “Measuring the adoption of ddos protection services,” in *Proceedings of the 2016 Internet Measurement Conference*. ACM, 2016, pp. 279–285.