WHITE PAPER

Reliable Reputation

Authors:

Renée Burton, Laura da Rocha, Brent Eskridge

Table of Contents

Introduction

The reputation, risk, or likelihood of abuse of Internet infrastructure is an important factor in evaluating and prioritizing potential threats. In particular, certain nameservers, top-level domains (TLDs), domain registrars, and autonomous system numbers (ASNs) for routing IP traffic, are more likely to be leveraged by threat actors for a variety of reasons, including monetary, oversight, and political ones. Unfortunately, there is no standard method in the literature for creating or normalizing such scores, making it difficult to compare the results within a dataset, across data types, and across vendors.

Infoblox has developed a method for scoring that is interpretable and consistent regardless of the underlying type of reputation, for example whether we are scoring TLDs or nameservers.[1] This algorithm relies only on the number of events, or observations, making it easy to implement and translate into different environments. Further, it creates a normalized score that is statistically optimal based on those counts. As a result, it is widely applicable. Members of a security operations center (SOC), for example, can utilize the score to reliably prioritize potential threats in their environment, while threat hunting and data science teams can use this approach to automate threat detection in a way that withstands change over time. Researchers can compare trends in threats and the impacts of different network visibility on reputation scores.

This paper describes the algorithm, discusses some specific use cases and the impact of data type on the score distribution, and highlights some limitations of the method. We provide a detailed description of how to interpret the results for users, and introduce additional enrichment that can help inform the decision-making process. Our goal is to provide enough detail for other organizations and researchers to replicate the work with their own data, and to provide users with information necessary to interpret the results.

We will use top-level domains (TLDs) to demonstrate the algorithm and produce results, and later compare this with results for name servers. The term score, or reputation score, means a numeric value associated with potential risk of abuse or threat. Specifically, we will show how to determine “which TLD has the worst reputation?”, as well as answer other questions including:

- How does one TLD’s reputation compare to all other TLDs' reputations?
- What TLDs have an expected level of abuse?
- How does a TLD’s reputation change over time?

Background

When evaluating the potential threat of a domain name or IP address, a common technique is to assess related aspects, such as registration and hosting information. Threat researchers and analysts have acquired, over time, a sense of untrustworthy domain registrars, abused top-level domains, and unscrupulous hosting providers. Translating that knowledge into a repeatable, defensible score that can be used by people and automated processes alike is more challenging.

[1] We use the term reputation score in this paper to mean reputation, abuse, or risk score, equivalently.

There are a number of approaches to reputation scoring, including graph theoretic and machine learning algorithms. In contrast, this work focuses on a different approach: using only the count of observed events. We consider this case because it is simple to maintain and explain to users, but also because in many cases only count information is available, and because in our experience, count-based algorithms are quite effective.

The Power Law Challenge

To create a reputation score for TLDs, suppose we have a collection of registered domain names and we have labeled some of these as malicious.[2] We can group these domains by their TLD and count the total number of domains, as well as the number of malicious domains per TLD. The simplest count-based approach is to take the ratio of these two numbers, that is,

$$Score(x) = M_x / T_x$$

is the reputation score for the TLD, x, where M_x is the number of malicious domains for the TLD x and T_x is the total number of domains for the TLD x in our observation set. This provides a score in the range of 0 to 1 and indicates the relative frequency of malicious domains in the TLD.

Unfortunately, there are a number of limitations to this approach that make it difficult to use in practice. One of the problems is that by the nature of this data, the distribution of the scores will follow a power law, meaning that the vast majority of TLDs will have scores near zero that vary by only hundredths, and are therefore difficult to distinguish in a meaningful way (see Figure 1 below).[3] For example, if the TLD `cyou` has a score of 0.95, we can interpret that as less reputable than a score of 0.19 for the `com` TLD. But how do we interpret the difference between the `com` TLD and the `net` TLD, which have scores of 0.19 and 0.23, respectively? Is the difference between a score of 0.19 and 0.23 substantially significant? It is impossible to answer these questions without further information about what is considered an “expected” and “unexpected” score.

[2] For simplicity, we use the term malicious throughout this paper to include suspicious and verified malicious domains.

[3] Like most natural data, these ratios will follow a distribution often referred to as Zipf’s Law.

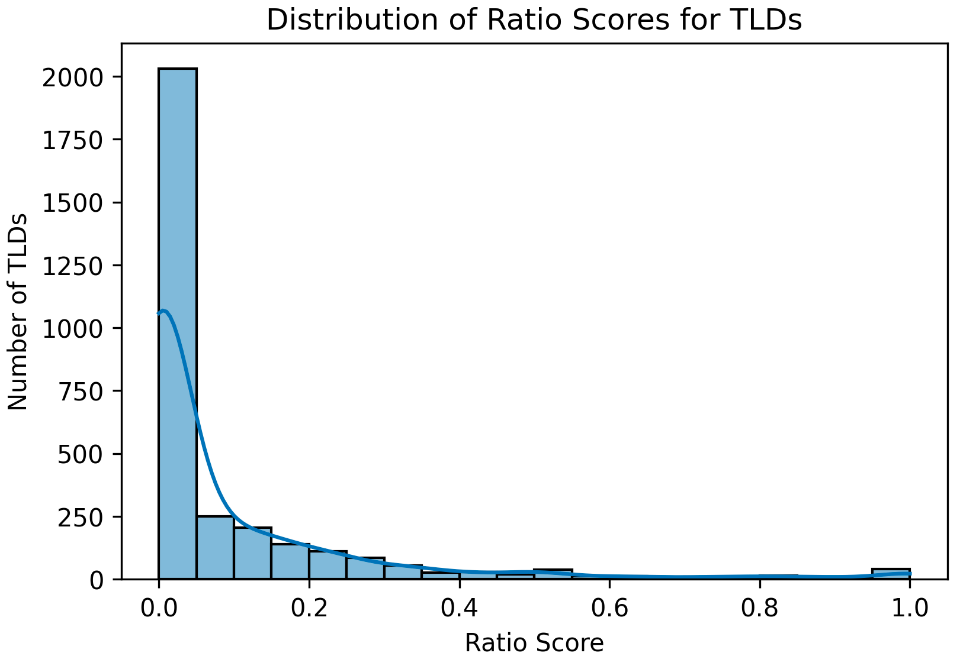


Figure 1. Using a simple ratio score, the distribution of scores across TLDs is heavily skewed and difficult to interpret. Variation across time and data sets make it impossible to set standards for consistently identifying bad TLDs.

Spamhaus' Attempt to Answer the Power Law Challenge

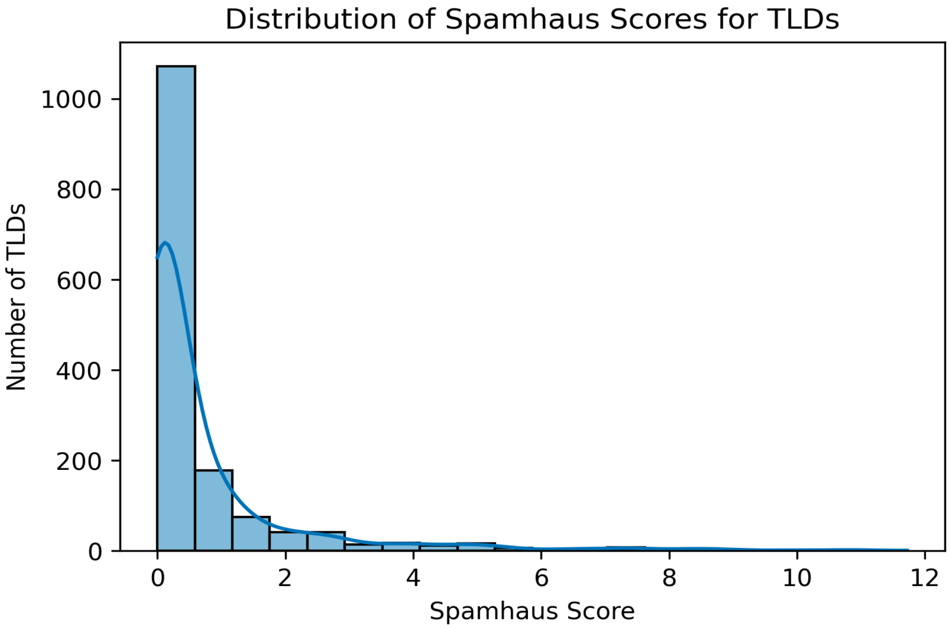
Spamhaus published their solution to this challenge in the context of scoring domain registrars, which we can apply to TLDs as well for comparison.[4] To further separate scores, they multiply the ratio by a log of the number of malicious total counts. As a result, the Spamhaus score for reputation is

$$SpamhausScore(x) = (M_x / T_x) * log(M_x)$$

This approach allows them to separate items with a similar overall ratio by emphasizing the number of malicious observations. For example, if two TLDs had the same ratio, but one had twice as many malicious observations overall, the Spamhaus score would differ by $log(2)$. This score has no upward bound, but transforms the distribution of scores to be somewhat wider than a simple ratio distribution. The Spamhaus score applies a non-linear map to the ratio data, emphasizing the total number of malicious observations, as shown in Figure 2 below.

[4] https://www.spamhaus.org/statistics/registrars/

Figure 2. The Spamhaus score for the same TLD data as Figure 1. In this sample, the Spamhaus score is bound by 12, but could have different maximum scores in different samples or at a different time.



We can see the impact in Table 1 below. Under the ratio score, the difference between the net and com scores is 0.04 and the difference between the cyou and buzz scores is 0.02 respectively. But under the Spamhaus score, although the ordering of the scores remains, the delta is quite different. The difference between cyou and buzz is 0.26, much larger than the 0.06 seen between net and com. This is a result of weighting the score by the malicious count alone, causing the cyou to increase significantly, altering the delta between scores. Essentially, the numbers have changed but the challenge of interpreting the difference between the results still remains. Is net much more abused than com? Is the relative abuse between net and com more or less than that of cyou and buzz? Neither of these systems gives a good answer.

Table 1. Comparison of reputation values using ratio scores versus Spamhaus' score.

TLD	Ratio Score	Spamhaus Score
com	0.19	2.70
net	0.23	2.76
buzz	0.93	9.73
cyou	0.95	9.99

The Interpretability Challenge

Both the ratio and Spamhaus scores suffer from a lack of interpretability. In either case, we have no way to understand how to make sense of the results in a consistent manner. If the com TLD has a ratio score of 0.49, what does that mean? Is it better or worse than we expect? And, if we consider it relative to another TLD with a score of 0.51, how much "worse" is the latter? Without a reliable mechanism to compare a score to all the others, the score has questionable value. While transforming the data by multiplying, as Spamhaus does, spreads the scores out, it provides no anchor and has no limit, so it does not fix the data skew. Furthermore, if we calculate a different type of reputation score, we have no means to understand the scores independently. Specifically, a score of 0.6 in the context of TLD reputation may have a completely different meaning than in the context of name server reputation.

Our Algorithm

Our algorithm is designed to address these issues. This method creates an optimal score from the count data, where optimal means that statistically no other algorithm can provide a better distinction of malicious and benign behavior. We then normalize the score in such a way that it can be interpreted consistently, over time, and regardless of the type of reputation being evaluated. As a result, both people and processes can determine the reputation of an item, e.g., a TLD, relative to all others, allowing them to make more informed decisions.

To accomplish this, we use the same count data described earlier, where $M_x$is the number of malicious items for x and T_x is the total of all items for x. We let r_x be the ratio M_x/T_x. Then the score

$$s_x = r_x / (1 - r_x)$$

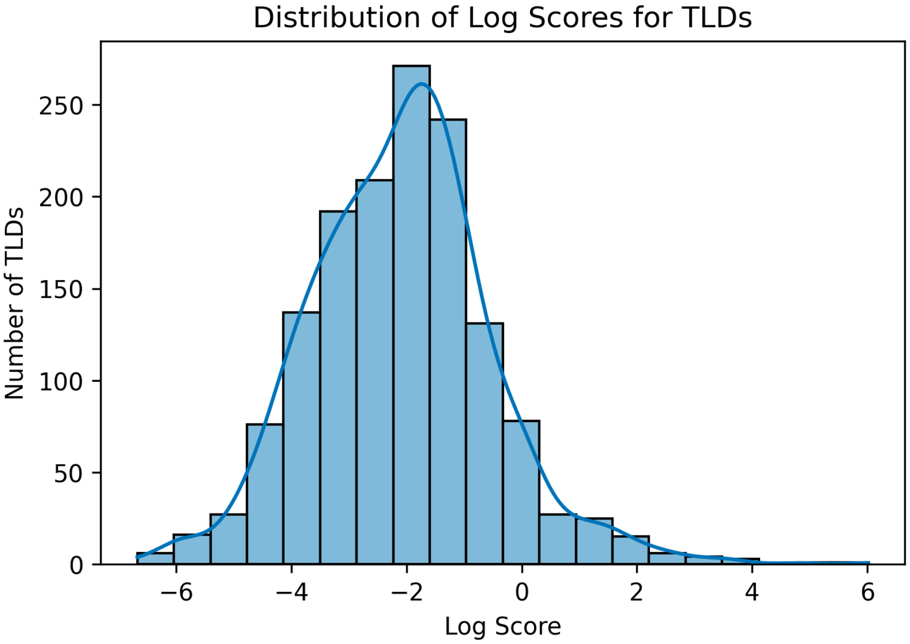


Figure 3. Distribution of the non-infinite log score approaches a normal distribution with a bell-shaped curve.

is optimal.[5] This score, like the simple ratio score above, will be heavily skewed in distribution. To correct for this, we take the log of this value; $log(s_x)$ will be an

[5] This is a result of the Neyman-Pearson Lemma. We don't provide a complete derivation of the scoring algorithm in this paper, but we begin with probability distributions that a domain within a given TLD, e.g., will be malicious.

approximately normal infinite distribution (see Figure 3 below).[6] We will call this the log score in the paragraphs that follow.[7]

The log score allows us to make powerful interpretations of the results. By calculating the mean and standard deviation of the finite log scores, we can interpret the log score as a deviation from the mean. The mean of the distribution is also referred to as the expected score; items near the mean have expected, or average, behavior relative to the entire group. As a result, we know how the risk associated with a TLD varies from the expected behavior and we can quantitatively compare the difference in risks between two TLDs.

The log score distribution includes infinite values; TLDs with no observed malicious behavior will have a score of negative infinity, while any that contained only malicious observations would have a score of positive infinity. These values are not included in our calculation of the mean or standard deviation; the finite log score is interpreted relative to other finite scores. The infinite values are outliers generally associated with limitations in the observations, such as low number of samples for a particular TLD.

Creating Scores for Users

For the convenience of human users, we create an ordinal score from the log score to simplify the results. We do this by dividing the log scores into fixed width bins, centered on the mean. In our products, we use a score range of 0-10 and a bin width of 1 standard deviation, as shown in Figure 4 below. The mean of our log scores will have an ordinal score of 5, as will those within 0.5 standard deviation of the mean. We can now infer, for example, that a TLD with an ordinal score of 7 is between 1.5 and 2.5 standard deviations above the mean log score of all other TLDs, and an ordinal score of 10 is at least 4.5 standard deviations above the mean. This gives users the means to interpret the score of one TLD relative to another, and relative to all others, using well established statistical measures. Moreover, this interpretation is the same regardless of the type of reputation score: a name server with a reputation score of 7 will also be 1.5-2.5 standard deviations above the mean log score of all other name servers.

In this mapping, the negative and positive infinity log scores fall into the outside bins. A negative infinity log score means there were no malicious observations and is assigned an ordinal score of 0, while a positive infinity log score means that all observed events were malicious and is assigned an ordinal score of 10. The resulting score distributions have a bell-shaped curve centered on 5 with fat tails at the endpoint 0 and 10. While the exact shape of this distribution may vary by data type, the interpretation is the same.

We use standard deviation to create our ordinal scores, where *mean* and *stddev* are the mean and standard deviation of the finite log scores, and we label them using interval ranges of the log score:

- $[-infinity, mean - 3.5 * stddev)$ is a score range of 0-1 and a very low risk
- $[mean - 3.5 * stddev, mean - 1.5 * stddev)$ is a score range of 2-3 and a low risk

[6] This is a result of Wilks Theorem. We are not utilizing the theorem's test statistics properties which have restrictions on the probability distributions.

[7] Formally, this is a log likelihood score.

- $[mean - 1.5 * stddev, mean + 1.5 * stddev]$ is a score range of 4-6 and expected, or moderate, risk
- $[mean + 1.5 * stddev, mean + 3.5 * stddev)$ is a score range of 7-8 and a high risk
- $[mean + 3.5 * stddev, +infinity]$ is a score range of 9-10 and a very high risk

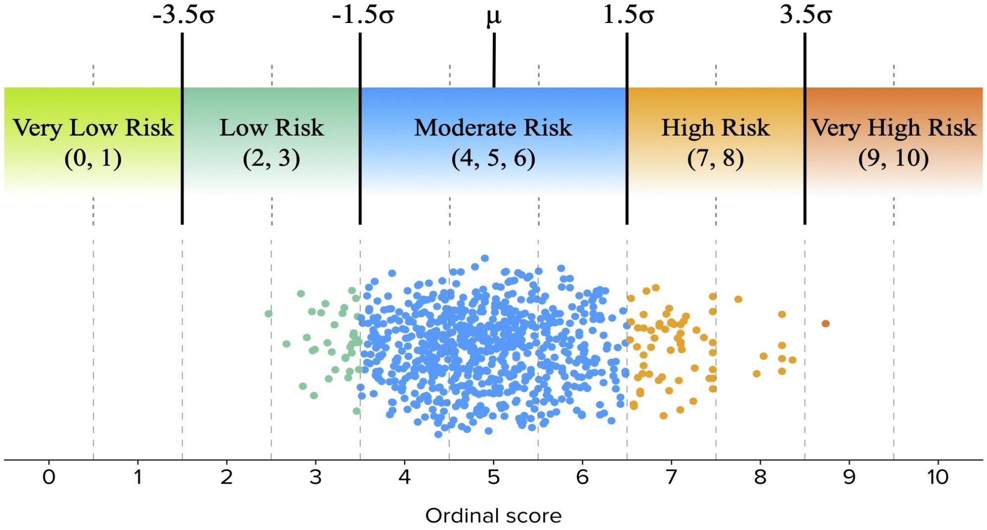


Figure 4. Illustration of risk score "bins" based on standard deviation, for non-infinite scores.

Under this normalization, the `com`, `net`, `buzz`, and `cyou` TLDs have scores of 5, 6, 8, and 8 respectively, corresponding to moderate and high risk. The distribution of risk scores, or the reputation, of our TLD sample set is shown below in Figure 5. There are a large number of TLDs with a risk score of 0 that have a low number of observations. To compensate for this case, we calculate a confidence level, described below.

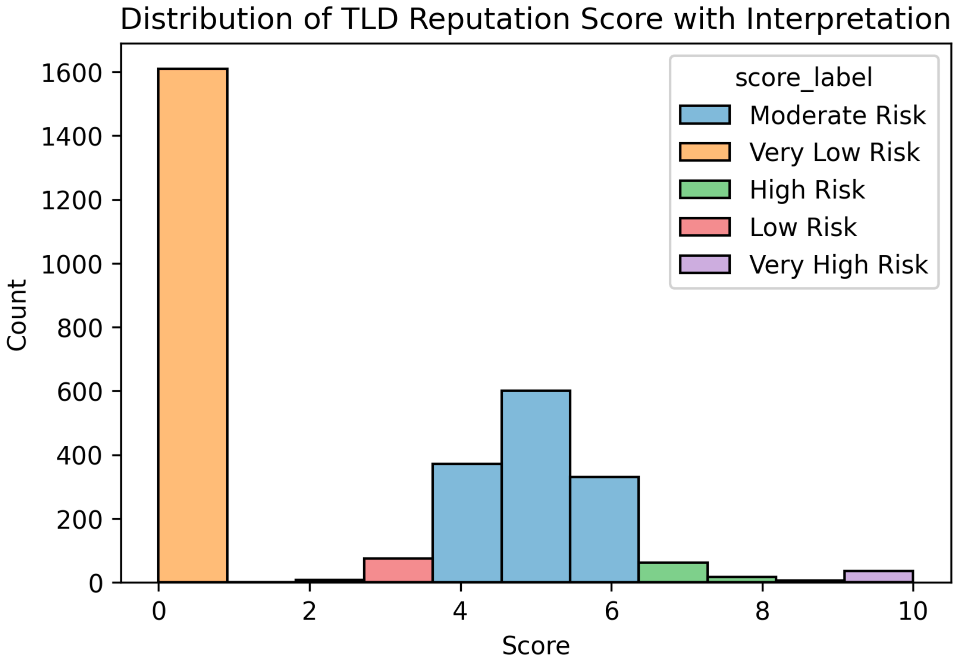


Figure 5. Each of the ordinal scores is a bin of one standard deviation in width. These can be interpreted as risk relative to the other TLDs and given a risk label. Those TLDs with no malicious observations have a negative infinity log score which becomes a 0 in the ordinal mapping.

Confidence and Popularity

In many cases, the number of observations may be fairly low due to bias in the sampling or extreme variance in the original population. For example, if the collection of TLD data is from an organizational network, that network may not observe a large number of TLDs, particularly those associated with languages or countries not related to the users of that network. Further, as new TLDs are added to the domain name system, they may not be well used for some time. The recently added country code TLD in Hebrew, `xn--4dbrk0ce`, is a case in point for both of these possible situations.

One option for reputation scoring is to exclude those samples, but this will reduce the overall information that can be conveyed to users. Instead, we have chosen to identify thresholds above which we have a high confidence in the reputation score, and below which we consider low confidence. The exact threshold for confidence is subjective and in making our choices we consider both the type of sampling data and the overall volume for our sample sets. By including confidence we provide users meaningful enrichment to distinguish reputation even when the observations are low. As shown in Figure 6 below, the low confidence reputation scores for TLDs in this sample are well distributed across the range of scores, except for score 0 — likely due to the low number of observations previously described.

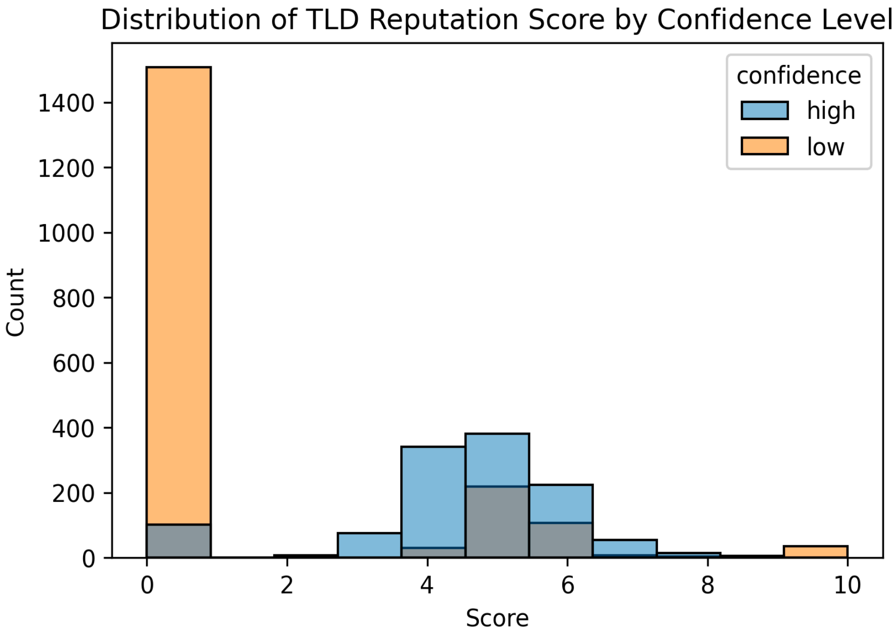
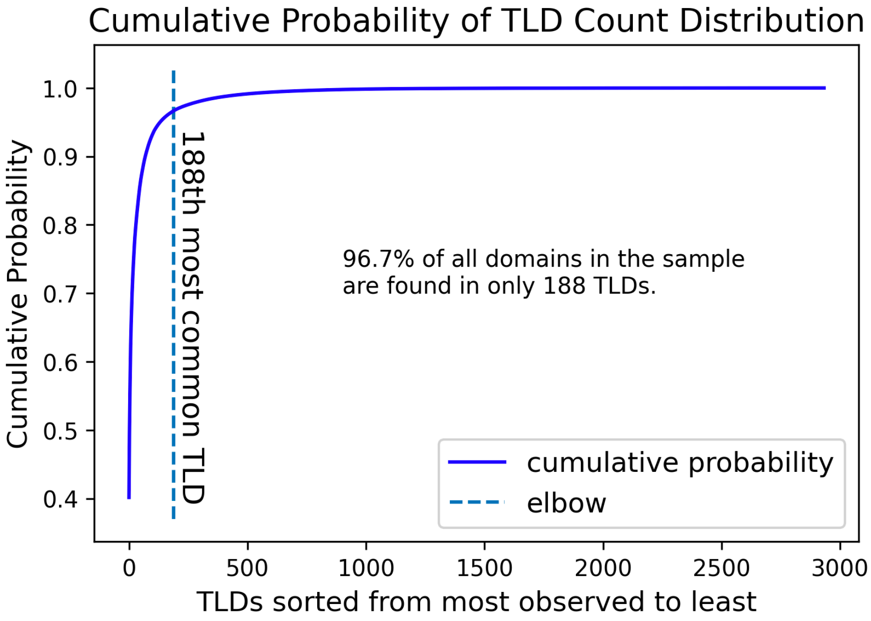


Figure 6. When there are a low number of observations overall for a TLD, we can include it in our scoring but indicate the result as low confidence. We can see here that there are a large number of TLDs with relatively few observations, none of which are malicious, resulting in a large number of TLDs with a risk score of 0 but also a low confidence. Similarly, only low confidence TLDs have a risk score of 10 — the fact that all observed events were malicious could be due to the low number of observations.

We have found it useful to also consider the popularity, or dominance, of the scored items. Popularity can be computed in a number of different ways, for example, one might consider the 100 most-used TLDs popular. However, we elected to use a statistical method to determine popularity (the “elbow computation method”) rather than setting a fixed threshold such as that. In our formulation, popular TLDs are a fairly small set that account for the vast majority of domains in our sample set. Specifically, we know the count of domains by TLD follows a power law distribution, often referred to as Zipf's Law. This phenomenon is widely observed in natural systems and is one we have studied extensively in the context of the domain name system (DNS).[8]

If we sort our TLDs in order of the observed count and calculate the cumulative sum of these counts, or the cumulative probability of the sample, the resulting distribution will be one that sharply rises and has a very long tail (see Figure 7 below). The point at which the slope of this curve turns is often referred to as the 'elbow', or 'knee', of the distribution; from that point forward, each additional TLD has a very small number of domains observed in it. The TLDs that occur prior to the elbow threshold are considered to be popular.

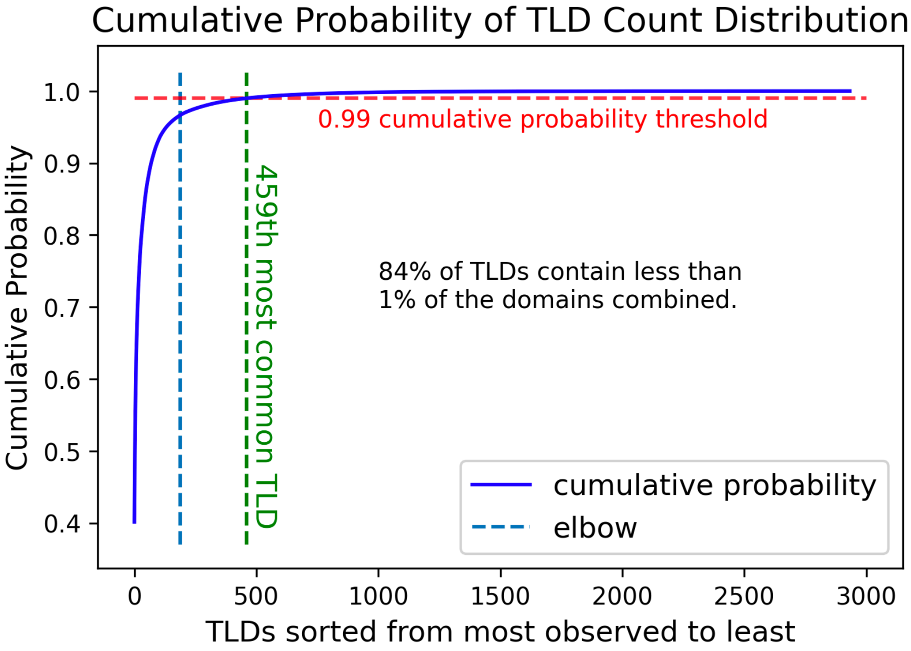
Figure 7. The vast majority of the domains in our sample are within a small fraction of the TLDs. These TLDs are considered popular.



Because the tail of this distribution is so long, we often find it useful to consider another threshold for rare items. These are the large majority of the TLDs that when combined, represent a very small percentage of the total domains. For example, in our TLD sample set, we find that 459 TLDs account for 99% of all domains in the sample, and that the remaining TLDs combined contain less than 1% of the domains - this is the cumulative probability threshold illustrated in Figure 8 below. We consider this large set of remaining TLDs rare within our observations.

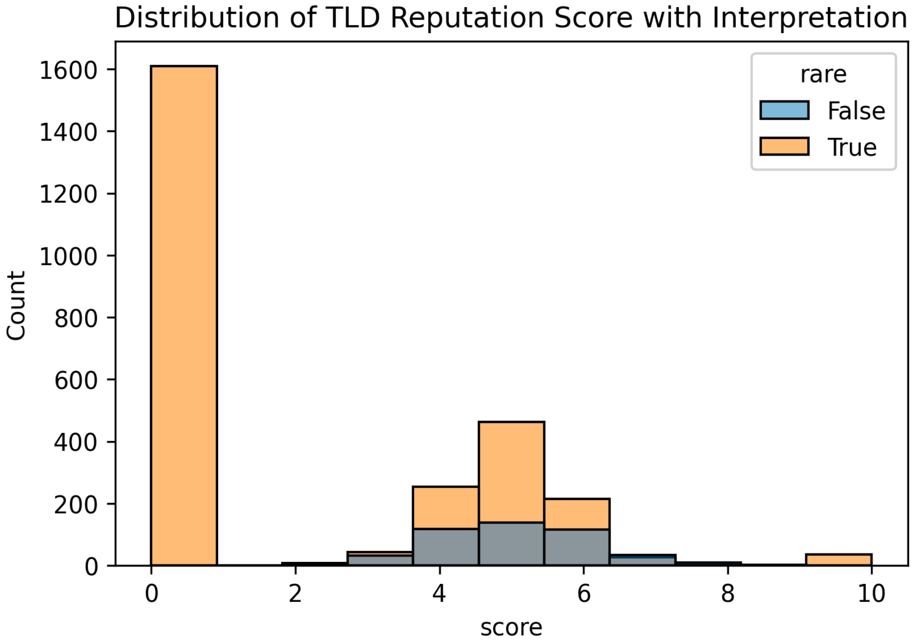
[8] https://www.infoblox.com/wp-content/uploads/infoblox-whitepaper-inforanks-infoblox-ranking-service.pdf, https://insights.infoblox.com/resources-whitepapers/infoblox-whitepaper-no-ranking-list-is-perfect-a-top-domains-list-comparison

Figure 8. The plot demonstrates how the use of a second threshold for rareness can highlight the long tail and provide additional context.



The combination of confidence, popularity, and rareness can provide extra context to the decision-making process to help prioritize work and further utilize the reputation scores. Since confidence is based on the total count of observations, there is a correlation between confidence, popularity, and rareness, but each provides a slightly different vantage point of the data. We might, for example, want to be suspicious of rare TLDs (graphed below in Figure 9), regardless of their reputation score.

Figure 9. The distribution of reputation scores, separated by rareness, meaning TLDs with a low number of observed domains. In this chart, rareness represents a cumulative probability of less than 1%.



In combining our reputation score algorithm with additional context for confidence, popularity, and rareness, we have developed a powerful mechanism to evaluate potential threats and understand the overall threat landscape based on categories like TLDs, name servers, and registrars. In the sections that follow we will lay out these results in further detail for TLD and nameserver reputation.

Applications

The main advantages of this algorithm are that it can be easily applied to any type of underlying reputation, and that its interpretation will be consistent across applications. Different aspects of a network can help with identifying and prioritizing potential threats, and in this section we give two examples of how this algorithm was applied to TLD and registrar reputation. We present the results of both implementations and showcase how they can be used for threat hunting.

TLD Reputation

In this section we discuss in more detail our results for calculating TLD reputations and what are the high and very high risk TLDs per Infoblox's reputation algorithm. Figure 10 shows the ordinal score distribution after its calculation for the month of August, and shows the count of TLDs for each score. We can observe that the data is approximately normally distributed, as expected, and there were a total of 66 TLDs scored as high and very high risk for that particular month.

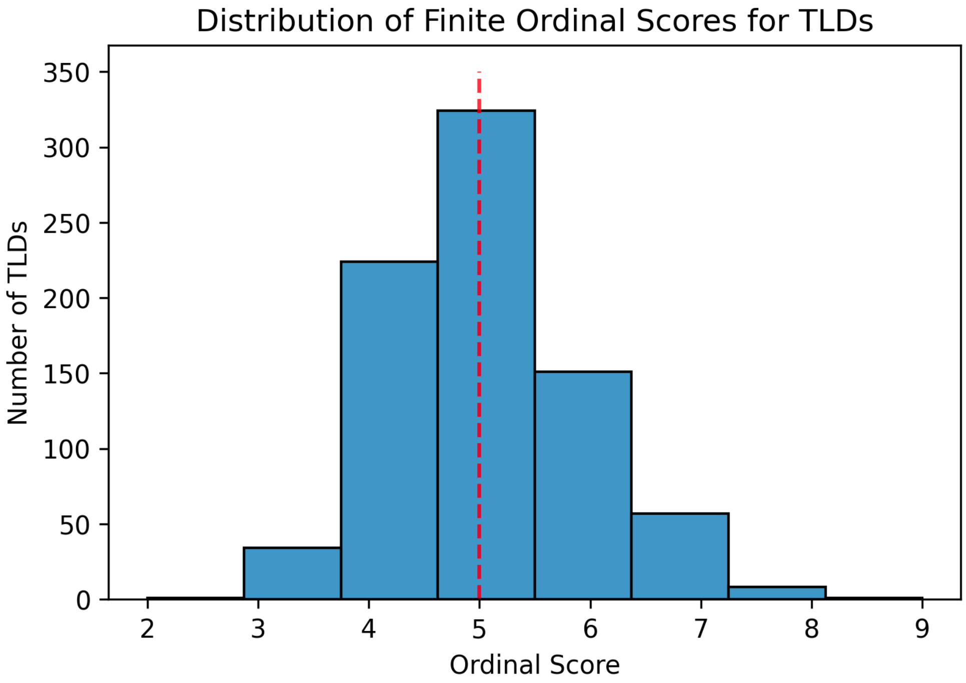


Figure 10. Ordinal score distribution for finite high and low confidence TLDs. The data follows an approximately normal distribution.

To provide the reader with insights on high risk TLD trends over time, we share the sixteen TLDs that were consistently observed as high or very high risk of abuse over the three consecutive months of the evaluation, for high confidence values only. In general, they all align with threat researchers' experiences and observations of what are commonly abused TLDs. For contrast, we also included the 27 TLDs that are popular and were consistently observed as having an expected level of abuse during the same time period (again, only the high confidence values).

Table 2. Comparison of high risk and moderate risk TLDs over the period of June, July and August 2022; low confidence TLDs are not included in the table.

TLDs that are high / very high risk (ordinal score of 7 or above)				
bid	click	icu	my.id	top
buzz	ga	ml	quest	ws
cam	gq	monster	sbs	xyz
cf				
TLDs that are popular and have an expected level of abuse (ordinal score of 4, 5, or 6)				
ae	com	com.vn	org	ru
app	com.ar	hr	org.uk	tech
cl	com.br	in	pro	uk
co	com.cn	me	ro	us
co.in	com.co	net	rs	vn
co.th	com.my			

Some TLDs did not appear as high or very high risk for all of the months evaluated. The table below shows how the high risk TLDs change over time. Some of them are only seen as high risk for one or two of the months, and as previously described, sixteen TLDs were consistently scored as high or very high risk across months.

Table 3. Evaluation of high risk and confidence TLDs per Infoblox TLD reputation algorithm across three months.

TLDs that were high risk (score of 7-10) for 3 out of 3 months				
bid	click	icu	my.id	top
buzz	ga	ml	quest	ws
cam	gq	monster	sbs	xyz
cf				
TLDs that were high risk (score of 7-10) for 2 out of 3 months				
autos	casino	cyou	tk	
beauty	cc	pw	vip	
TLDs that were high risk (score of 7-10) for 1 out of 3 months				
ac.ke	cn life	md.ci	ne.pw	rest
asso.ci	lol	mobi.tt	pics	skin
cfd		mom	presse.ci	ug

The table below shows a sample of domains for five of the high risk TLDs. The domains were randomly sampled from a separate, independent source of data than that used for creating the algorithm. From a quick overview we can observe that there seems to be a high number of domains from a domain generation algorithm (DGA), and highlights how we can use our scoring algorithms for threat hunting and to prioritize items to review in a network.

Table 4. Sample of domains from five high risk TLDs

High risk TLD	Sample associated domains	Ordinal score
buzz	klcjbtcrogyjvkj[.]buzz tryillpizza[.]buzz ltdzocvadhipecq[.]buzz clwiki[.]buzz kmcninzhouptwwj[.]buzz dpwlnjmxmtjzqnz[.]buzz	8
top	rocktechvpn1[.]top ghhrh[.]top fuzhu33[.]top 0ruua5nrbppmifdo6ne7ccifvf76fumh[.]t hurenvhol93cp9slu7udlqte599621cj[.]t updateaz[.]top	8
click	giadungthongminh24h[.]click beritabumi[.]click radioalcyber[.]click yv74d3uze75m3[.]click mobileayuda[.]click mostafasajjadifard[.]click	7
gq	leforhirsnusuc[.]gq outadtatuvanwi[.]gq kannvifirabase[.]gq densomemalo[.]gq' fuddberniticonta[.]gq cromamordiapos[.]gq	7
xyz	civ-ar61[.]xyz fishyfaamnft[.]xyz maffeo[.]xyz gioitren01[.]xyz ryzodee1[.]xyz felole[.]xyz	7

The results from the algorithm returned TLDs that aligned with our expectations of highly abused TLDs. To avoid confirmation bias with the results, we also evaluated the consistently high risk TLDs compared to other lists of abused TLDs. In general, we observed that the majority of the consistently high risk TLDs also had a bad reputation in other lists. On the other hand, there are also variations between the different lists, which is expected since the results ultimately depend on the data and samples used for scoring. In addition, we expect some variation of TLDs over time, as threats across different TLDs may vary due to threat actors' activities.

Table 5. Comparison of TLD reputations from different sources. Cells with an "x" indicate that the TLD was present in the source.

Infoblox high risk TLDs	Spamhaus top 10[9]	Palo Alto Networks analysis top 10s by threat type [10]	SURBL top 30 [11]
bid		x	
buzz			
cam	x	x	
cf		x	x
click			x
ga			x
gq	x	x	
icu	x	x	x
ml	x	x	x
monster			
my.id			
quest		x	
sbs		x	
top	x		x
ws		x	
xyz		x	x

Nameserver Reputation

To demonstrate a different application of the reputation algorithm, we computed scores for nameserver domains using a sample of approximately five million registered seed domains. For each of these, from the list of associated name servers we extracted the second level domain for the nameservers, referred to as the nameserver domain. For example, the domain badguy[.]com might have a nameserver ns1[.]badnameserver[.]com; the nameserver domain in this case is badnameserver[.]com. We extracted 177,000 associated, unique

[9] https://www.spamhaus.org/statistics/tlds/ — data for as September 21, 2022

[10] https://unit42.paloaltonetworks.com/top-level-domains-cybercrime/, Table 3 with top 10 TLDs by malicious, phishing, malware, grayware, C2 (total of 32 unique TLDs) — blog released on November 11, 2021

[11] https://www.surbl.org/tld — data for October 18, 2022

nameserver domains. We then counted the number of original sample domains that were marked as malicious in our threats database for each nameserver domain. In other words, for each nameserver domain, we have the percentage of the original set that is malicious and has a nameserver in the nameserver domain. From there, we compute an ordinal score for each nameserver domain.

Nameserver domains fundamentally differ from TLDs in a few ways. First, while a bad actor can choose which TLD they register domains within, they do not control the TLD. Some TLDs are more abused than others, but in the wild (beyond our sample set) it seems unlikely that every domain within a TLD is malicious. This is not the case with nameservers; a bad actor can operate their own nameserver and every domain associated with that nameserver can be considered malicious. For this reason, a significant number of nameserver domains may have an infinite log score, that is, the number of malicious domains observed equals the total number observed for that nameserver domain. In our experiment, over 2,000 malicious nameserver domains had an infinite log score.

At the same time, for the same reason, the number of nameserver domains is much larger than that of TLDs. The distribution of registered domains to nameservers follows the same power law curve described earlier, and given the long tail, there are also a very large number of nameserver domains for which there is no malicious observation. This results in a log score of negative infinity for a much larger set; in this case 141,000 non-malicious nameserver domains had a log score of negative infinity due to the lack of observation of malicious activity. Of the set of scored domains, approximately 6700 nameserver domains have a finite log score.

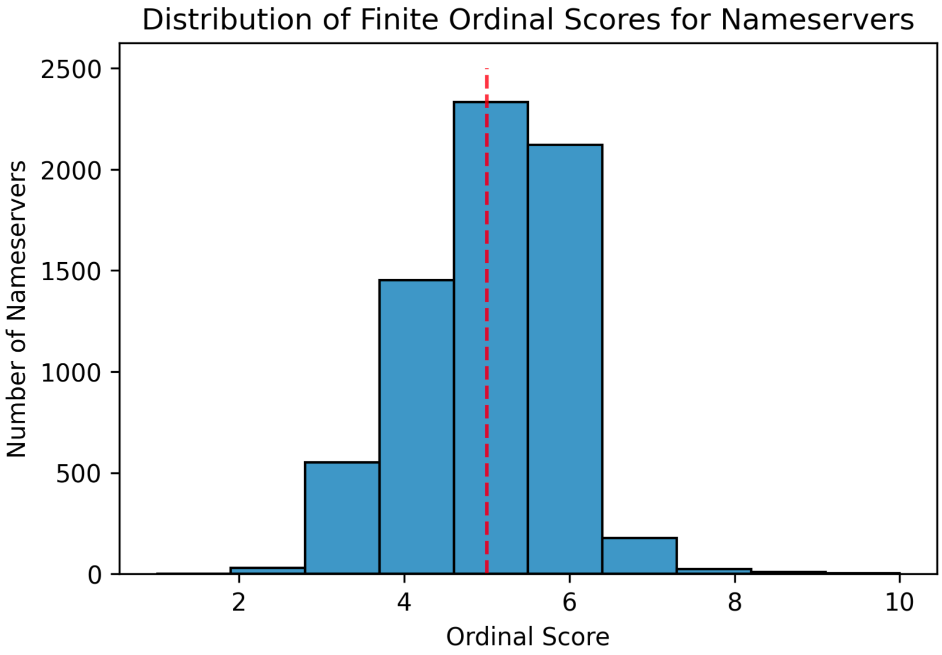


Figure 11. Ordinal score distribution for finite high and low confidence nameservers. The data follows an approximately normal distribution.

If we consider only high confidence scores, that is those for which we have observed at least 30 domains using the nameservers, only 4200 nameserver domains remain. This filter results in a trimodal distribution with a large peak at a score of 0, the bell curve peak at 4, and another small peak at 10, as shown in Figure 12 below. In this particular data set, there are a large number of nameservers with no malicious observations (the spike at 0). A large percentage of the mid-scoring nameservers are low confidence, creating a peak at the score 4 instead of 5. Comparing the distributions between all the scores, and only those with a high confidence, shows the impact of a confidence threshold on the results.

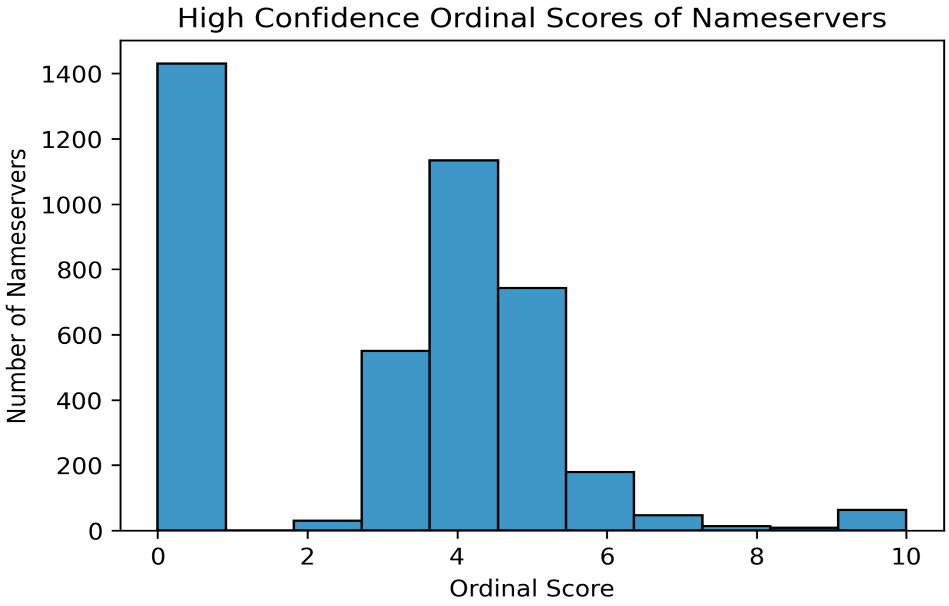


Figure 12. The ordinal score distribution of nameserver domains with high confidence scores.

If we examine the scores of the most commonly observed nameservers, we see that they are of moderate risk. This is unsurprising as large services are often utilized both for their affordability and the ability to hide within the noise of large volumes of DNS traffic. These scores support the fact that while many threat reports, for example, reference Cloudflare nameservers used by malicious actors, Cloudflare also serves a large number of legitimate domains. However, registrars that offer very cheap domain registration are often particularly favored by criminal actors, and we can see this in the results as well. We listed some of the more popular services and their reputation scores in Table 6 below.

Table 6. A sample of well-known nameserver domains, their associated commercial entity, and reputation score.

Service	name server Domain	Reputation Score
Name Cheap	`namecheap[.]com`	7
Cloudflare	`cloudflare[.]com`	6
GoDaddy	`domaincontrol[.]com`	5
Google	`googledomains[.]com`	4
OVH	`ovh[.]net`	4

For threat hunting, we are interested in identifying nameservers that are significantly more likely to be associated with malicious activity than average. Our scoring algorithm allows us to do this. Some of these nameservers may be actor controlled and others may be highly abused services. In our experiment, nearly 30,000 nameserver domains had scores greater than 6; that is, they were either high risk or very high risk. Of these, over 130 are high confidence scores. To investigate these nameservers further, we identified domains that were using them, independent of the original set. A sample of these results is shown in Table 7 below.

Table 7. A sample of high risk nameservers, their commercial association, and some representative high risk domains that utilized the nameserver.

High risk name server domains	Association	Sample associated domains

`supersonicdns[.]com`	Sav[.]com	`heathad[.]top`
		`laytoit[.]top`
		`anownbuy[.]top`
		`atsoonus[.]top`
		`atyardan[.]top`
		`bedwhook[.]top`
		`diemyadd[.]top`
		`fundohot[.]top`
		`letdorun[.]top`
		`lotsitit[.]top`

`anonsecuredns[.]com`	BS Corp (an Internet domain service)	`esp-apple[.]com`
		`icloud-ke[.]com`
		`found-maps[.]com`
		`icloud-sms[.]com`
		`maps-cloud[.]com`
		`mms-lcloud[.]com`
		`applefmi-id[.]com`
		`icloud-lock[.]com`
		`local-apple[.]com`
		`lost-founds[.]com`
		`appleld-find[.]com`
		`appleld-maps[.]com`
		`icloud-share[.]com`

`dnstechnoprovider.com`	dropcatch[.]com	`anbebut[.]top`
		`diedois[.]top`
		`dooknow[.]top`
		`endofdo[.]top`
		`ifbigis[.]top`
		`newdoas[.]top`

`thinkingfastdns[.]com`	unknown (registered June 2021 with me Silo)	`betweenpathask[.]top`
		`bigsouthsilver[.]top`
		`birdrecordwind[.]top`
		`ayehenmil[.]live`
		`agnameship[.]buzz`
		`abletaipan[.]live`

`floatingpointdns[.]com`	unknown (registered uary 2022 with Name	`oxygenseaseed[.]xyz`
		`pagedearquite[.]xyz`

		partytalkblow[.]xyz passsmilefact[.]xyz passtallclimb[.]xyz

Given the large number of nameserver domains, focusing threat hunting on those with a very high risk score helps prioritize and quickly identify suspicious activity. The results in the above table give us an example of the different types of behavior we can see using nameserver reputation. All of these nameserver domains have a very high risk score. We see nameservers associated with obvious lookalikes, services that attempt to 'catch' expiring domains, and fairly anonymous nameservers registered in the last year. In all of these cases, the domains being served appear to be questionable on their surface: either seemingly similar to well known services or potential DGAs.

Conclusion

The algorithm we have described can be replicated by organizations so that they can apply it to their own data, and use the results in any number of ways: from assessing risk to making policy decisions, to threat hunting. The algorithm provides a consistent, interpretable scoring methodology that can be applied to multiple types of data sets, such as TLDs, nameservers, and registrars.

We have also shown how this algorithm can be used in the decision-making process when evaluating the risk of a domain based on both the nameserver and the TLD. In particular, using our scoring algorithm, we are able to quickly identify suspicious domains from high risk name servers, as shown earlier in Table 7. We can combine that knowledge with most abused TLDs, shown earlier in Table 5, to gain further confidence that the associated domains are likely used for malicious activity. The example domains shown in Table 7 highlight that many are both associated with risky nameservers and abused TLDs.

Appendix

Infoblox high risk TLDs	Spamhaus top 10	Palo Alto Networks top 10s by threat type	SURBL top 30
bid	live	bid	live
buzz	info	info	info
cam	cam	cam	com
cf	tk	cf	cf
click	casa	casa	click
ga	surf	ga	ga
gq	gq	gq	jp
icu	icu	icu	icu
ml	ml	ml	ml
monster		support	cn
my.id		email	link
quest		quest	shop
sbs		sbs	biz
top	top	stream	top
ws		ws	de
xyz		xyz	xyz
		cyou	cyou
		su	tk
		uno	in
		cm	ru
		tokyo	co
		help	org
		rest	me
		win	cc
		best	app

		zw	net
		cd	uk
		pw	br
		date	fr
		am	site
		ke	
		bd	

Top abused TLDs from different sources as of the date used for the comparison. Spamhaus for September 21, 2022; Palo Alto Networks for November 11, 2021; SURBL for October 18, 2022. Blue shading indicates TLDs that overlap with another source's abused list; purple shading indicates TLDs that appear in multiple lists but not in our list of TLDs scored as consistently high risk during our three-month observation period (i.e. they may appear during one or two of those months).

Infoblox unites networking and security to deliver unmatched performance and protection. Trusted by Fortune 100 companies and emerging innovators, we provide real-time visibility and control over who and what connects to your network, so your organization runs faster and stops threats earlier.

Corporate Headquarters
2390 Mission College Blvd, Ste. 501
Santa Clara, CA 95054

+1.408.986.4000
www.infoblox.com