On the Potential of Proactive Domain Blacklisting

Mark Felegyhazi
mark@icsi.berkeley.edu

Christian Kreibich
christian@icir.org

Vern Paxson
vern@icir.org

International Computer Science Institute
Berkeley, California, USA

Abstract

In this paper we explore the potential of leveraging properties inherent to domain registrations and their appearance in DNS zone files to predict the malicious use of domains *proactively*, using only minimal observation of known-bad domains to drive our inference. Our analysis demonstrates that our inference procedure derives on average 3.5 to 15 new domains from a given known-bad domain. 93% of these inferred domains subsequently appear suspect (based on third-party assessments), and nearly 73% eventually appear on blacklists themselves. For these latter, proactively blocking based on our predictions provides a median headstart of about 2 days versus using a reactive blacklist, though this gain varies widely for different domains.

1 Introduction

One of the primary techniques for protecting people from financial scams, malicious web pages, and other nuisances on the Internet is the use of *blacklists*: continuously updated lists that enumerate known-bad entities that systems can check before potentially harmful interaction with an entity takes place. Upon finding the entity on a blacklist, the system prevents access and/or generates a warning indicating the danger. A large number of organizations maintain such blacklists, listing entities such as the IP addresses of senders of spam,[1] domain names or IP addresses involved in scams,[2] and URLs leading to malicious web pages.[3] Substantial filtering machinery exists throughout the Internet (for example in mail user/relay agents and web browsers) that queries these lists to recognize and treat accordingly entities known to be dangerous.

Blacklists provide the benefit of lookup efficiency: systems can conduct lookups quickly and precisely. However, blacklists have the major drawback of operating in an overwhelmingly *reactive* fashion: blacklist maintainers learn of malicious entities only after these entities have become active (e.g., due to messages appearing in a "spam trap" account, or a crawled web page returning malicious code). Thus, a window of vulnerability remains during which users can suffer from malicious exposure because an active entity has not yet appeared on a blacklist. Since the perpetrators of Internet crime operate their scam campaigns on infrastructures of substantial scale, however, once we have detected an initial seed entity of badness, we might have an opportunity to *predict* pending badness by other as-of-yet inconspicuous entities if we find these associated with the same perpetrators. Such *proactive blacklisting* would offer the major benefit of diminishing the window of exposure, thus often preventing malicious infrastructure from functioning before its operators even put it to use. On the other hand, the prediction mechanism must work with high accuracy to avoid causing "collateral damage" due to errors.

In this paper we take a first look at the potential of proactive blacklisting in the context of domain names. We observe that miscreants frequently register domains used in Internet scams in bulk, and operate them using related sets of name servers. We propose a method for inferring sets of malicious, not-yet-blacklisted domains based on initial "seed" domains that we observe used maliciously through their appearance on non-proactive blacklists. For our inference we draw upon DNS zone file data along with limited "WHOIS" domain registration data. We measure the accuracy of our predictions using a combination of several popular blacklists plus services that themselves make predictions about future misuse. We find that from a fairly modest set of initial seeds we can predict a large set of additional malicious domains, with arguably quite low false positives.

We next provide background on domain registration procedures and existing work on blacklisting (§ 2). In § 3 we describe our methodology in detail and follow with an evaluation of it using real-world blacklisting data (§ 4). We discuss our findings in § 5 and briefly conclude in § 6.

[1] E.g. CBL, SBL, SpamCop, and SORBS.
[2] E.g. ivmURI, JWSDB , SURBL, and URIBL.
[3] E.g. PhishTank, the SafeBrowsing API, and IE 8's SmartScreen service.

2 Background

Domain Registration. To register a domain, a customer interacts with a domain *registrar* accredited by ICANN to lease domains as permissible by the relevant top-level domain (TLD) *registry*, such as VeriSign for `.com`, or DENIC for `.de`. The registry is mainly responsible for coordinating the registration procedure for a given TLD and maintaining the corresponding domain registration database. When a domain becomes active, the registry includes its DNS information in the corresponding DNS *zone file*, which lists for each domain its authoritative name servers.

For our study we focus on `.com`, the largest collection of Internet domains. Its zone file lists authoritative name servers for each `.com` domain, along with the "glue" records for each name server that cannot be independently resolved. The zone file currently contains $\approx$ 80M domains, with $\approx$ 70–100K domains added and 70K domains deleted each day. We have obtained a daily snapshot of the `.com` zone file from VeriSign since May 2009, and hence can retrieve past associations between domains and name servers.

Related Work. Several studies have examined IP address blacklists that enumerate abusive senders, mostly to assess their effectiveness. Jung and Sit characterized spam traffic to an academic institution, finding lists covered up to 80% of spam senders, while 14% of the DNS lookups at the site were blacklist queries [1]. Ramachandran et al. developed techniques for leveraging blacklist queries in order to identify botmasters checking the listing status of their own bots [8], and presented evidence that while address-based blacklists may have limited coverage of a botnet's members, those bots that are detected are generally listed quickly [7]. Sinha et al. compared four prominent blacklists and found false negatives ranging from 35% to 98.4% and false positives between 0.2% and 9.5% [10].

Similar effectiveness studies exist for phishing URL blacklists. Sheng et al. found that two thirds of phishing campaigns last at most two hours before being listed, although coverage at the appearance of a campaign is generally poor [9]. By contrast, Ludl et al.'s comparison of the effectiveness of Google and Microsoft's phishing URL blacklists found that 90% of the campaigns studied were covered by Google's list at the time of the authors' initial query [3]. Makey compiled membership comparisons of sender blacklists from MTA logs of a large academic institution, finding that large blacklists generally provide broad coverage, while smaller ones frequently filter specific sender sets with high accuracy [5].

Another line of work aims to improve the accuracy of blacklists by distinguishing global and local "badness" information. Zhang et al. proposed Highly Predictive Blacklisting (HPB), a mechanism to customize global blacklists taking into account the relevance of different entries for local targets [13]. HPB strikes a balance between globally compiled blacklists (likely to contain irrelevant entries) and locally compiled ones (likely incomplete) by computing relevance scores for individual users of the blacklist. Soldo et al. expanded HPB by also factoring temporal considerations into the prediction algorithm [12]. Neither approach is truly proactive—they narrow the existing global offender list to one relevant for a particular blacklist subscriber, but do not predict novel arrivals on the blacklist. Sinha et al. proposed a similar approach but with the addition of proactive blacklisting of notorious sender networks unless they exhibit high (positive) relevance to a receiver network [11], leveraging the observation that spam senders frequently appear co-located in narrow network prefixes.

To avoid the reactive nature of blacklists, Ma et al. proposed a classifier leveraging host-based features exposed in URLs (such as IP addresses, WHOIS records, and geography) as well as their lexical structure. Based on a training corpus of URLs leading to malicious content they achieve 95–99% classification accuracy [4]. Prakash et al. likewise observed common lexical properties of URLs, and proposed a proactive filtering mechanism for phishing URLs by constructing likely URLs from known instances [6]. These approaches are complementary to ours.

Closest to our work is the "gold list" published by URIBL.[4] The list consists of domains predicted to appear on blacklists in the future. It contains 1,000s to 10,000s of domains, though from the statistics on the web site not many of these appear to indeed cross over to the regular URIBL blacklist.

3 Methodology

We base our approach on the insight that in order to operate scams in an ongoing fashion, miscreants must employ a sizable number of domains to avoid ready blacklisting. They can obtain large numbers of domains by registering in bulk with a given registrar. (In a previous study of spam campaign orchestration [2], we witnessed bulk registration of hundreds of domains at a time.) Leveraging this observation, we take known-bad domains as input and derive from them associated domains likely to see employment in related scams in the future. We call the set of initial known-bad domains the *seeds* and the prediction result the *inferred domain set*. Thus, our method operates in a proactive fashion given an initial reactive component to drive the prediction algorithm. Note that our approach is complementary to reactive domain assessments, such as employed in the

[4] http://www.uribl.com/gold.shtml

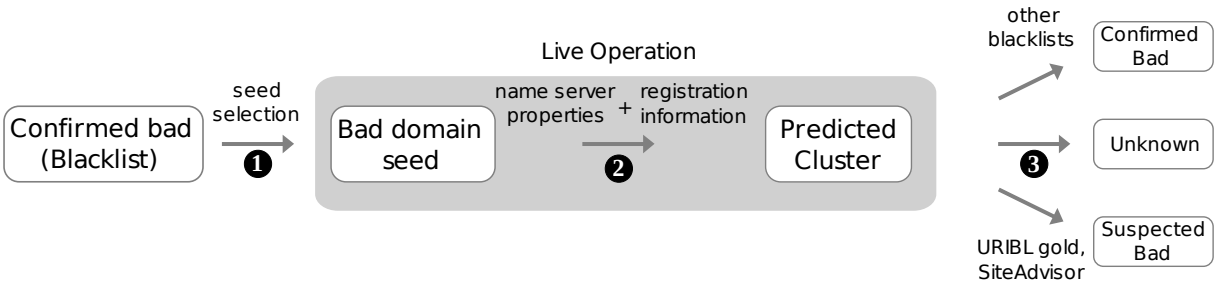

Figure 1: Experimental setup. ❶ Blacklisted entries in JWSDB are selected as seed domains. ❷ Clusters of predicted domains are produced using zone file and WHOIS information. ❸ Using additional sources, we quantify *correct*, *likely correct*, and *potentially incorrect* predictions. The shaded area indicates machinery required for live blacklist operation.

upcoming domain blocklist (DBL) of spamhaus.org—our approach could extend the set of domains evaluated by such assessments.

At a high level, our experimental setup operates in three stages, summarized in Figure 1. First, using a source of known-bad domains we select initial blacklisted domains as seeds. From these domains, we predict clusters of related domains likely to be blacklisted in the future based on nameserver features and registration information. We can apply our procedures for analyzing this features in either order: we only infer likely future malicious behavior for domains that have both the requisite name server features and the requisite registration features. Finally, we evaluate the accuracy of the predicted clusters using additional blacklists. We now discuss these phases in more detail.

3.1 Obtaining Bad Domains For Seeding

We seed our domain inference with a set of domains viewed as definitely malicious. For our study, we selected domains that appear on the blacklist provided by joewein.net (JWSDB) in January 2010. The JWSDB feed consists of a daily blacklist of malicious domains extracted from URLs seen in emails sent to mailboxes operating the spam filter software *jwSpamSpy*. JWSDB adds on the order of 500 new domains each day. We chose the JWSDB feed because it provides historic data on registration times.

We focused on the `.com` TLD for two reasons. First, it still dominates scams: over the past two years, it has accounted for 44% of all domains blacklisted in the JWSDB, followed by `.cn` (38%) and `.info` (8%). Second, we can obtain `.com`'s zone file, enabling access to historic name server information.

3.2 Name Server Features

Initially, we intended to infer bad domains based on common registration information. By itself, this lacks power in two ways. First, because registrars do not provide bulk listing of domains registered with them, we cannot readily extrapolate a bad seed to the full set of associated miscreant domains. Second, even if we could obtain such listings, if we only have the limited top-level WHOIS information then the fact that *benign* actors will also register domains with the same registrar on the same day makes it difficult to determine which domains in the listings indeed reflect the same miscreant.

We can address both of these considerations by leveraging domain zone information, when available. Among the information a zone file provides is an exhaustive list of all subdomains in the zone as well as their authoritative name servers (NSs). In addition, a domain's date of activation is implicitly provided by the domain's appearance in the list. We can thus use zone files to leverage our observation that miscreants manage their domains in batches, not only during registration, but also by serving multiple domains from the same NS.

We use the `.com` zone file to identify all authoritative NSs that have in the past resolved a JWSDB domain between May 2009 and January 2010. Figure 2 plots the distribution of the number of distinct name servers serving a given JWSDB domain. We observe that the majority of domains have only a few name servers during their lifetime, but some change name servers several times. Moreover, the domains that employ new or self-resolving name servers are likely to encounter more name servers than those domains that do not match any of our NS features. We hypothesize that these changes between new NSs reflect double-fluxing, i.e., the owner quickly changes the name server to avoid outages due to blacklisting of the NS itself.

We initially considered all such NSs as a potential source for inference, but this did not lead to satisfactory results: some of the NSs belong to major hosting companies, which host large numbers of legitimate domains as well. To avoid this problem, we observe that NSs for malicious domains tend to satisfy two criteria:

1. *Freshness:* The domain of the NS itself was registered only recently. For example, for NS

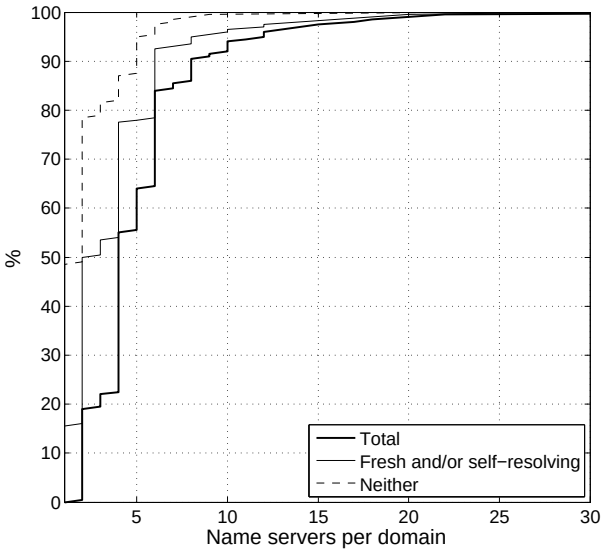

Figure 2: Distribution of the number of distinct NSs resolving a JWSDB domain over the course of its lifetime: total number of NSs (thick), fresh and/or self-resolving ones (thin), and those that are neither (dashed).

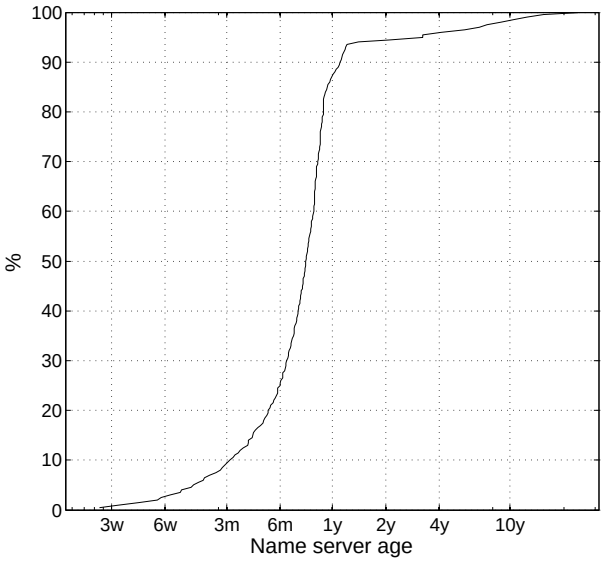

Figure 3: Distribution of all name server ages for domains blacklisted in the JWSDB between May 2009 and January 2010.

`ns.example.com`, the age of `example.com` is low. We use an age of less than one year to indicate youth—as shown in Figure 3, almost 90% of NSs involved in hosting malicious domains are younger than a year.

2. *Self-Resolution:* The NS resolves its own domain name. For example, `example.com`'s name server is `ns.example.com` rather than `ns.thirdparty.com`.

We leverage these two features as follows. If a bad domain switched to a new NS at time T, then we search for all domains that switched to the same NS at time T. Note that our NS-based inference is conservative, as there could be other pending-malicious domains that switch to the same NS but at a different time. If a domain switches to self-resolution at time T, then we search the entire zone file for all domains that switched to self-resolution at time T and with the same registration profile.

Figure 4 shows the distribution of NS features, grouped by the number of NSs employed by the seed domains in the JWSDB dataset. Our two criteria dominate all NS usage patterns, from a single NS up to 44, with the exception of a set of domains using 5 NSs (we discuss this case in § 5). 82.2% of all blacklisted domains encounter at least one new NS during their lifetime. Furthermore, many bad domains switch to a self-resolving NS at some point in time. Thus, the NS features of freshness and self-resolution hold promise for finding companion domains associated with known-bad domains.

3.3 Registration Information

Using WHOIS, we obtain registration information for the entire set of domains inferred using the two NS features. Our goal here is to narrow down the inferred set of domains to those that are co-registered with one of the seed domains. We call this remaining set of inferred domains the *inferred clusters*.

Before proceeding, we can double-check our basic assumption that miscreants register domains in groups. We performed WHOIS queries to obtain the registration information for all domains in the JWSDB blacklist from May 2009 through January 2010. Generally, a domain's WHOIS record provides the registrar's name and WHOIS server, the domain's authoritative name servers, the domain's current status, and the dates of domain registration, update, and expiration. One can further explore registration information by contacting the registrar's WHOIS server to obtain the name, address, phone number and email of the registrant, and the domain's administrative, technical, and billing contacts. However, registrars rate-limit queries to their WHOIS service, so for our assessment we only drew upon the initial set of general WHOIS information.

The majority of these registration groups are small: 50% contain only one domain and 10% have more than 25 members. We also find that the majority of domains do not belong to these small registration groups: 93% of the domains in JWSDB were jointly registered with another domain, on the same day, and using the same registrar and 80% of the JWSDB domains were registered in batches of at least 10 domains. Figure 5

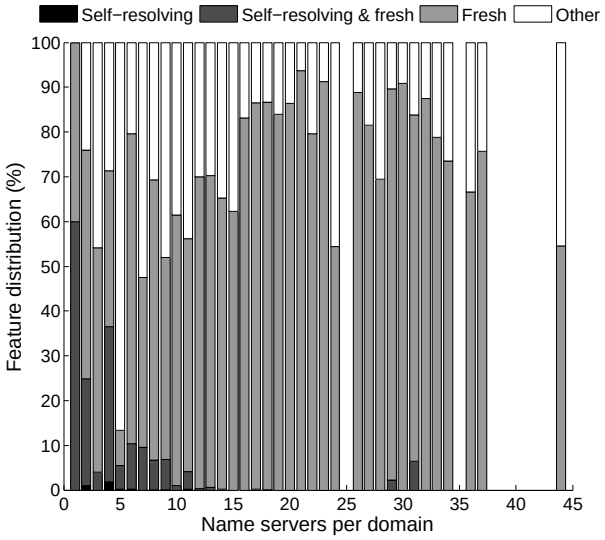

Figure 4: Distribution of NSs that are fresh and/or self-resolving, grouped by the number of name servers per known-bad seed domain.

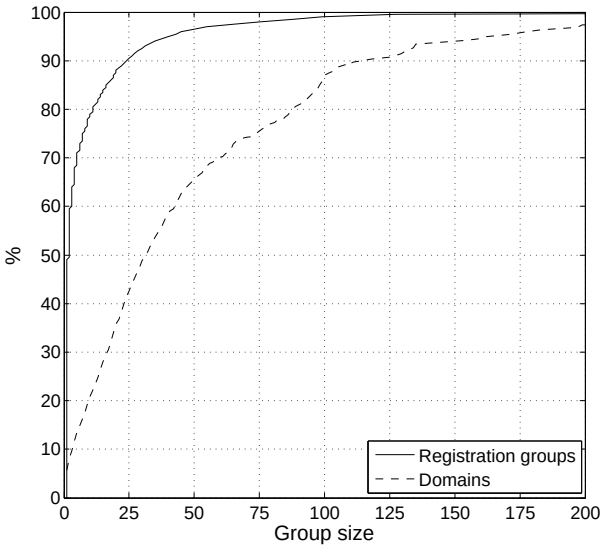

Figure 5: Cumulative distribution of number of registration groups vs. total number of domains.

compares the distributions.

3.4 Validation of Malice

In the final stage, we evaluate the accuracy of our inferences using sources of known and suspected bad domains.

To verify known-bad behavior, we test inferred domains for membership in any of (1) the original JWSDB blacklist and (2) the URIBL blacklist. As we maintain historical data for all of these, we can retrieve the historical behavior of malicious domains. In addition to these blacklists we also test the domains using McAfee's SiteAdvisor[5] domain reputation service. SiteAdvisor provides a "threat level" in its reports of green/yellow/red, for which we consider red as reflecting known-bad.

To assess "likely but unconfirmed" bad domains, we use two sources: (1) historical data from the URIBL "gold list" mentioned above and (2) SiteAdvisor reports indicating a "yellow" threat level, or that multiple users have reported the domain as malicious.

Any remaining domains have unknown maliciousness and may potentially present false positives. Here, the possibility arises that URIBL might use the same sort of inference procedure as we do for constructing their "gold list", which would make evaluating against it unsound. However, we note that URIBL reports that only a small proportion of their gold list eventually appears on their regular blacklist, while many of our inferred domains do. In addition, we find that we frequently are able to infer malice considerably earlier than is done on the "gold list". These disparities suggest that URIBL's "gold list" candidate selection methodology differs from ours.

Finally, we note that we have hand-checked a number of the potential false positives and find circumstantial evidence that the domains are in fact malicious. For example, we frequently observethe use of two seemingly unrelated English nouns together to form a single domain name—widely employed in various online scams. As we lack a systematic way to determine definitively that these domains are benign, we assume they are in fact false positives.

4 Evaluation

We now present an evaluation of our approach. We discuss the characteristics of the inference process, assess the correctness of the inferences, and examine the potential time savings afforded by the proactive nature of our method.

4.1 Inference Characteristics

Using 41,159 domains in the JWSDB blacklist from May 2009 through January 2010, we find that they cluster into 4,875 groups of common registrations (same day and same registrar). Table 1 compares the world's ten largest domain registrars to those registering the JWSDB domains. The difference suggests that miscreants find most of the world's largest registrars difficult to work with, either because they employ successful abuse-tracking mechanisms or have requirements that render them harder to register with in the first place.

To examine patterns of name server and registrar commonality further, we look at differing sets of seeds taken

[5] http://www.siteadvisor.com

Registrar	Country	Domains	%
Godaddy Inc.	US	32.6M	29.7
eNom Inc.	US	9.1M	8.3
Tucows Inc.	CA	7.4M	6.8
Network Sol. Inc.	US	6.5M	5.9
1&1 AG	DE	4.7M	4.3
Melbourne IT	AU	4.5M	4.1
Wild West Domains	US	3.1M	2.9
Moniker Inc.	US	2.8M	2.5
Register.com	US	2.5M	2.3
ResellerClub.com	IN	2.4M	2.2
Planet Online Corp.	US	6.6K	16.1
Webzero Inc.	US	6.0K	14.7
China Springboard	CN	4.9K	11.9
eNom Inc.	US	4.4K	10.7
Xin Net Corp.	CN	2.9K	6.9
Ename Corp.	CN	1.5K	3.6
Moniker Inc.	US	1.3K	3.2
Bizcn.com Inc.	CN	1.2K	2.9
OnlineNIC Inc.	US	0.9K	2.2
Hupo.com	CN	0.8K	1.9

Table 1: Top 10 registrars worldwide (top, from webhosting.info) vs. those registering domains in the JWSDB (bottom).

from JWSDB. First, we explore inference based on using a large set of seeds: all domains blacklisted by the JWSDB in January 2010. There were 3,653 such seed domains, for which the `.com` zone files show a total of 16,690 NSs, of which 2,730 are distinct. 88% of this distinct set were "fresh" by our definition (registered in 2009 or later), and all self-resolving domains were hosted on new NSs.

Our inference method based on NS features (§ 3.2) and registration commonalities (§ 3.3) predicts 12,799 domains based on the 3,653 bad seeds. This reflects an overall expansion factor of 3.5 of our inference algorithm. We deem these domains malicious and likely to be used in the future in a spam campaign or other malicious activity.

A basic next question concerns to what degree we can obtain effective inference using a more modest set of initial seeds rather than an entire month's worth of data. Starting with a smaller sample set, we are more likely to choose domains that are in distinct inference clusters. To assess this effect, we selected random seed domains of increasing sample size from the total set of JWSDB domains in Jan 2010 and computed the size of the inferred cluster, performing 5 runs for each sample size. Table 2 shows the inference algorithm's results for seed sample sizes ranging from 25 domains at a time to the entire month's dataset. The inference suggests a large set of new domains when using a small number of seeds, and we discover new, potentially malicious domains even if the seed sample contains many domains.

Sample	Cl. size	Multip.	TP	FP?
25	443.0	17.7	74.1	1.3
50	649.7	13.0	81.4	2.3
100	1,178.6	11.8	80.4	1.4
200	1,997.2	10.0	78.0	3.5
400	2,816.7	7.0	78.0	2.4
800	3,536.0	4.4	78.8	2.9
3653	11,053.0	3.0	73.7	6.6
3653*	12,799.0	3.5	63.7	19.2

Table 2: Inference productivity averaged for different seed sample sizes, JWSDB dataset, January 2010. The second column shows resulting cluster sizes followed by multiplication factors from initial seed sets to cluster sizes, true positive rates, and potential false positive rates. The "Albanian outlier cluster" is excluded in all but the last row, marked with an asterisk, which repeats the results for the entire January dataset to allow for comparison.

Of special interest is a single, large inference cluster containing 1,746 domains, roughly five times bigger than the second-largest cluster. During the evaluation, we could not confirm that domains in this cluster are indeed malicious, but we find considerable circumstantial evidence that in fact they are. They exclusively belong to a huge group over 80,000 domains registered under a single name in Albania in January and February 2010. Since this outlier has decisive impact on the sampling results, we excluded it from the evaluation for any sample size smaller than the whole January 2010 dataset.

4.2 Inference Accuracy

Figure 6 summarizes the outcome of each of the inferred registration clusters. The inferences generally work very well: based on a small number of seed domains we unearth large clusters of associated domains, with an average of 42 domains in a group, and reaching up to 389 domains (excluding the outlier cluster of 1,746 domains). In Figure 6, 10% of the clusters contain only a single domain, hence for these clusters our inference is ineffective. Two-thirds of the time, a seed from JWSDB leads us to additional domains not seen in JWSDB itself, and often we obtain dozens of such additions. Thus our approach can amplify modest observations of bad behavior in the wild to numerous new candidates for proactive blacklisting.

Of the domains we inferred, we find 73% subsequently appeared on one of our evaluation blacklists. (Recall that the URIBL "gold list" only claims a rate of around 4%.) Using the URIBL gold list and McAfee SiteAdvisor to flag potentially suspect (but not confirmed) domains, as discussed above, we find that 93% of the inferred domains are either known-bad or suspected to become so. Note that 84% of our clusters con-

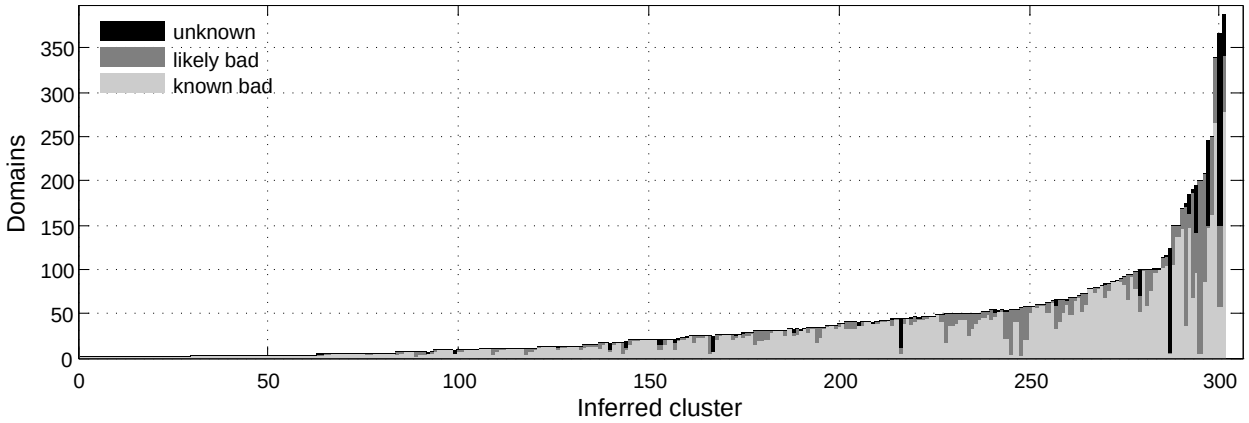

Figure 6: Predictions for each of the inferred registration clusters. The bars show the proportion of domains in our inferred registration clusters: the number of domains confirmed bad in JWSDB, URIBL, or SiteAdvisor (light gray), number of domains suspected bad in URIBL "gold list" or in McAfee SiteAdvisor (medium gray), and number of domains that are potential false positives (black).

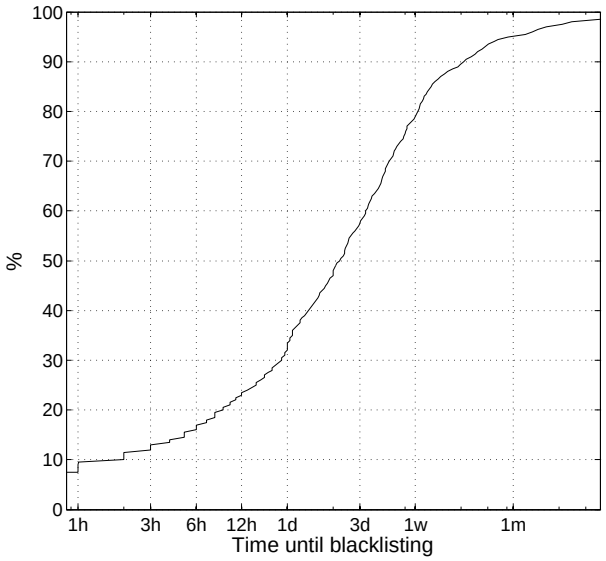

Figure 7: Distribution of time saved by proactive blacklisting.

tain only known-bad and suspected bad domains.

In addition, almost all of the potential false positives lie in 10 top "missed" clusters. Besides the major outlier cluster mentioned earlier, we visually inspected several of these "missed" clusters and assert that many of them are likely to be true positives.

4.3 Time to Blacklisting

Proactive blacklisting does not provide a benefit unless it enables a head-start over regular, reactive blacklisting. To quantify the savings in time, we study for each inferred domain its temporal difference, i.e., the timespan from the earliest domain blacklisted from its cluster until the domain itself is eventually blacklisted. We find proactive blacklisting immediately worthwhile: 75% of the domains are spread over more than 6 hours, 60% over at least one day, and 12% over more than a week (see Figure 7). We also observe that in 8% of the cases, proactive blacklisting does not help as domains are blacklisted at the same time as the earliest seed domain in a cluster. Even halving these time frames to weaken the assumption that we identify clusters at the beginning of their lifespan provides substantial benefit.

5 Discussion

Registration clustering. We define registration clusters based on the registrar's name and the time at which the domain was registered. This is an optimistic definition, as domains in the same campaign could be registered over time or at several registrars. To allow the possibility of miscreants registering domains at the same registrar over a period of time, our approach would require a different and less specific definition of registration cluster.

NS heuristics. We base our cluster inference on two insights, namely that malicious domains and their NSs are (*i*) likely to be new and (*ii*) typically managed together, increasing the chances of overlap and reuse. These assumptions need not always hold. Some domains are hosted on more established name servers that our heuristics do not cover. In particular, a substantial set of domains in our dataset registered with eNom Inc. did not switch to new NSs, but rather kept the eNom name servers as the authoritative NS. This effect is visible in Figure 4 in the set of domains with 5 NSs, which is exactly the number of NSs eNom Inc assigns to new domains.

Available information. We largely drive our methodology by NS information in the zone file. Hence, we

can only execute our method if we have access to the zone file for the specific TLD. For most of the major gTLDs, this is the case, and thus we have ample opportunity as currently the `.com`, `.info`, and `.net` TLDs cover 55–70% of the domains in major blacklists. For ccTLDs, however, only their registries have access to the given zone file, and availability can be difficult (such as for `.ru`). Another potential bottleneck is access to the WHOIS database. Fortunately, VeriSign makes registry records for `.com` and `.net` available, but many ccTLDs enforce an impractically low query rate limit.

Evasion techniques. Any defense technique needs to consider evasive maneuvers by the opponent. Two such strategies come to mind: distribution of registration over time and registrars, and distribution of name resolution over a large number of NSs or well-established NSs. The former is feasible, but would substantially increase the effort required to operate a large number of domains. Note that miscreants likely prefer some registrars due to their tolerant or negligent domain registration procedures. Another reason for selecting a registrar could be bullet-proof hosting as a service, in collaboration with the miscreants. Either way, forcing miscreants to change registrars frequently would likely increase their operational costs. The latter would likewise increase operational costs, while still providing zone file information to discover additional sets of bad domains. Alternatively, the miscreants could operate their scams from well-established name servers at major hosting companies, which would expose them to the detection mechanisms at those companies.

6 Conclusion

Our results present an initial exploration of the potential of domain-based proactive blacklisting. Starting from a relatively small set of known bad domains we are able to infer a large set of other bad domains with only a small number of false positives. Our methodology is based only on registration and name server information and leverages the key observation that Internet miscreants require substantial numbers of domains to maintain their scams in an ongoing fashion. We believe that this direction of defense holds great promise, particularly since parties central to the domain registration lifecycle and infrastructure operation (such as domain registries, registrars, and major hosting companies) could employ methodologies such as ours comparatively easily and comprehensively.

7 Acknowledgements

We thank VeriSign for zone file feeds and WHOIS records, Rick Wesson at Support Intelligence for improved WHOIS query services, and both joewein.net and URIBL.com for their blacklist feeds.

This work was supported in part by the National Science Foundation under grants NSF-0433702 and CNS-0905631, and by the Office of Naval Research under MURI Grant No. N000140911081.

References

[1] J. Jung and E. Sit. An Empirical Study Of Spam Traffic And The Use Of DNS Black Lists. In *Proceedings of IMC'04*, pages 370–375, Taormina, Sicily, Italy, October 2004. ACM SIGCOMM.

[2] C. Kreibich, C. Kanich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage. Spamcraft: An inside look at spam campaign orchestration. In *Proceedings of LEET'09*, Boston, USA, April 2009.

[3] Ludl, C. and McAllister, S. and Kirda, E. and Kruegel, C. On The Effectiveness Of Techniques To Detect Phishing Sites. In *Proceedings of DIMVA'07*, Lucerne, Switzerland, July 2007.

[4] J. Ma, L. Saul, S. Savage, and G. Voelker. Beyond Blacklists: Learning To Detect Malicious Web Sites From Suspicious URLs. In *Proceedings of the 15th SIGKDD Conference*, pages 1245–1254. ACM, 2009.

[5] J. Makey. Blacklists Compared. `http://www.sdsc.edu/~jeff/spam/Blacklists_Compared.html`, February 2010.

[6] P. Prakash, M. Kumar, R. R. Kompella, and M. Gupta. PhishNet: Predictive Blacklisting to Detect Phishing Attacks. In *Proceedings of INFOCOM'10*, San Diego, California, USA, March 2010. IEEE.

[7] A. Ramachandran, D. Dagon, and N. Feamster. Can DNS-based Blacklists Keep Up With Bots? In *Proceedings of CEAS'6*, Mountain View, CA, USA, July 2006.

[8] A. Ramachandran, N. Feamster, and D. Dagon. Revealing Botnet Membership Using DNSBL Counter-intelligence. In *Proceedings of SRUTI'06*, San Jose, CA, USA, July 2006. ACM/USENIX.

[9] S. Sheng, B. Wardman, G. Warner, L. Cranor, J. Hong, and C. Zhang. An Empirical Analysis of Phishing Blacklists. In *Proceedings of CEAS'09*, Mountain View, CA, USA, July 2009.

[10] S. Sinha, M. Bailey, , and F. Jahanian. Shades of Grey: On the Effectiveness of Reputation-Based Blacklists. In *Proceedings of Malware'08*, pages 57–64, Fairfax, VA, USA, October 2008.

[11] S. Sinha, M. Bailey, and F. Jahanian. Improving Spam Blacklisting Through Dynamic Thresholding and Speculative Aggregation. In *Proceedings of NDSS'10*, San Diego, CA, USA, February 2010. Internet Society.

[12] F. Soldo, A. Le, and A. Markopoulou. Predictive Blacklisting as an Implicit Recommendation System. In *Proceedings of INFOCOM'10*, San Diego, California, USA, March 2010. IEEE.

[13] J. Zhang, P. Porras, and J. Ullrich. Highly Predictive Blacklisting. In *Proceedings of the USENIX Security Symposium '07*, San Jose, California, USA, July 2008.