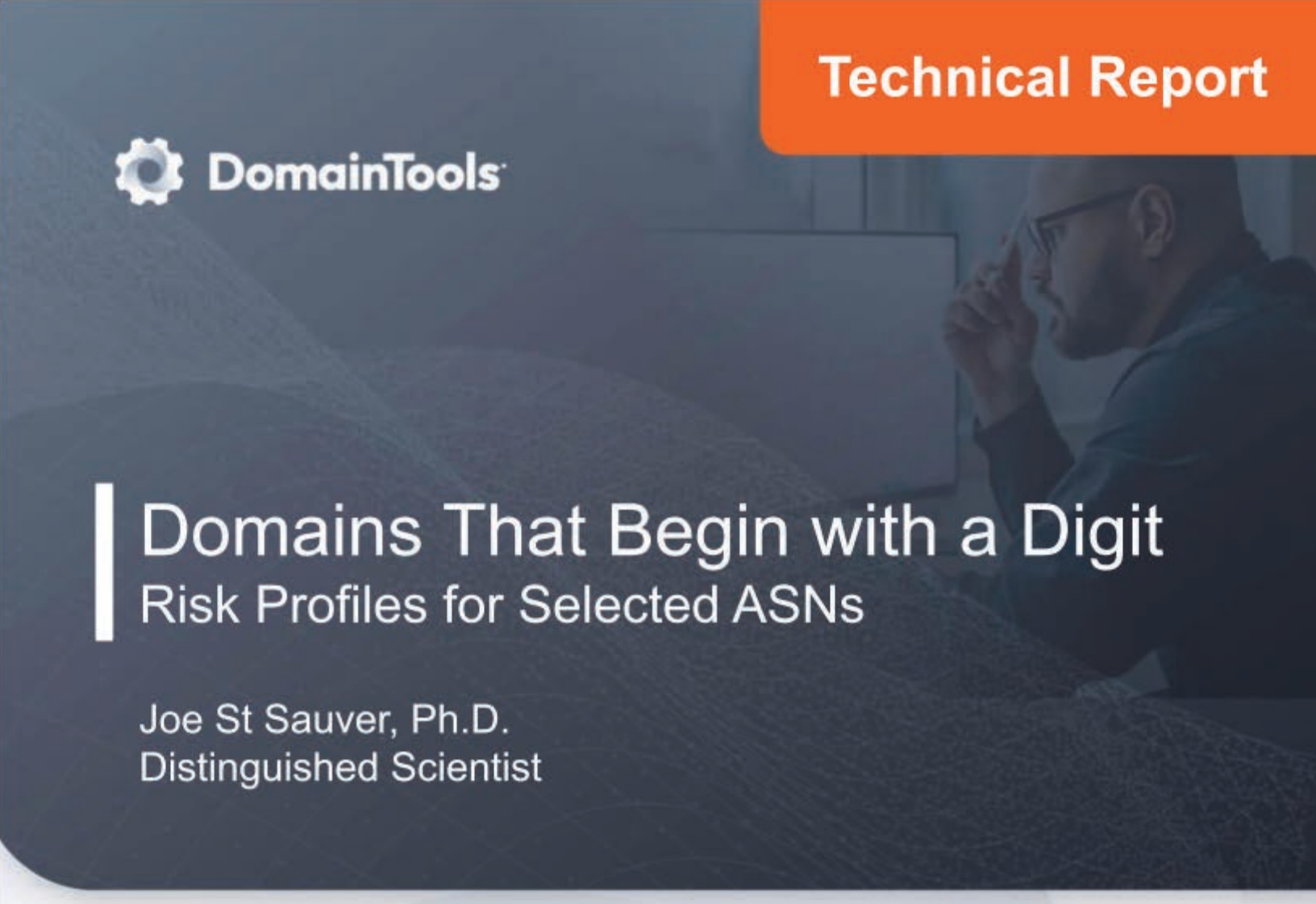
Technical Report

DomainTools



Domains That Begin with a Digit

Risk Profiles for Selected ASNs

Joe St Sauver, Ph.D.
Distinguished Scientist

Table of Contents

Executive Summary

DomainTools currently computes proprietary risk scores for registered domains, giving each domain a score from 0 to 100. Those risk scores can help users decide if a given domain is safe or too risky. Some may wonder if similar risk scoring could be extended to a larger aggregate, such as entire autonomous systems – perhaps there are "safer" autonomous systems and "riskier" autonomous systems, just as there are safer and riskier registered domains? We decided to empirically compute risk score for an *ad hoc* selection of autonomous systems, focusing specifically on registered domains that begin with a digit. Our process was as follows:

- Identify large (or otherwise potentially interesting) autonomous systems ("ASNs")
- Leverage routing data to get a list of the Classless Inter-Domain Routing (CIDR) prefixes originated by each of those ASNs
- Look up each prefix in DNSDB, getting the fully qualified domain names (FQDNs) using those IPs
- Condense the discovered FQDNs to just registered domains
- Exclude registered domains that don't begin with a digit
- Exclude ASNs that don't yield at least 2,000 registered domains beginning with a digit
- Look up a DomainTools risk score for each of the remaining registered domains
- Graph the resulting risk scores

Our findings from this work include the following:

- While it has been possible to register domains that begin with a digit since at least 1989, 611 of the candidate autonomous systems considered for this 2022 study still have fewer than 2,000 identifiable registered domains of that sort. Those ASNs were screened from this study upon detection.
- Looking at the remaining 174 ASNs (where there were at least 2,000 domains that started with a digit), ASNs can and do vary widely in their median (50th-percentile) risk score, running from a median risk score of 0 to a median risk score of 100. We found that to be an unexpectedly broad range.
- Violin plots produced with the Python3 Seaborn graphics package provide a compelling way to show the distribution of risk scores seen, and make it clear that many risk score distributions are multimodal.
- Differences in risk scores exist across ASN "home countries". While the Asia Pacific Network Information Centre (APNIC) region is often known for domains starting with digits, RIPE had some of the best and worst of these.
- We'd expected some "notorious" ASNs to score poorly (based on popularly-accepted lore about those ASNs), but some of those ASNs actually returned mundane scores for domains that begin with a digit.
- We also encountered some generally-well-regarded ASNs that scored unexpectedly poorly, prompting us to report on risk *sub*-score (“component risk score”) data (phishing, malware, spam, sinkhole and zerolisting subscores).

- In some cases, disappointing median risk scores were largely the result of hosting one or more sinkholes within that ASN. Sinkhole domains get a score of 100 by definition (and user systems attempting to communicate with a sinkhole ARE exhibiting signs of compromise), BUT sinkholes are essential to security research, don't represent active "threats" per se, and arguably ASNs hosting them shouldn't be "penalized" with 100 score'd domains for allowing those sinkholes to exist on their IPs.
- The DomainTools Risk Score API, having been designed for integrations and enrichment use cases, is suboptimal for studies of this sort involving millions of queries at once. Preferable alternatives include either using Iris APIs, which returns risk data and can handle up to 100 domains per chargeable query, or downloading the risk score data in bulk.

Part A. Introduction

In this report, our focus is on DomainTools risk scores for registered domains that start with a digit (rather than a letter), breaking that data down by autonomous system. Because familiarity with these concepts may vary, we begin by proving backfill to ensure that we're all working from a common conceptual foundation.

DomainTools Risk Scores

DomainTools routinely provides risk scores for individual registered domains. As stated in "Technical Brief: Domain Risk Score" (see https://www.domaintools.com/content/domain-risk-score-technical-brief.pdf):

> *One can think of domains with a high risk score as belonging on a "domain watchlist"—they are domains which the algorithms indicate may become dangerous in the near future.*

DomainTools scores range from a low of 0 to a high of 100, as determined by DomainTools-proprietary machine-learning-based algorithms as described in the just-cited Technical Brief. Domains get a worst-case risk score of 100 only if:

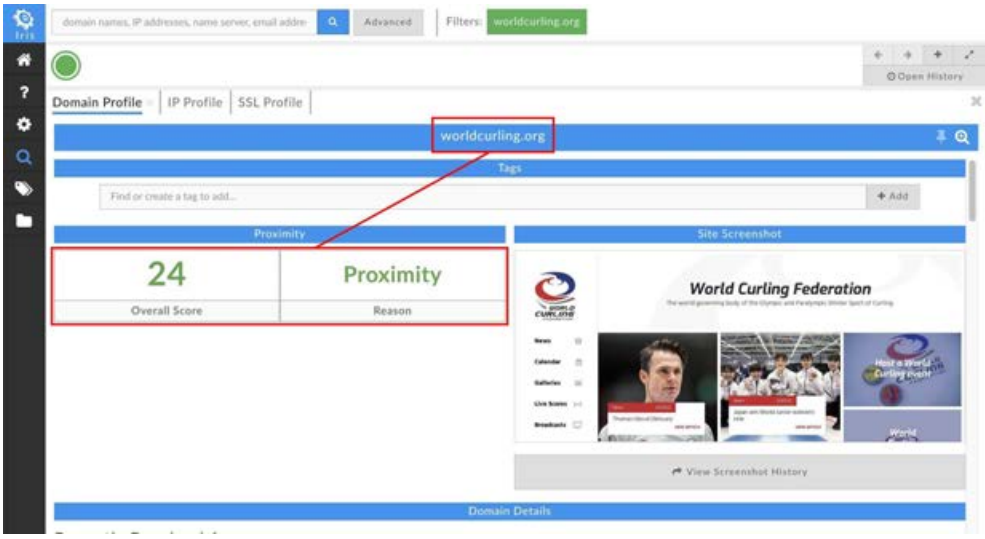
> (A.) they're already listed on at least one of the 3rd-party intel feeds that DomainTools ingested, or
> (B.) the domain has been redirected to a known sinkhole.

At the other end of the risk scale, critical/highly popular domains (such as Google.com, Microsoft.com, Apple.com, Twitter.com, etc.) will routinely be "zerolisted" (e.g., forced to have a risk score of zero), thereby precluding any possibility of highly disruptive blocking since zerolisting (or a forced 100 score) overrides any other score that might otherwise be associated with a domain.

Most domains receive a risk score somewhere between those two extremes. The previously mentioned Technical Brief recommends interpreting Domain Risk Scores as follows:

- *50+, suspicious*
- *70+, our recommended threshold for indicating malicious intent*
- *90+, strong confidence in near-term weaponization*

Let's consider a concrete (if somewhat contrived) example. Perhaps you were curious about the sport of curling. After reviewing https://en.wikipedia.org/wiki/Curling, you want to dig in further. You discover the World Curling Federation's site at `https://worldcurling.org/`, but having never visited that site before, you check DomainTools Iris before going to that site. Iris conveniently provides an "overall risk score:"



In this case, the overall risk score of 24 is based solely on "proximity." As mentioned in the DomainTools blog article entitled "Introducing Our Innovative Domain Risk Score": (see),

> *Proximity tells us how closely connected a given domain is to domains that are on well-vetted industry blocklists. This has been a very effective way to identify dangerous domains, and it is a component of our new Risk Score.*

Older, well-established domains (such as the World Curling Federation site) get scored *solely* based on proximity. Younger domains (e.g., domains no more than 30 months old), have a broader range of scores, including specific scores for spam, phishing, and malware risks. A registered domain's overall risk score will be the maximum value from the set of the domain's (1) proximity score, (2) spam score, (3) phishing score, and (4) malware score, unless the domain has been zerolisted (or received a forced 100 override). Given its comparatively low risk score of 24, the example World Curling Federation's website is obviously fine to visit (as we'd expect).

Iris Investigate, as shown above, is a graphical user interface. If we prefer (and we have a suitable DomainTools subscription), we can also get risk scores via the DomainTools API For example, using the DomainTools command line interface tool, we see equivalent JSON-format risk score results that look like:

```
$ domaintools risk worldcurling.org
{
    "response": {
        "domain": "worldcurling.org",
        "risk_score": 24,
        "components": [
            {
                "name": "proximity",
                "risk_score": 24
            }
        ]
    }
}
```

Domains That Begin with a Number

Our research focuses specifically on risk scores for domains that begin with a number. Registered domains that begin with a number have been permitted since October 1989, (see "Requirements for Internet Hosts," https://www.rfc-editor.org/rfc/rfc1123 at section 2.1). Therefore, by now, there SHOULDN'T be anything unusual about that subset of domains, yet those domains remain (perhaps paradoxically) uncommon.

If domains were uniformly distributed across potential starting characters, we'd expect that domains beginning with a digit would constitute (10/(26+10))*100 = 27.7% of all domains. Reviewing Appendix A, however, we can see that the fraction of registrable domains that begin with a digit is actually quite low, with a median of just 4.2% and an average of 7.9% computed ASN-by-ASN across our 174 studied autonomous systems.

Autonomous systems located in the APNIC region tend to have some of the highest rates for domains beginning with a digit (perhaps for cultural or linguistic reasons as discussed in "The Secret Messages Inside Chinese URLs,").

Speaking of ASNs, What Are They?

Autonomous system numbers (or "ASNs") were originally created and assigned for use by network engineers for wide-area-routing-related purposes.

Over time, however, ASNs have also come to be used by *non*-network engineers as a succinct way of referring to a group of network prefixes originated by an organization. Most large organizations/ISPs will have at least one ASN, and some may have multiple ASNs (perhaps used by different business units or different geographical regions with different routing policies).

There are over 110,000 assigned ASNs, too many for us to meaningfully analyze in this limited pilot study. Therefore, we began by selecting some of the largest/most important ASNs as analysis candidates, augmented with other ASNs which had been mentioned elsewhere. Sources consulted included (in alphabetical order):

- CAIDA's ASRank site
- CIDR Report
- Hamachek's Bad ASN List
- Hurricane Electric's BGP Toolkit
- Spamhaus's "World's Worst Botnet ASNs"
- Tor's Good/Bad ISPs

Ultimately, we considered a total of 785 ASNs for the study. To be included, we imposed a minimum threshold of at least 2,000 unique registered domains beginning with a digit. The 174 ASNs that met this threshold are listed in Appendix A; the remaining 611 ASNs (which we excluded) are in Appendix B.

Finding the Unique Registered Domains Associated with an ASN

DomainTools risk scores are currently computed for registrable domain names, not ASNs.

ASNs originate IP address prefixes, not registered domains.

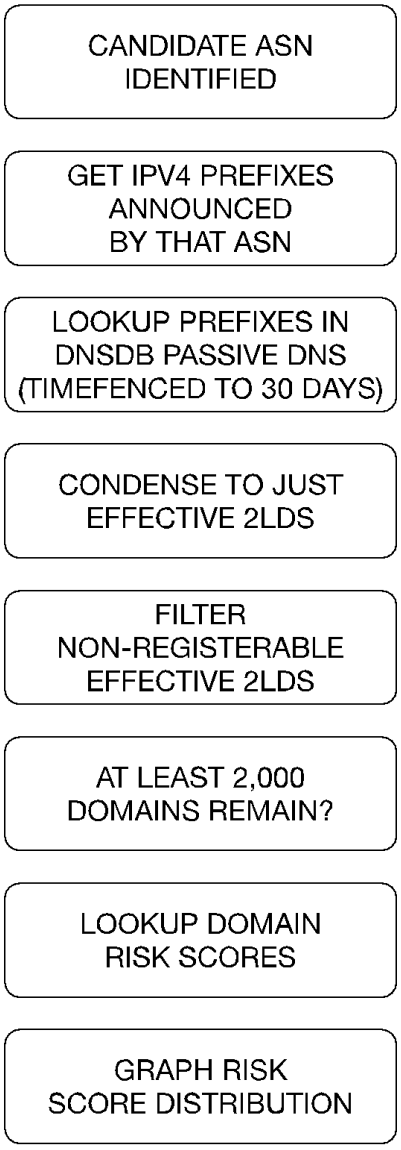
How to bridge that chasm? Fortunately, we can leverage 3rd party routing data and Farsight DNSDB (now part of DomainTools) to make the required connection.

3rd party routing data can provide us with a list of the CIDR prefixes from which an ASN originates. The Oregon Routeviews project is one excellent potential source for ASN --> prefix mapping data. For an easy-to-use GUI interface, look up an ASN of interest on Hurricane Electric's excellent prefix tab.

Once we have a list of IP address prefixes, we can look them up in DNSDB. This allows us to find the FQDNs using those IPs.

We can then reduce those FQDNs to unique registered base (“apex”) domains, and finally query DomainTools for their corresponding risk scores. The full "processing pipeline" looks like:

Processing Pipeline



Part B. Challenges

Gathering Risk Scores

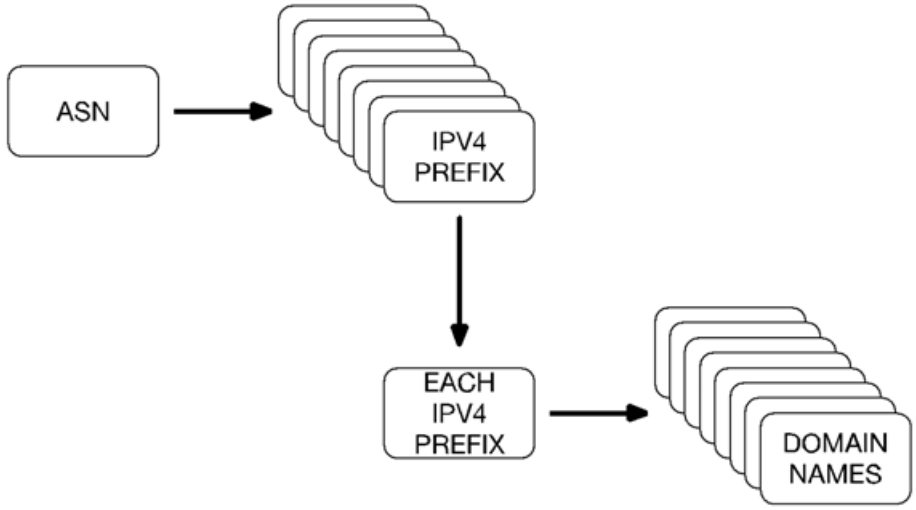
There are three different methods for collecting risk score data at scale and all of them come with benefits and tradeoffs. The first method is to use the DomainTools Domain Risk Score API. The second method would be to use one of Iris family APIs, such as the Iris Enrich API or Iris Investigate API, which include the Domain Risk Score. The third method is to download and use the DomainTools Risk feed. We will review each of these methods briefly and discuss their benefits and limitations for performing this research.

The Domain Risk Score API was designed for enrichment and triage of domain names within custom tools or one of the DomainTools SIEM/TIP integrations. It is focused on providing risk scores quickly and efficiently. With its design of providing 120 queries per minute, It is not ideal for doing this sort of large scale analysis. Given a specified rate limit, we can compute the ceiling for how many risk scores we could look up per-day. For example, in our case:

> (120 checks/min) * (60 min/hr) * (24 hrs/day) = a theoretical maximum of 172,800 checks/day/account

At that theoretical rate, checking even a million registered domains would take nearly six days. That's a LONG time if you've got a bunch of ASNs to explore:

- Even a single ASN may originate dozens/hundreds/thousands of CIDR prefixes (each spanning thousands or even millions of IPv4 addresses), and
- Each of those network address blocks may in turn host dozens, hundreds, or even thousands of domains.



Nevertheless, this is the method we used for this research. Given some pragmatic limitations we completed a little over 100,000 domain queries per day.

The Iris APIs are far more performant than the standalone Domain Risk API we used for this study, in part because you can now pass a list of up to a hundred domains per call. Moreover, "the API considers a batch request as a single query for accounting and rate limiting purposes, so batches are encouraged anytime you have more than a single domain name to profile." The Iris API is a real bargain given this special treatment!

The Iris API will also report if a domain is "active" (e.g., has been satisfactorily actively resolved within the last month), or "inactive." If inactive, the Iris API will still report the last-known risk score; the Domain Risk API will simply report that the domain was not found.

Last, when using the Iris API, you'll get a more inclusive response, returning many potentially interesting attributes associated with the queried domain rather than just a risk score, but this means that you may need to devote attention to extracting just the bits and pieces of interest from the JSON results. For example:

```
$ domaintools iris_investigate --domain "worldcurling.org" | jq -r
'"\(.response.results[].domain)
\(.response.results[].domain_risk.components[])"'
worldcurling.org {"name":"proximity","risk_score":21}
```

Finally, there is the Domain Risk feed. This feed is generated daily and contains a list of domains and risk scores for all domains who scored 70 or greater. This feed is designed for customers who need rapid access to all domains which DomainTools considers suspect and possibly registered with malicious intent. The obvious advantage to using the file is that you have all of the risks available locally for your own deeper analysis. For example, high-throughput analysis using this feed can be accomplished using MTBLs, see "Efficiently Accessing a Moderately-Large Sorted and Uniquely-Keyed CSV File in Python3 with MTBL". This feed, unfortunately, was not applicable for this research because we needed all scores for all domains, not just those scored at 70+.

What Constitutes “Registererable Domains”?

In addition to working through our throughput limits, we also worked through issues around "effective 2nd-level domains" (as defined in the Public Suffix List, see https://publicsuffix.org/list/public_suffix_list.dat) vs "registered domains" (as reflected in Whois and used for DomainTools data).

For example, consider FQDNs ending "myfritz.net" – that suffix is listed on https://publicsuffix.org/list/public_suffix_list.dat. This means that the Public Suffix List would have us treat different myfritz.net hosts as unique entities.

However, if you check <anything>.myfritz.net in DomainTools Iris, ALL myfritz.net FQDNs get collapsed to JUST myfritz.net for risk scoring purposes. This means that there's no point in repeatedly asking for tens of thousands of unique myfritz.net domains if they're all just going to be reduced down to myfritz.net (and all those queries will return the same risk score).

Appendix C contains a list of domains that we condensed for risk scoring purposes. That list will likely continue to expand over time.

Timing and Inclusion Limitations of This Pilot Study

Data was largely collected and analyzed during May and June 2022, and passive DNS was timefenced to 30 days preceding the date of collection. This is both a protracted window over which to pull data, and simultaneously a short study period. While we have no reason to believe that May and June 2022 were in any way "special" months when it comes to registered domains beginning with a digit, neither can we guarantee that they weren't – one hot-running DGA with a preference for domains beginning with a digit could obviously have had a material impact.

Remember that this is just a pilot study, and should ideally be replicated at greater scale.

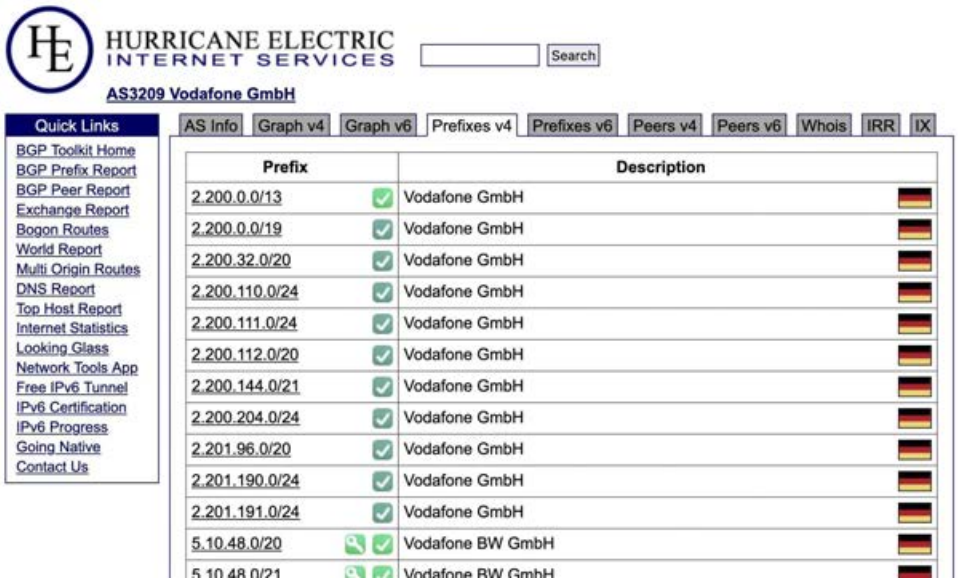
We also want to re-emphasize that while we screened 785 ASNs and studied 174, the ASNs that we picked were selected by us, were subject to a 2,000 domain minimum threshold, and are NOT a representative "random sample." For those reasons, we urge you to avoid attempting to draw broad general inferences based on this pilot study.

Part C. A Worked Example

Checking A Sample ASN – Vodafone (AS3209)

To this point, we've largely been talking in generalities. Let's make this concrete by looking at an example.

According to https://bgp.he.net/AS3209#_asinfo, Vodafone, the multinational telecom company, originates 2,042 IPv4 CIDR prefixes, representing a total of 12,766,976 IPv4 addresses as of the time the data for this report was collected. A portion of those look like:



Those prefixes include both "more specific" prefixes (such as 2.200.0.0/19 and 2.200.32.0/20) and larger "covering" prefixes (such as 2.200.0.0/13), but since we'll eventually deduplicate our domains, we're not going to worry about any overlapping CIDR prefixes that may be present.

We look up each of the Vodafone prefixes in DNSDB by making a series of scripted queries of the form:

```
dnsdbq -i 109.192.0.0/15 -l0 -A30d -j -T datefix >> dnsdb-output.jsonl
```

Each such query will find domain names associated with the specified prefix. The only thing that will vary from query to query is the network prefix, shown here in red. Decoding the specific options used for that command:

- **`dnsdbq`** is our command line DNSDB client, see https://github.com/dnsdb/dnsdbq
- **`-i 109.192.0.0/15`** is how we'll specify the CIDR prefix we want to look up in DNSDB
- **`-l0`** (dash ell zero) asks for up to a million results, if that many results are available
- **`-A30d`** timefences our results to entries that have been seen at least sometime in the last 30 days
- **`-j`** requests output in JSON Lines format
- **`-T datefix`** asks for datetime fields in human readable form rather than Un*x seconds
- **`>> dnsdb-output.jsonl`** asks for our output to be appended to the specified file

Imagine a command file with several thousand `dnsdbq` lines of that sort, one for each identified prefix (it is easy to copy and paste a list of prefixes and build a script of similar commands using vim or your favorite text editor).

Having created that file, assuming it was called `dnsdbq-queries.sh`, we'd run it by saying:

```
$ bash dnsdb-queries.sh
```

When that script finished, we'd probably have a fair-sized output file. In this case:

```
$ wc -l dnsdb-output.jsonl
3387175 dnsdb-output.jsonl          <-- nearly 3.4 million lines of output
```

The lines in that output file contain a LOT more data than we need. Therefore, as a first step in processing that output, we can ignore the RRtype, Rdata and other fields, and extract just a list of unique FQDNs using jq (see https://stedolan.github.io/jq/):

```
$ jq -r '.rrname' < dnsdb-output.jsonl | sort -u > fqdns-only.txt
$ wc -l fqdns-only.txt
2283438 fqdns-only.txt
```

That single step removes about 1/3rd of our original results. Now let's condense those FQDNs to effective 2nd-level domains as defined by the Public Suffix List. We'll do that with a little script called `2nd-level-dom-large` that leverages the Public Suffix List (for a copy of that script, see Appendix D). Sample usage:

```
$ 2nd-level-dom-large < fqdns-only.txt | sort -u > 2lds-only-sorted-uniqued.txt
$ wc -l 2lds-only-sorted-uniqued.txt
417669 2lds-only-sorted-uniqued.txt
```

We've dropped from nearly 3.4 million FQDNs to just over 400,000 delegation points, but "just over 400,000" delegation points is still a lot of domains to look up. **We need to remember that we're only interested in domains that begin with a digit.** We'll use egrep to do additional filtering:

```
$ egrep "^[0-9]" 2lds-only-sorted-uniqued.txt > num-only.txt
$ wc -l num-only.txt
38942 num-only.txt
```

Much better! We're now down to under 39,000 domains. Looking at what's left, we note that over 37,000 of the remaining 38,942 lines are associated with just ONE single 2nd-level domain, myfritz.net, including:

```
00miprgyr7s8d7ie.myfritz.net
00mkxfrx2lwehm5w.myfritz.net
[etc]
```

myfritz.net is an example of a domain that DomainTools collapses to a single registrable domain, so we'll exclude the myfritz.net domains (and the other domains listed in Appendix C):

```
$ egrep -v -f stuff-to-filter.txt num-only.txt > num-only-filtered.txt
$ wc -l num-only-filtered.txt
732 num-only-filtered.txt
```

We're now down to just 732 domains. Normally we wouldn't even bother looking any further at this ASN (treating it as being "too small"/irrelevant for the purposes of our study, since it is under our 2,000 domain threshold), but we'll "keep going" for the sake of this example.

Now that we have a list of candidate domains to risk score, let's build a set of queries to retrieve those scores. Just as we modified a list of prefixes announced by an ASN to create a script of `dnsdbq` commands, we'll modify our list of effective 2nd-level domains to create a script of `domaintools risk` queries.
A representative script line looks like:

```
$ domaintools risk 000111000.eu >> risk-score-output.jsonl ; sleep 0.25
```

Decoding that command:

- **`domaintools`** is the company's Python and CLI API access tool (see https://github.com/DomainTools/python_api)
- **`risk`** is one of multiple API endpoints the tool can access (see `$ domaintools -h` for a listing of other options)
- **`000111000.eu`** is one domain we want to "risk score"
- **`>> risk-score-output.jsonl`** appends the output of the query to this file
- **`; sleep 0.25`** is a delay to keep us within our queries per hour rate target (adjust as appropriate)

The output from running that command looks like:

```
{
    "response": {
        "domain": "000111000.eu",
        "risk_score": 1,
        "components": [
            {
                "name": "proximity",
                "risk_score": 1
            }
        ]
    }
}
```

Because this domain is a well-established one, it does NOT include a full threat profile with multiple subscores; Domaintools only provides that full risk profile for domains that are **less than 30 months old.** As a result, this domain has a very minimalist domain report. Newer domains are more detailed, such as this report (its for a domain created in May 2021):

```
{
    "response": {
```

```
            "domain": "0441.store",
            "risk_score": 17,
            "components": [
                {
                    "name": "proximity",
                    "risk_score": 1
                },
                {
                    "name": "threat_profile",
                    "risk_score": 17
                },
                {
                    "name": "threat_profile_phishing",
                    "risk_score": 17
                },
                {
                    "name": "threat_profile_malware",
                    "risk_score": 11
                },
                {
                    "name": "threat_profile_spam",
                    "risk_score": 1
                }
            ]
        }
    }
```

Once we've run our script for all of the domains starting with a digit that are associated with this ASN, we can then extract just the bits we need with a jq command such as the following:

```
$ jq -r '"vodafone-as3209,\(.response.domain),\(.response.risk_score)"'
< risk-score-output.txt | sort -u > vodafone-as3209.csv
```

The resulting CSV lines looks like the following:

```
vodafone-as3209,000111000.eu,1
```

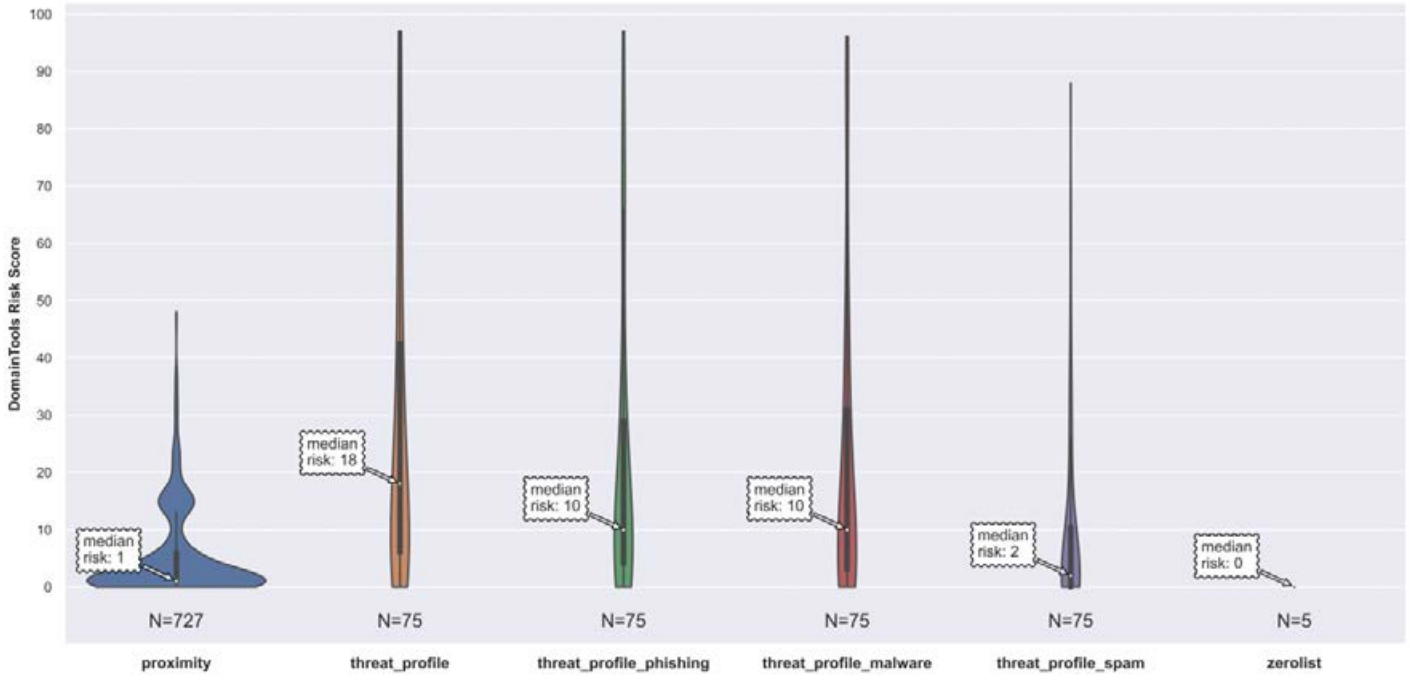
We can then combine that CSV file with CSV files for other ASNs of interest (plus a stub header file) and run it through the multi-ASN violin plot code shown in Appendix E. (Don't bother looking for the Vodafone ASN in the plots in the next section – remember it had too few domains starting with a digit to bother including it as part of the final actual analysis).

Continuing to "play through" for the purposes of this example, however, let's move onto per-ASN sub-score plots. We need to do a little data re-formatting to be ready to make those graphs. The code to do that required reformatting can be seen in Appendix F. We'd run that code by saying something like:

```
$ ./read_jsonl.py < amazon-as16509.jsonl > all-risks-output.csv
```

Then, running code similar to that in Appendix G to show the subscores for Vodafone AS3209, we can see the risk subscores. Because we scaled our violin plots by count for the intra-ASN violin plots, it's obvious that most

of the observations must be based on proximity sub-scores (N=727) since only 75 observations even HAVE threat profiles, and only 5 observations were zerolisted.



Part D. Risk Scores Across the Studied ASNs

Tabulated Risk Scores

The ASN, home country code, median (50th-percentile) risk score, and number of domains beginning with a number for each of our studied ASN can be seen in the following table. The ASN names are shown in the table in an abbreviated form; for ASNs you may not recognize, you can look up the unknown ASN by its number:

```
$ whois as3320
[...]
org-name:       Deutsche Telekom AG
```

ASN	ASN Home Country	Median Risk Score	N
dtag-as3320	DE	0	2,332
neue-medien-as34788	DE	1	10,352
one-as51468	DK	1	10,076
strato-as6724	DE	1	25,894
namesco-as8622	UK	2	4,234
netcup-as197540	DE	2	4,012
ionos-as8560	DE	3	29,545
infomaniak-as29222	CH	4	2,898
host-eu-as20738	DE	5	16,320
host-eu-as34011	DE	5	5,940
host-eu-as8972	DE	5	9,129
iomart-as20860	UK	5	3,769
sk-broadband-as9318	KR	5	2,847
xneelo-as37153	ZA	5	3,828
lg-dacom-as3786	KR	6	5,223
dreamscape-as38719	AU	7	8,348
new-century-as9919	TW	7	2,383
ip-volume-as58110	CH	8	8,502
godaddy-as398101	US	9	17,128
host-eu-as20773	DE	9	35,654
host-eu-as21051	DE	9	22,427
squarespace-as53831	US	9	33,666
godaddy-as26496	US	10	50,536
hetzner-as24940	DE	10	45,299
ru-center-as48287	RU	10	7,005
selectel-as49505	RU	11	4,532
aptum-as13768	CA	13	7,536
att-as7018	US	13	2,197
comcast-as7922	US	13	2,578
regru-as197695	RU	13	18,656

ASN	ASN Home Country	Median Risk Score	N
weebly-as27647	US	13	7,506
google-as19527	US	15	11,317
host-eu-as21499	DE	15	2,260
inmotion-as22611	US	15	3,123
inmotion-as54651	US	15	3,299
new-dream-as26347	US	15	15,024
newfold-as29873	US	15	7,196
rackspace-as33070	US	15	6,654
sakura-as9371	JP	15	6,546
scaleway-as12876	FR	15	3,419
confluence-as40034	VG	16	35,952
gandi-as29169	FR	16	8,985
hk-comm-as140227	HK	16	57,538
singlehop-as32475	US	16	3,950
small-orange-as62729	US	16	2,325
timeweb-as9123	RU	16	9,315
cn-networks-ix-as4847	CN	17	2,195
hk-bb-as10103	HK	17	18,413
hk-bb-as9269	HK	17	18,561
korea-telecom-as4766	KR	17	13,330
china-telecom-as4812	CN	18	3,324
hz-alibaba-as37963	CN	18	58,528
oso-grande-as26337	US	18	3,550
baidu-as38365	CN	19	2,238
google-as15169	US	19	280,187
leaseweb-as60781	NL	19	14,750
ddos-guard-as57724	RU	20	4,033
unified-layer-as46606	US	20	73,156
china-mobile-as56048	CN	22	2,165
china-unicom-as4808	CN	22	6,386
huawei-as55990	CN	22	4,038
idc-chinanet-as23724	CN	22	5,935
sedo-as47846	DE	22	55,254
chinanet-as38283	CN	23	4,538
stackpath-as33438	US	23	4,345
china-mobile-as9808	CN	24	5,886
chinanet-as4134	CN	24	27,011
google-as396982	US	24	50,816
linode-as63949	US	24	64,515
ovh-as16276	FR	24	72,659

ASN	ASN Home Country	Median Risk Score	N
contabo-as51167	DE	25	5,949
jsc-iot-as29182	RU	25	5,649
leaseweb-as30633	US	25	11,757
tencent-as45090	CN	25	29,213
vultr-as20473	US	25	31,512
amazon-as14618	US	27	34,468
china-unicom-as4837	CN	27	18,252
fastly-as54113	US	27	14,425
interserver-as19318	US	27	2,464
ehostic-as45382	KR	28	2,469
enzu-as18978	US	28	5,887
hkbn-as17444	HK	28	5,779
nforce-as43350	NL	28	2,720
tw-mobile-as9924	TW	28	3,104
uk-2-as13213	UK	28	5,537
level3-as3549	US	29	7,192
digitalocean-as14061	US	30	48,220
hinet-as3462	TW	30	13,538
hivelocity-as29802	US	30	4,793
level3-as3356	US	30	6,466
pvt-layer-as51852	PA	30	3,680
servihosting-as29119	ES	30	2,643
gmo-int-as7506	JP	31	28,471
limestone-as46475	US	31	3,188
it7-networks-as25820	CA	32	13,455
softlayer-as36351	US	32	25,985
sun-net-as38197	HK	32	20,987
amazon-as16509	US	33	231,523
baidu-as55967	CN	34	2,452
cloudie-as55933	HK	34	23,456
m247-as9009	UK	34	5,772
ntt-as2914	US	34	12,141
bharti-as45609	IN	35	2,296
facebook-as32934	US	35	13,916
twitter-as13414	US	35	13,722
webnx-as18450	US	35	3,751
hostinger-as47583	CY	36	14,789
liquid-web-as32244	US	36	17,183
cloudflare-as13335	US	37	167,148
gigabit-as55720	MY	37	15,340

ASN	ASN Home Country	Median Risk Score	N
sonder-cloud-as133199	HK	38	8,256
trellian-as133618	AU	39	11,911
anchnet-as137443	HK	41	10,232
wholesale-int-as32097	US	41	5,686
sharktech-as46844	US	42	23,784
tencent-as132203	CN	42	94,631
west263-as139021	HK	42	35,431
netsec-as45753	HK	43	15,793
psychz-as40676	US	44	24,352
alibaba-as45102	CN	47	109,849
huawei-as136907	HK	49	7,048
oracle-as31898	US	51	5,743
cnservers-as40065	US	52	95,566
io-flood-as53755	US	52	5,774
namecheap-as22612	US	53	98,539
krypt-as35908	US	56	13,499
rackspace-as19994	US	56	5,559
dedipath-as35913	US	60	11,674
nocix-as25695	US	61	25,695
leaseweb-as7203	US	62	35,206
zen-layer-as21859	US	63	14,085
cogent-as174	US	65	147,885
microsoft-as8075	US	71	67,781
peg-tech-as398478	US	71	7,149
bgpnet-global-as64050	SG	73	117,560
colo-crossing-as36352	US	75	6,205
multacom-as35916	US	75	283,327
peg-tech-as398823	US	75	30,217
power-line-hk-as132839	HK	75	111,142
layerhost-as46573	US	76	14,885
leaseweb-as395954	US	77	33,042
esited-as22552	US	78	14,384
int-hostspace-as399674	US	78	19,991
luogelang-as135097	HK	78	31,329
sun-net-as136800	HK	78	64,553
hurricane-as6939	US	79	11,882
cloud-inno-as134175	HK	80	62,341
take-2-host-as20248	US	80	10,381
peg-tech-as54600	US	83	131,657
quadranet-as8100	US	83	35,509

ASN	ASN Home Country	Median Risk Score	N
dxtl-as134548	HK	85	91,329
egihosting-as18779	US	87	175,389
ucloud-as135377	HK	87	14,791
hostus-as7489	US	89	3,335
eonix-as62904	US	90	16,184
clayer-as137951	HK	91	43,319
ddos-guard-as262254	BZ	92	14,387
frantech-as53667	US	92	17,134
intercontinental-as398968	US	92	5,445
sanren-as139330	HK	92	4,007
chernyhov-as202984	RU	93	2,965
china-mobile-as56046	CN	93	7,372
inmotion-as3842	US	93	16,082
tcloudnet-as399077	US	95	5,279
yisu-cloud-as136970	HK	95	42,791
quick-pck-as46261	US	97	19,904
verotel-as31624	NL	97	38,963
hz-zhiyu-as59037	CN	98	4,354
gorilla-as53850	US	99	46,945
sakura-as9370	JP	99	11,966
tier-net-as397423	US	99	10,320
microsoft-as3598	US	100	9,096
sprint-as1239	US	100	11,692
udomain-as23881	HK	100	3,022

Unfortunately, that's a dauntingly long table – it can be hard to "get a sense" of numeric data from simple tabulations.

Violin Plots for All 174 Studied ASNs (Ordered By Median Risk Score)

Let's create *violin plots* for that data. We know many people may not be familiar with violin plots, but we believe their advantages make it worth experimenting (instead of simply using the more-familiar box plots). In particular, violin plots do an excellent job of representing the "shape" of a distribution (while also having an embedded boxplot included as the "spine" of the violins). Being able to actually show the shape of the graph is particularly important for distributions that might be multimodal, having several significant peaks rather than a single point of central tendency, as we'll see is the case for the DomainTools risk score data.

It can be easy to miss the median risk score in the default violin plot graph format. To overcome that deficiency, we've added explicit annotations to each of the violin plots, calling out the median score and the number of valid observations for each violin to help with interpretation of the various graphs.

We've also included small flags representing the country associated with each ASN. These flags were built from wikimedia.org SVG flag files, scaled to approximately consistently-sized PNG using https://cairosvg.org/

For the violin plots themselves, we used Seaborn, a terrific Python3 graphics library based on Matplotlib. For more on Seaborn, see https://seaborn.pydata.org/. For a copy of the Python3 code used to create the violin plots, see Appendices E, F and G.

Note that the violin plots that compare risk data ACROSS ASNs are set up to have **constant width**. This means that an ASN with just a few thousand domains will have the same width as one with several 100,000 domains. (For more information on Seaborn options and their meaning, see https://seaborn.pydata.org/generated/seaborn.violinplot.html). This particular option ensures that all violins will be wide enough to be interpretable.

The per-risk subplots, shown in the section after this one, however, have had the width of the violin plots set using the **scale="count"** option instead, thereby facilitating intra-ASN comparisons.

We've also used a Seaborn option to clip the violin displays at bottom and top at 0 and 100, since that's the full permitted range of the risk scores.

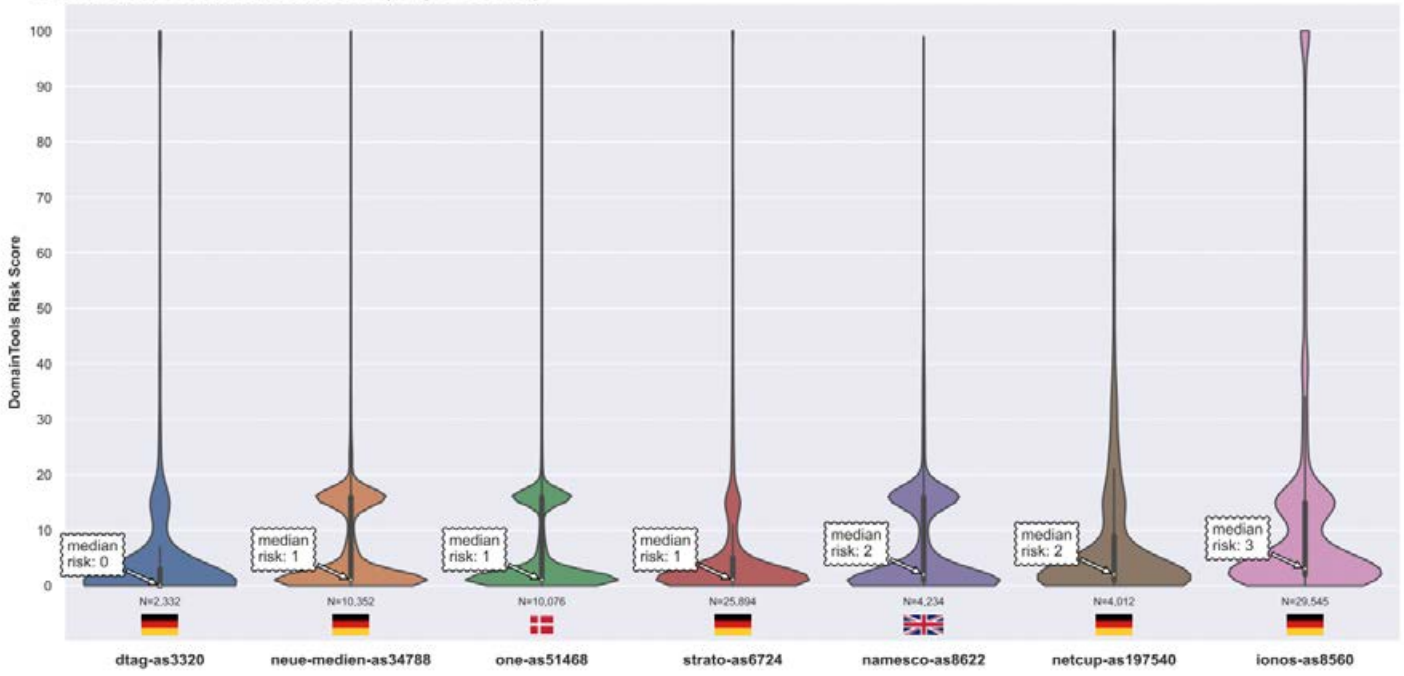
We'll now show you those violin plots for the 174 studied ASNs, ordered by median risk score.

This first set of graphs is broken up into 25 panels, with seven ASNs per panel. Order of ASN presentation within the graphs is by median risk score; in case of ties, ties are ordered alphabetically within the set of ASNs sharing the common score.

This graph series begins on the next page.

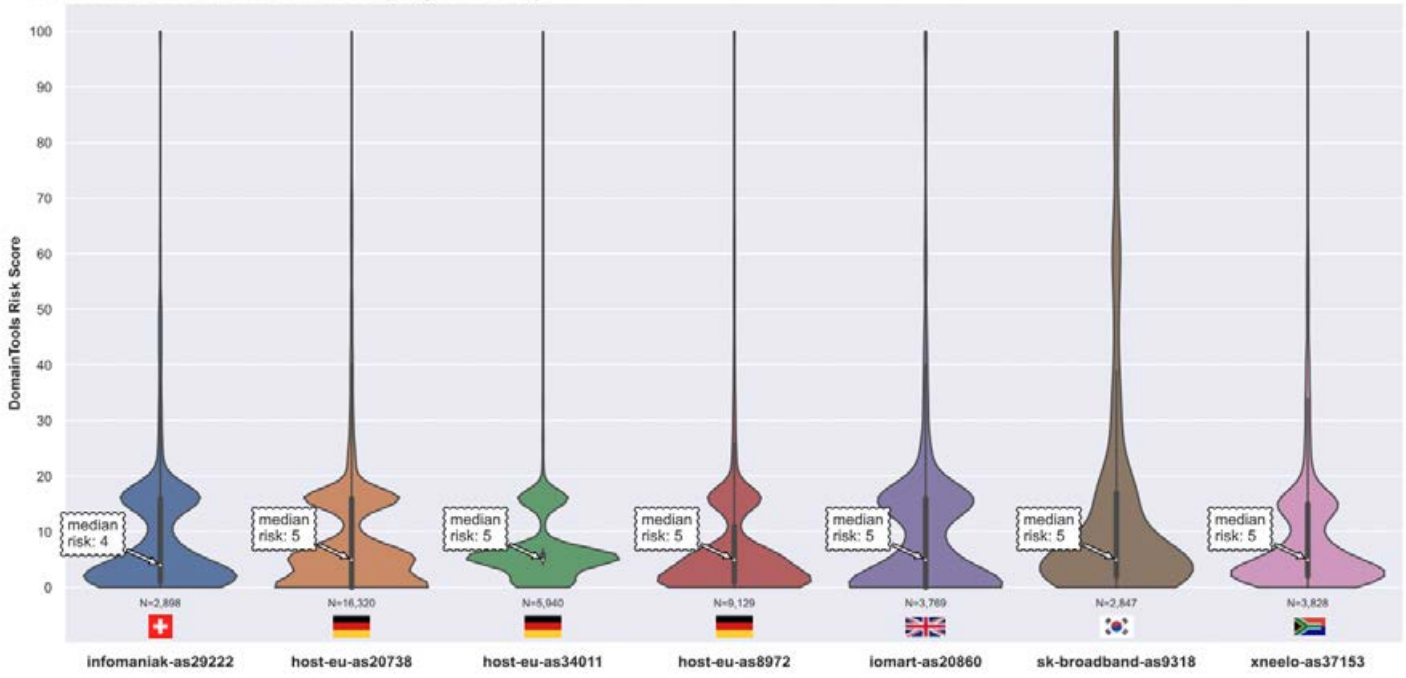
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 1 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



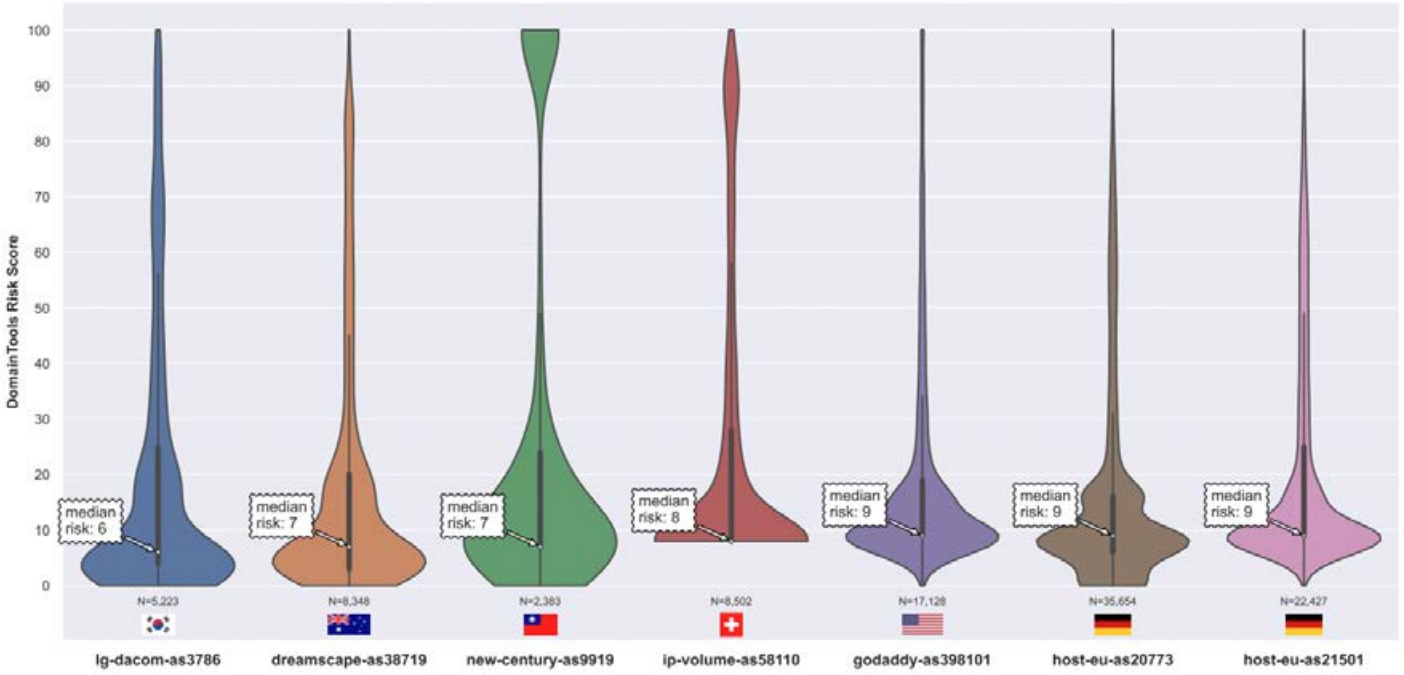
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 2 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



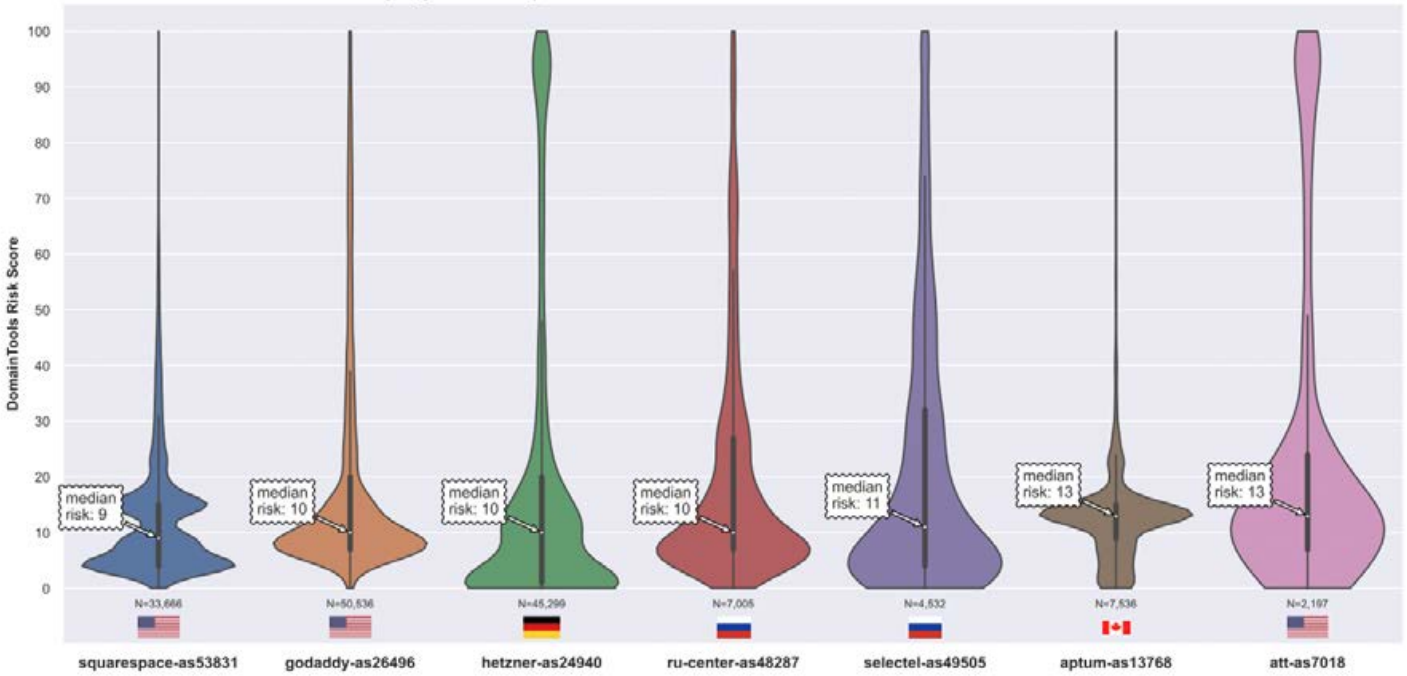
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 3 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



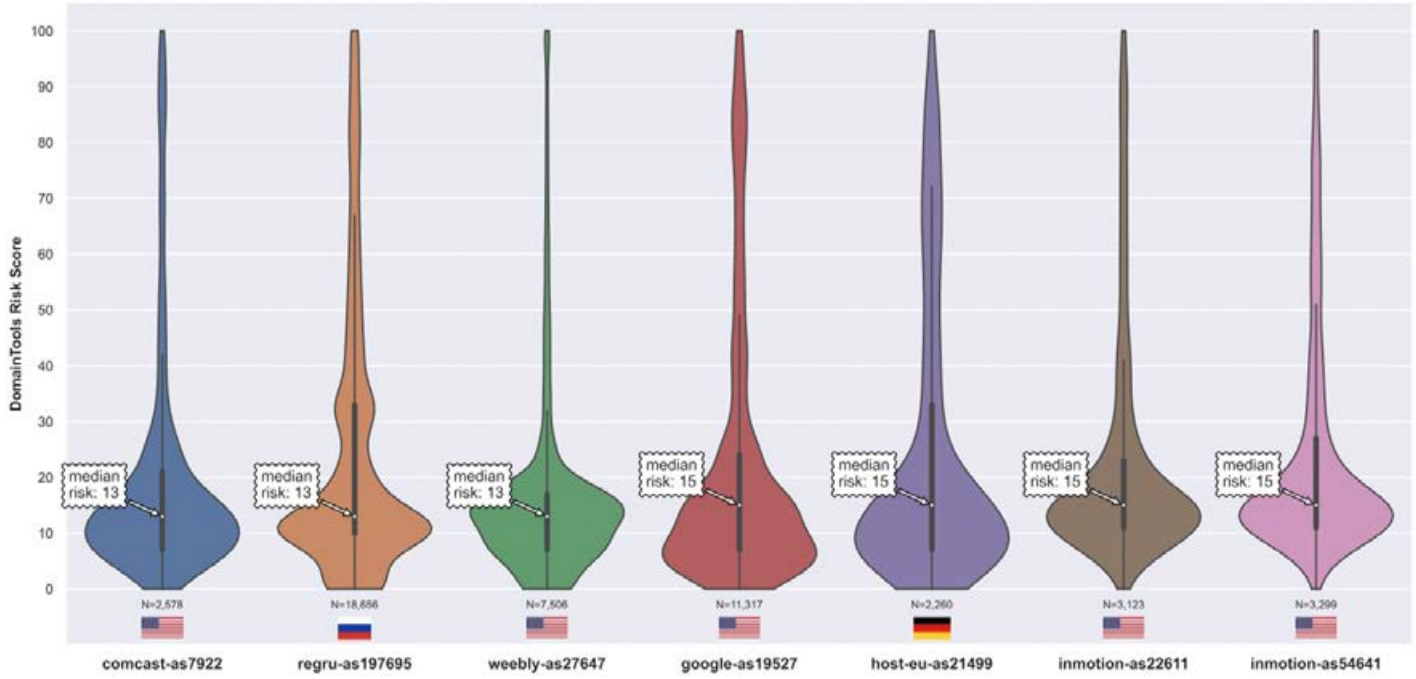
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 4 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



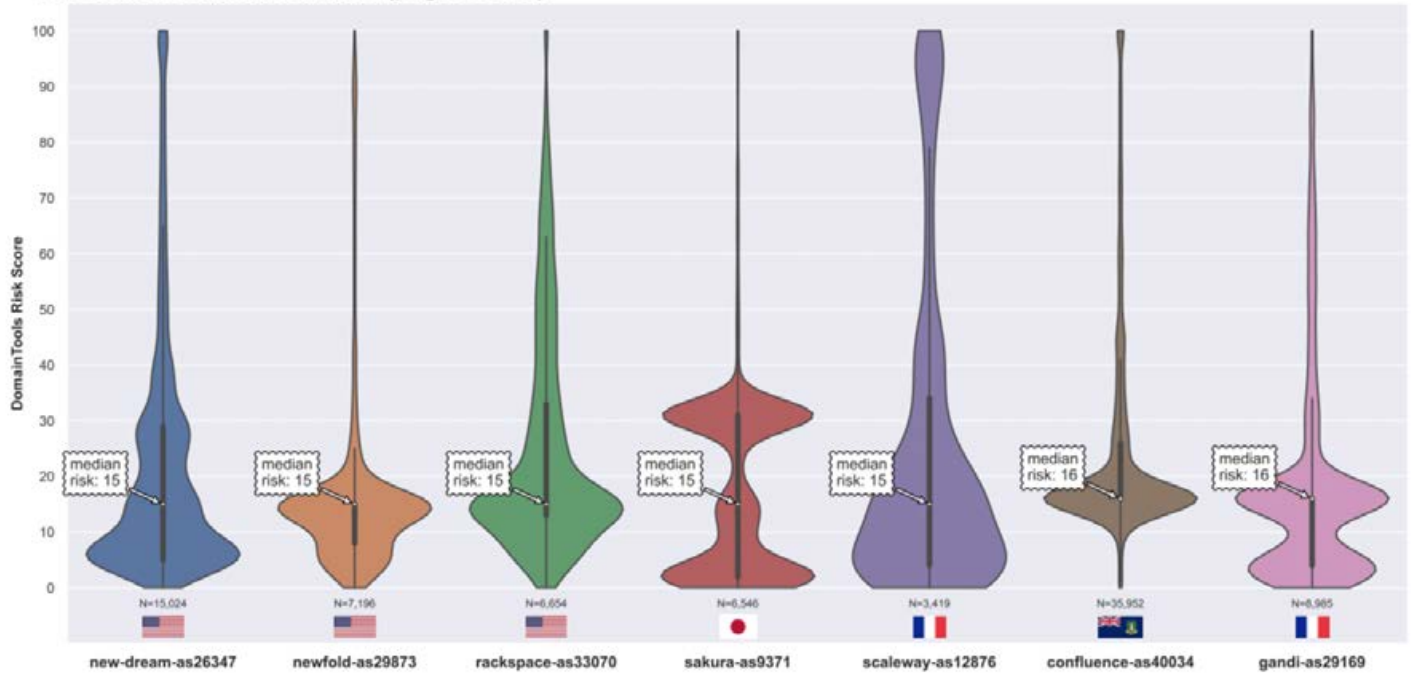
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 5 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



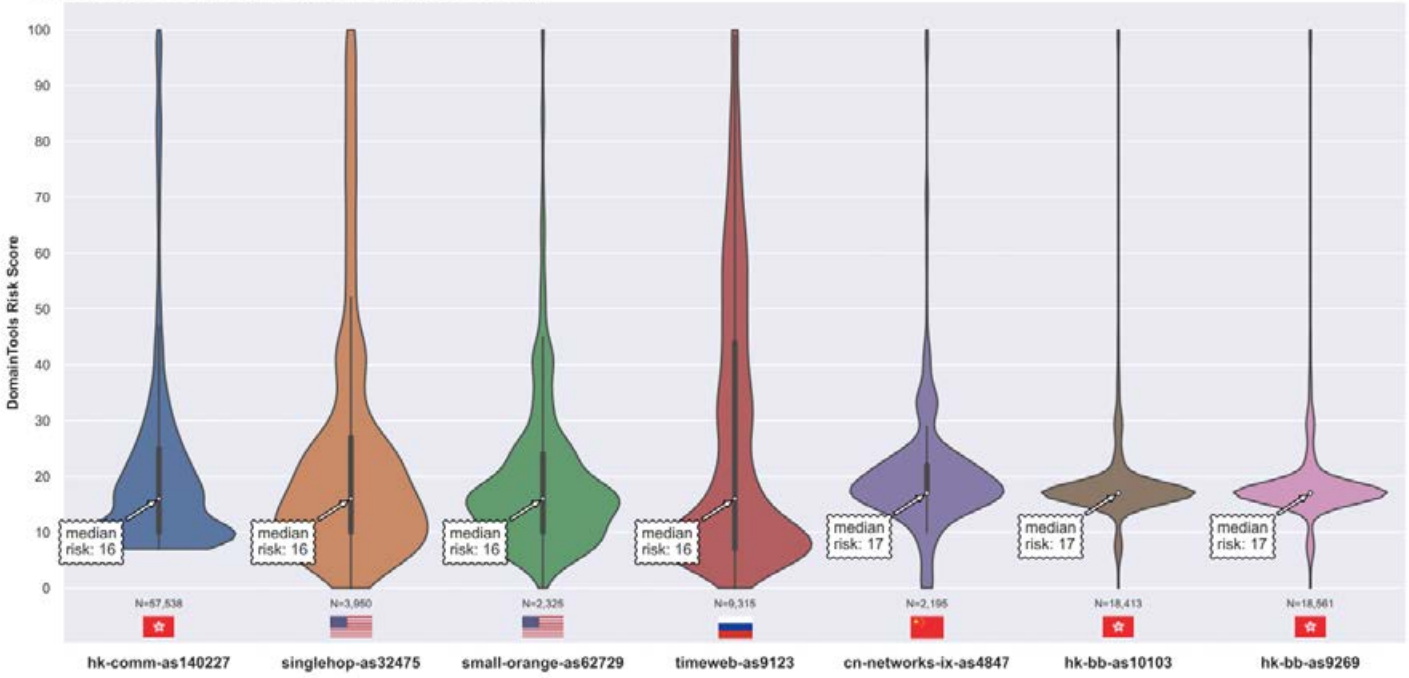
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 6 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



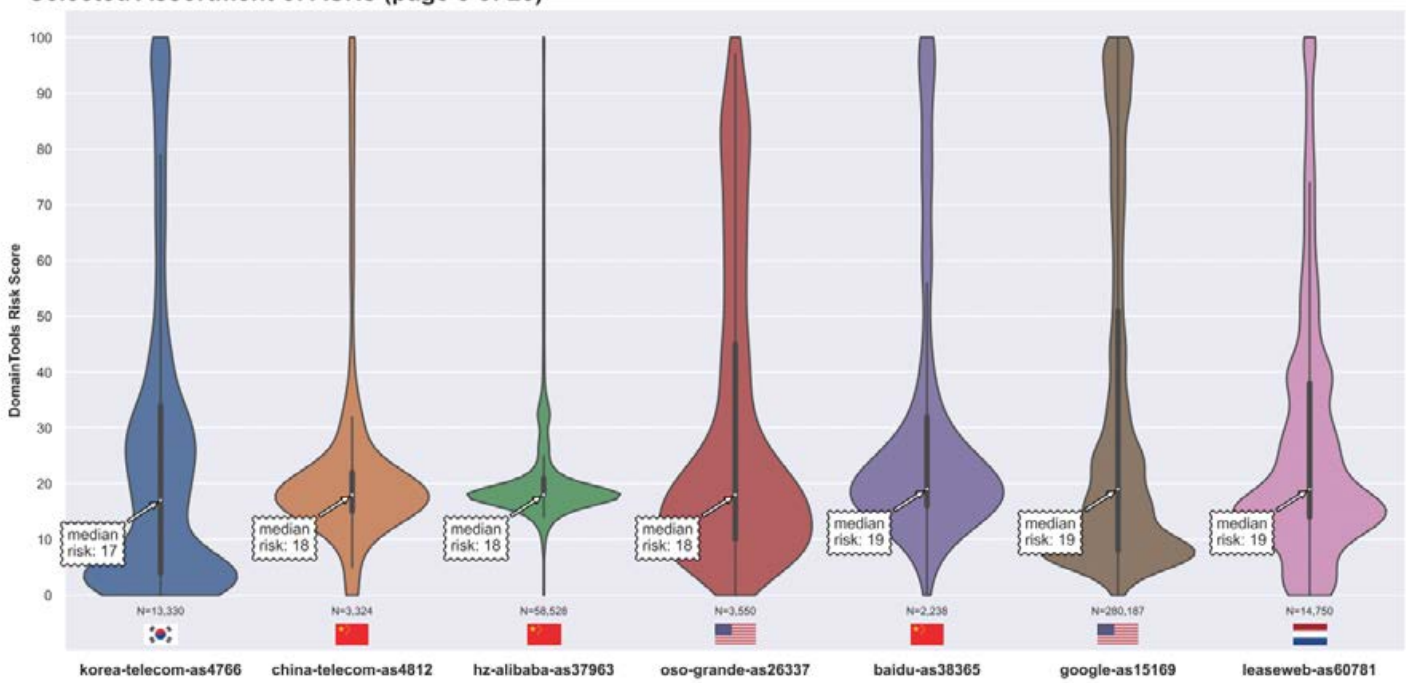
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 7 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



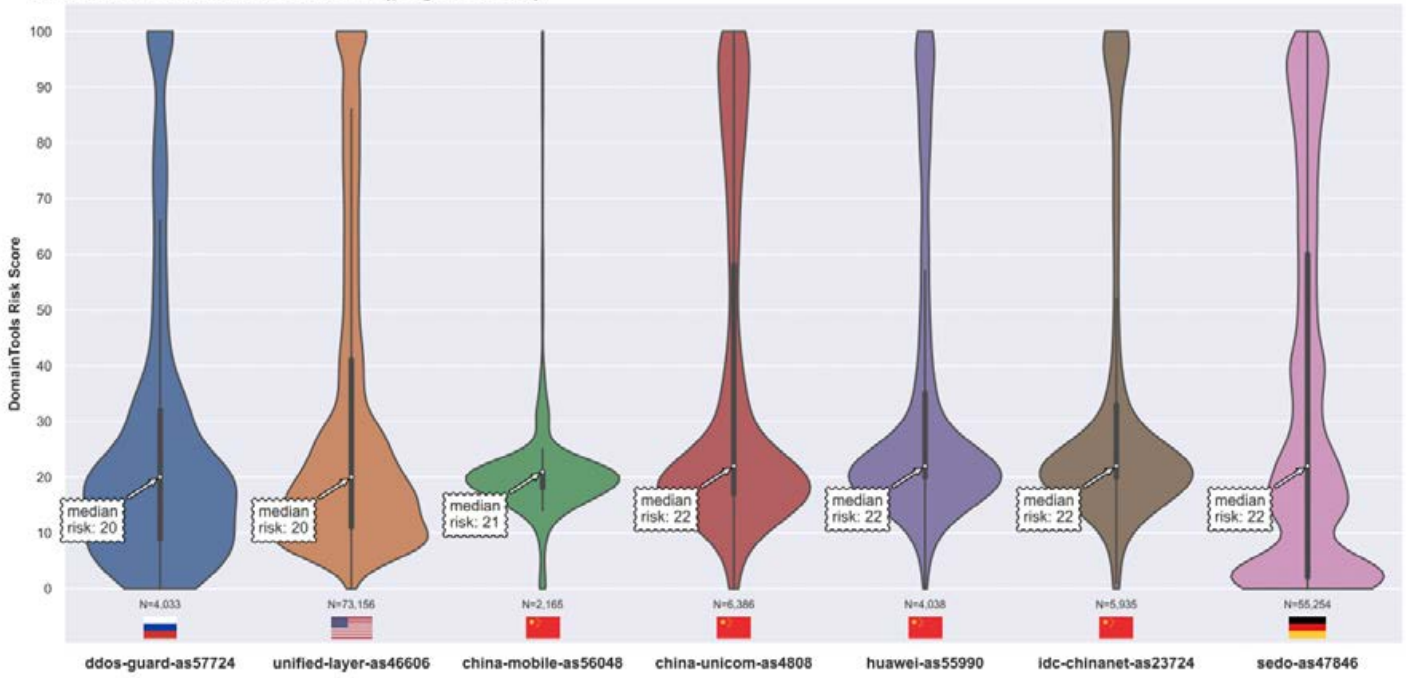
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 8 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



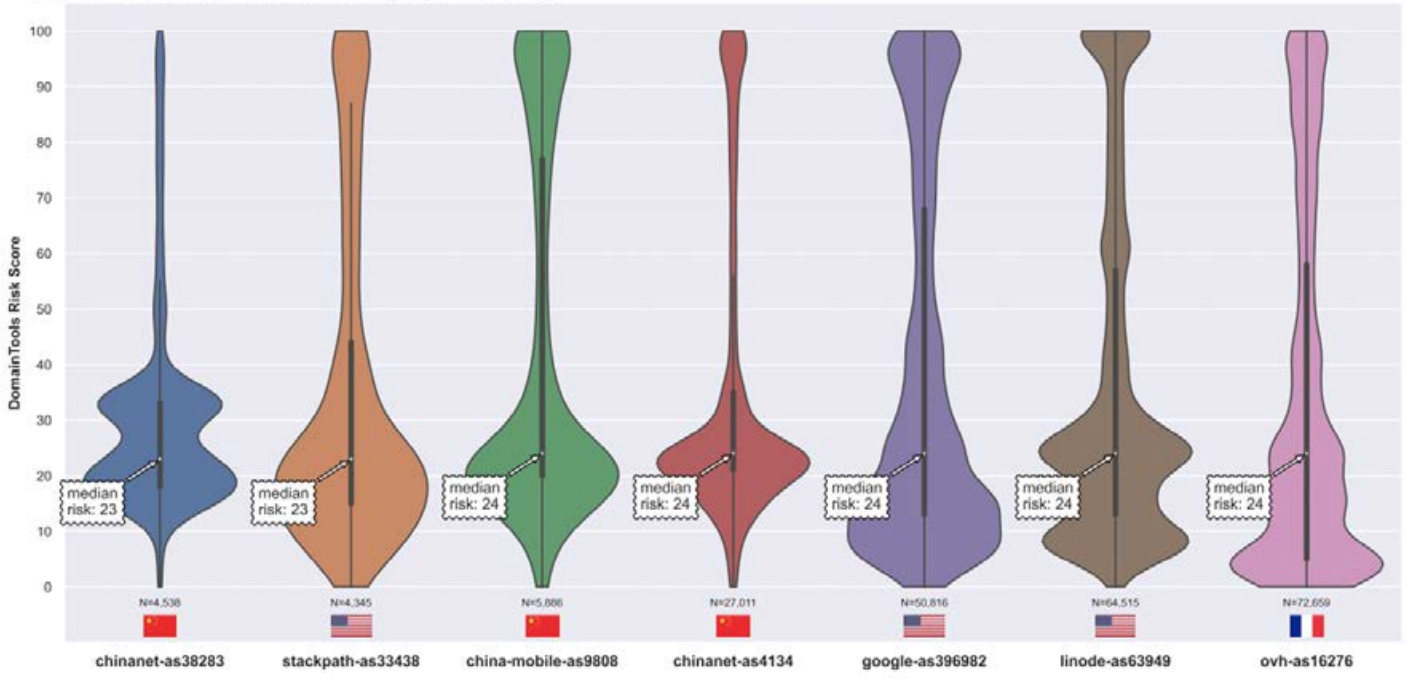
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 9 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



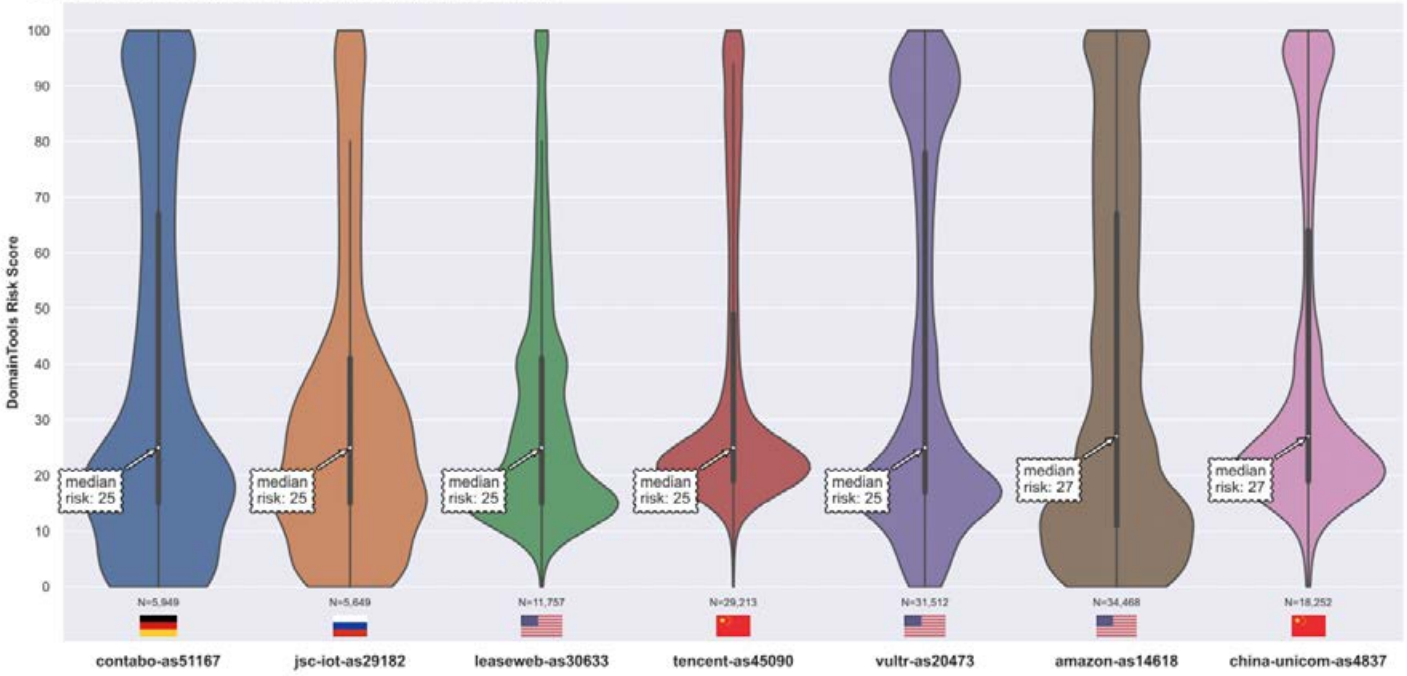
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 10 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



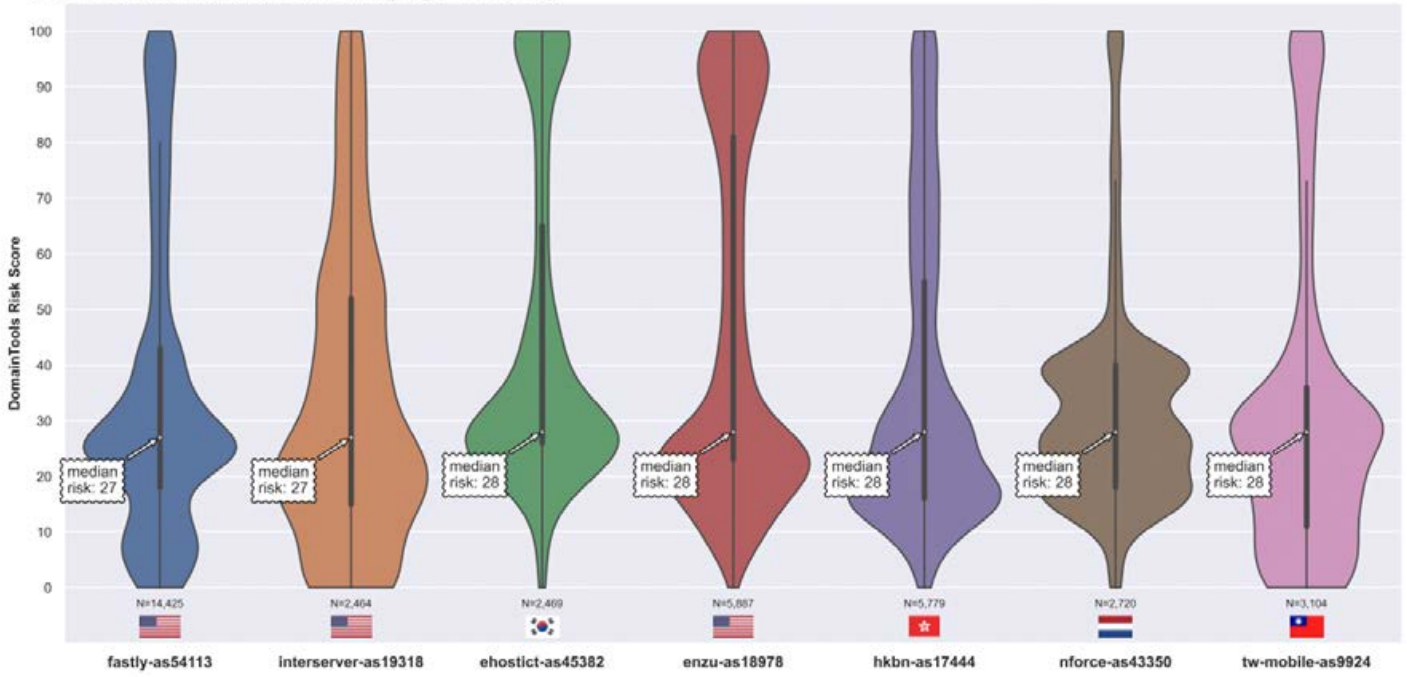
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 11 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



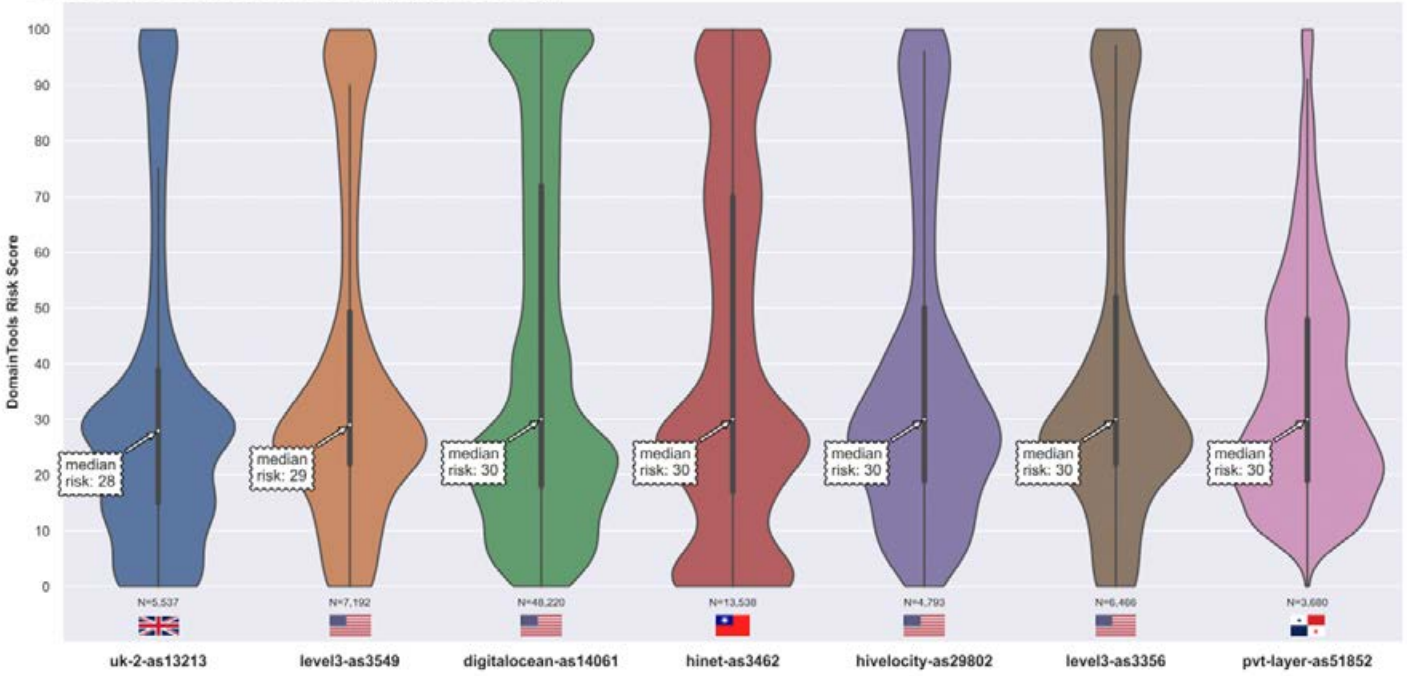
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 12 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



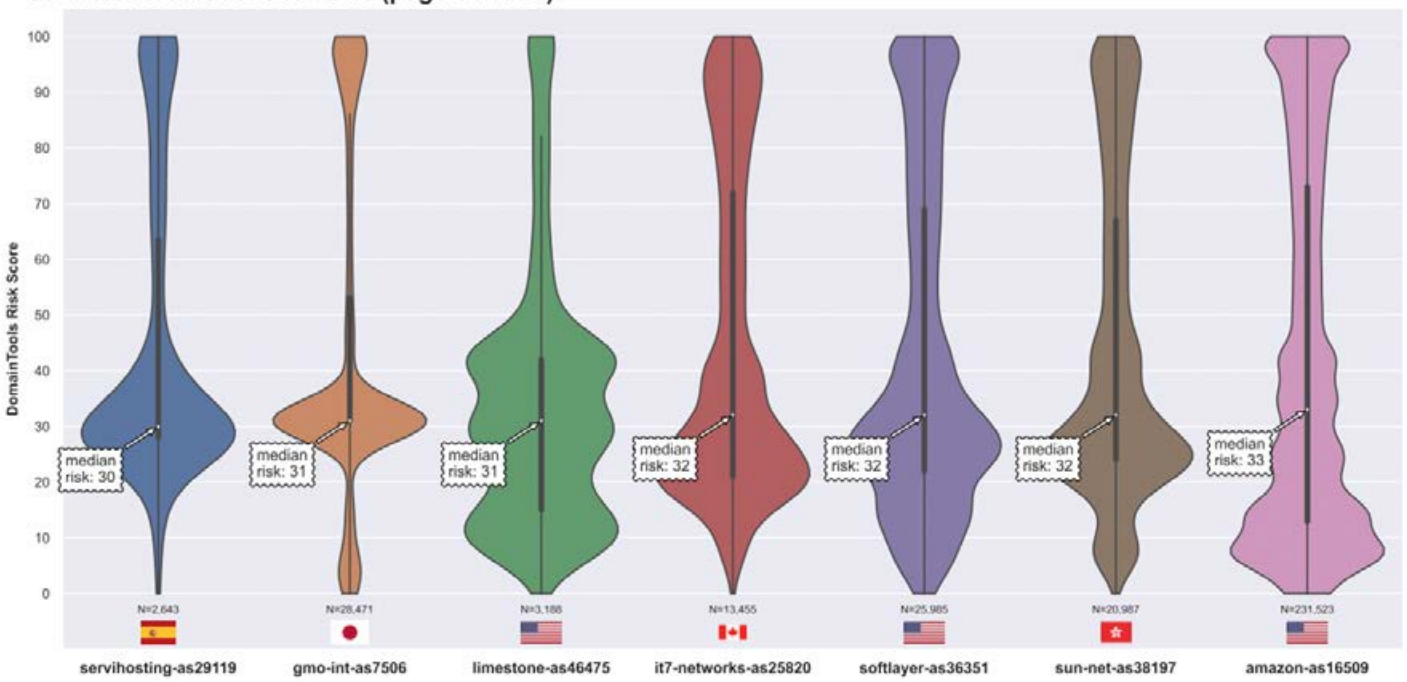
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 13 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



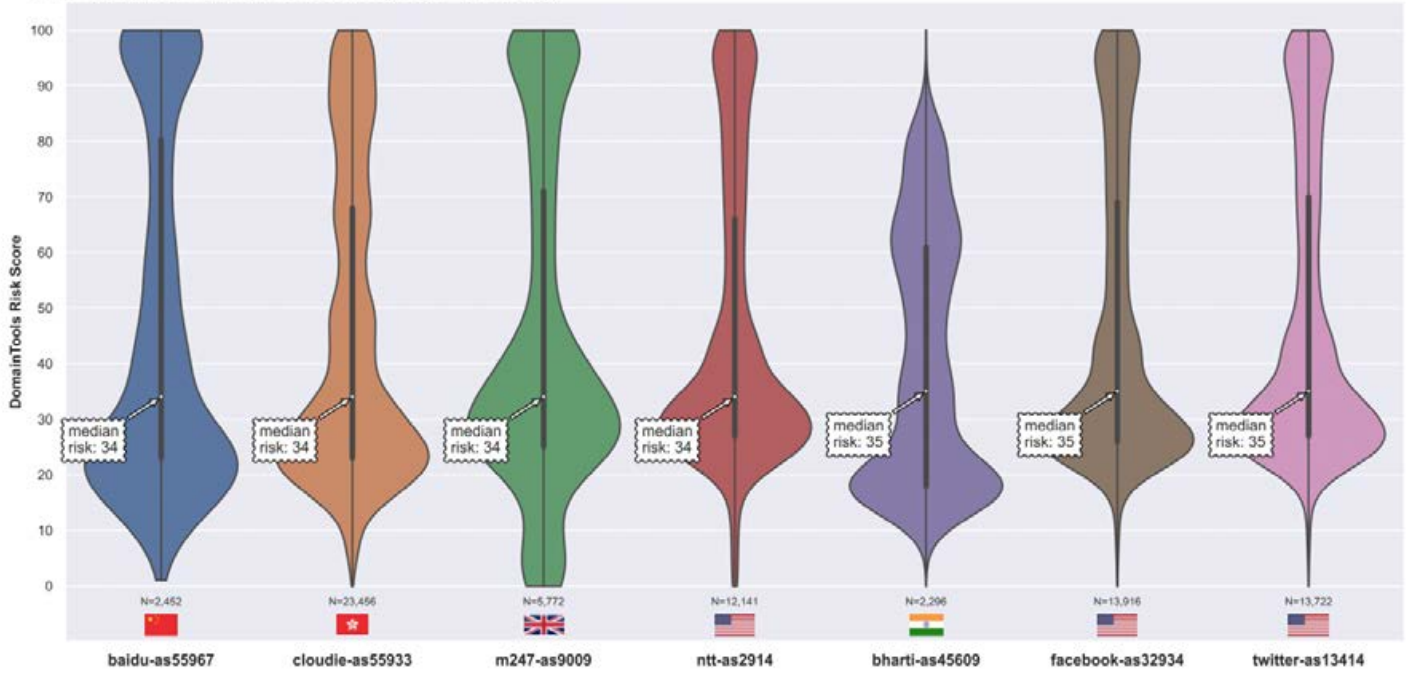
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 14 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



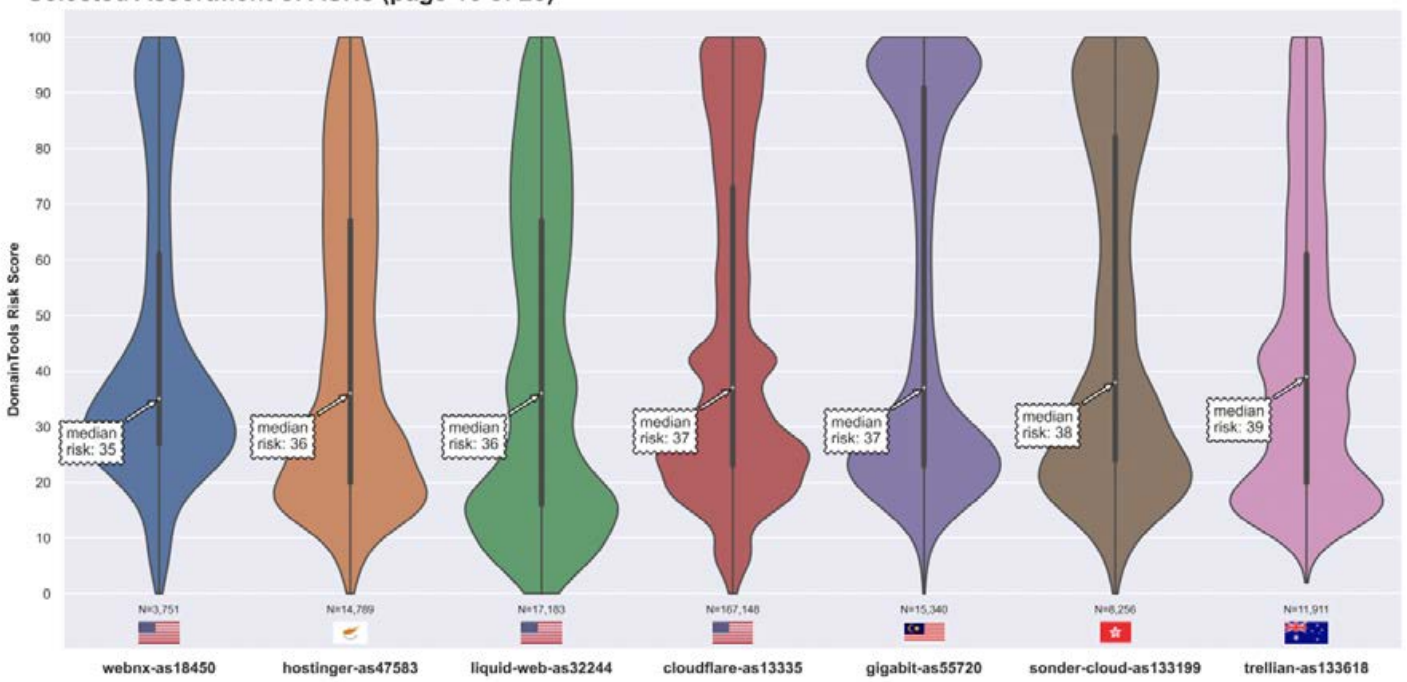
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 15 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



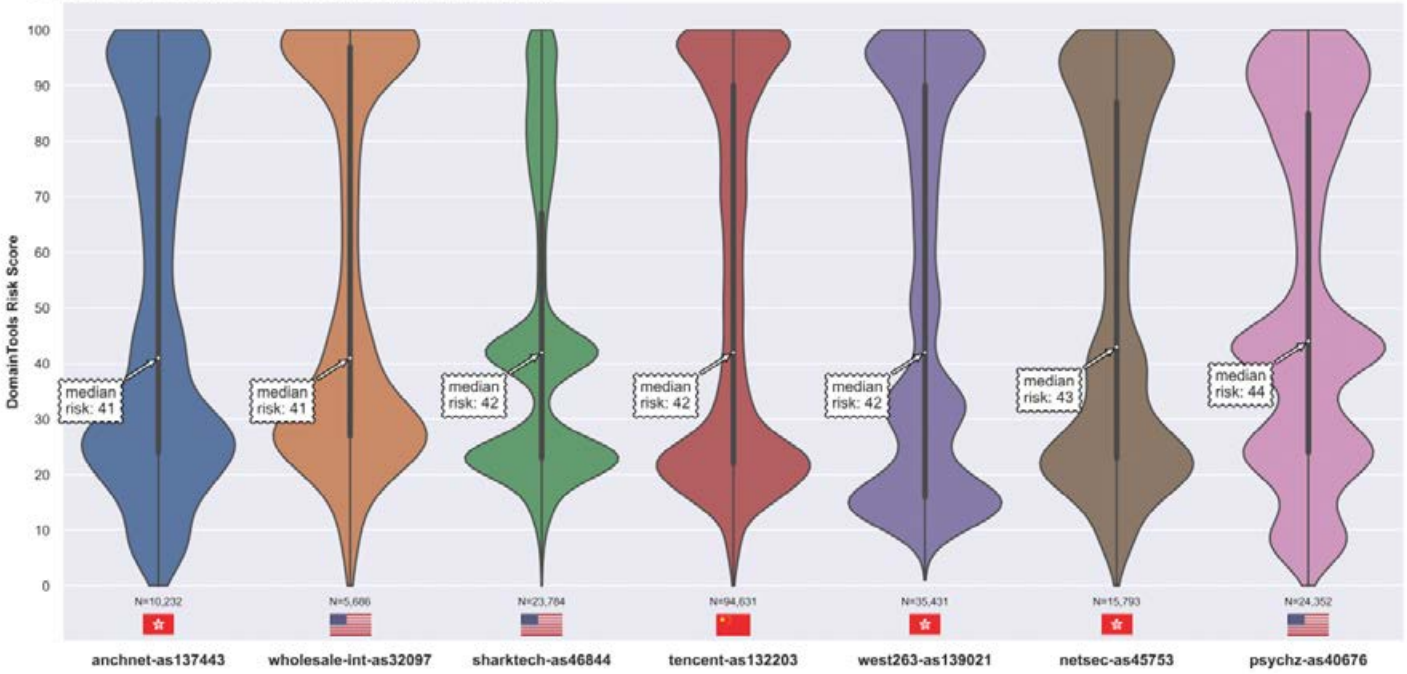
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 16 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



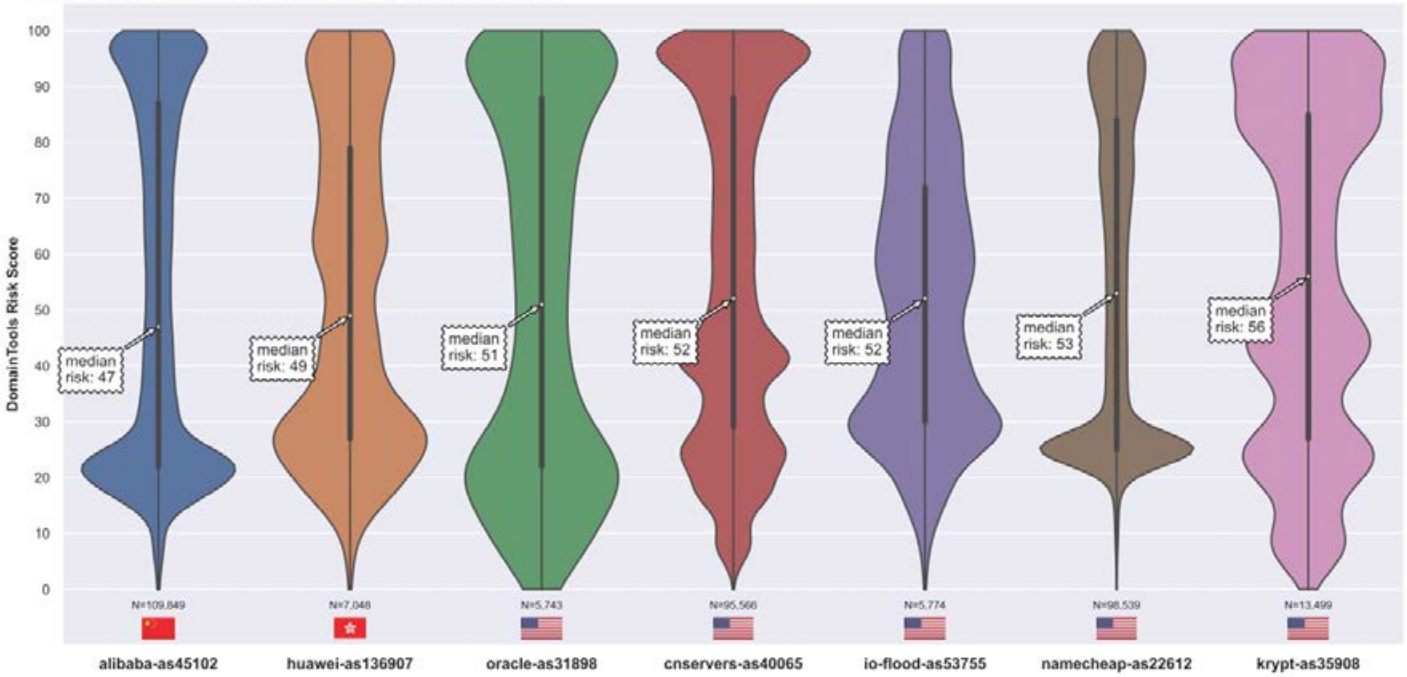
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 17 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



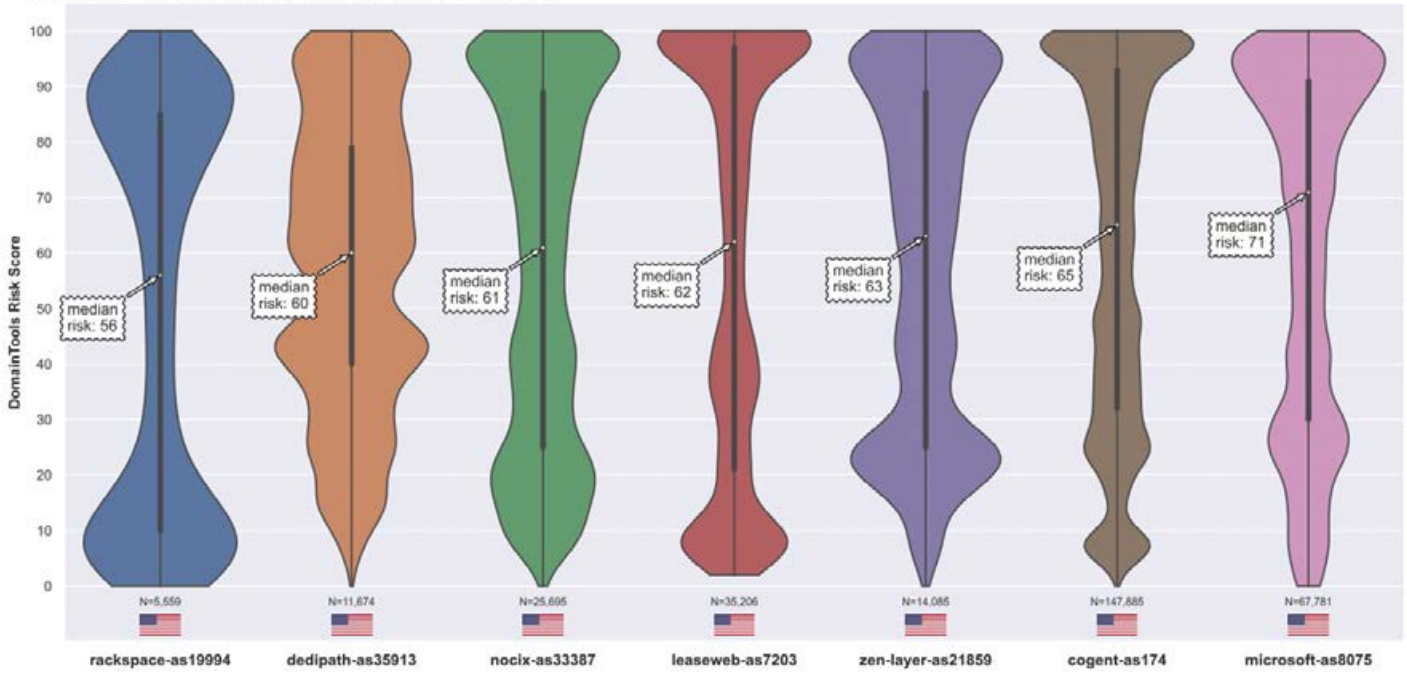
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 18 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



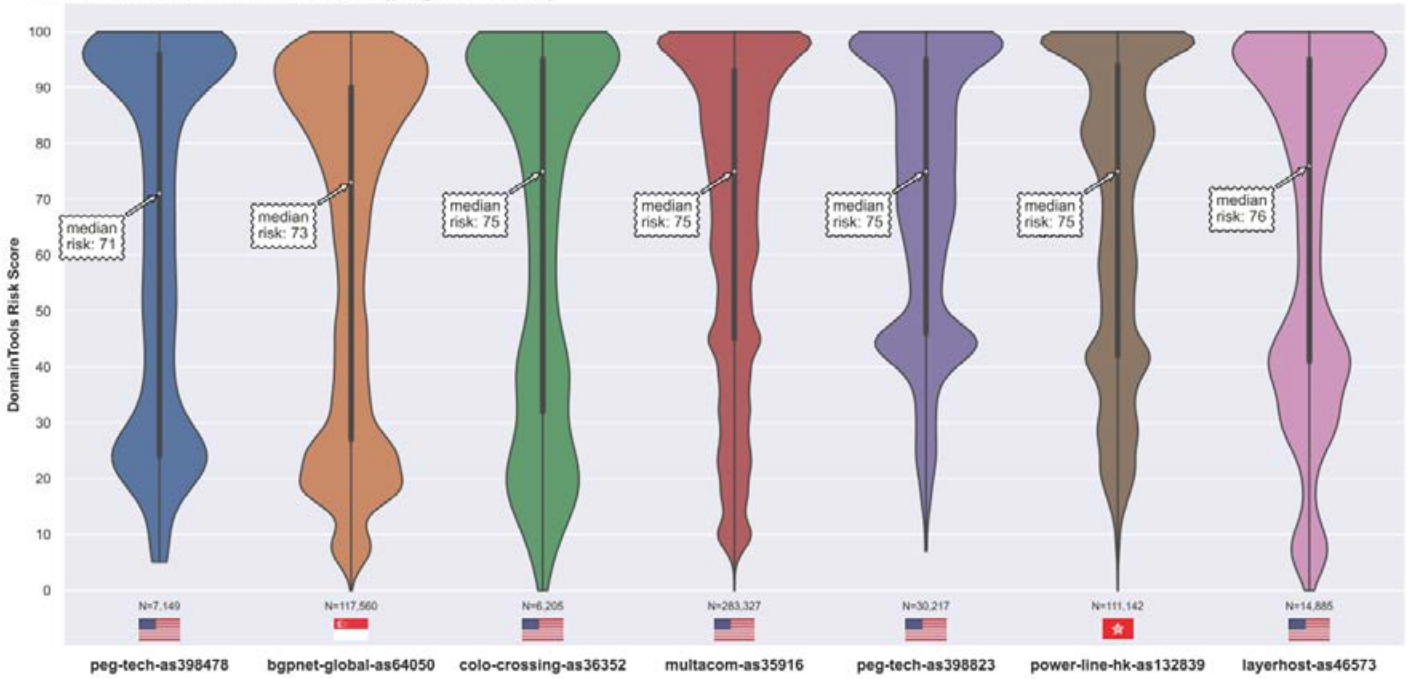
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 19 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



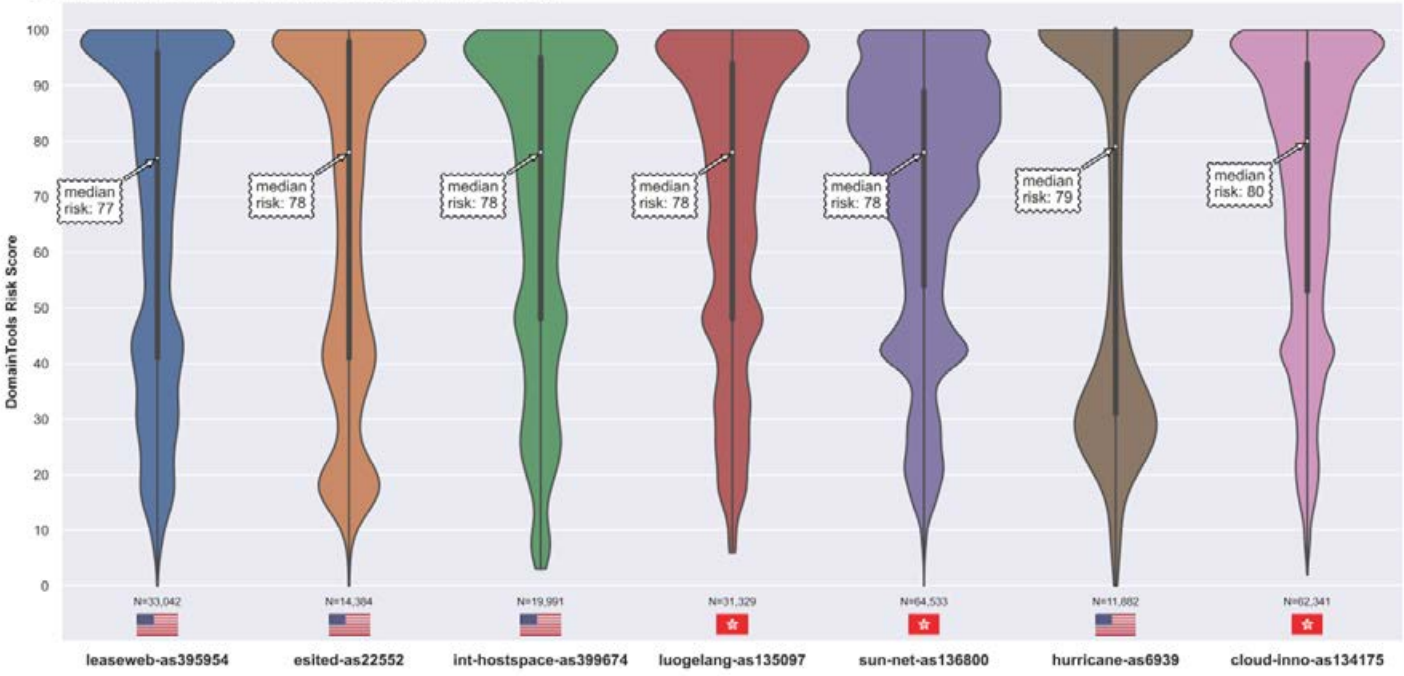
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 20 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



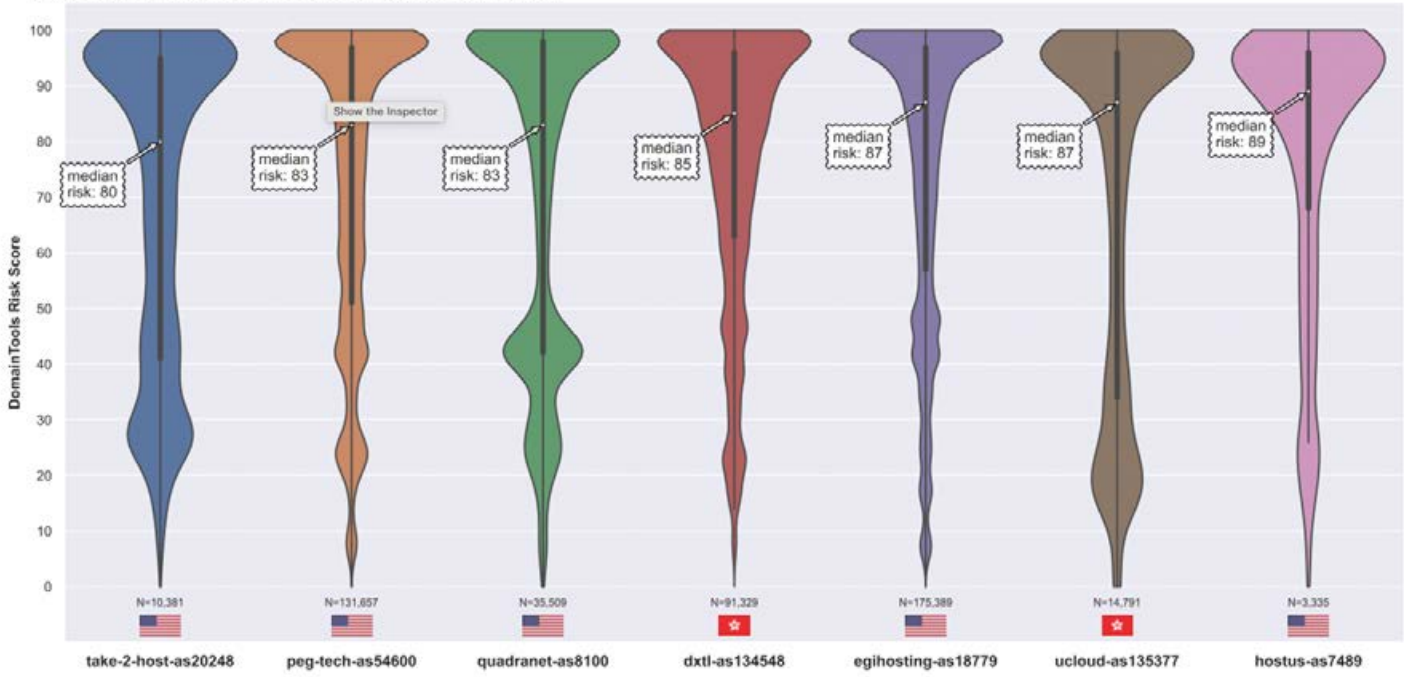
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 21 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



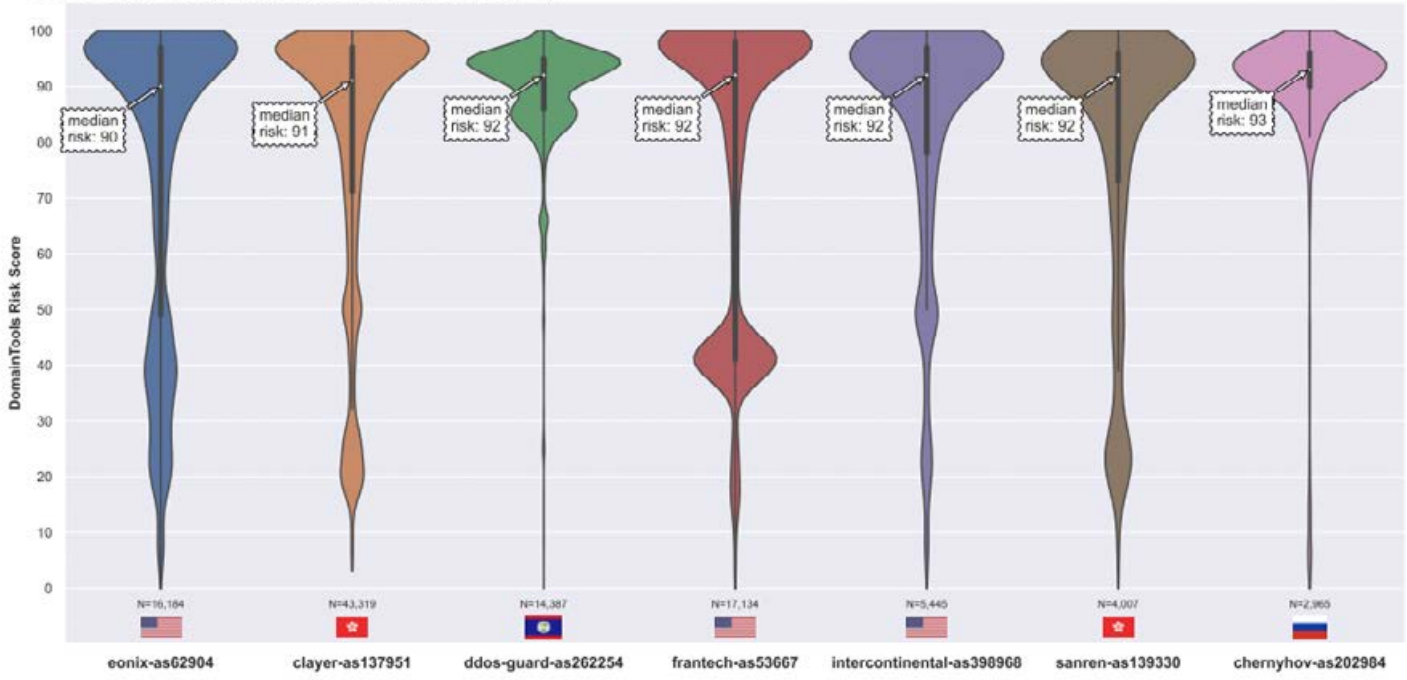
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 22 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



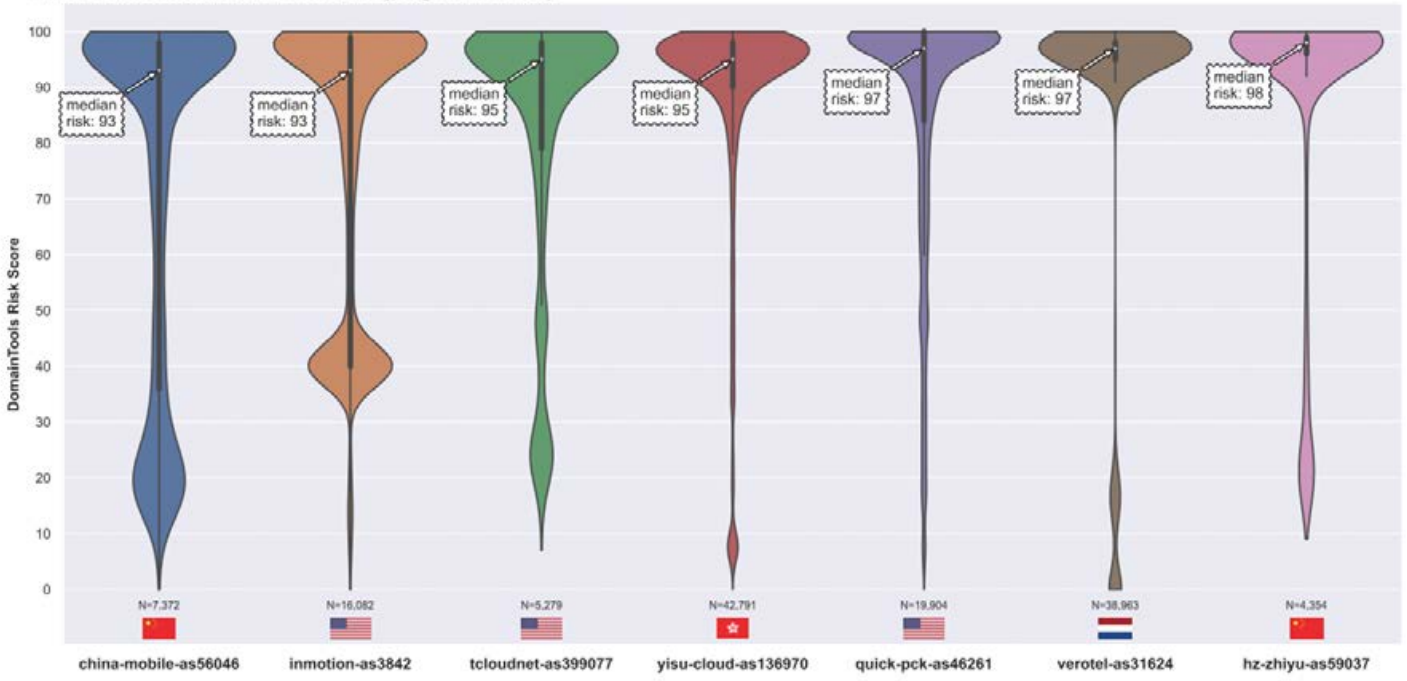
Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 23 of 25)

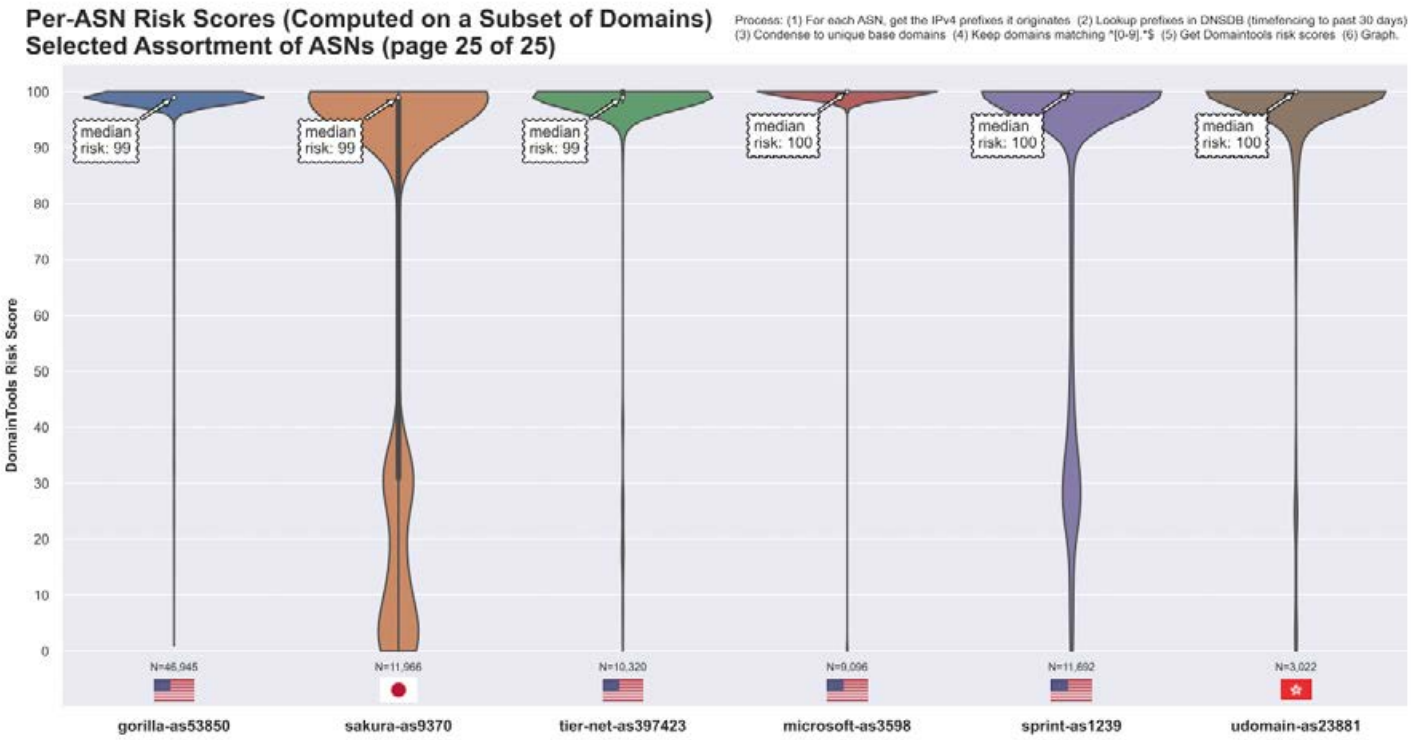
Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.



Per-ASN Risk Scores (Computed on a Subset of Domains)
Selected Assortment of ASNs (page 24 of 25)

Process: (1) For each ASN, get the IPv4 prefixes it originates (2) Lookup prefixes in DNSDB (timefencing to past 30 days) (3) Condense to unique base domains (4) Keep domains matching ^[0-9].*$ (5) Get Domaintools risk scores (6) Graph.





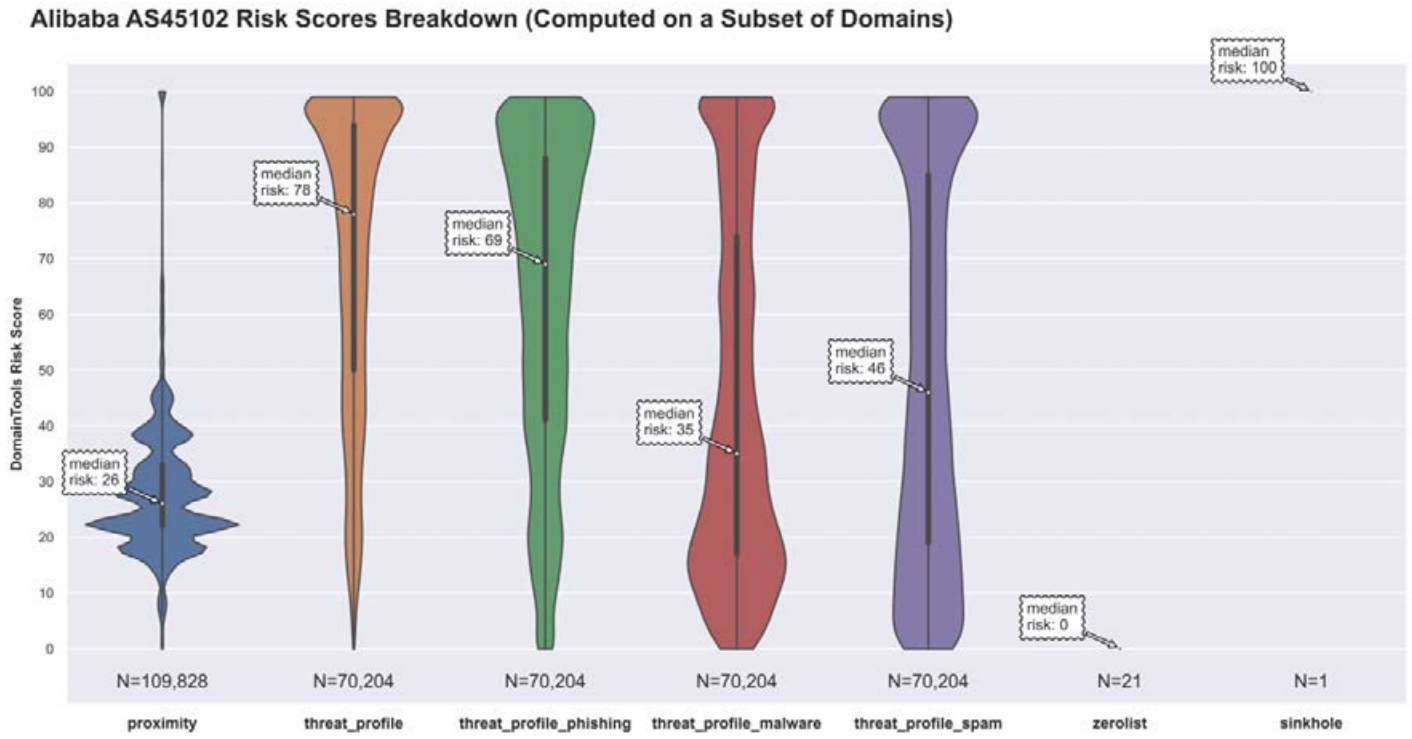
In some cases, as you look at the risk score distribution for a given ASN, you may find yourself wondering, "How the heck did that distribution come about? What subscores drove that?"

Part E. Per-ASN Subscore-Decomposed Risks

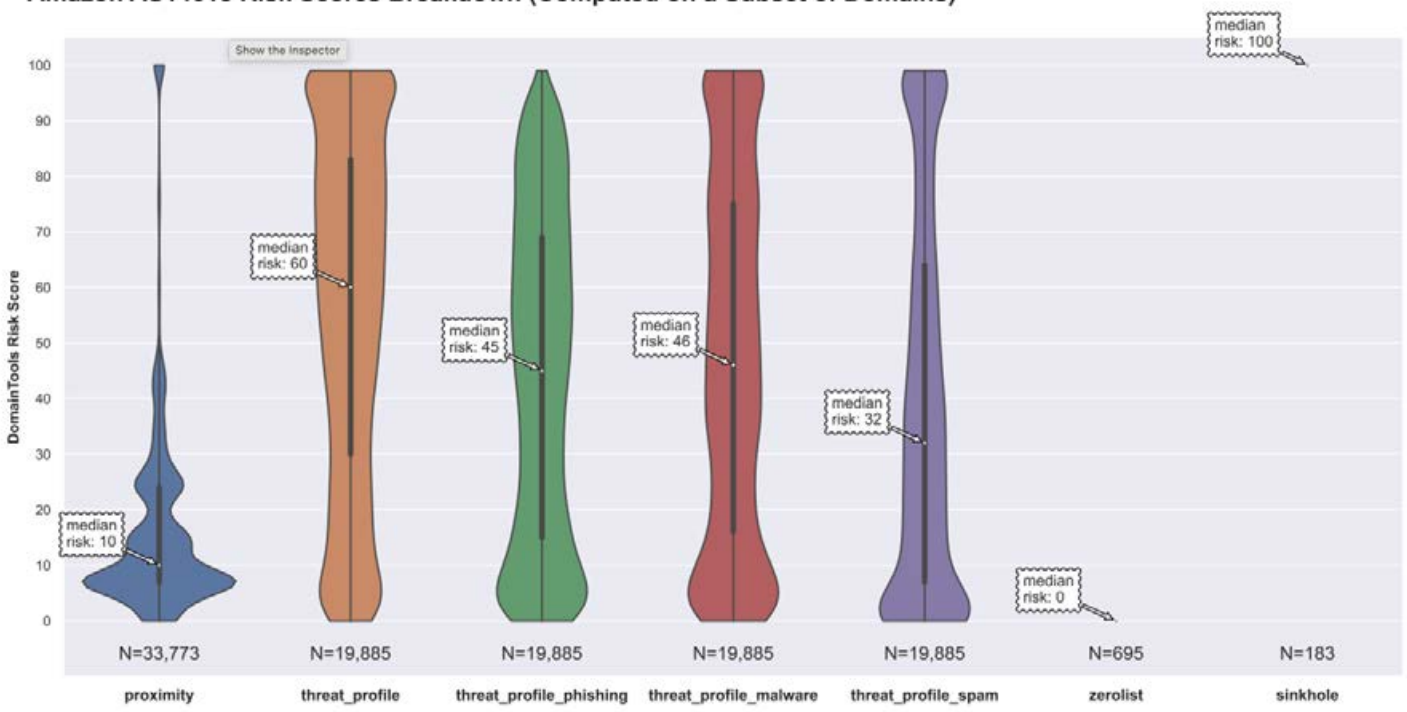
Per-ASN Subscore-Decomposed Risks

To help resolve the lingering questions from the previous section, we're providing per-ASN plots of the risk subscores in this section, including details about proximity, threat subscores, zero-listings and sinkhole listings. The ASN plots are listed alphabetically by abbreviated name.

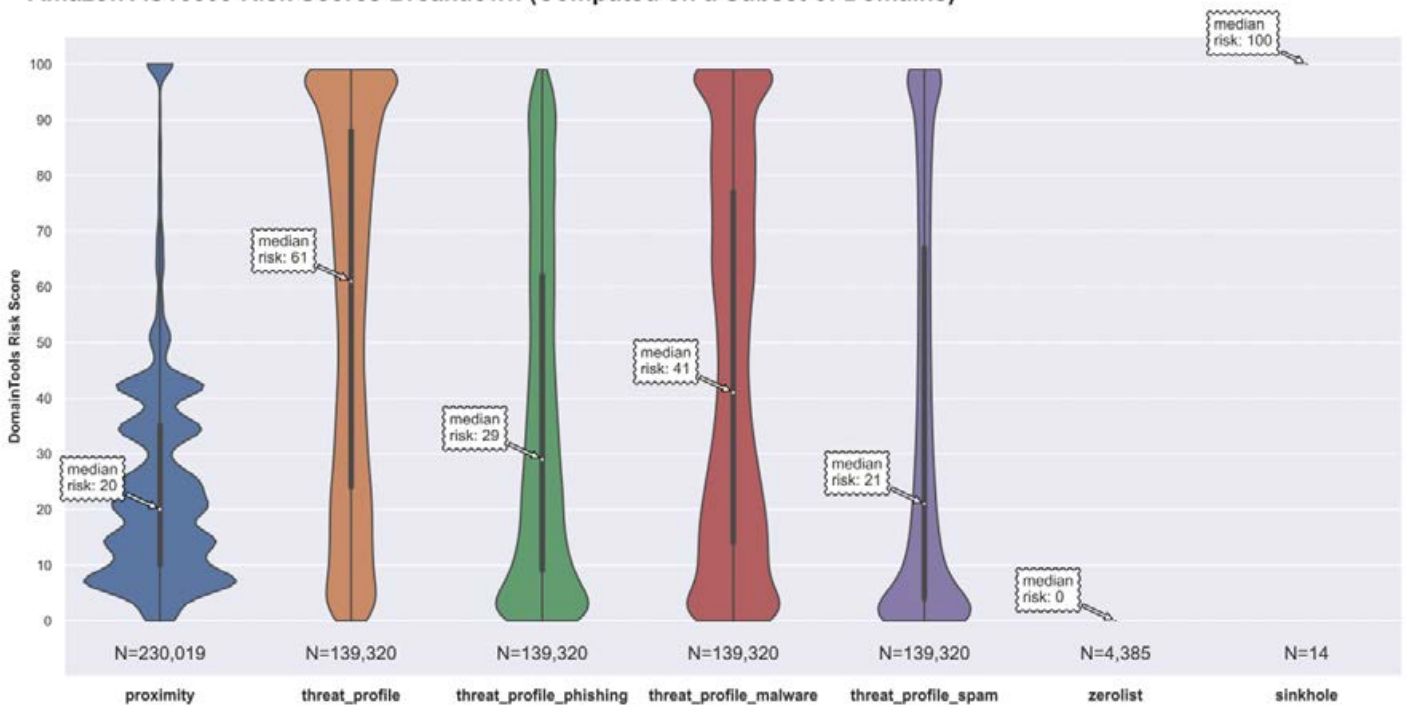
Zerolist and sinkhole listings, as a single score of 0 or 100 with zero dispersion, are shown as points rather than violins, if zerolistings or sinkholed domains are present.



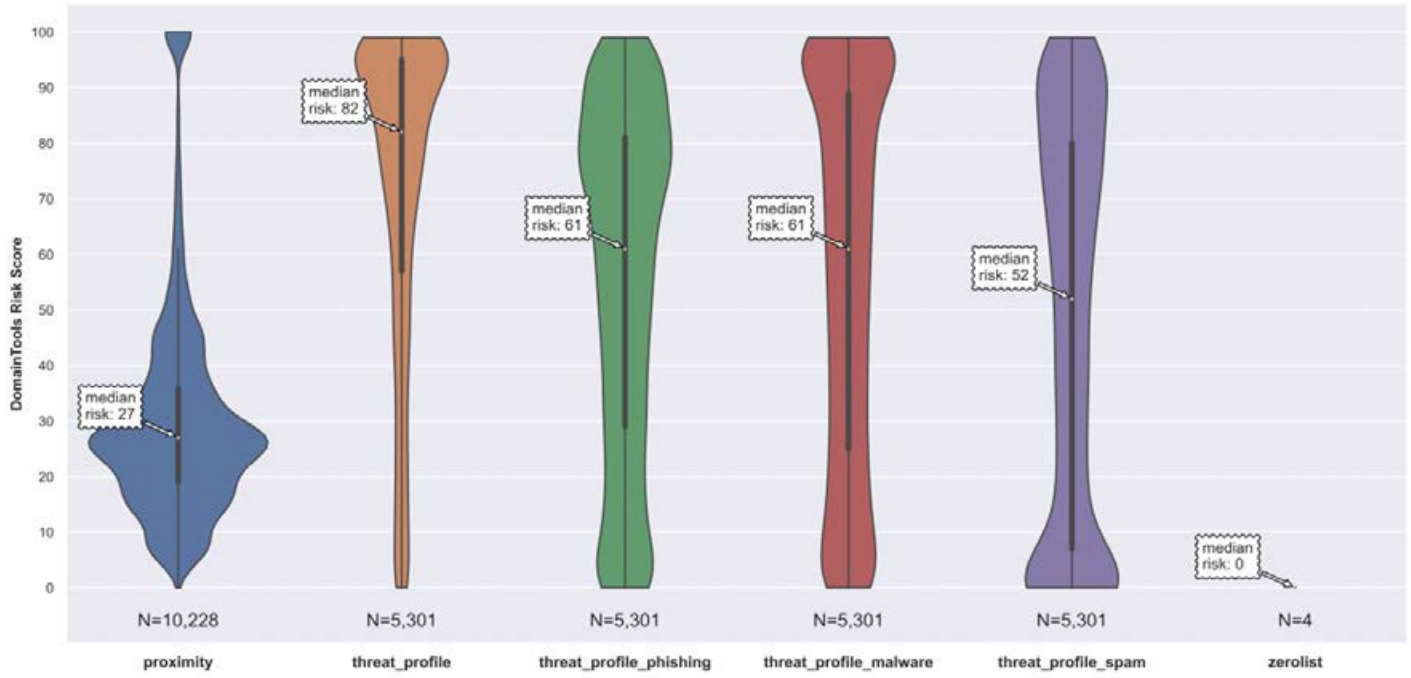
Amazon AS14618 Risk Scores Breakdown (Computed on a Subset of Domains)



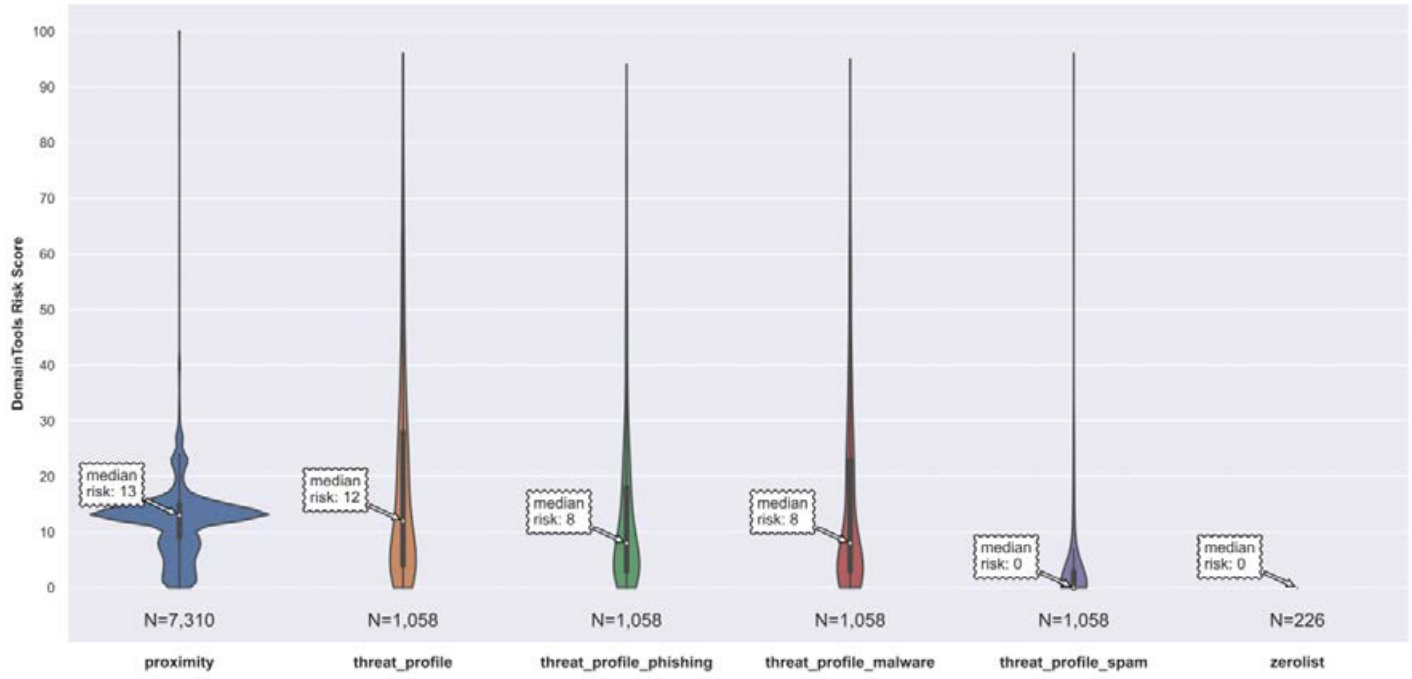
Amazon AS16509 Risk Scores Breakdown (Computed on a Subset of Domains)



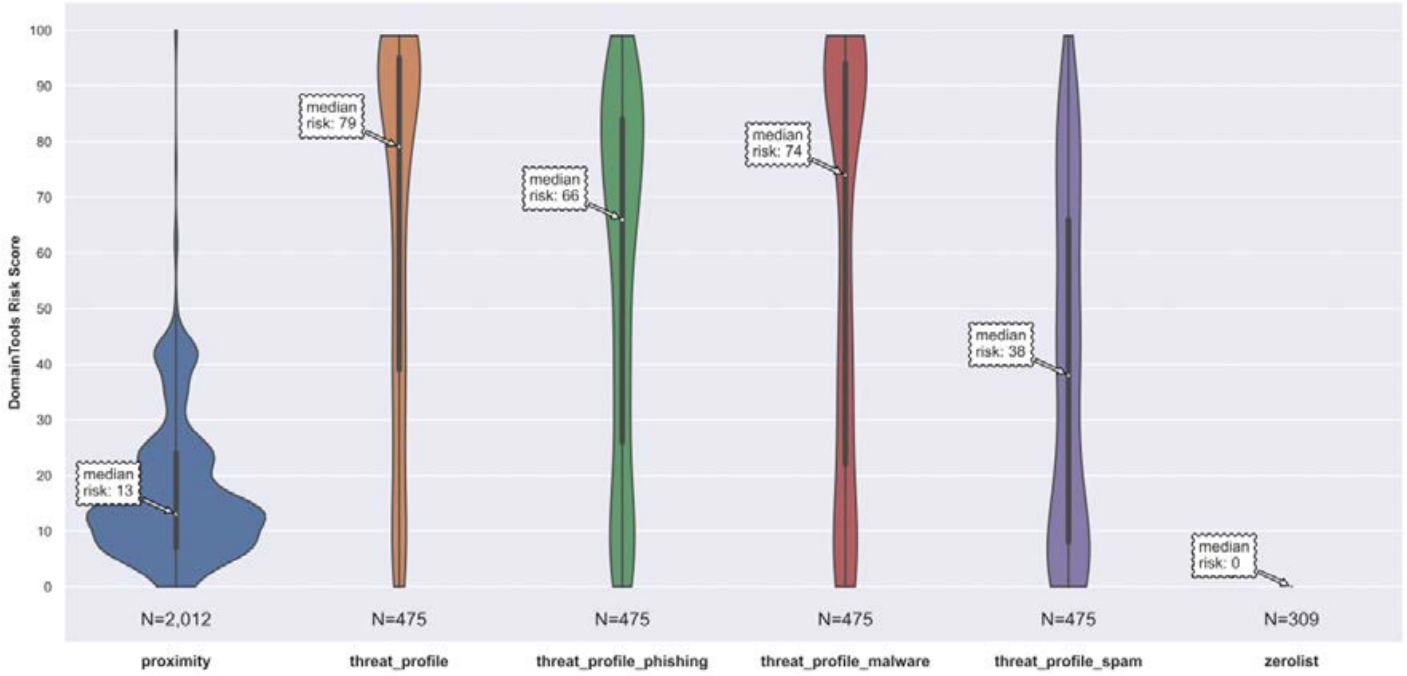
Anchnet AS137443 Risk Scores Breakdown (Computed on a Subset of Domains)



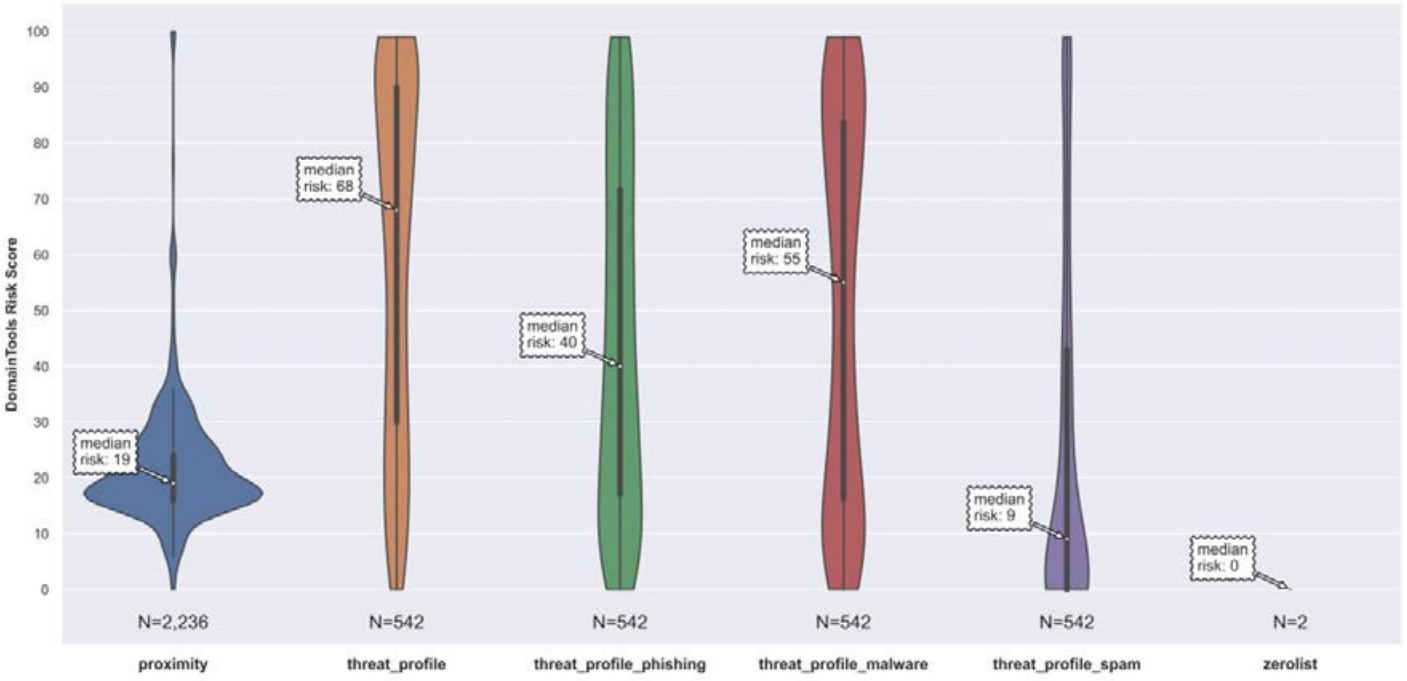
Aptum AS13768 Risk Scores Breakdown (Computed on a Subset of Domains)



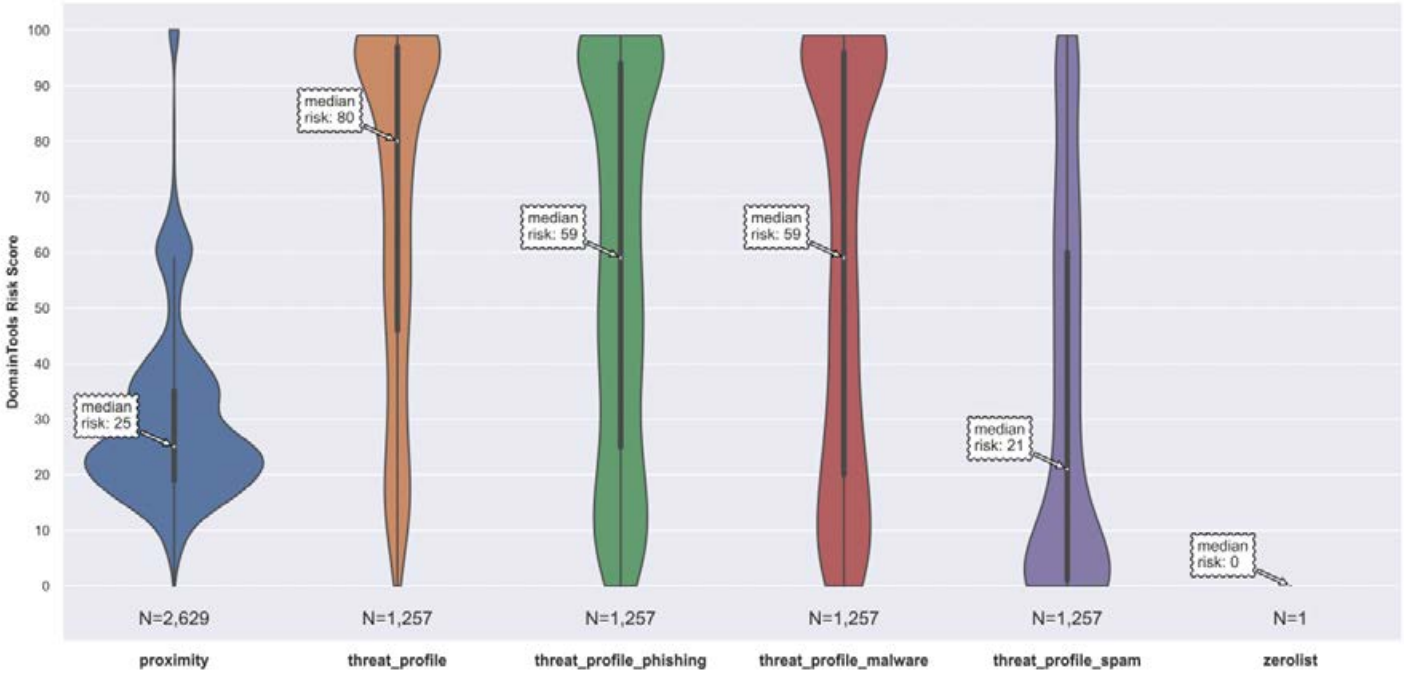
ATT-AS7018 Risk Scores Breakdown (Computed on a Subset of Domains)



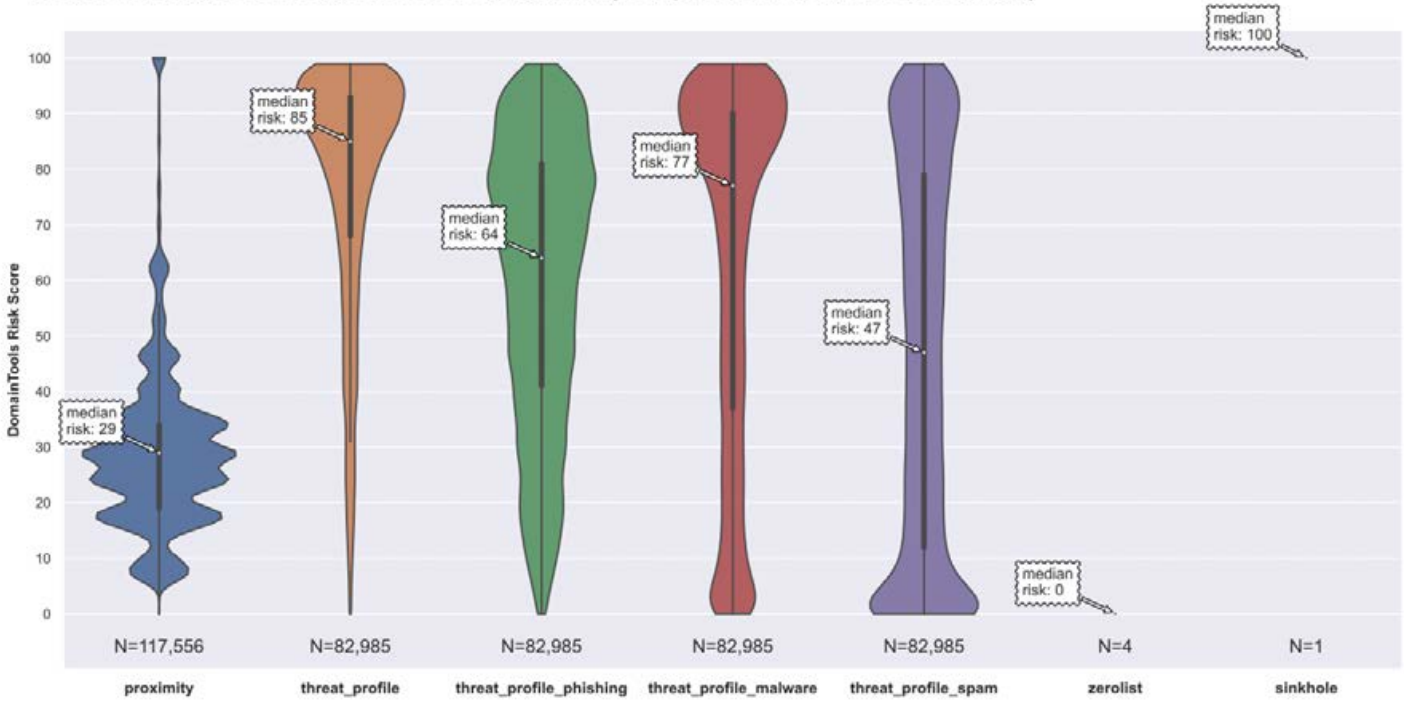
Baidu AS38365 Risk Scores Breakdown (Computed on a Subset of Domains)



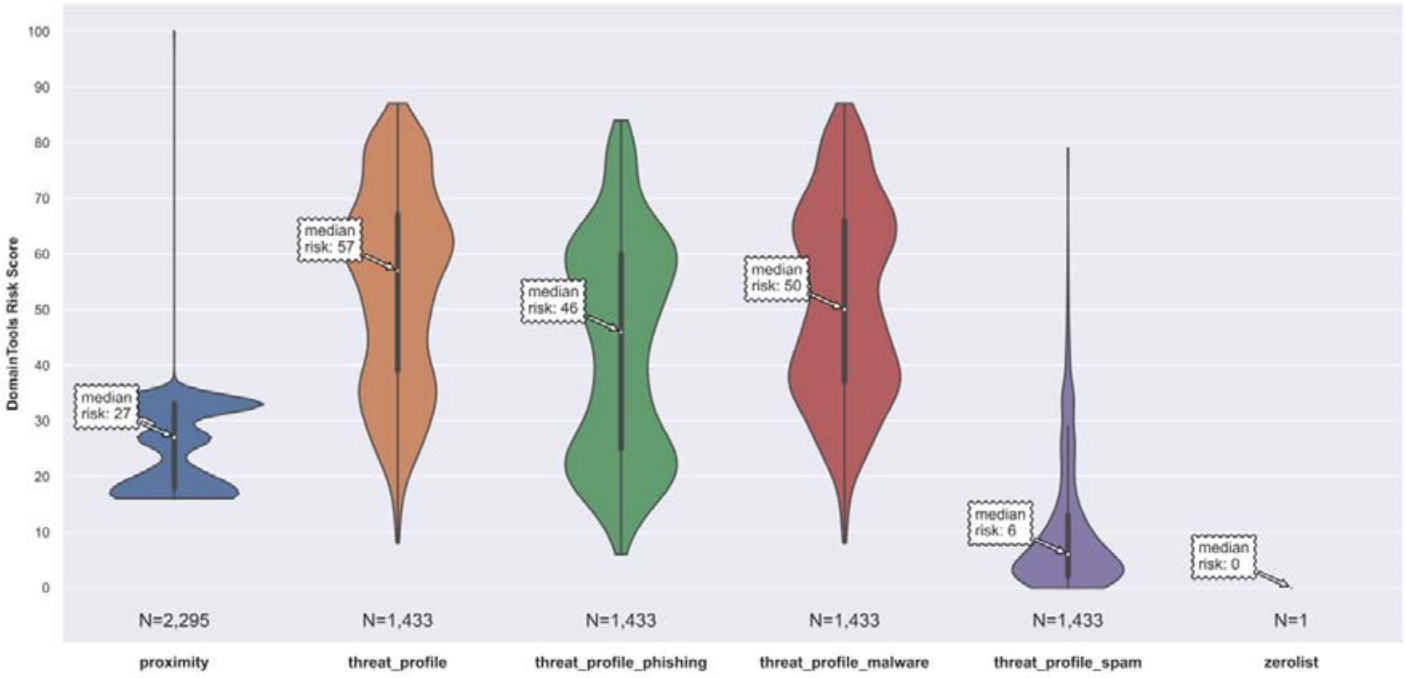
Baidu AS55967 Risk Scores Breakdown (Computed on a Subset of Domains)



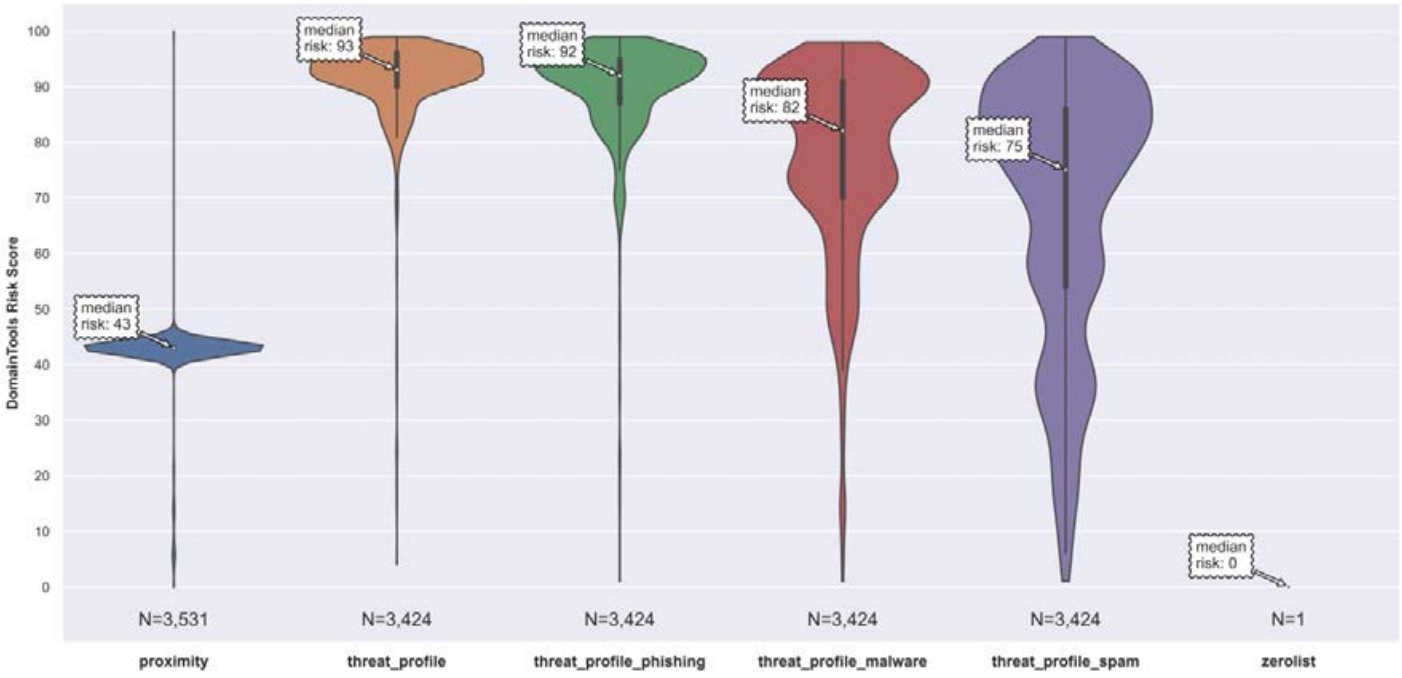
BGPNet Global AS64050 Risk Scores Breakdown (Computed on a Subset of Domains)



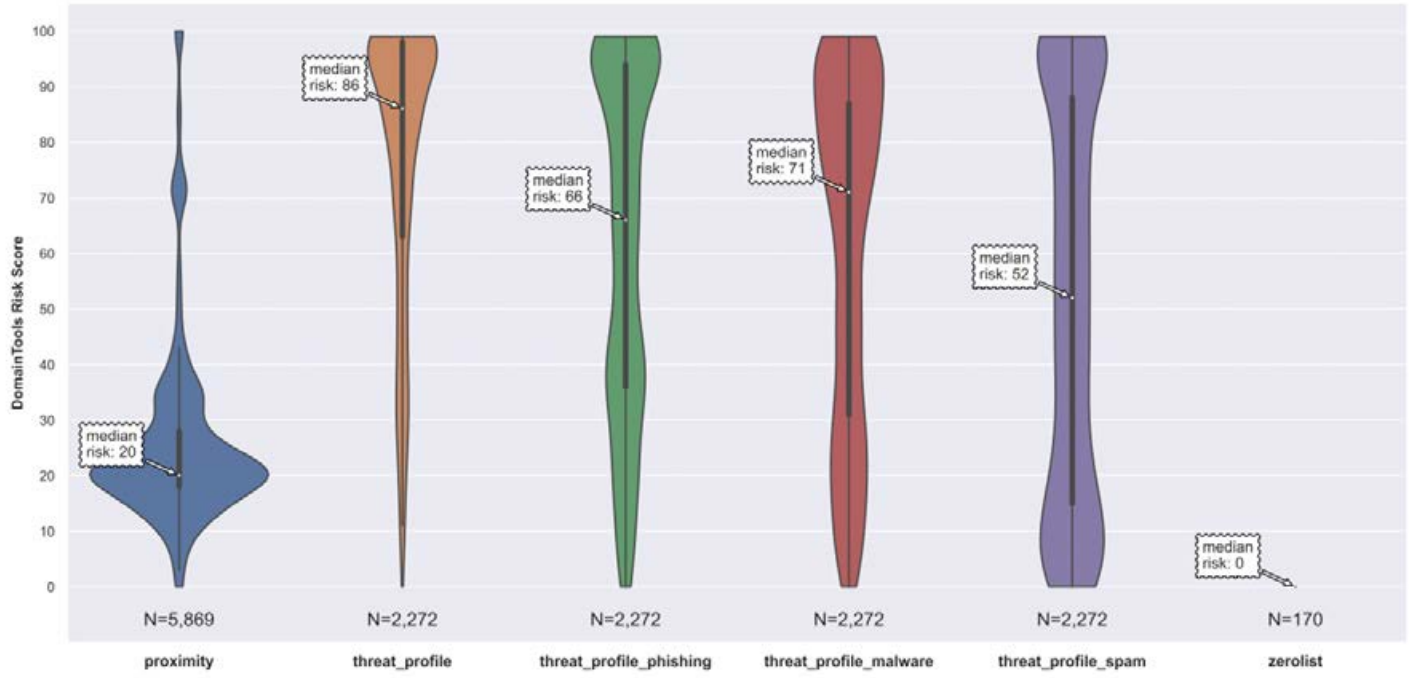
Bharti AS45609 Risk Scores Breakdown (Computed on a Subset of Domains)



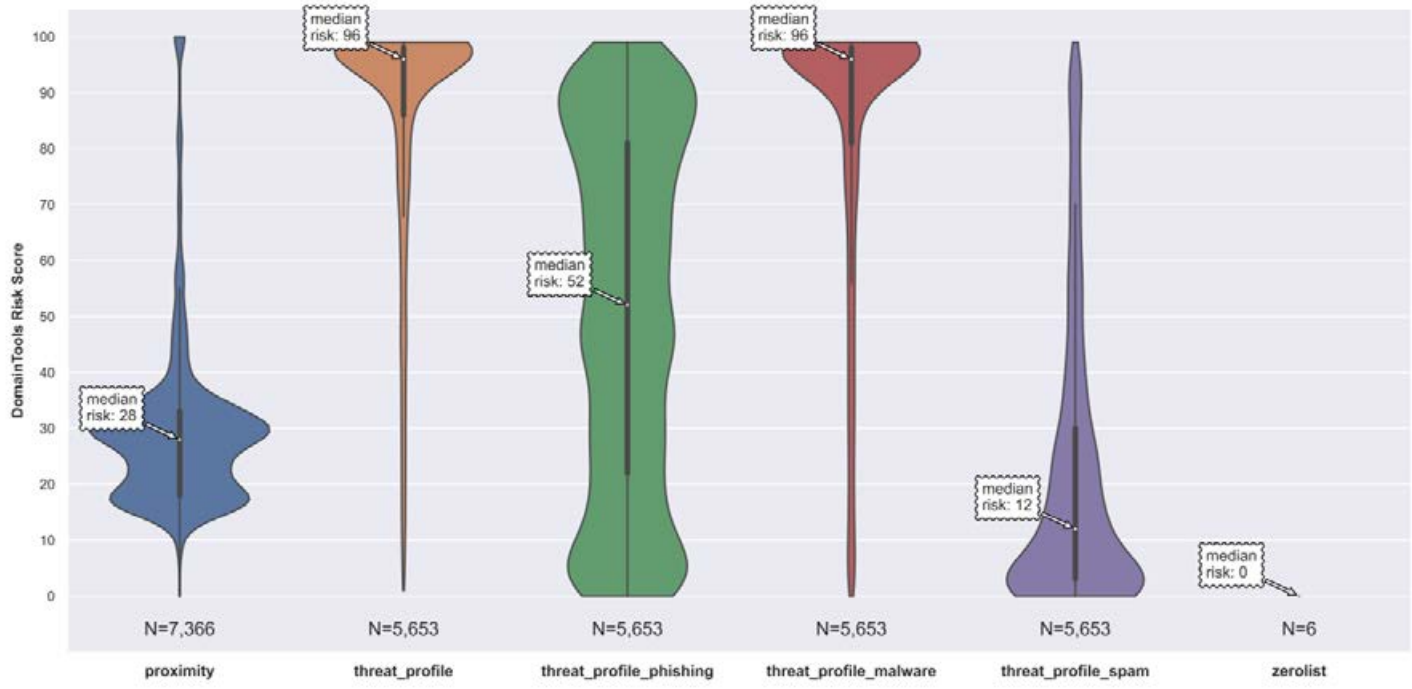
Chernyhov AS202984 Risk Scores Breakdown (Computed on a Subset of Domains)



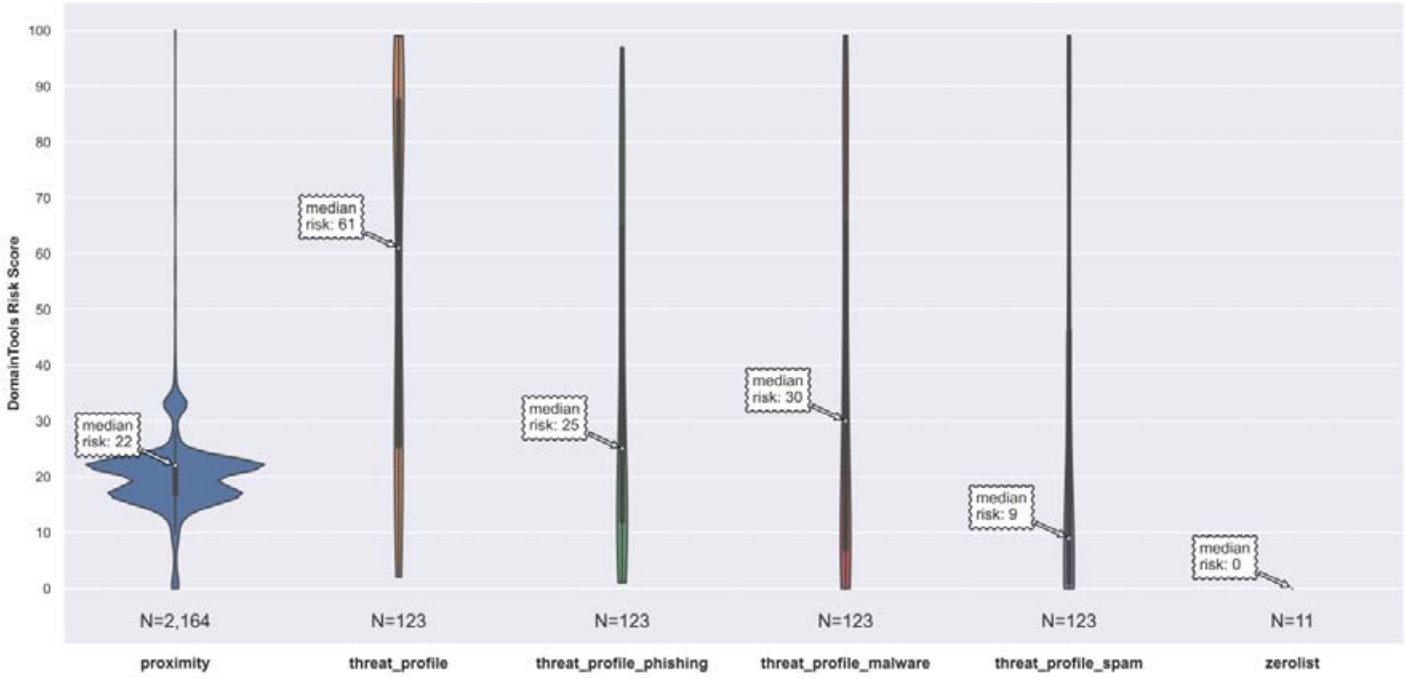
China Mobile AS9808 Risk Scores Breakdown (Computed on a Subset of Domains)



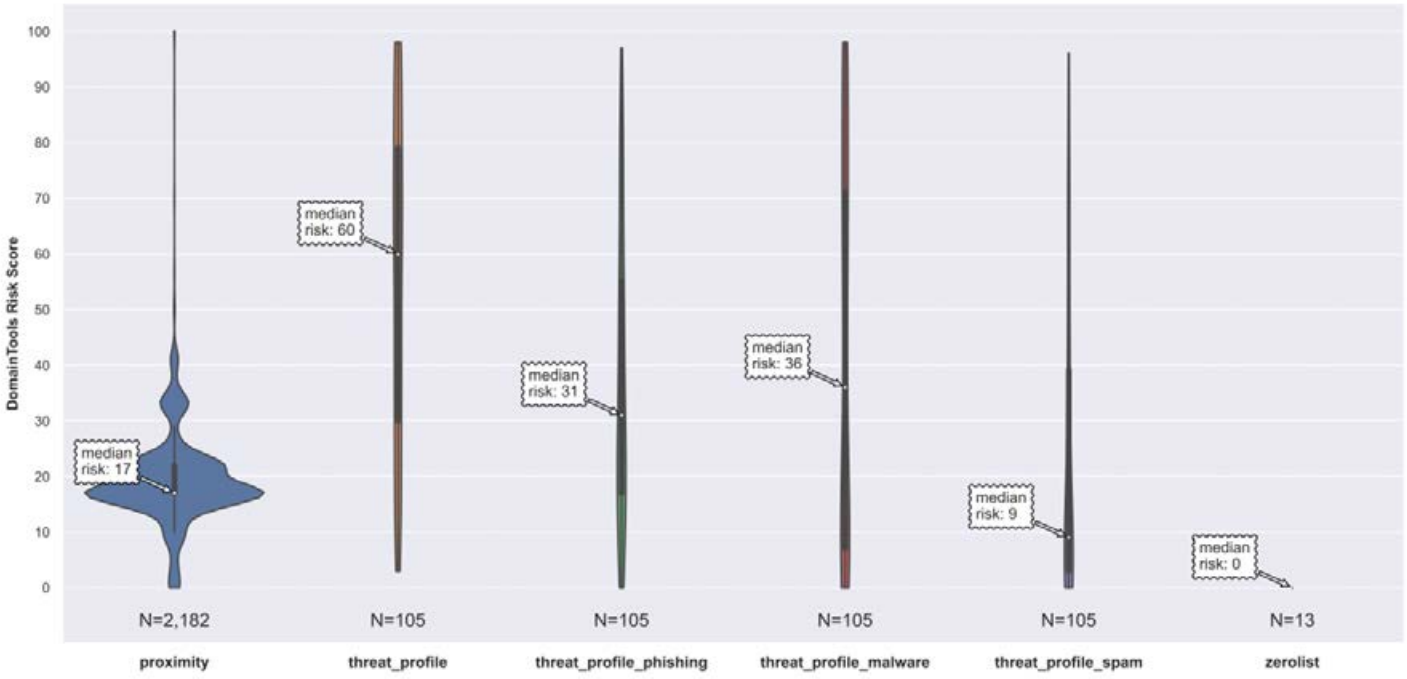
China Mobile AS56046 Risk Scores Breakdown (Computed on a Subset of Domains)



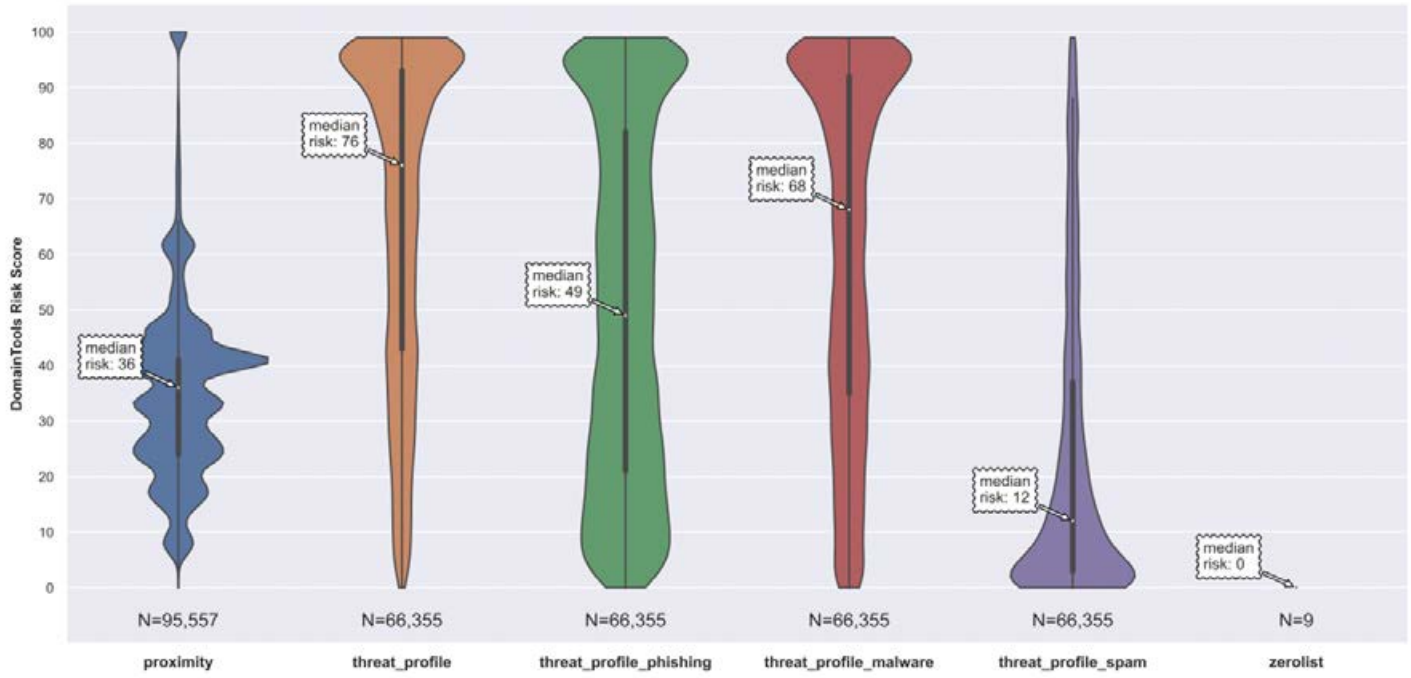
China Mobile AS56048 Risk Scores Breakdown (Computed on a Subset of Domains)



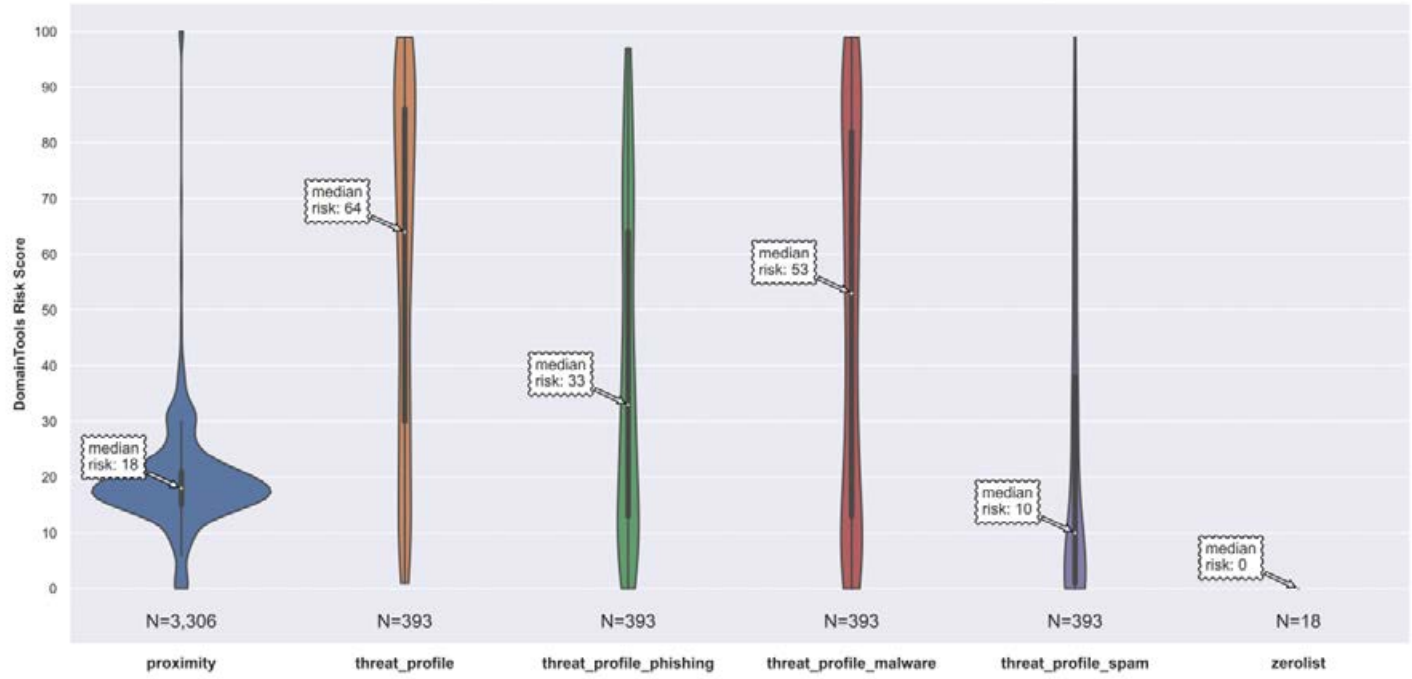
China Networks IX AS4847 Risk Scores Breakdown (Computed on a Subset of Domains)



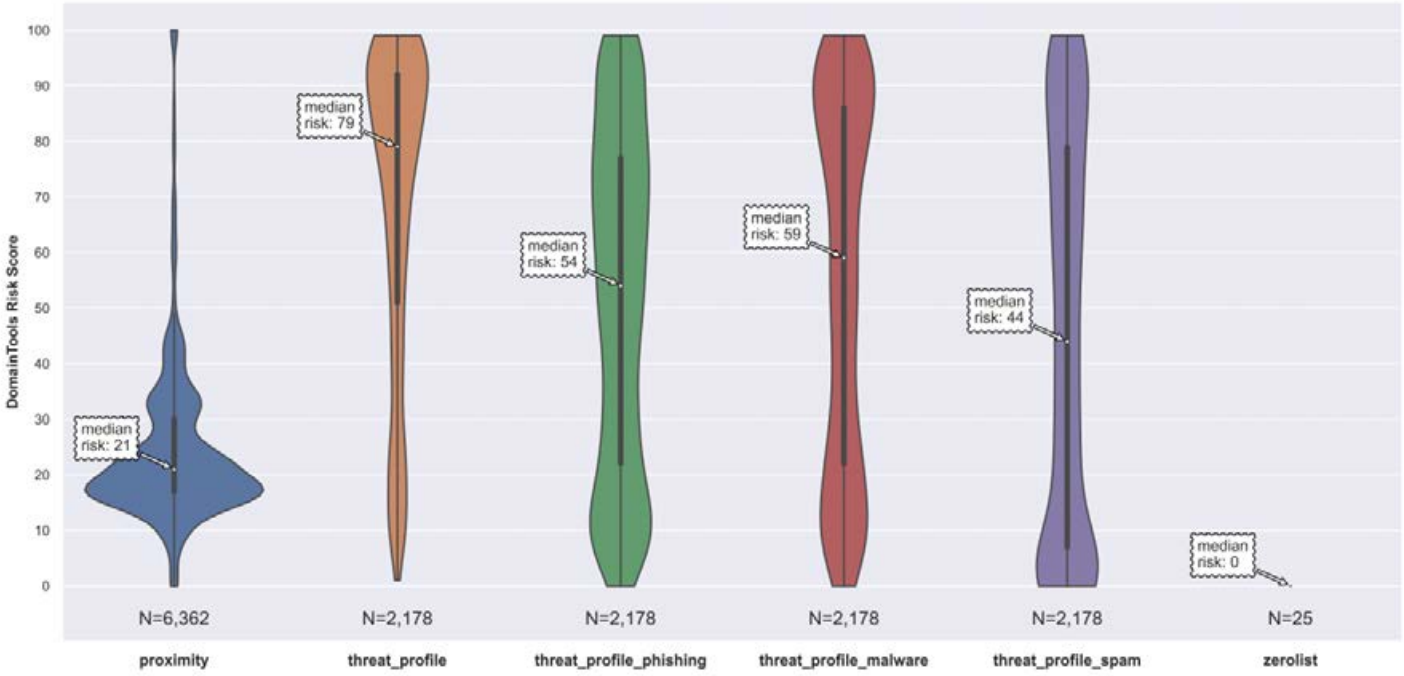
China Servers AS40065 Risk Scores Breakdown (Computed on a Subset of Domains)



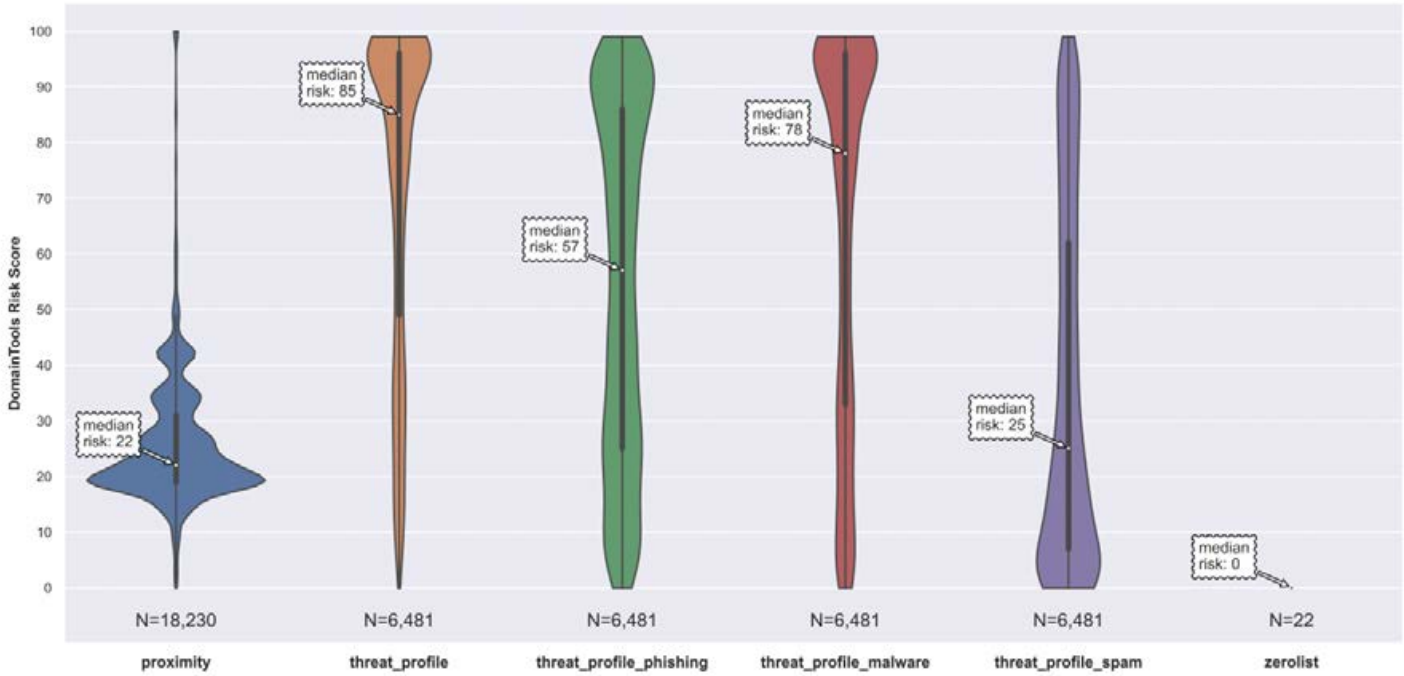
China Telecom AS4812 Risk Scores Breakdown (Computed on a Subset of Domains)



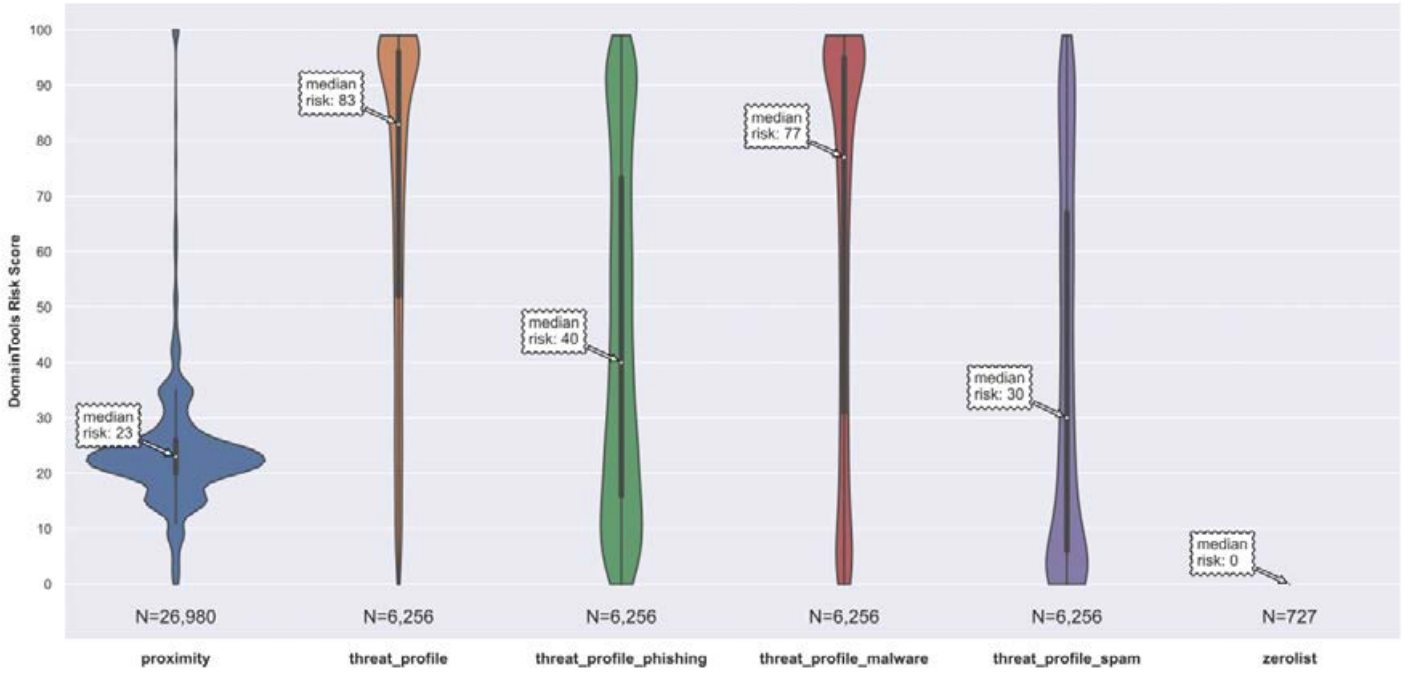
China Unicom AS4808 Risk Scores Breakdown (Computed on a Subset of Domains)



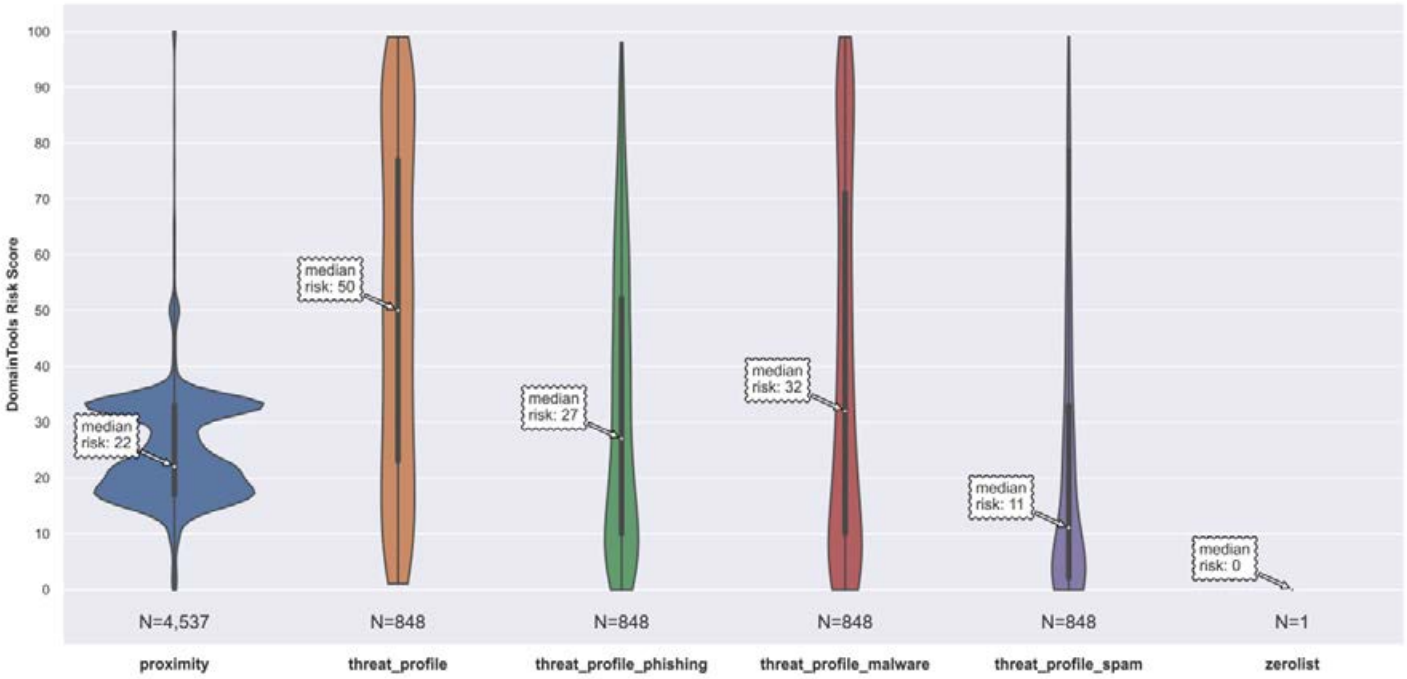
China Unicom AS4837 Risk Scores Breakdown (Computed on a Subset of Domains)



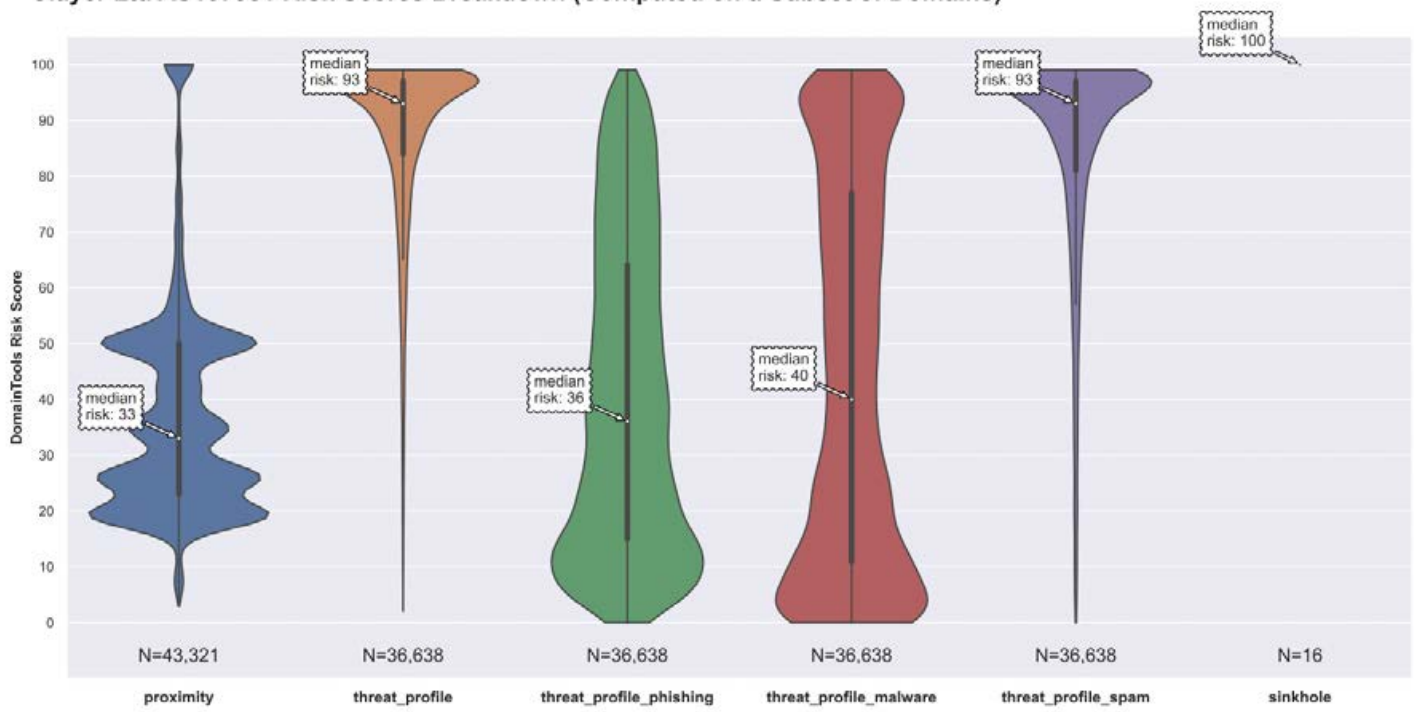
Chinanet AS4134 Risk Scores Breakdown (Computed on a Subset of Domains)



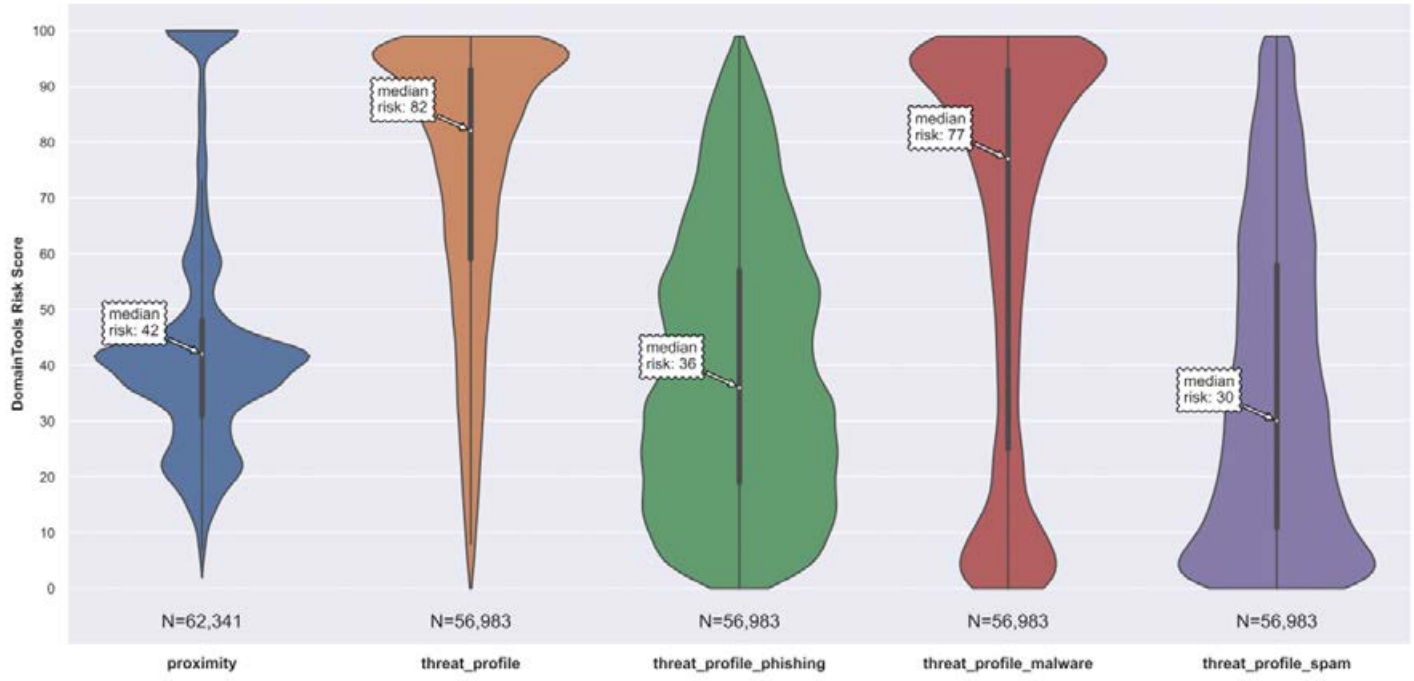
Chinanet AS38283 Risk Scores Breakdown (Computed on a Subset of Domains)



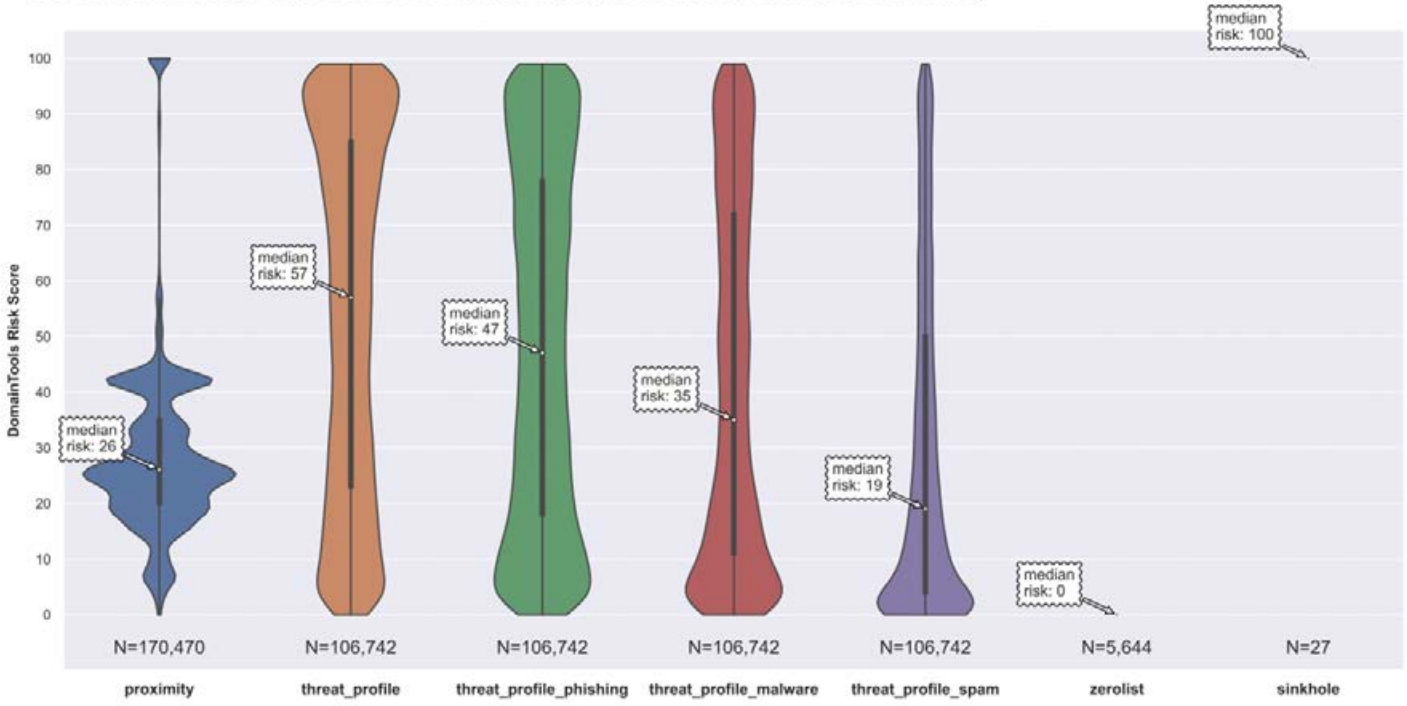
Clayer Ltd AS137951 Risk Scores Breakdown (Computed on a Subset of Domains)



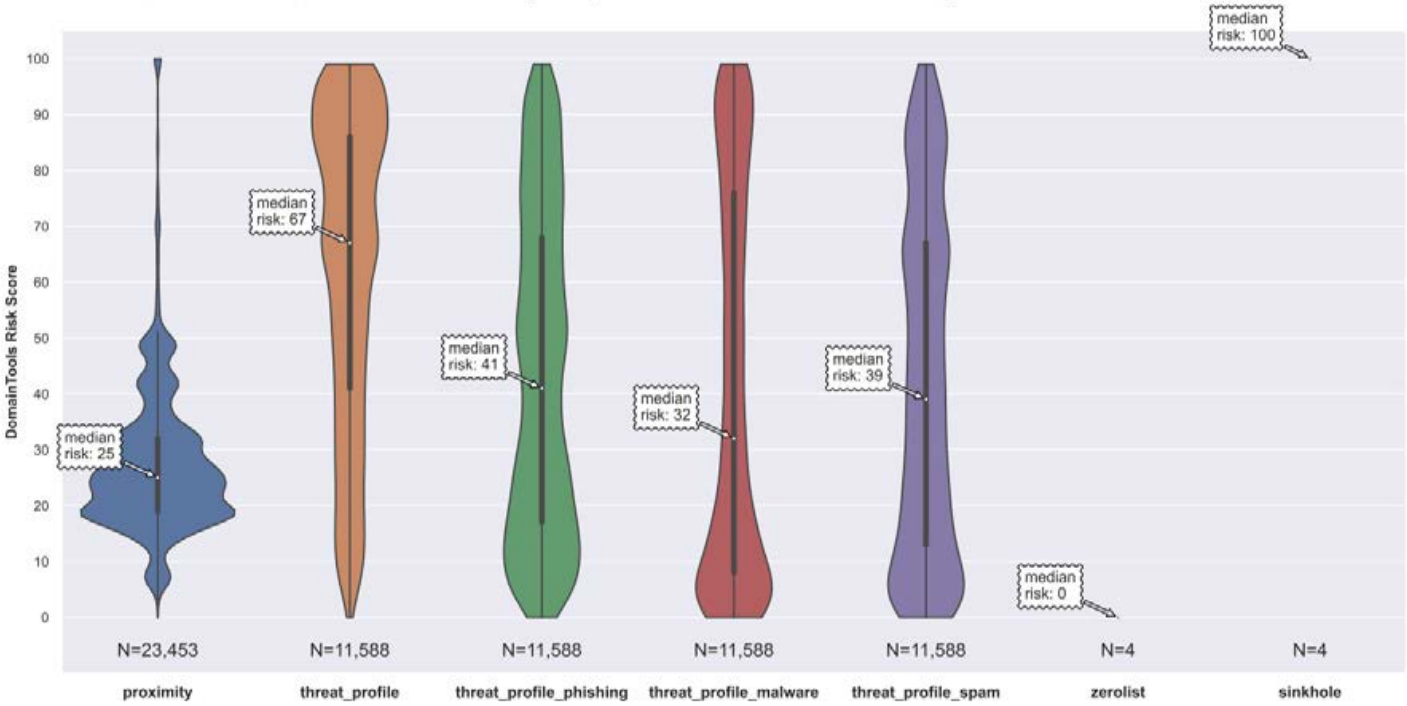
Cloud Innovation AS134175 Risk Scores Breakdown (Computed on a Subset of Domains)



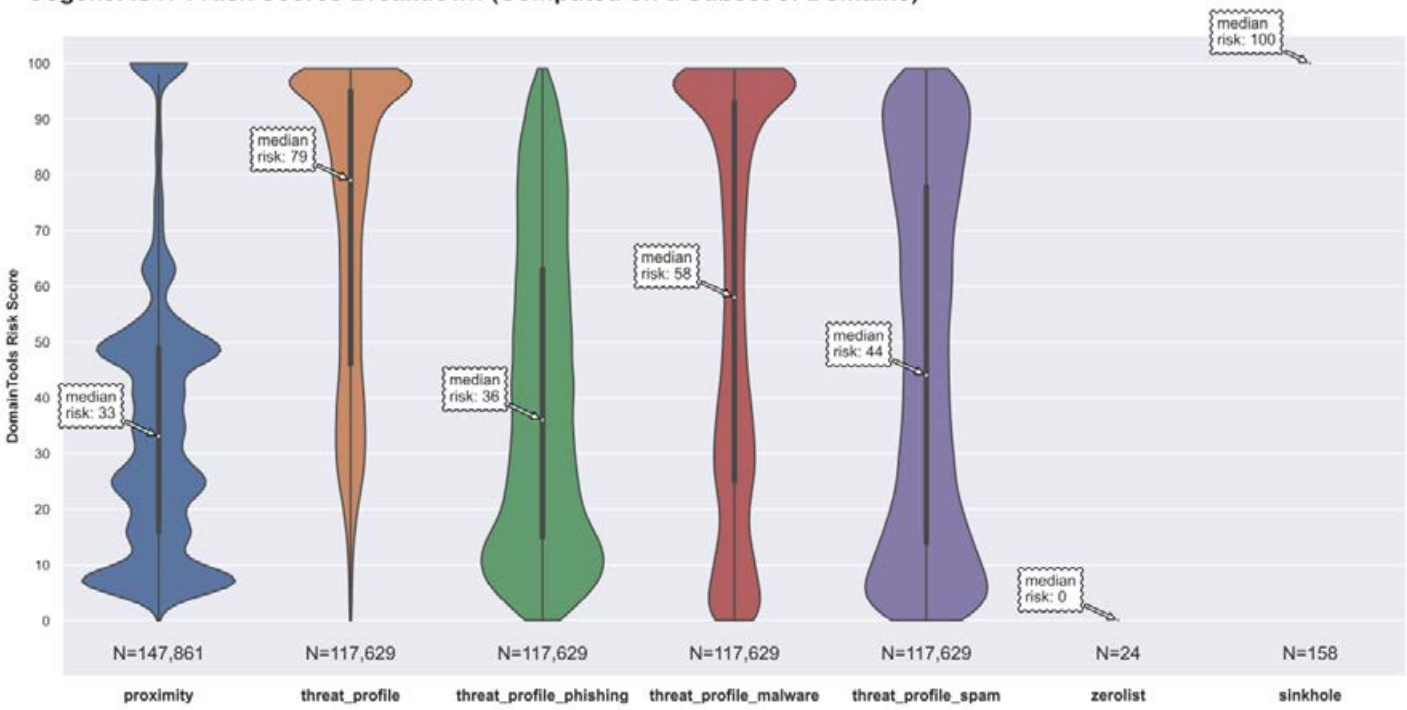
Cloudflare AS13335 Risk Scores Breakdown (Computed on a Subset of Domains)



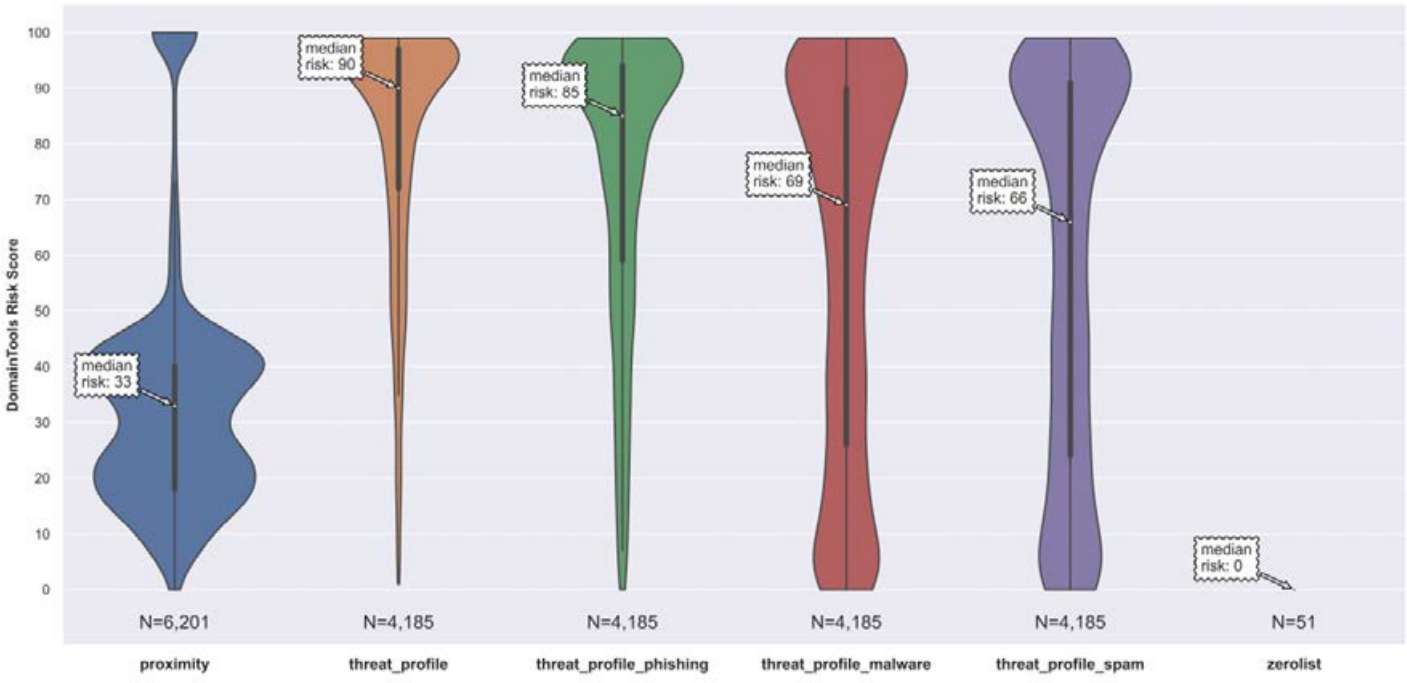
Cloudie AS55933 Risk Scores Breakdown (Computed on a Subset of Domains)



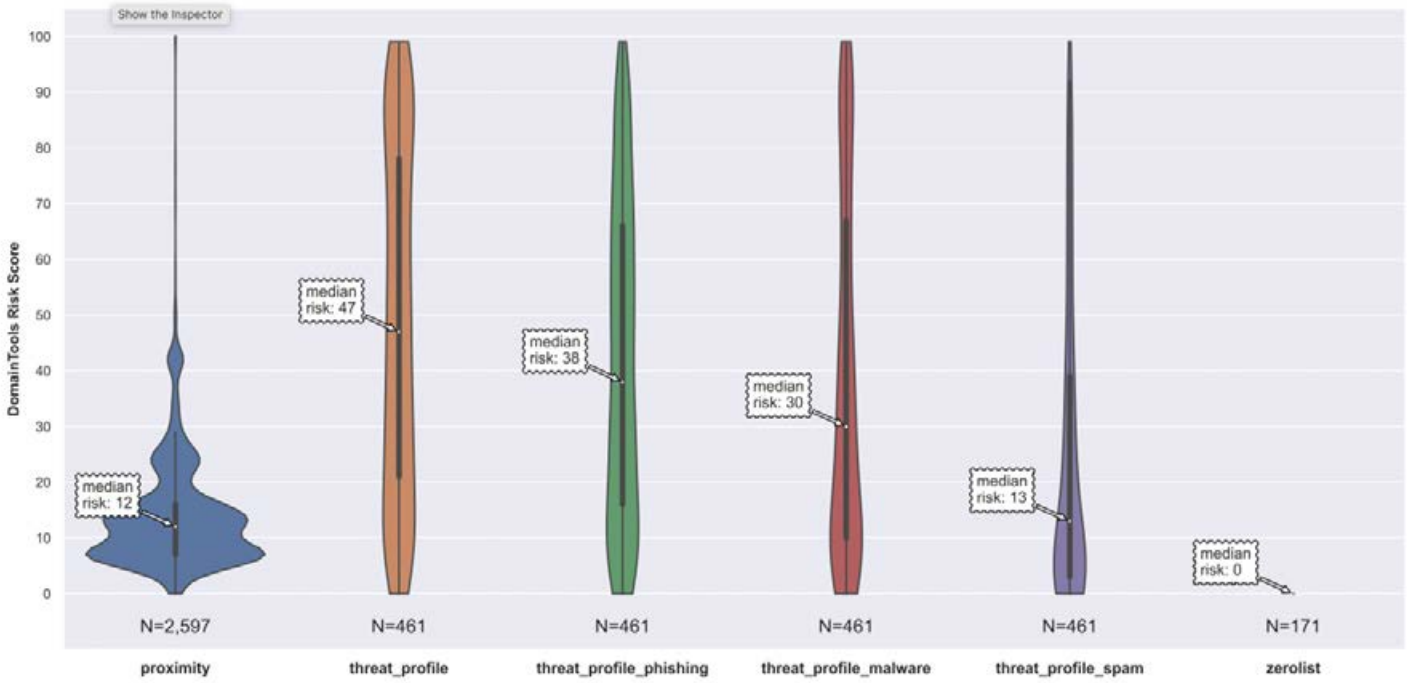
Cogent AS174 Risk Scores Breakdown (Computed on a Subset of Domains)



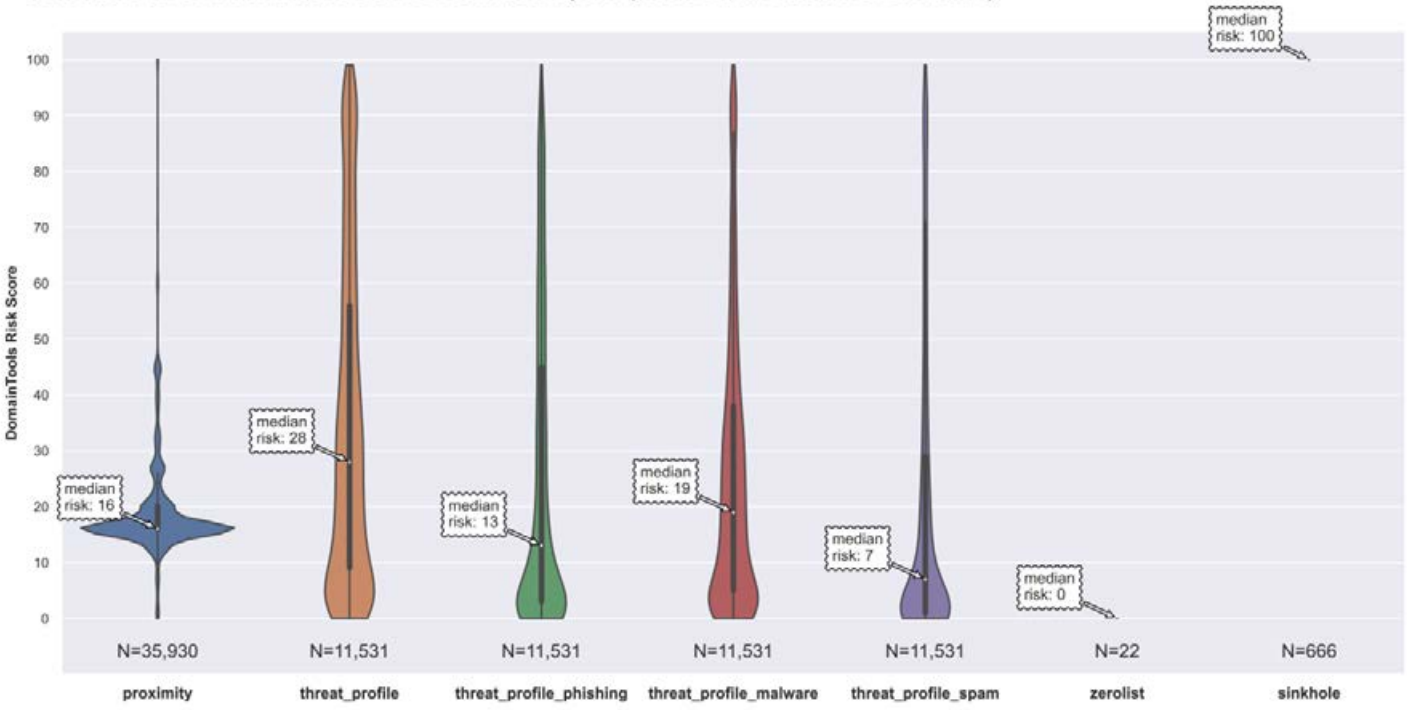
ColoCrossing AS36352 Risk Scores Breakdown (Computed on a Subset of Domains)



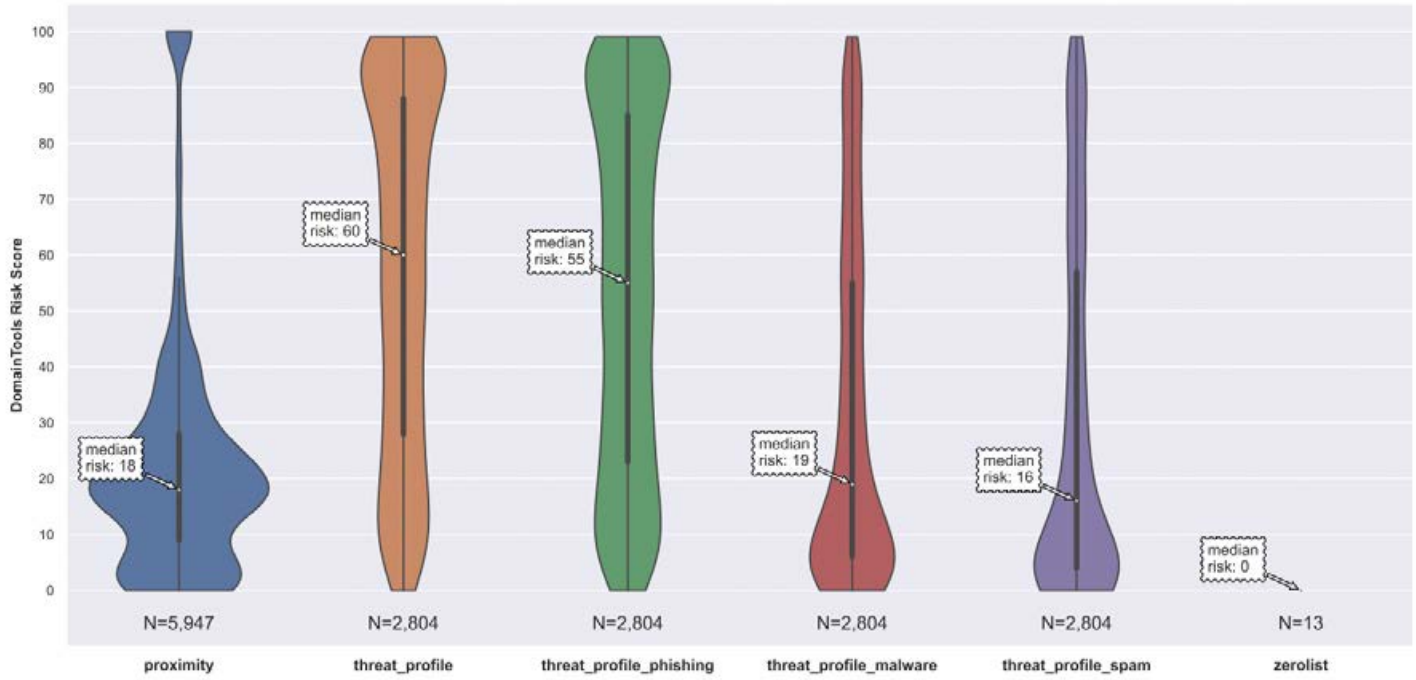
Comcast AS7922 Risk Scores Breakdown (Computed on a Subset of Domains)



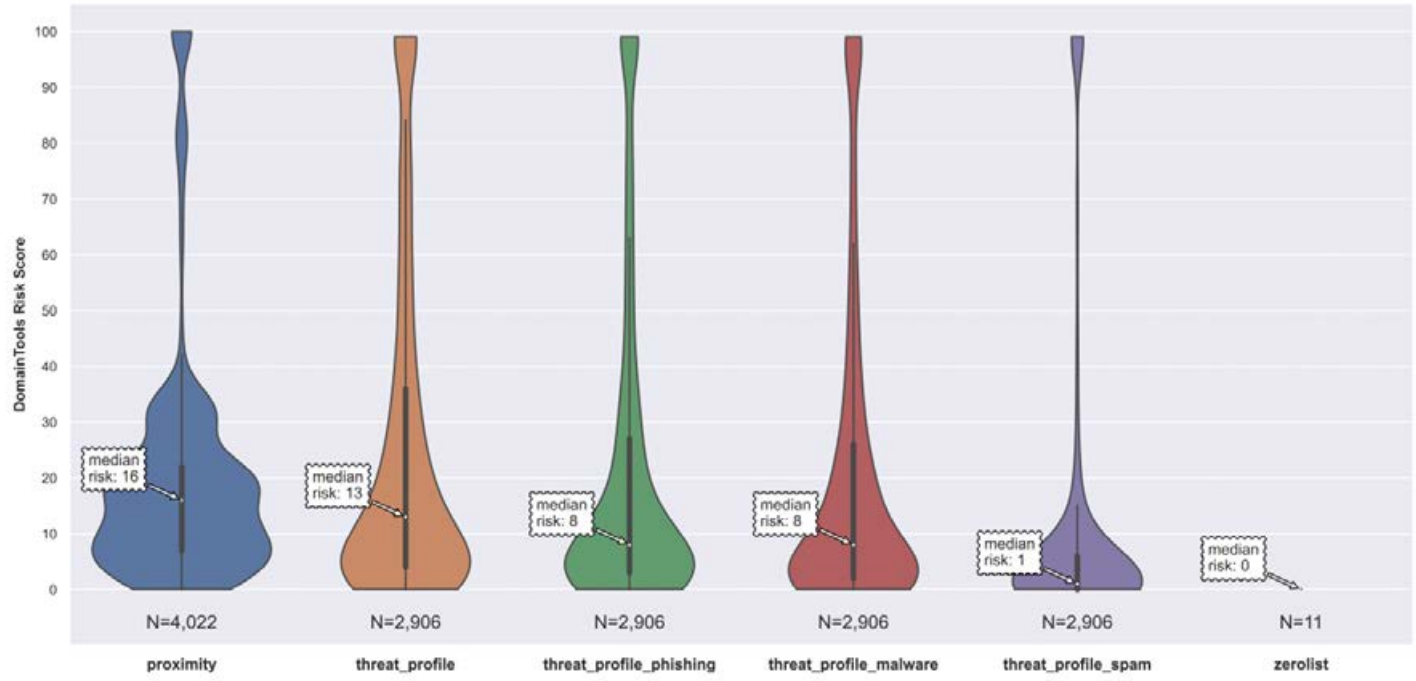
Confluence AS40034 Risk Scores Breakdown (Computed on a Subset of Domains)



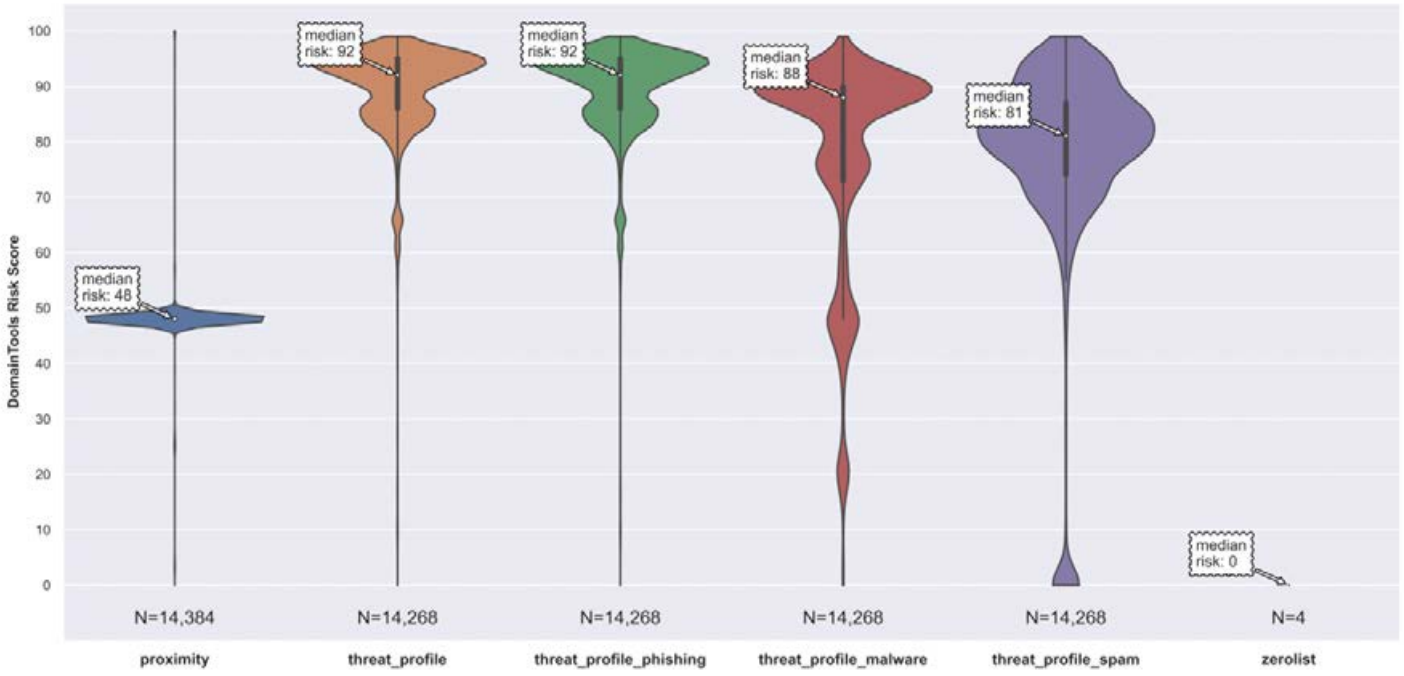
Contabo AS51167 Risk Scores Breakdown (Computed on a Subset of Domains)



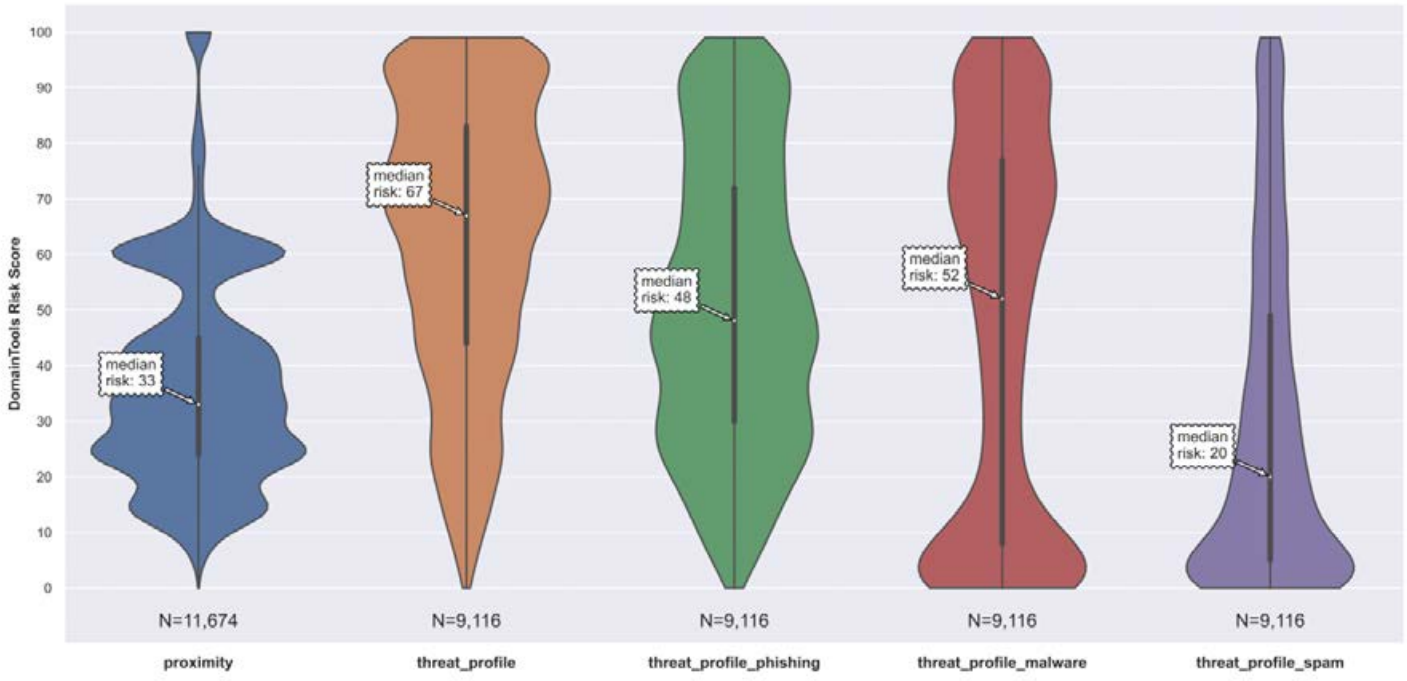
DDoS Guard AS57724 Risk Scores Breakdown (Computed on a Subset of Domains)



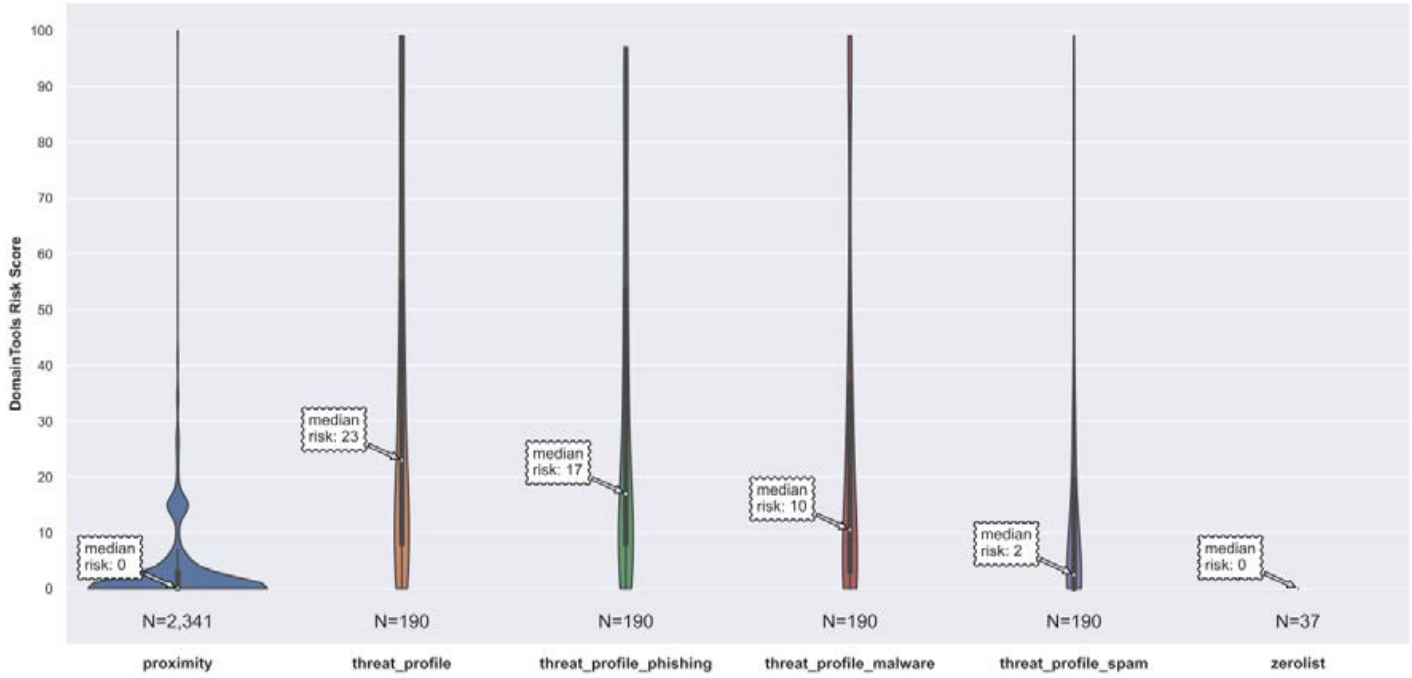
DDoS Guard AS262254 Risk Scores Breakdown (Computed on a Subset of Domains)



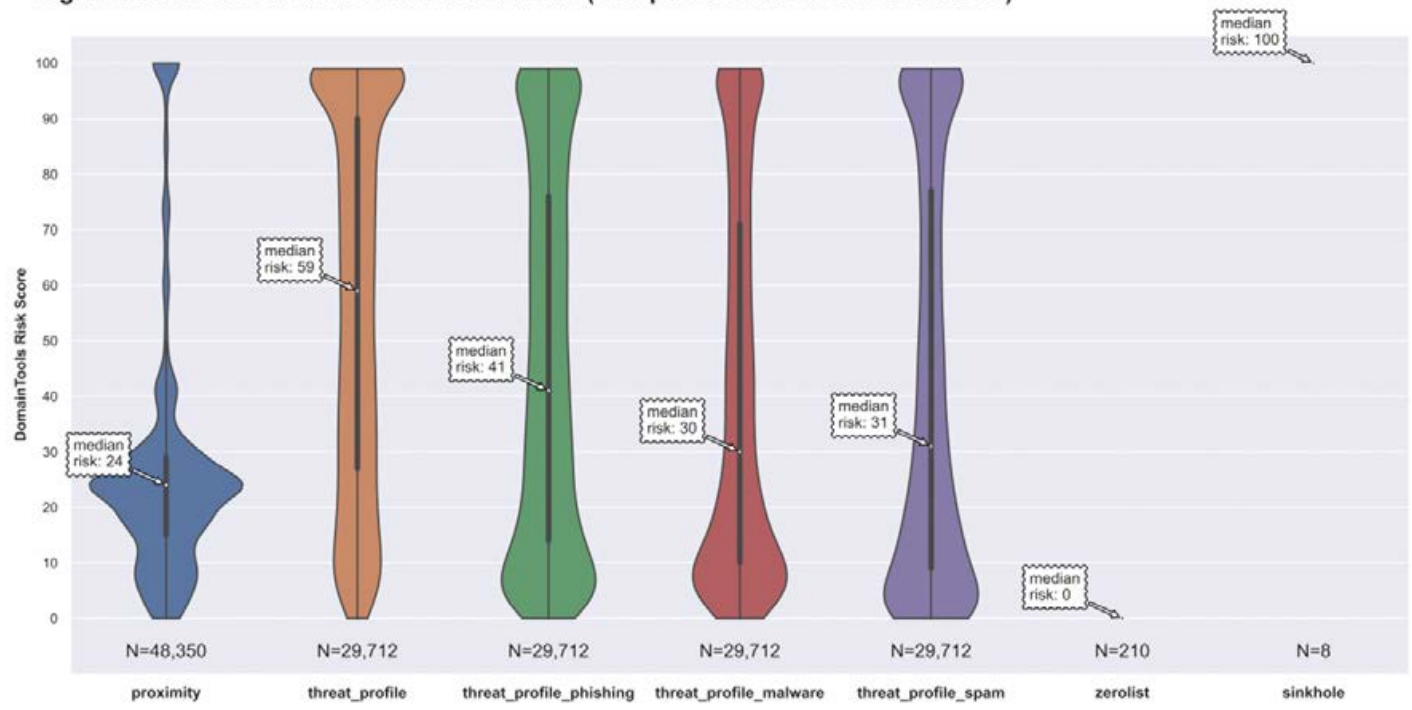
Dedipath AS35913 Risk Scores Breakdown (Computed on a Subset of Domains)



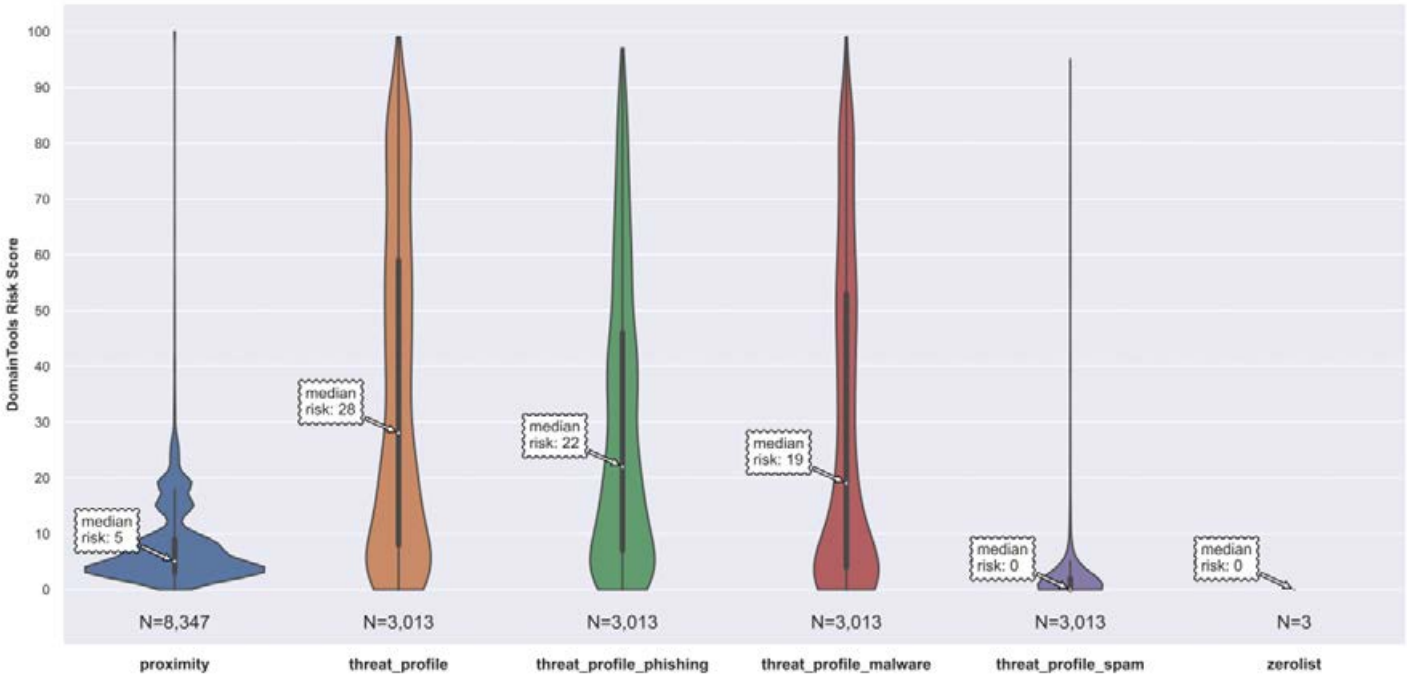
Deutsche Telecom AS3320 Risk Scores Breakdown (Computed on a Subset of Domains)



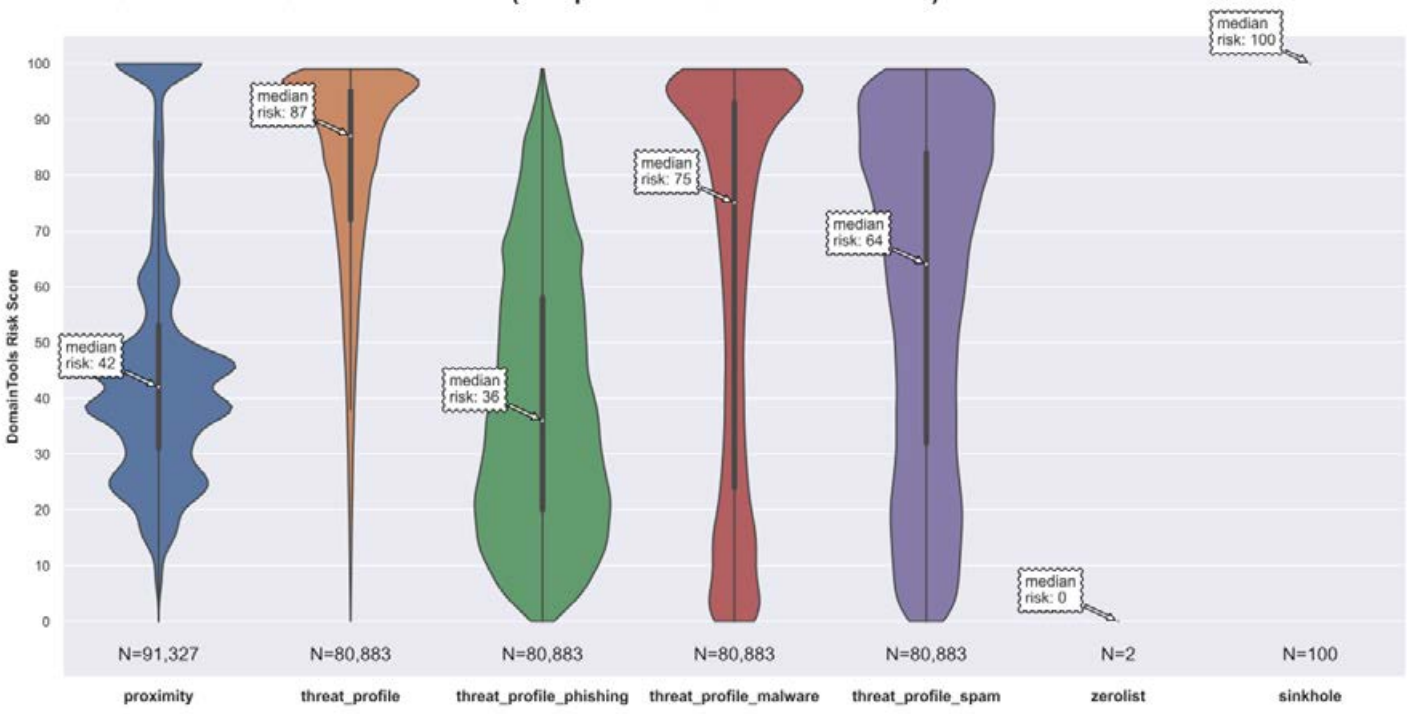
Digitalocean AS14061 Risk Scores Breakdown (Computed on a Subset of Domains)



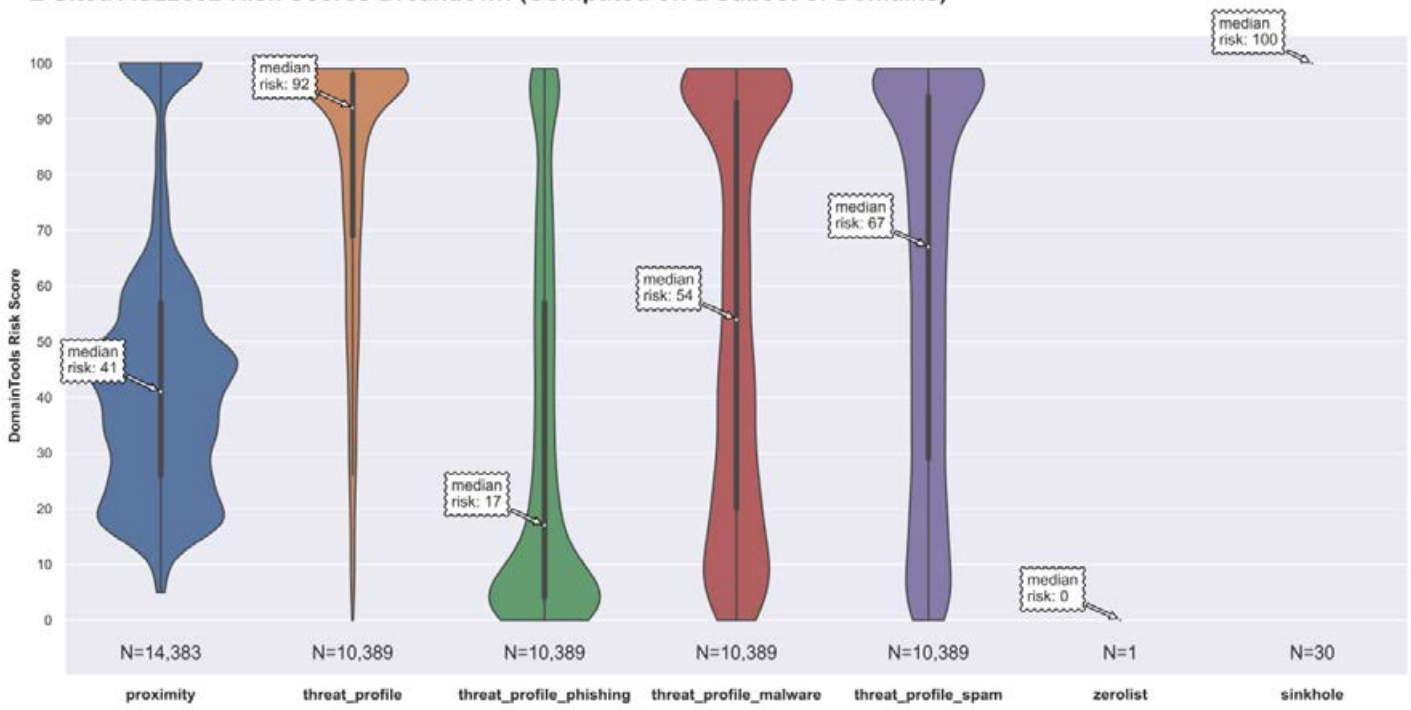
Dreamscape AS38719 Risk Scores Breakdown (Computed on a Subset of Domains)



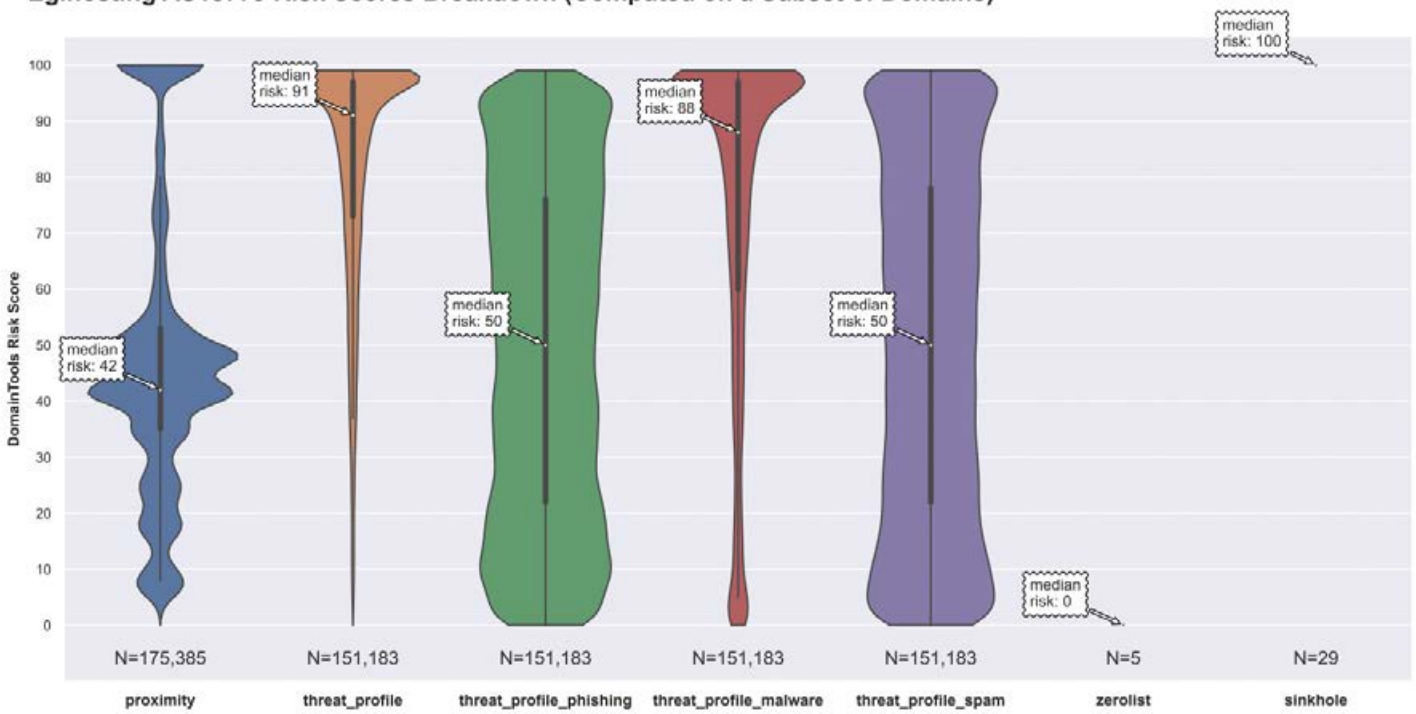
DXTL AS134548 Risk Scores Breakdown (Computed on a Subset of Domains)



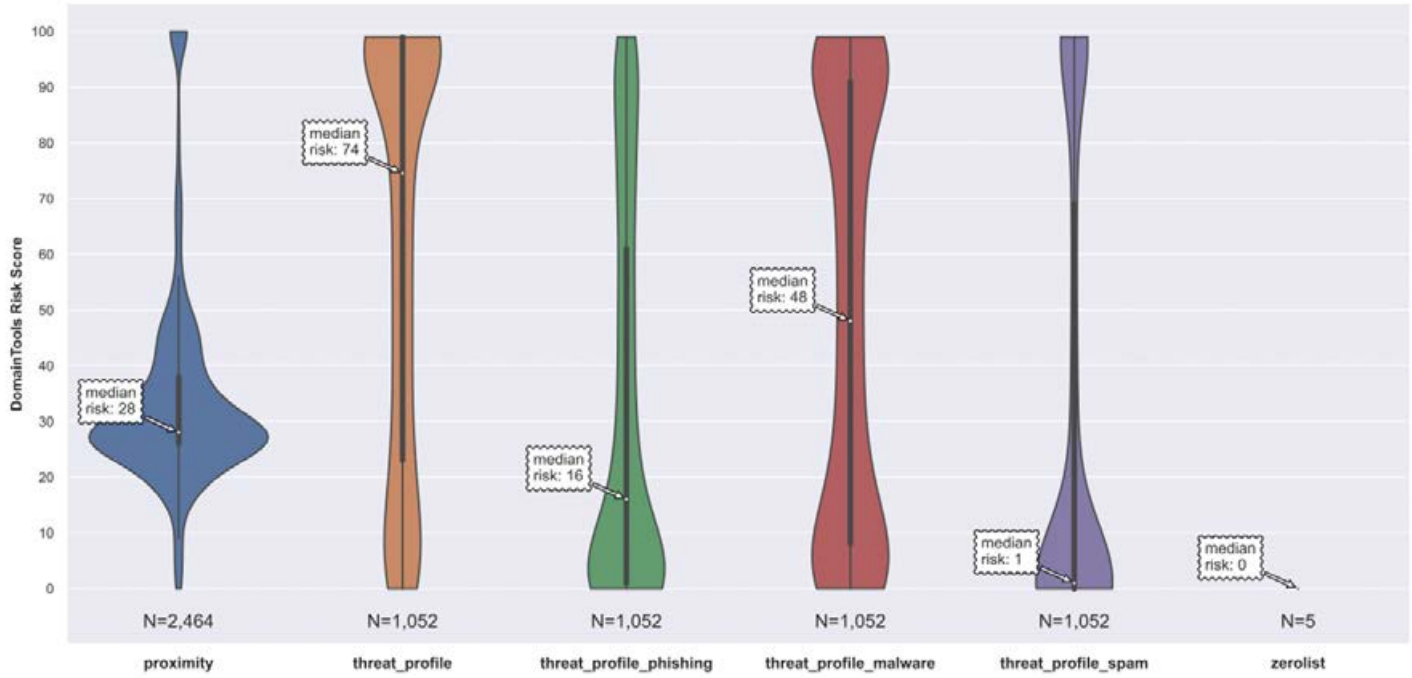
E Sited AS22552 Risk Scores Breakdown (Computed on a Subset of Domains)



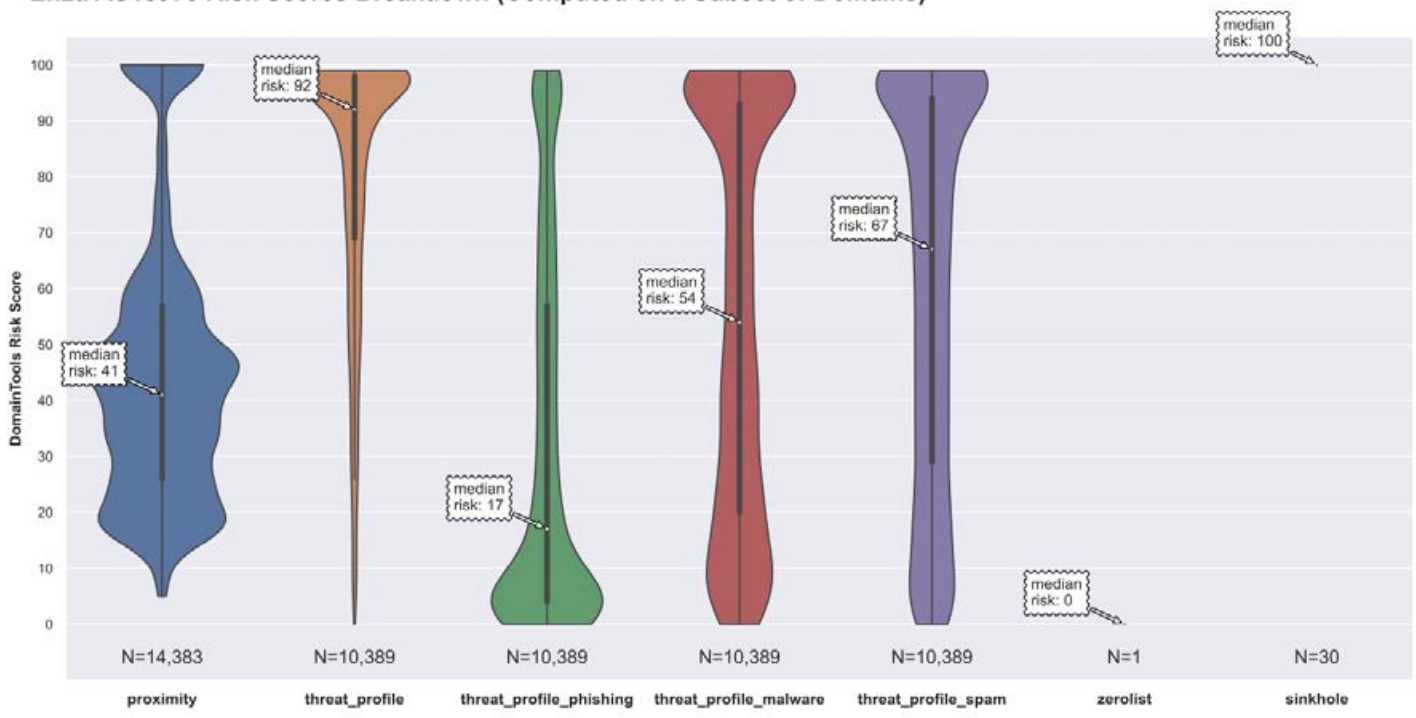
Egihosting AS18779 Risk Scores Breakdown (Computed on a Subset of Domains)



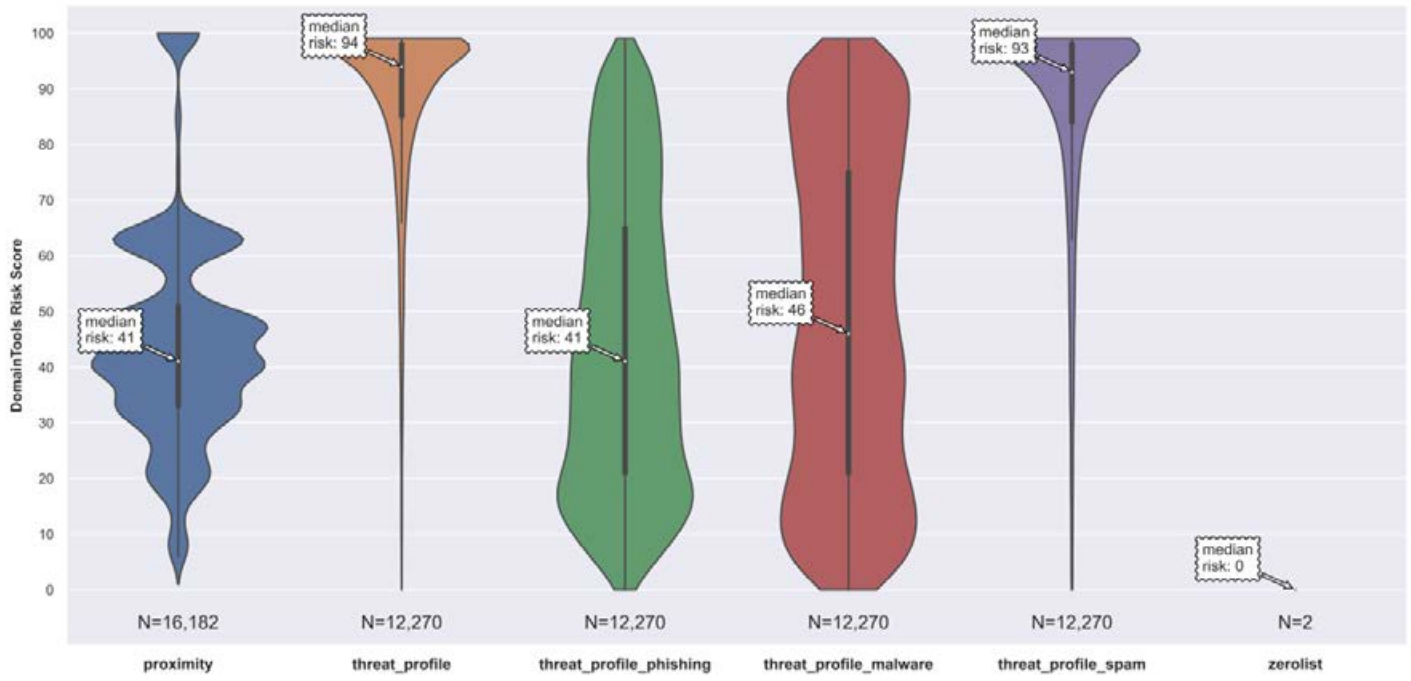
Ehostict AS45382 Risk Scores Breakdown (Computed on a Subset of Domains)



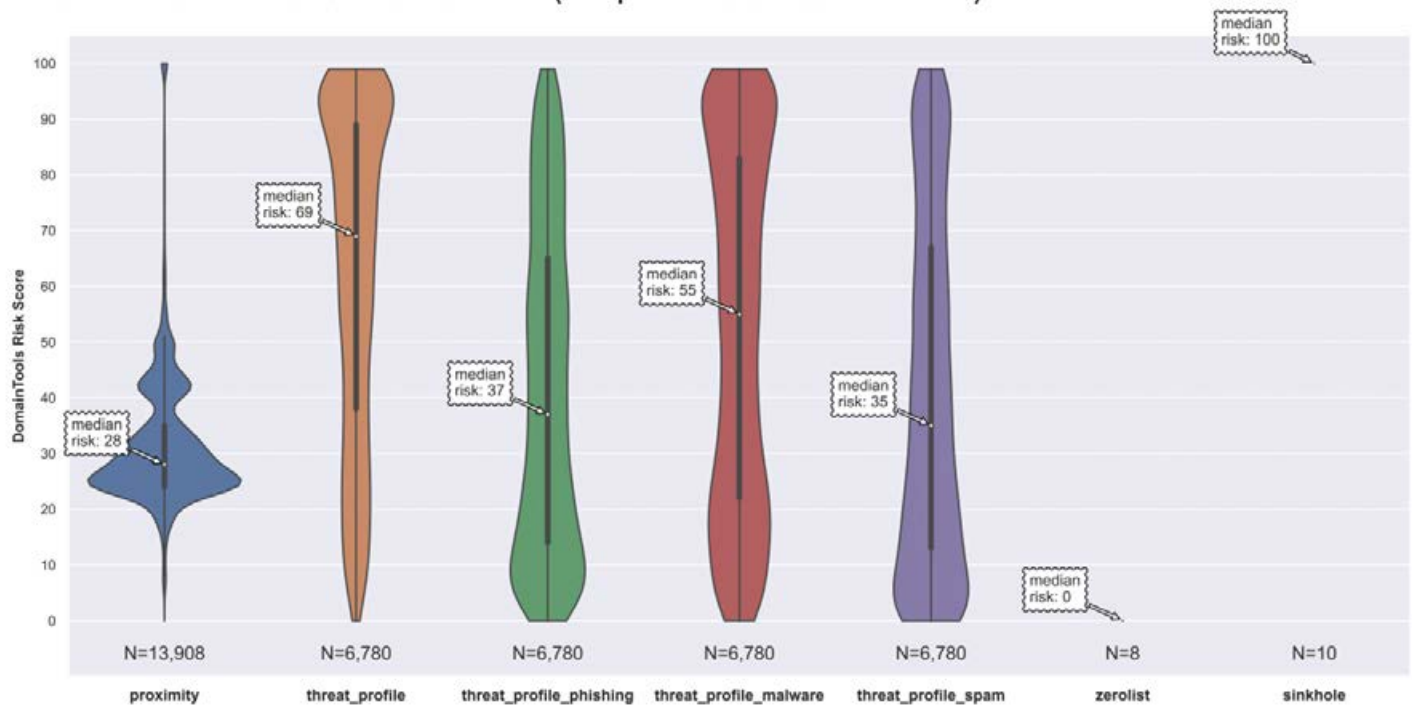
Enzu AS18978 Risk Scores Breakdown (Computed on a Subset of Domains)



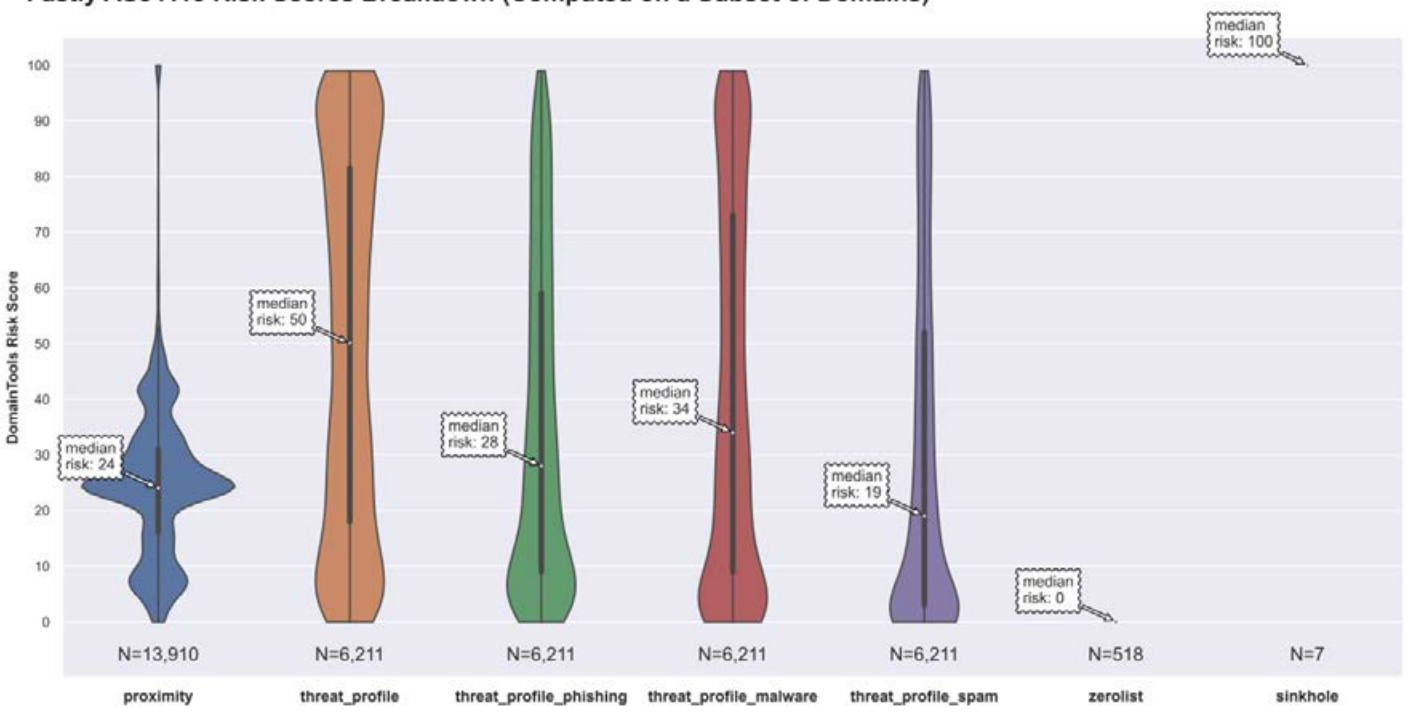
Eonix AS62904 Risk Scores Breakdown (Computed on a Subset of Domains)



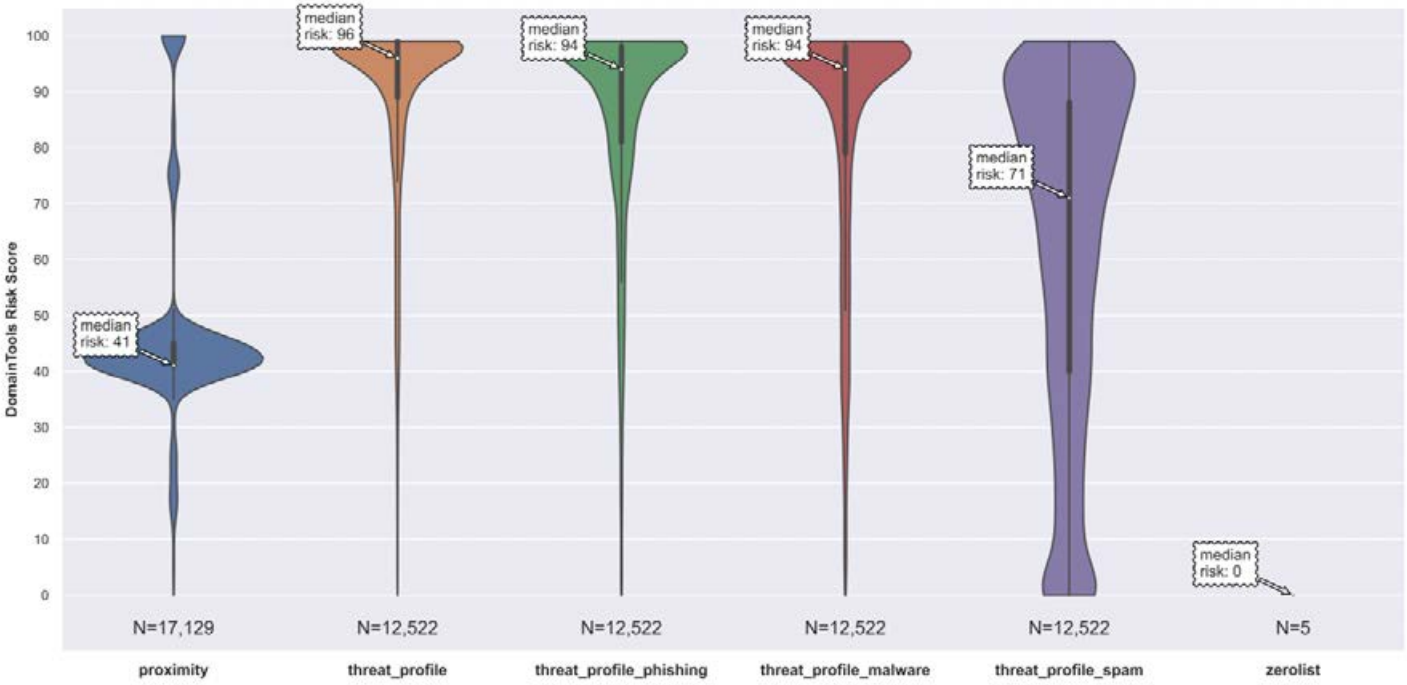
Facebook AS32934 Risk Scores Breakdown (Computed on a Subset of Domains)



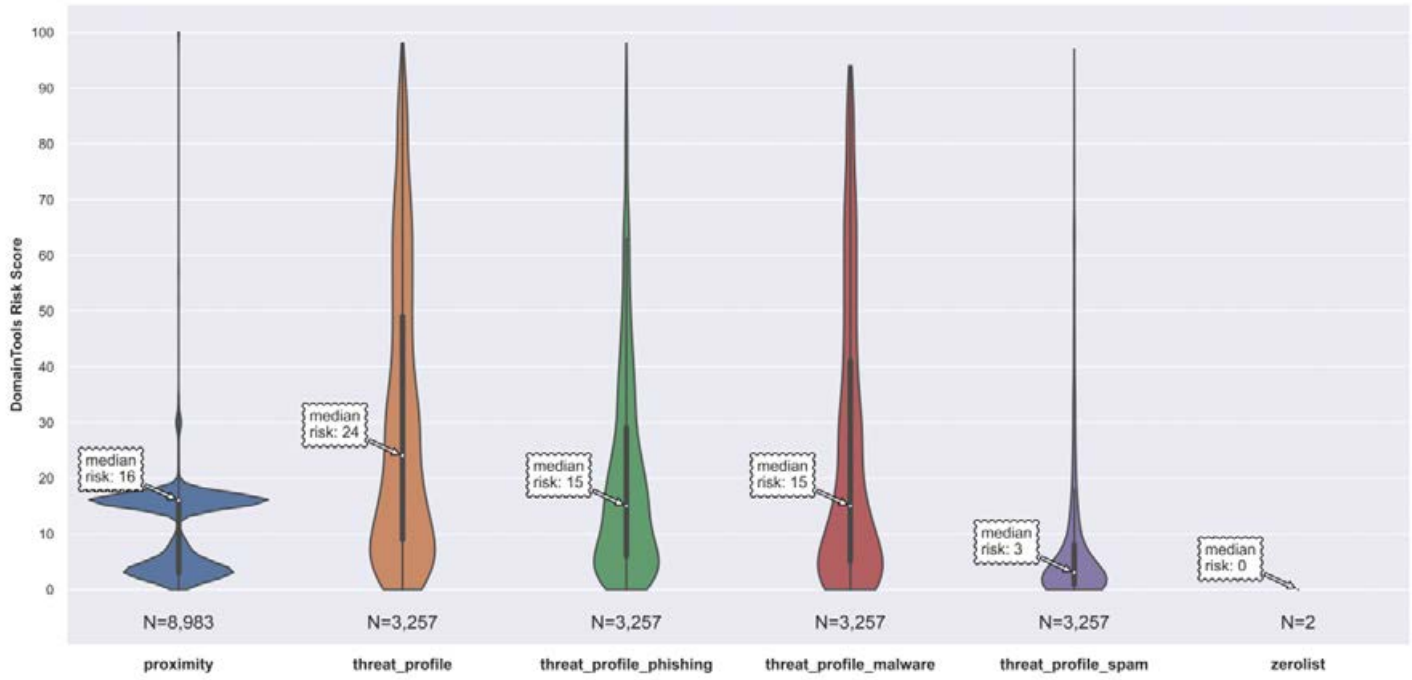
Fastly AS54113 Risk Scores Breakdown (Computed on a Subset of Domains)



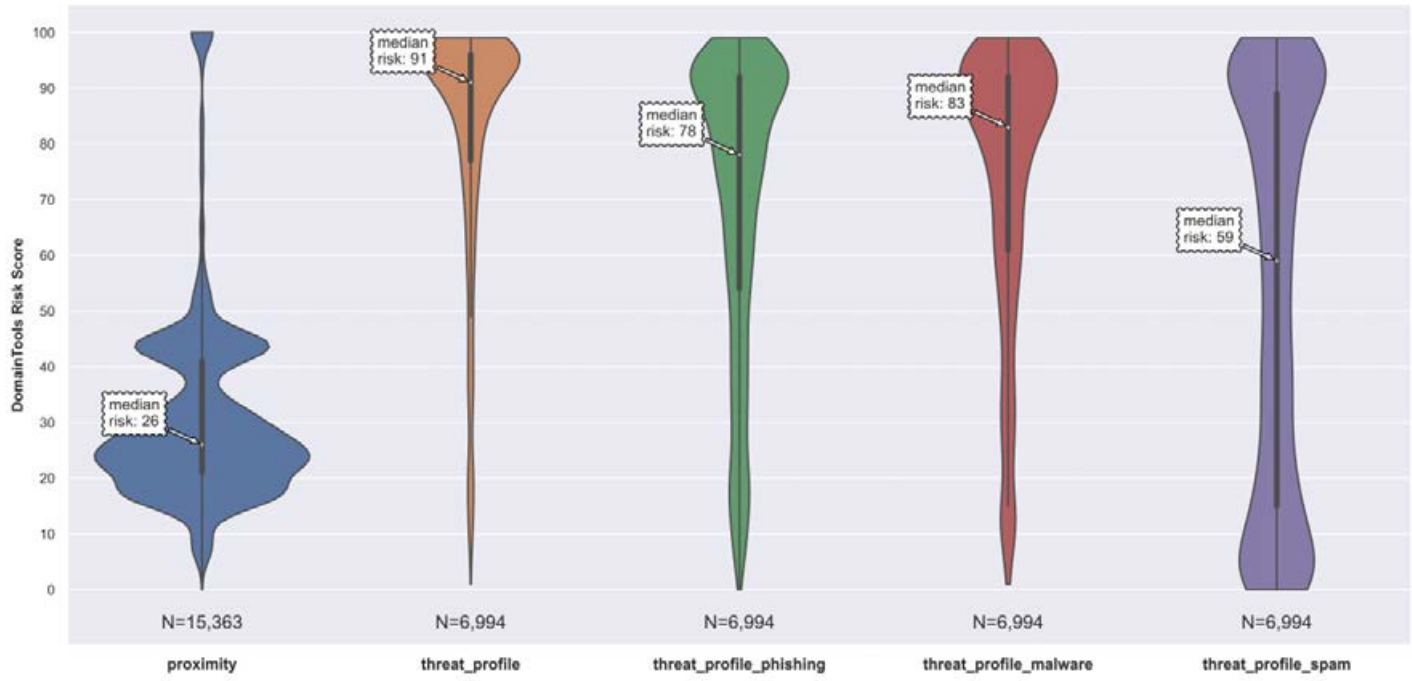
Frantech AS53667 Risk Scores Breakdown (Computed on a Subset of Domains)



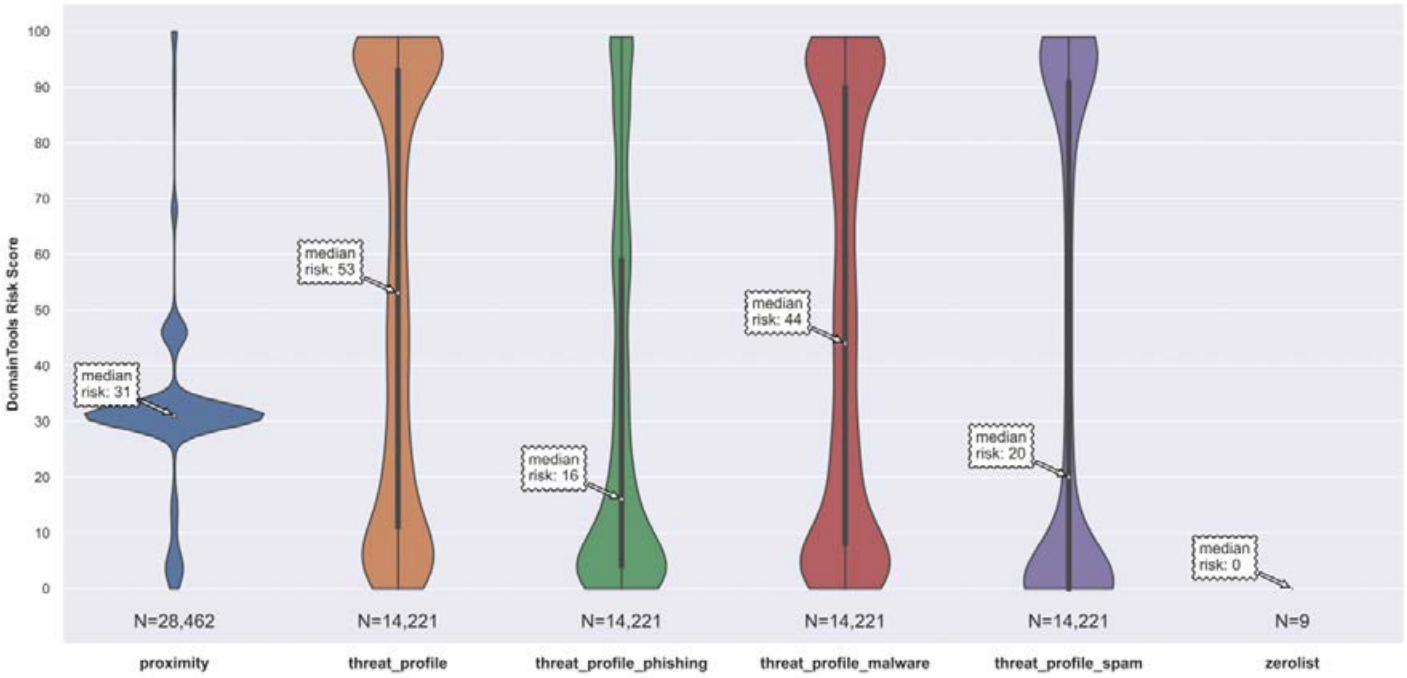
Gandi AS29169 Risk Scores Breakdown (Computed on a Subset of Domains)



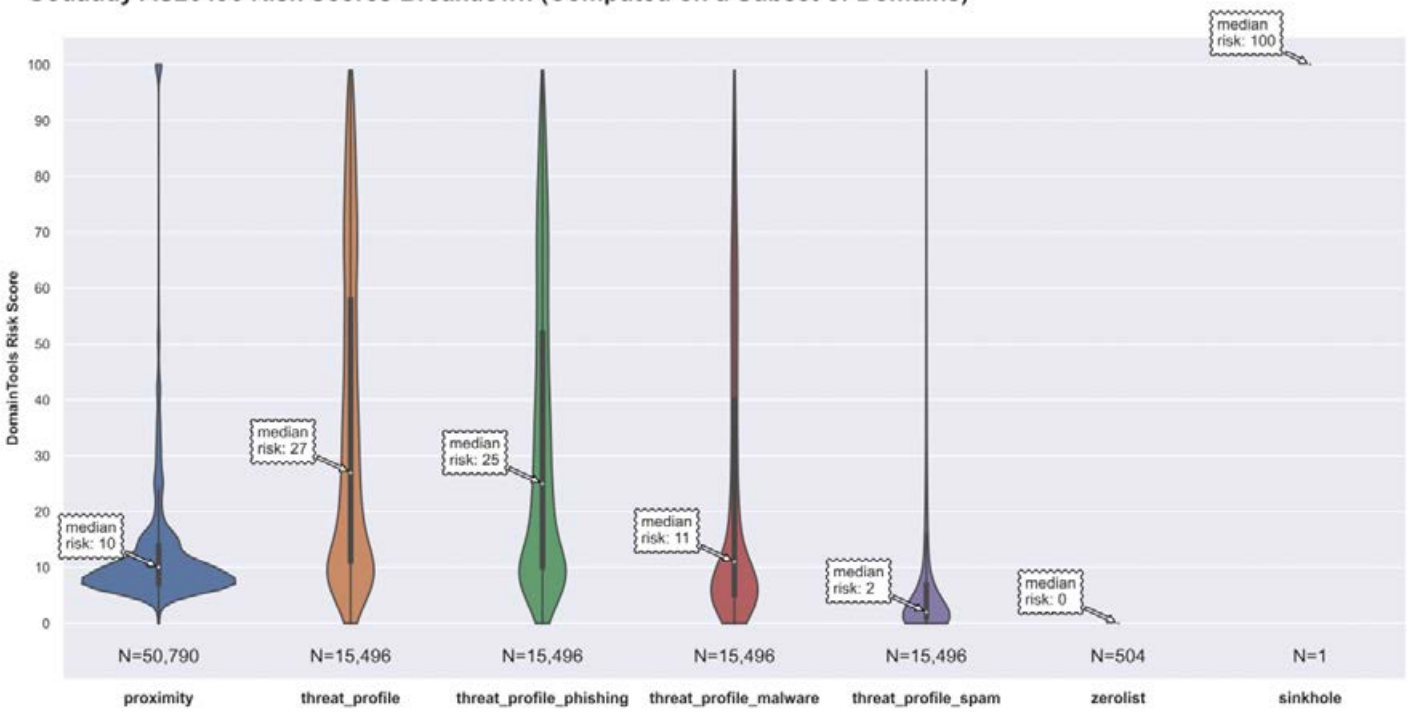
Gigabit Hosting AS55720 Risk Scores Breakdown (Computed on a Subset of Domains)



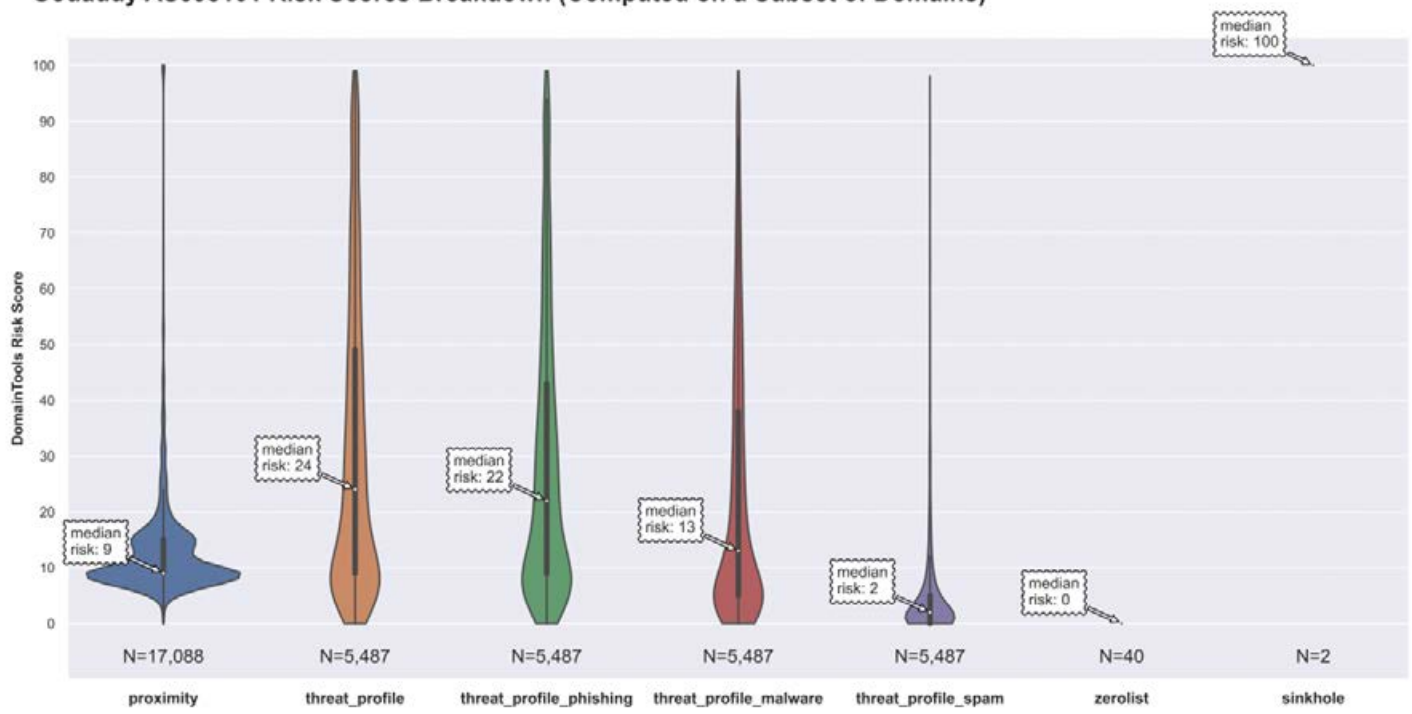
GMO Int AS7506 Risk Scores Breakdown (Computed on a Subset of Domains)



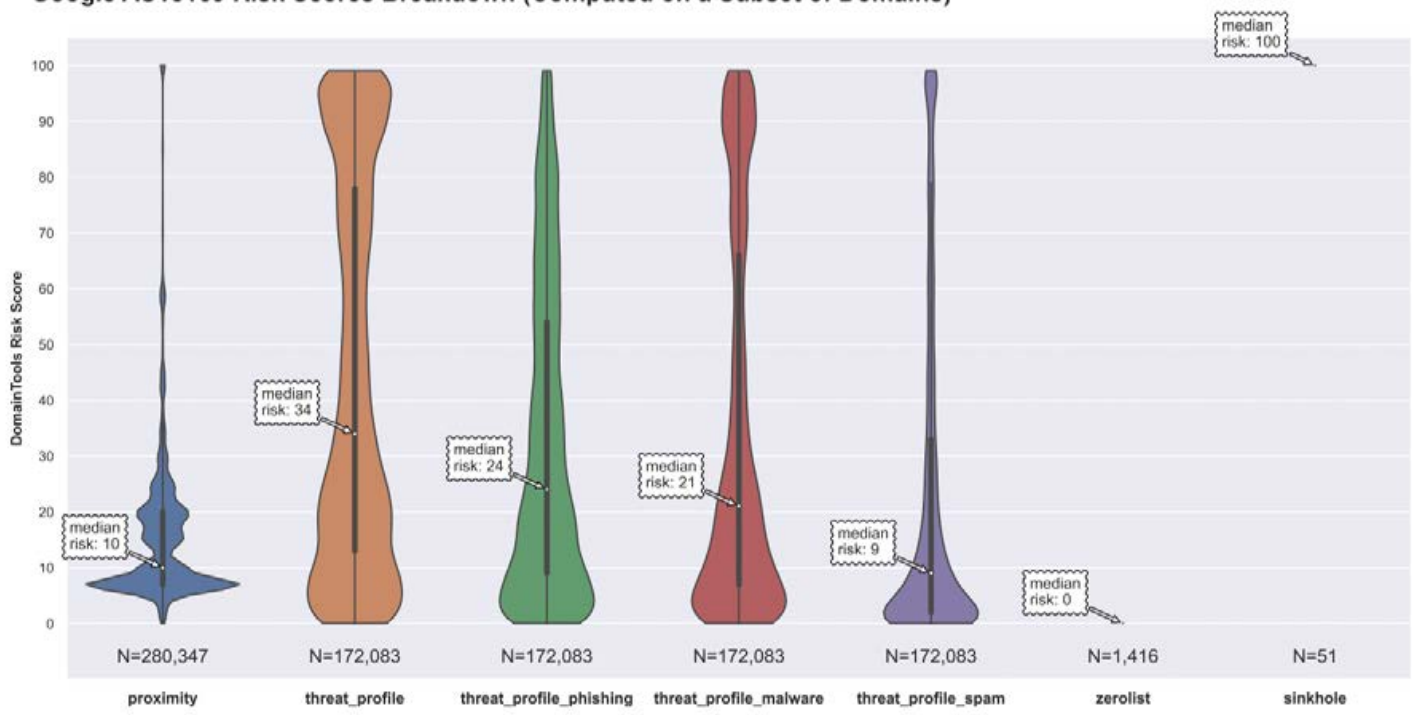
Godaddy AS26496 Risk Scores Breakdown (Computed on a Subset of Domains)



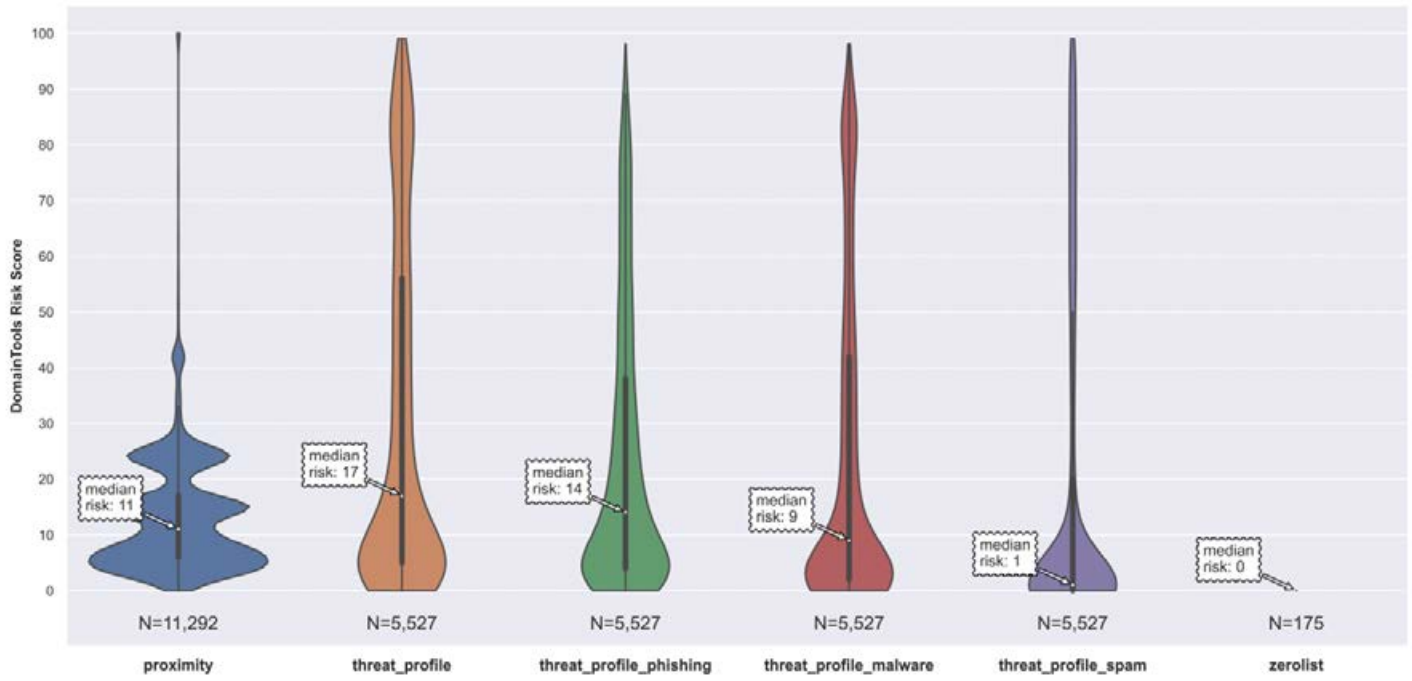
Godaddy AS398101 Risk Scores Breakdown (Computed on a Subset of Domains)



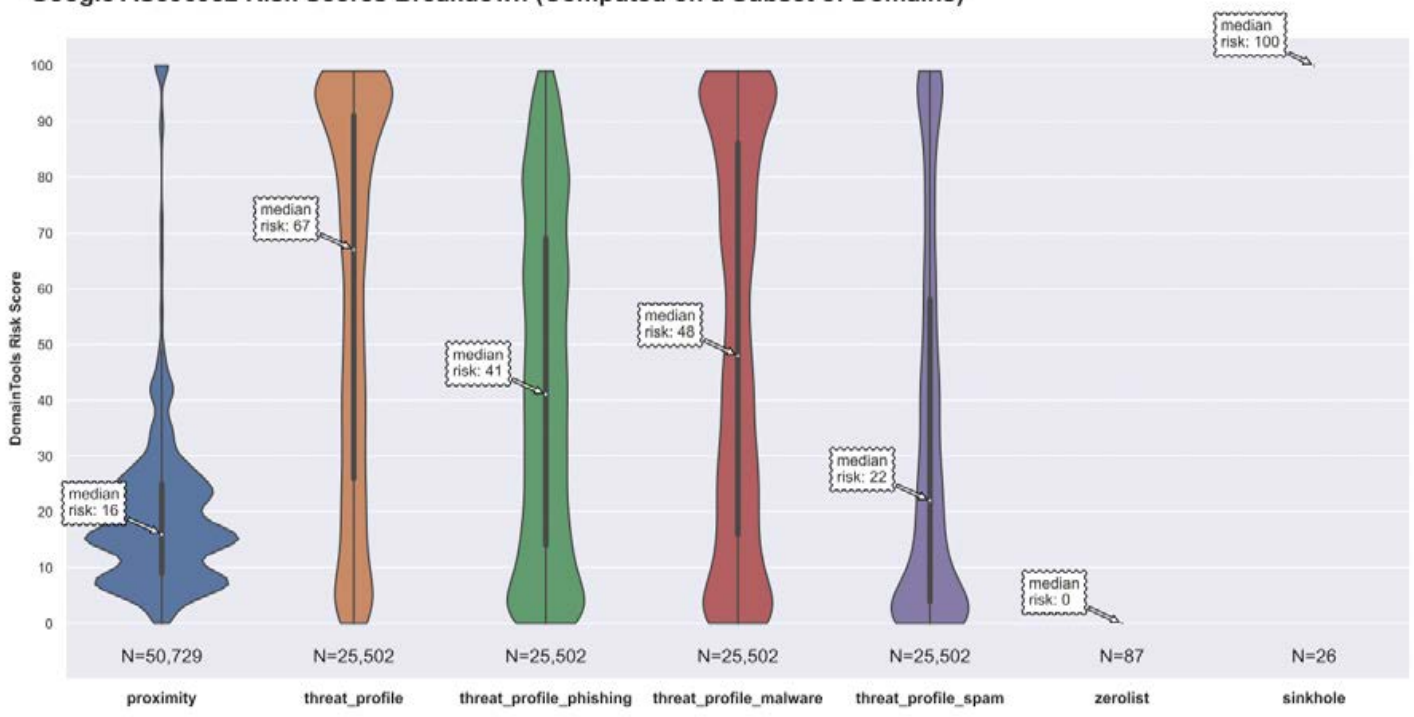
Google AS15169 Risk Scores Breakdown (Computed on a Subset of Domains)

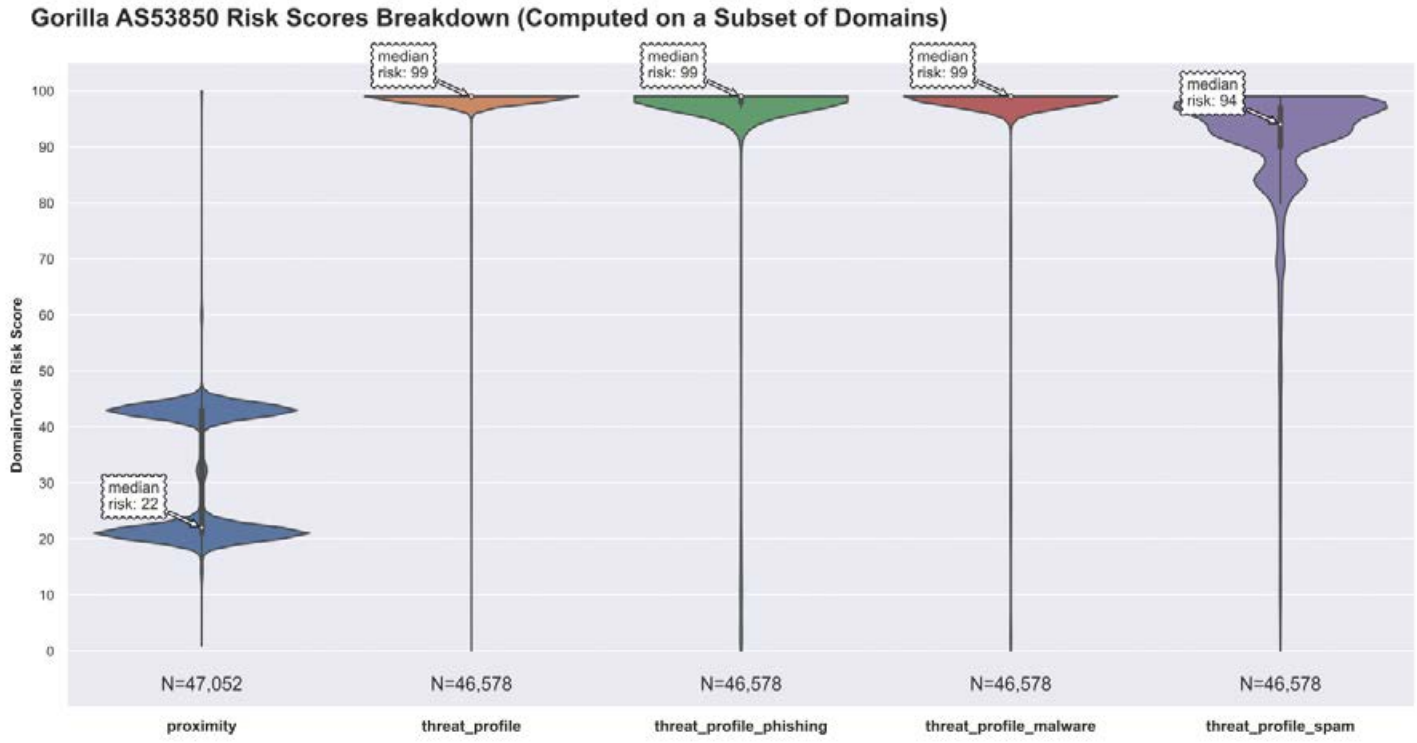


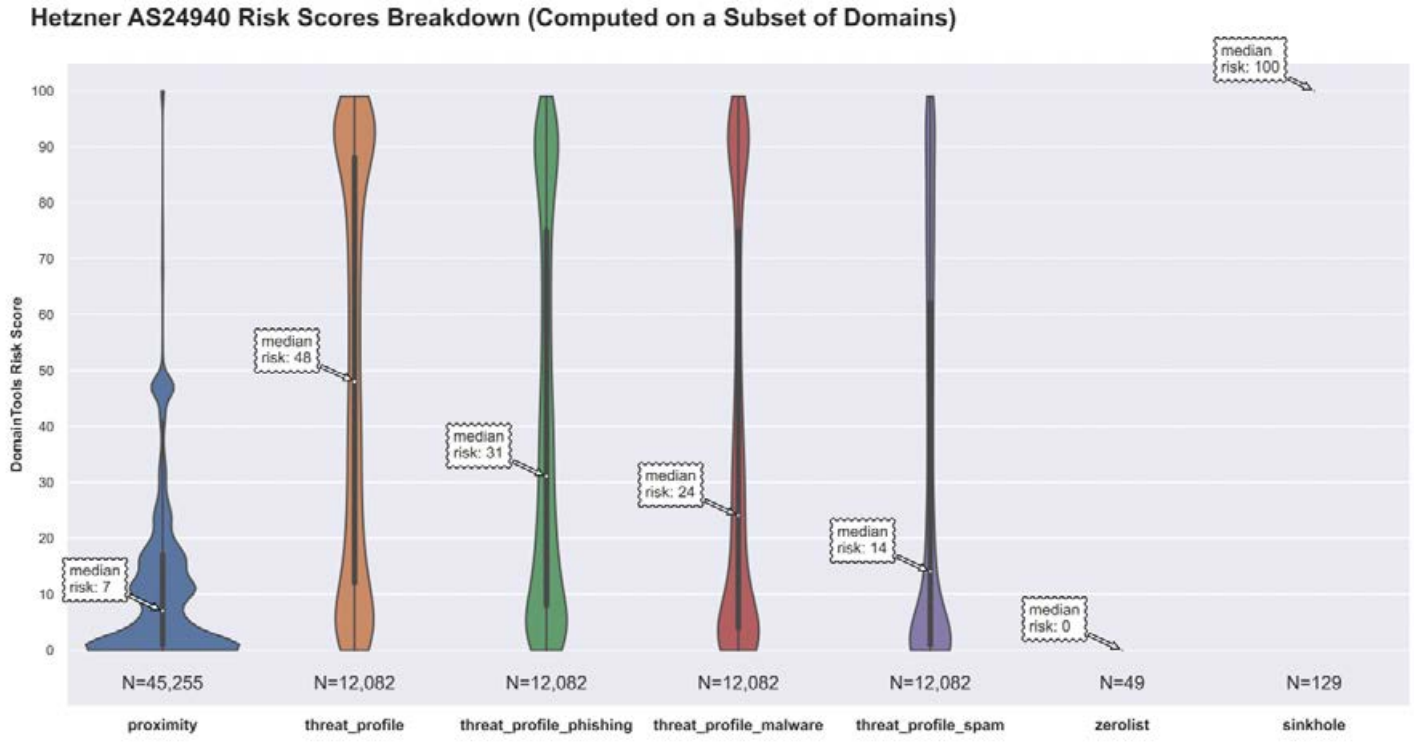
Google AS19527 Risk Scores Breakdown (Computed on a Subset of Domains)



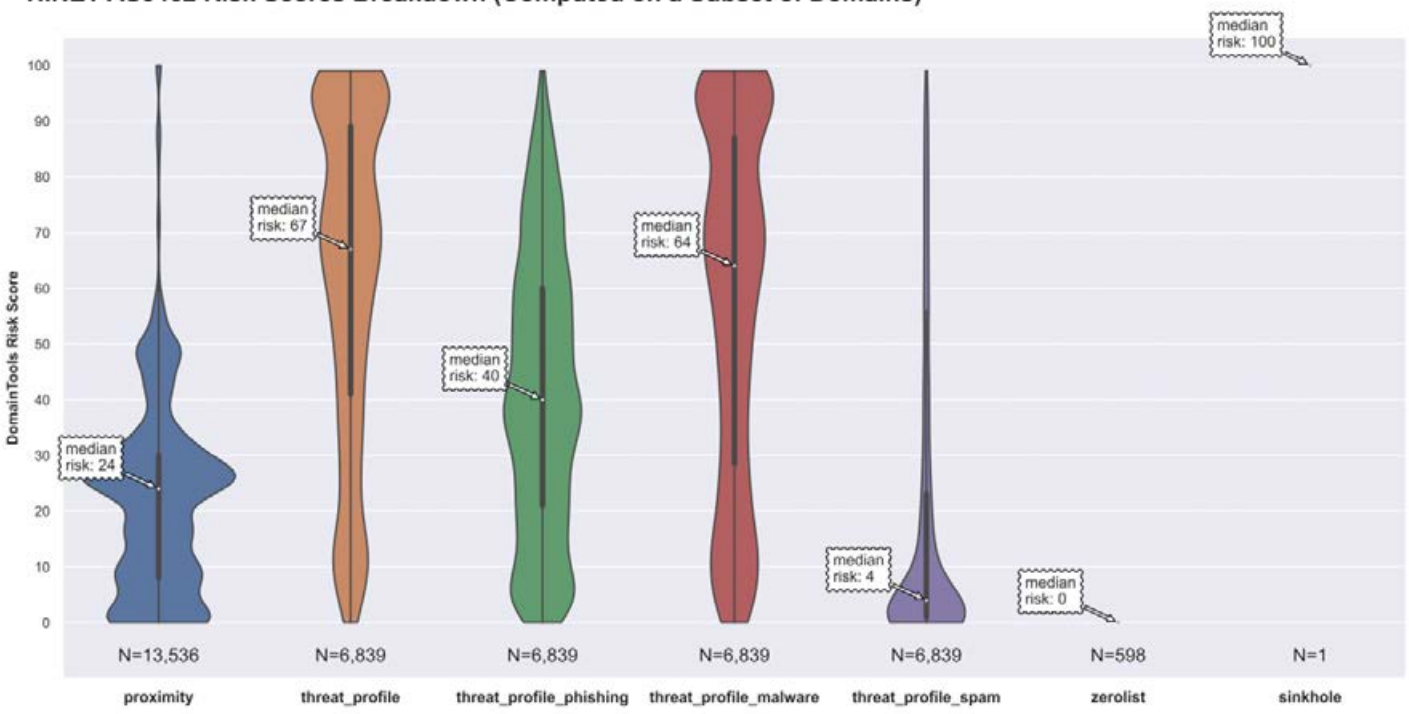
Google AS396982 Risk Scores Breakdown (Computed on a Subset of Domains)



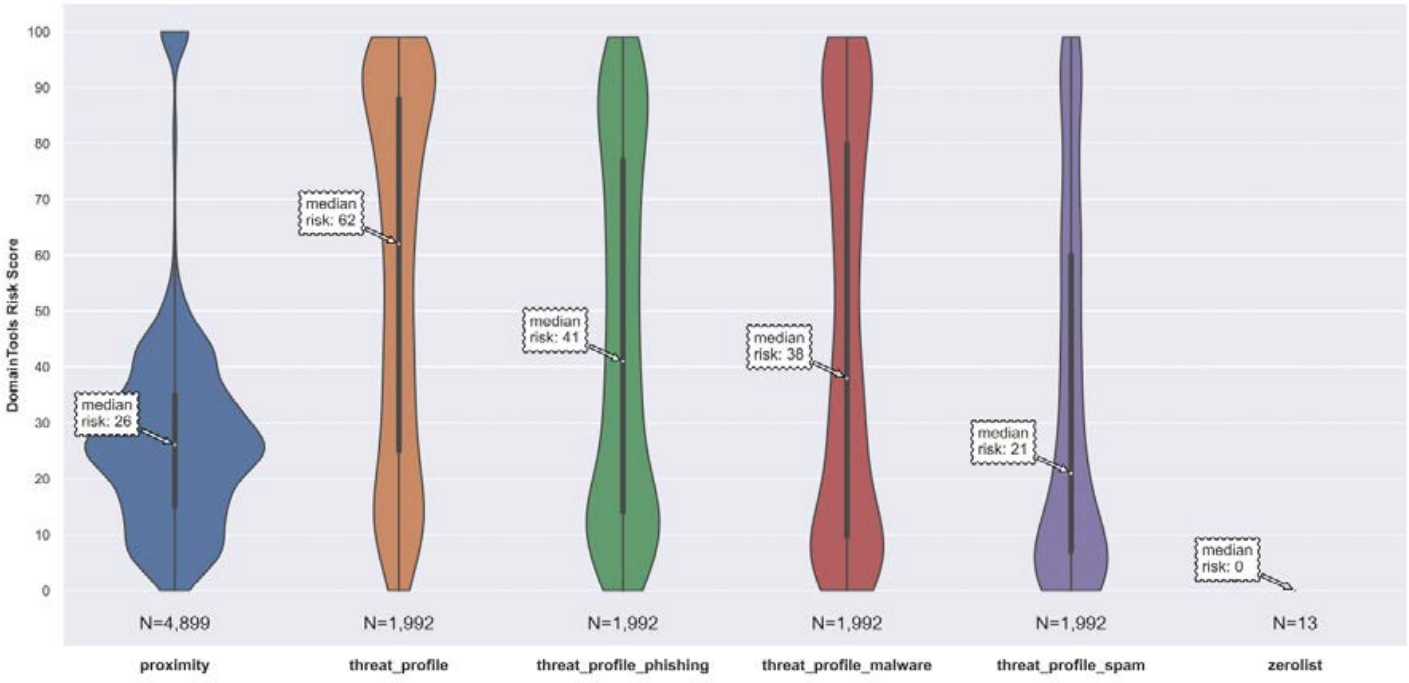

Gorilla AS53850 Risk Scores Breakdown (Computed on a Subset of Domains)
DomainTools Risk Score
median risk: 22
median risk: 99
median risk: 99
median risk: 99
median risk: 94
N=47,052
N=46,578
N=46,578
N=46,578
N=46,578
proximity
threat_profile
threat_profile_phishing
threat_profile_malware
threat_profile_spam


Hetzner AS24940 Risk Scores Breakdown (Computed on a Subset of Domains)
DomainTools Risk Score
median risk: 7
median risk: 48
median risk: 31
median risk: 24
median risk: 14
median risk: 0
median risk: 100
N=45,255
N=12,082
N=12,082
N=12,082
N=12,082
N=49
N=129
proximity
threat_profile
threat_profile_phishing
threat_profile_malware
threat_profile_spam
zerolist
sinkhole

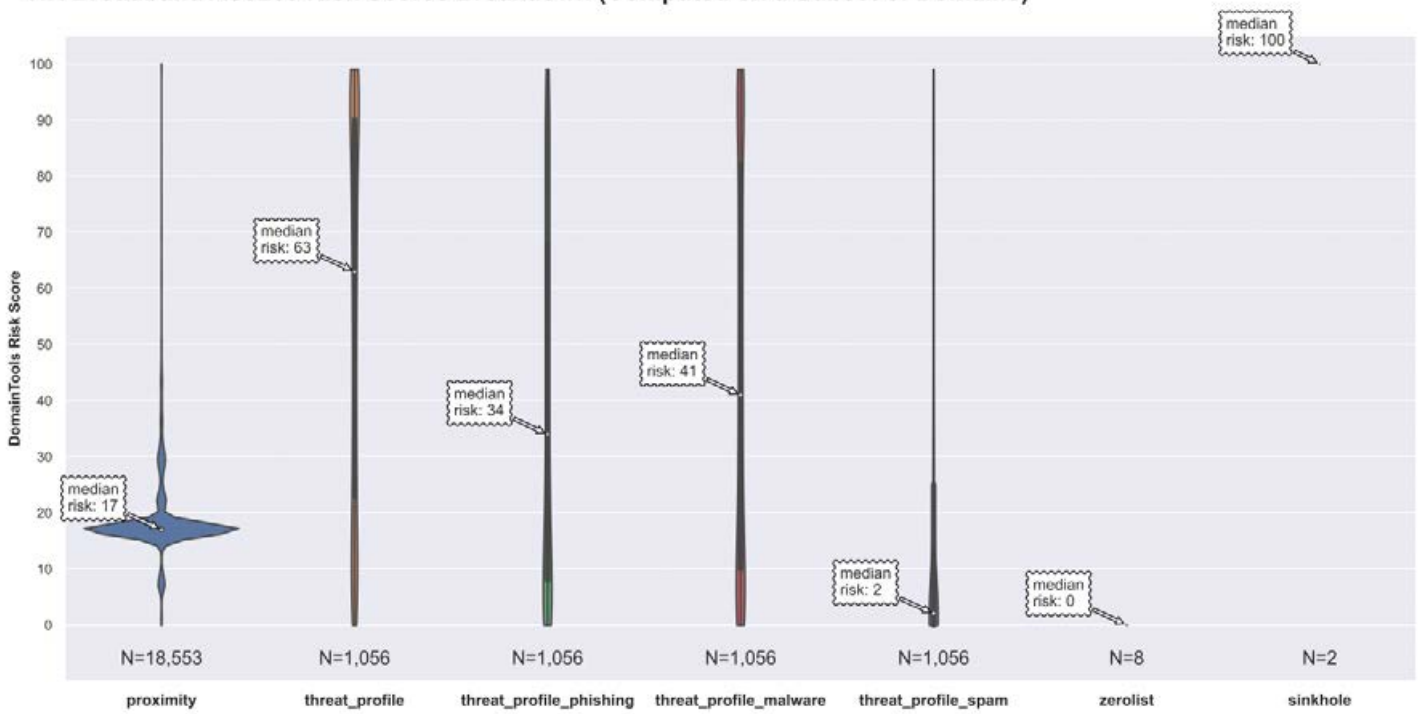
HINET AS3462 Risk Scores Breakdown (Computed on a Subset of Domains)



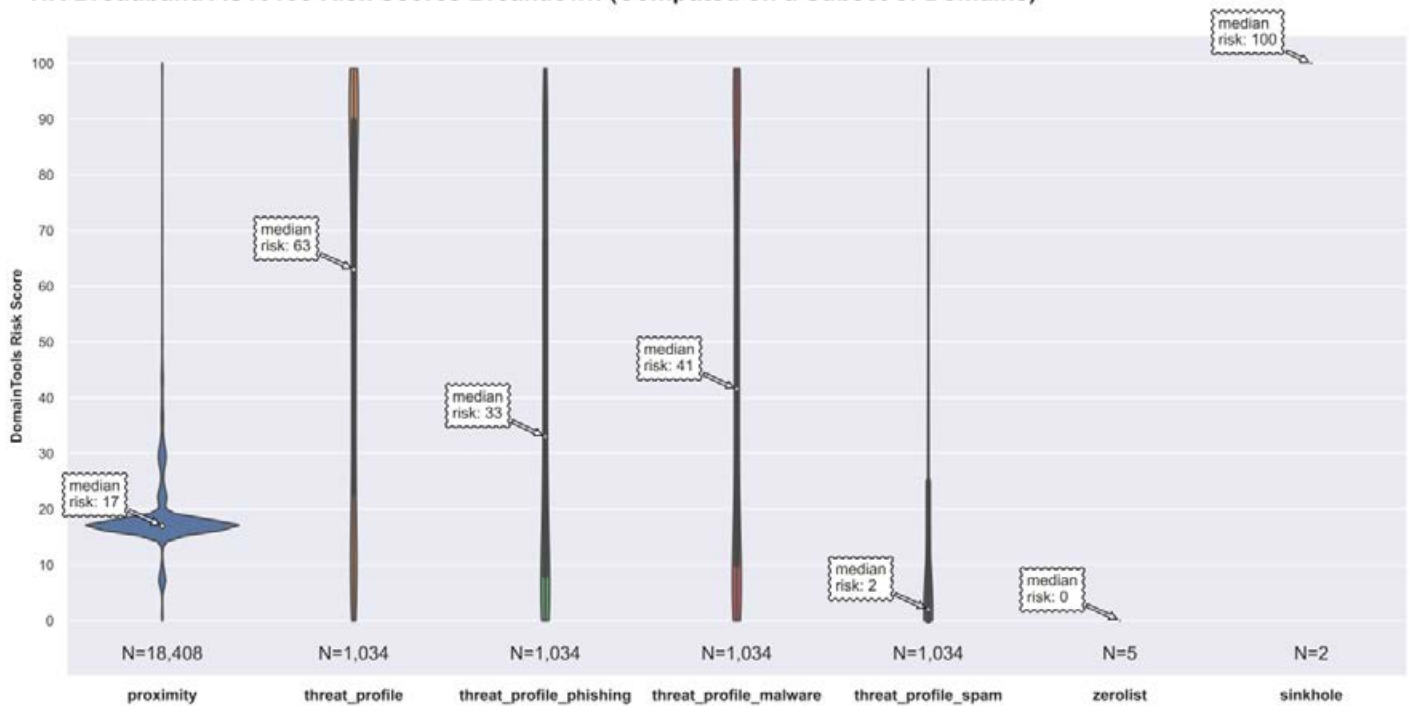
Hivelocity AS29802 Risk Scores Breakdown (Computed on a Subset of Domains)



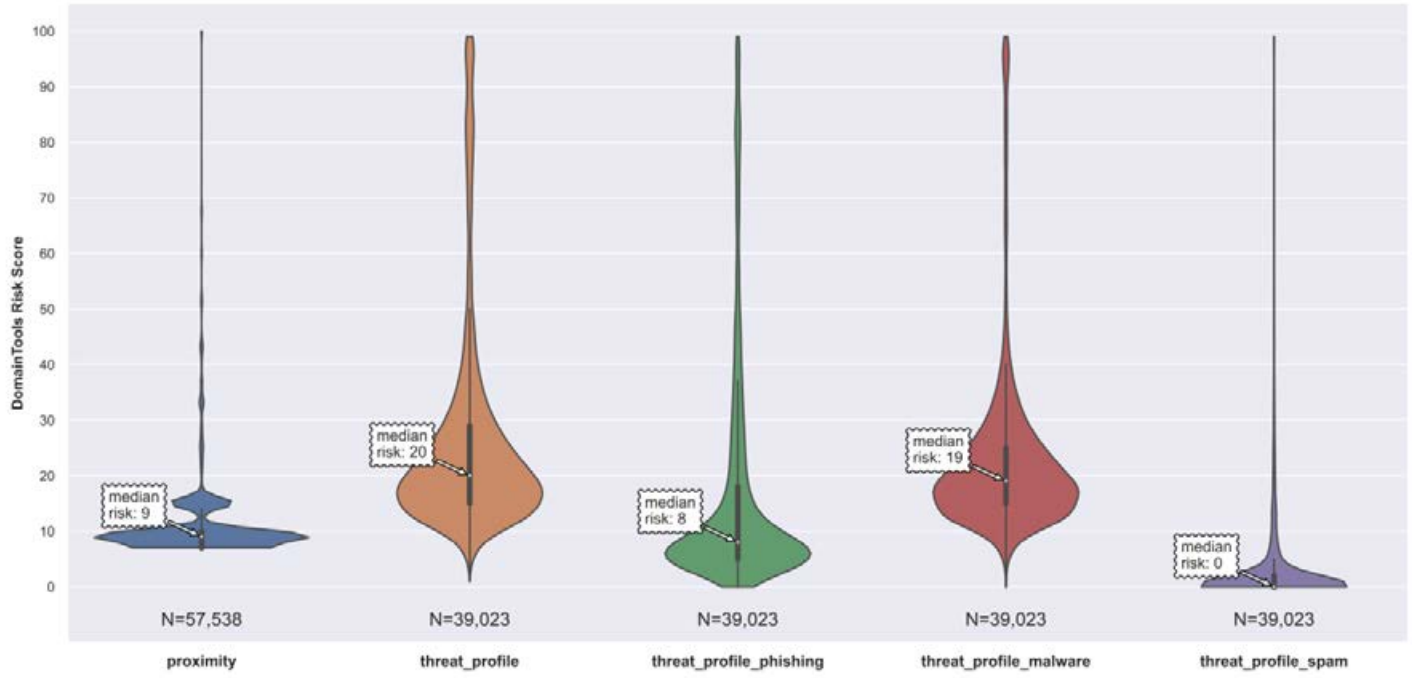
HK Broadband AS9269 Risk Scores Breakdown (Computed on a Subset of Domains)



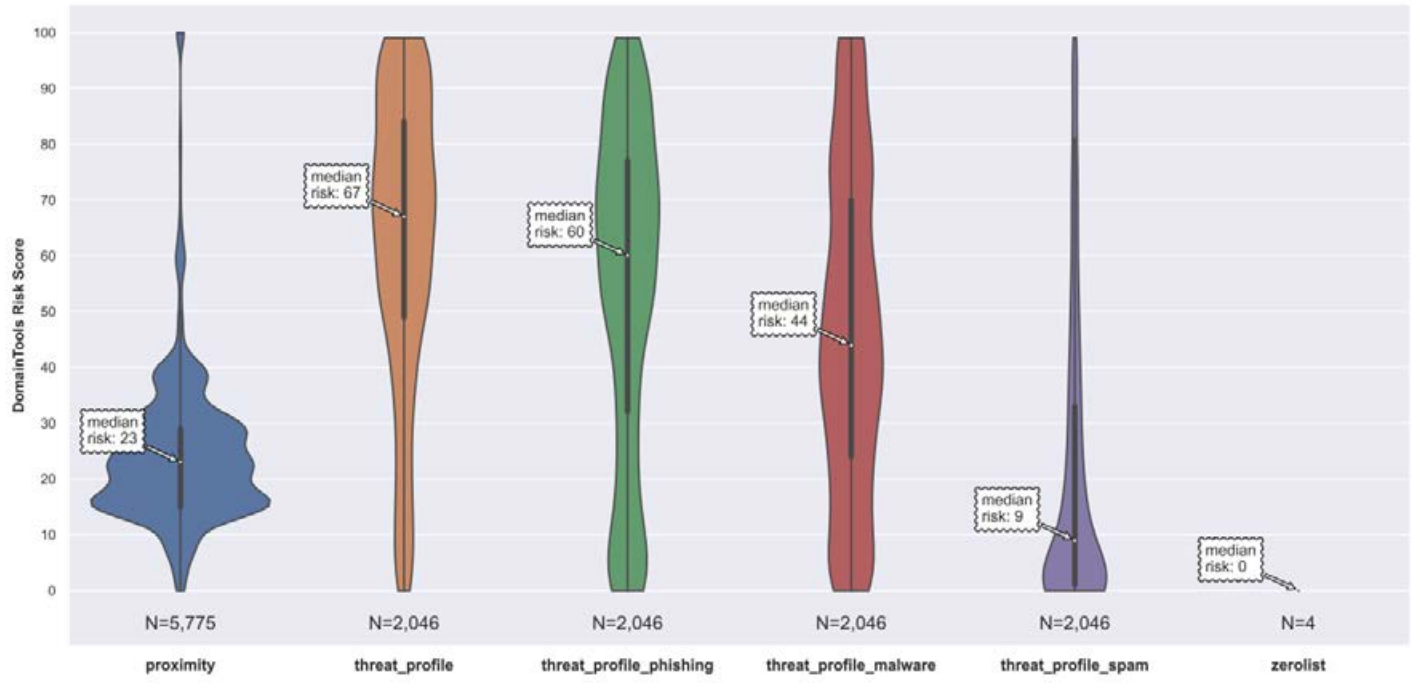
HK Broadband AS10103 Risk Scores Breakdown (Computed on a Subset of Domains)



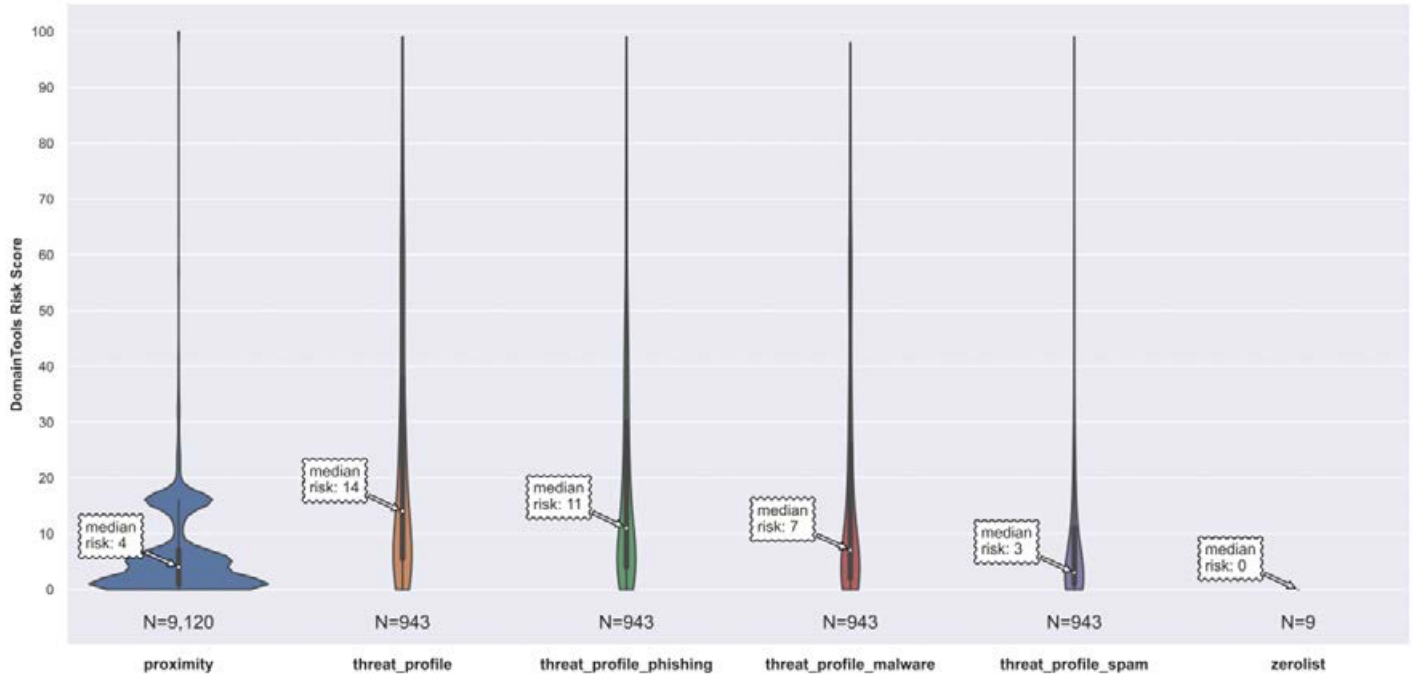
HK Comm AS140227 Risk Scores Breakdown (Computed on a Subset of Domains)



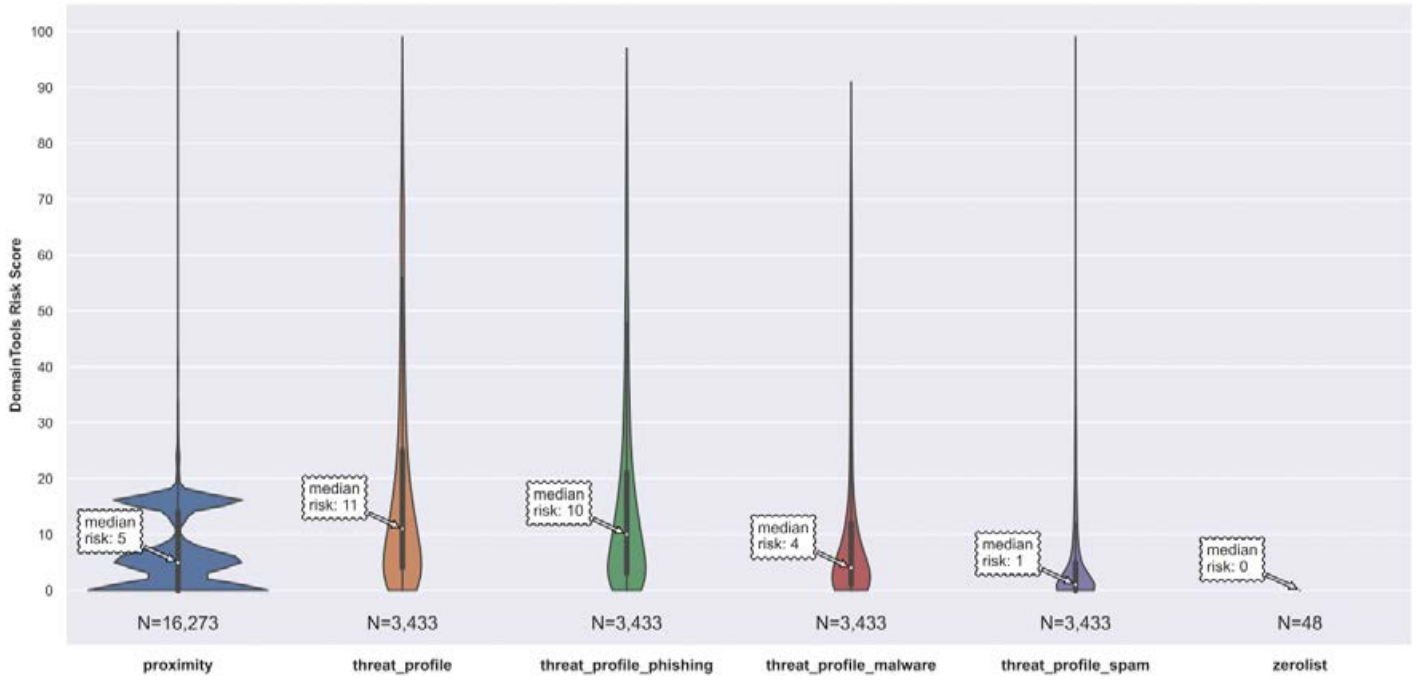
HKBN AS17444 Risk Scores Breakdown (Computed on a Subset of Domains)



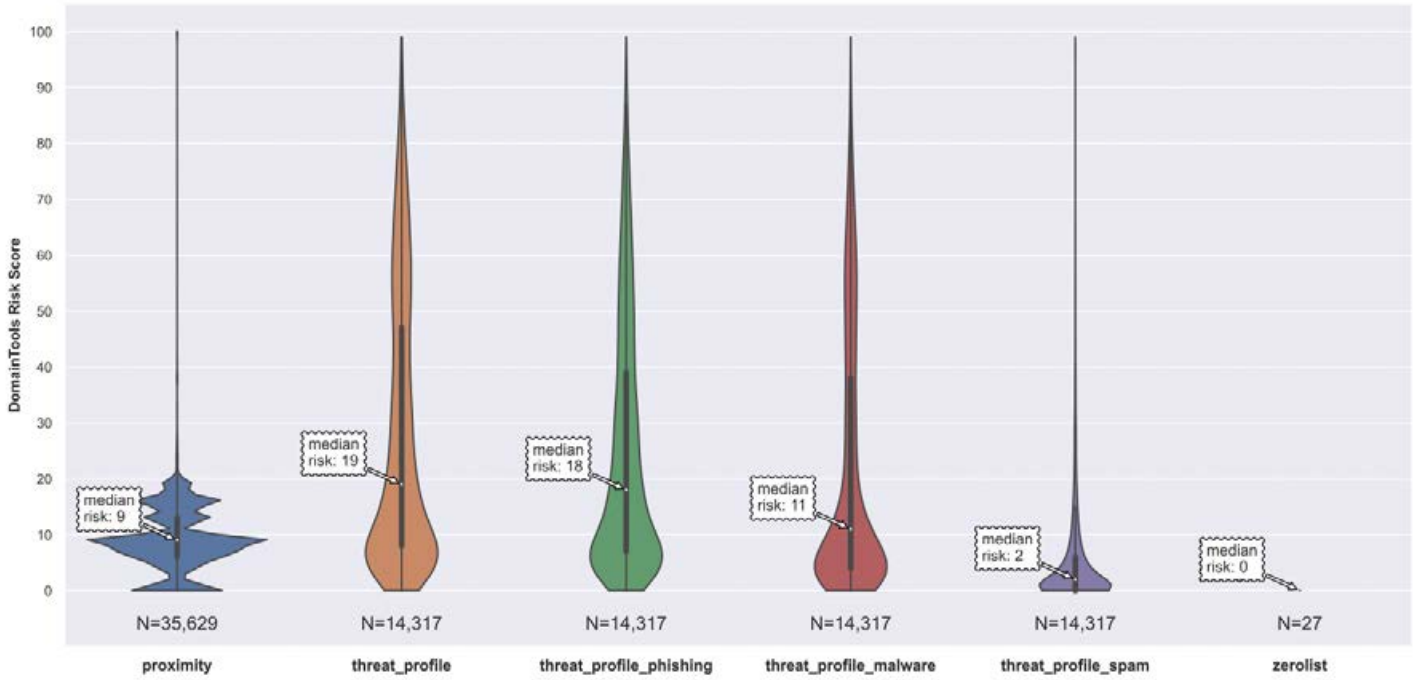
Host EU AS8972 Risk Scores Breakdown (Computed on a Subset of Domains)



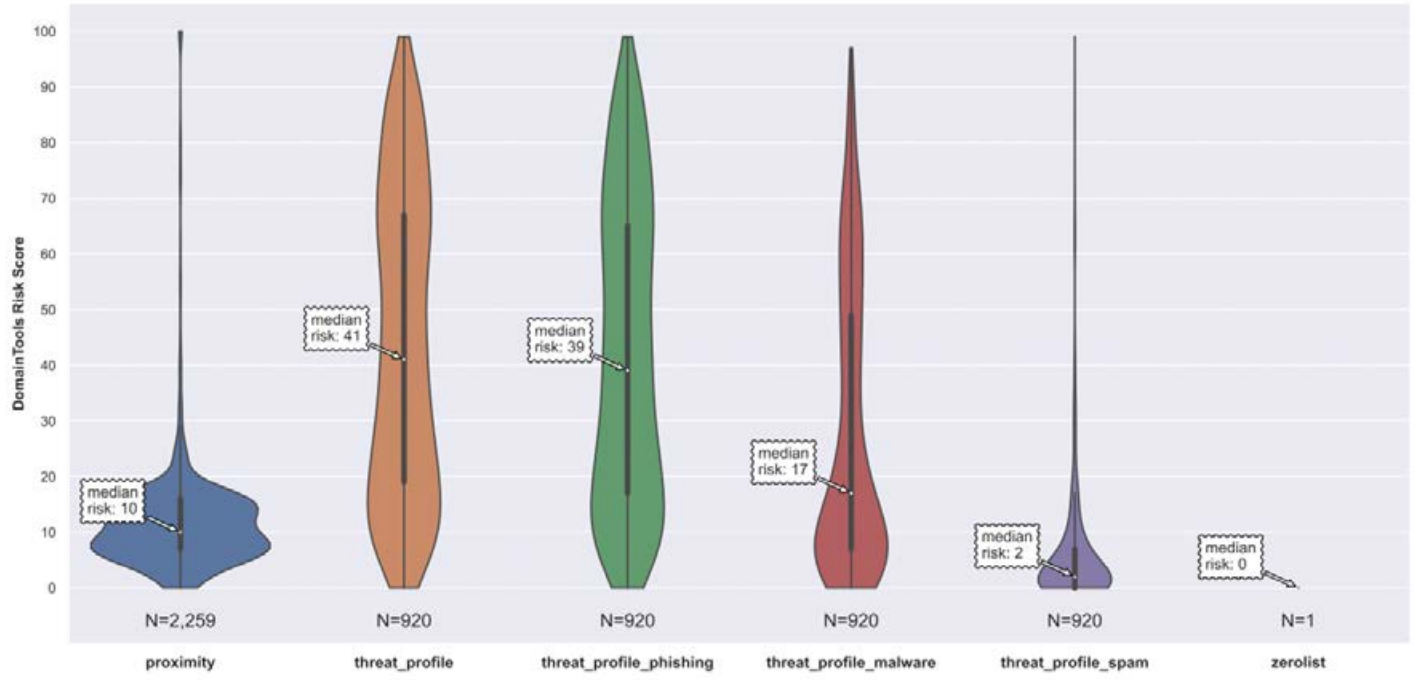
Host EU AS20738 Risk Scores Breakdown (Computed on a Subset of Domains)



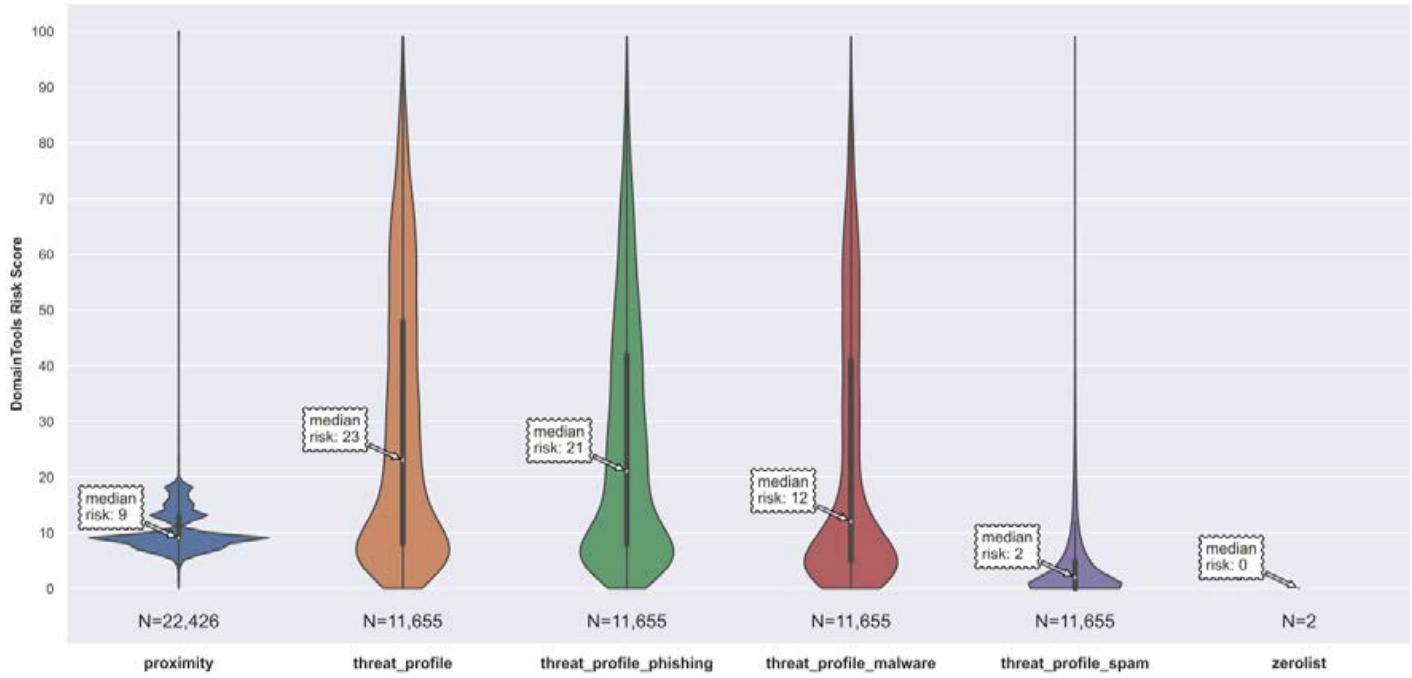
Host EU AS20773 Risk Scores Breakdown (Computed on a Subset of Domains)



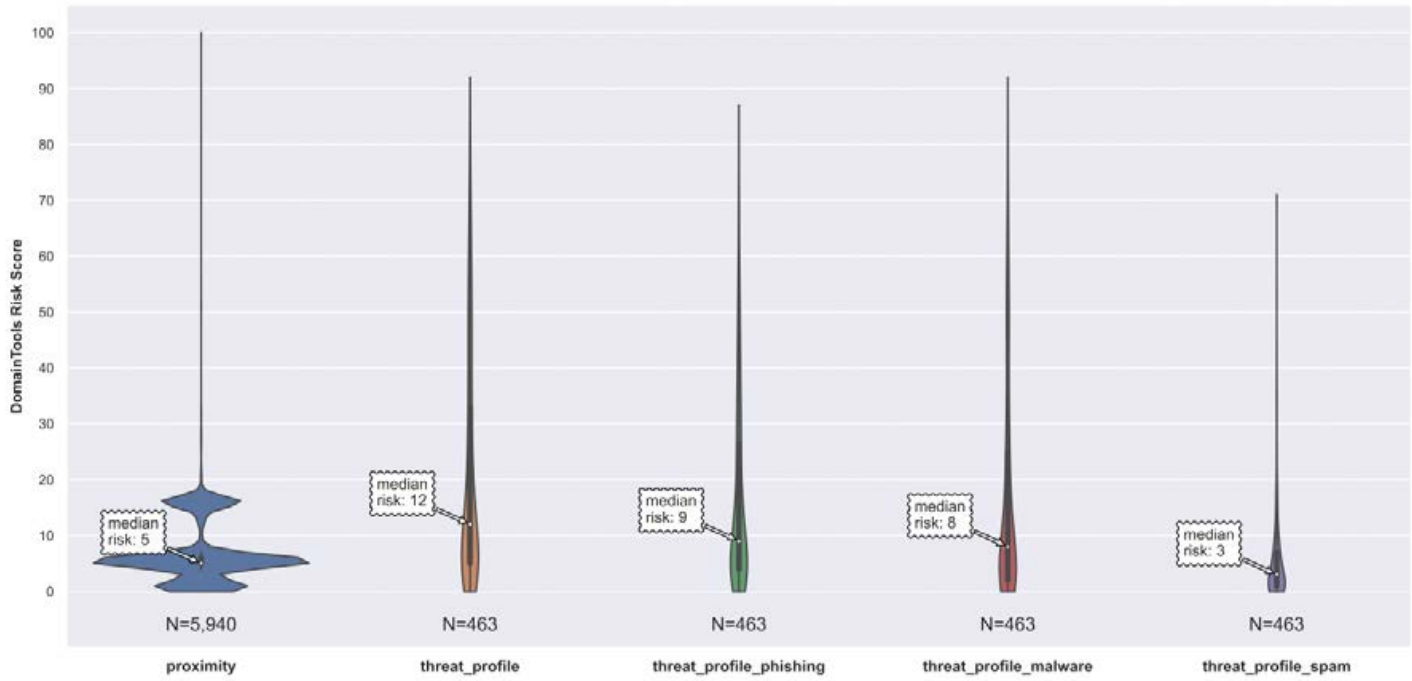
Host EU AS21499 Risk Scores Breakdown (Computed on a Subset of Domains)



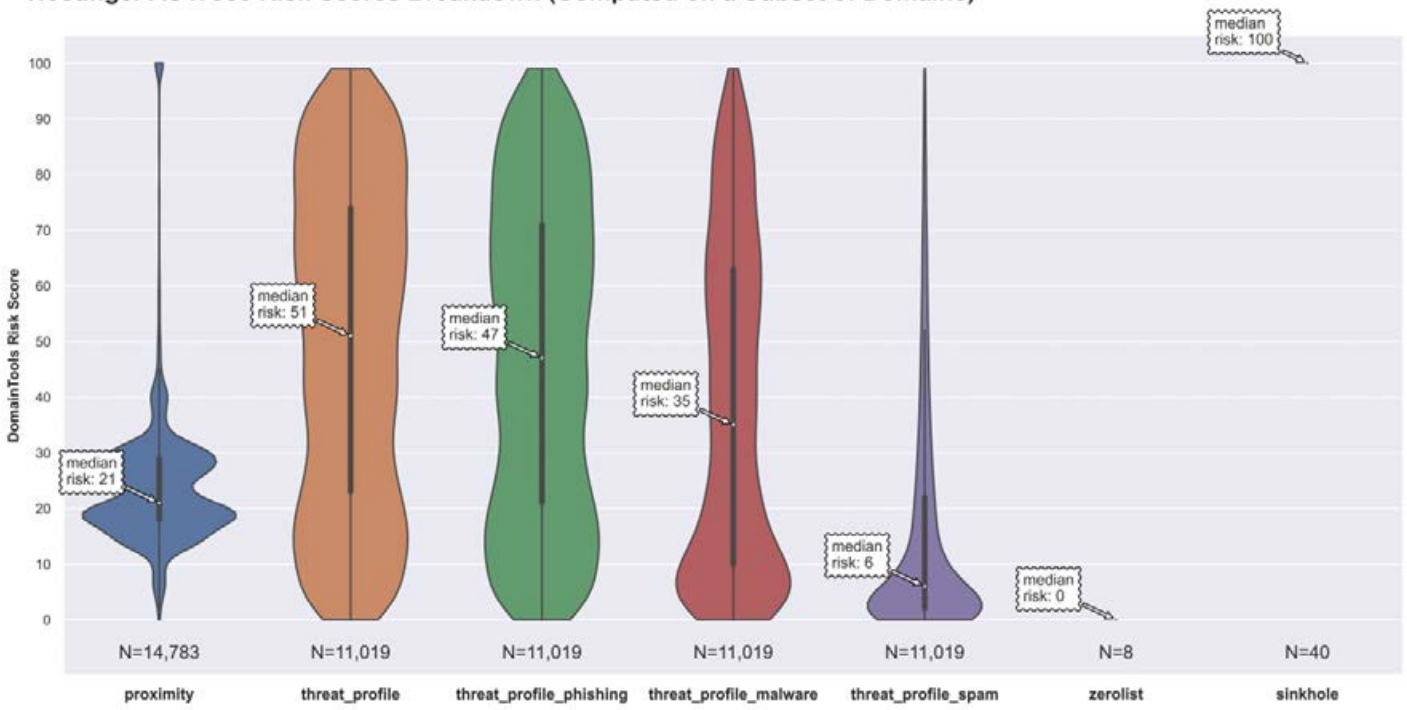
Host EU AS21501 Risk Scores Breakdown (Computed on a Subset of Domains)



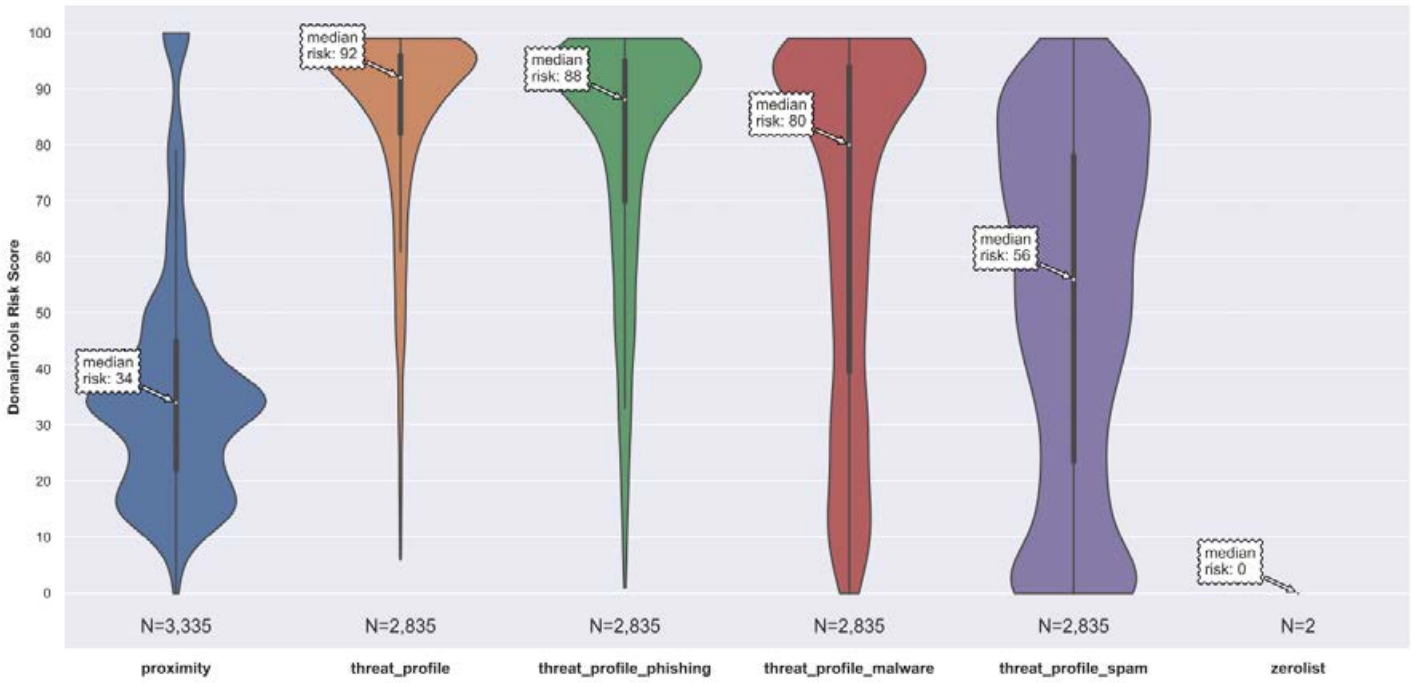
Host EU AS34011 Risk Scores Breakdown (Computed on a Subset of Domains)



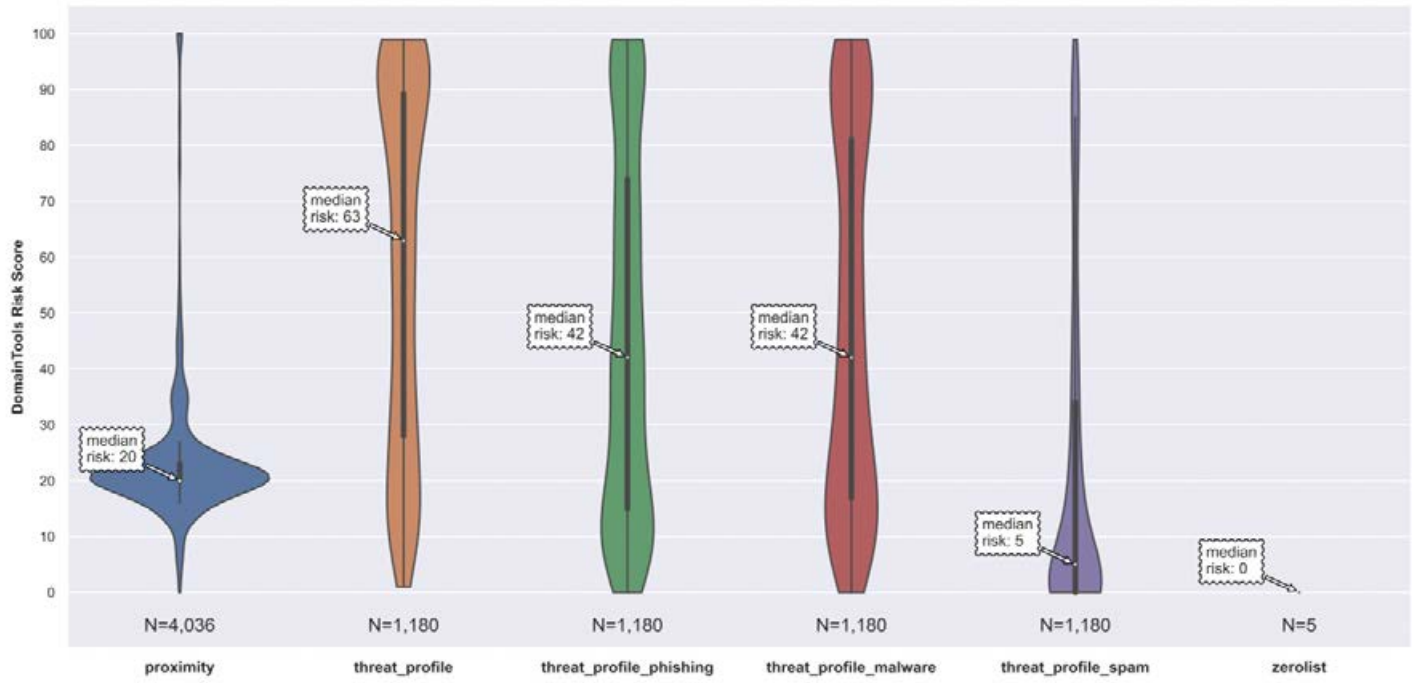
Hostinger AS47583 Risk Scores Breakdown (Computed on a Subset of Domains)



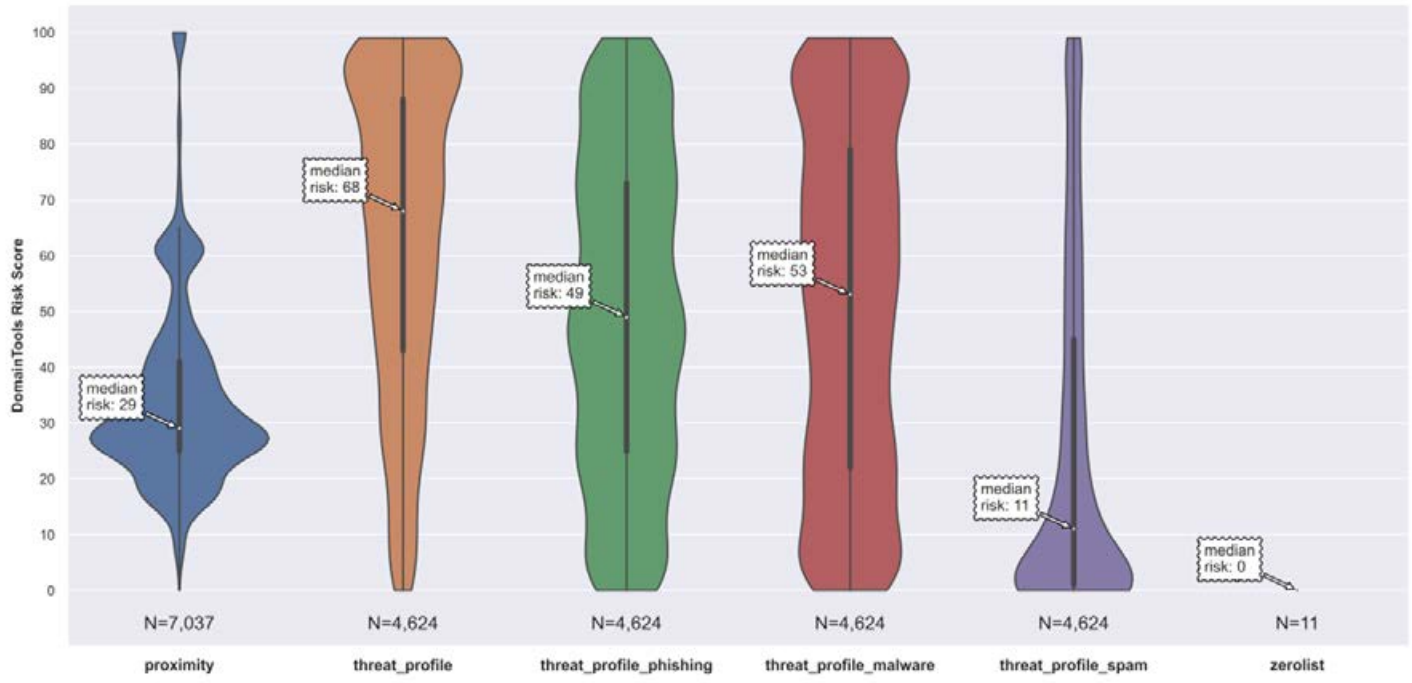
HostUS AS7489 Risk Scores Breakdown (Computed on a Subset of Domains)



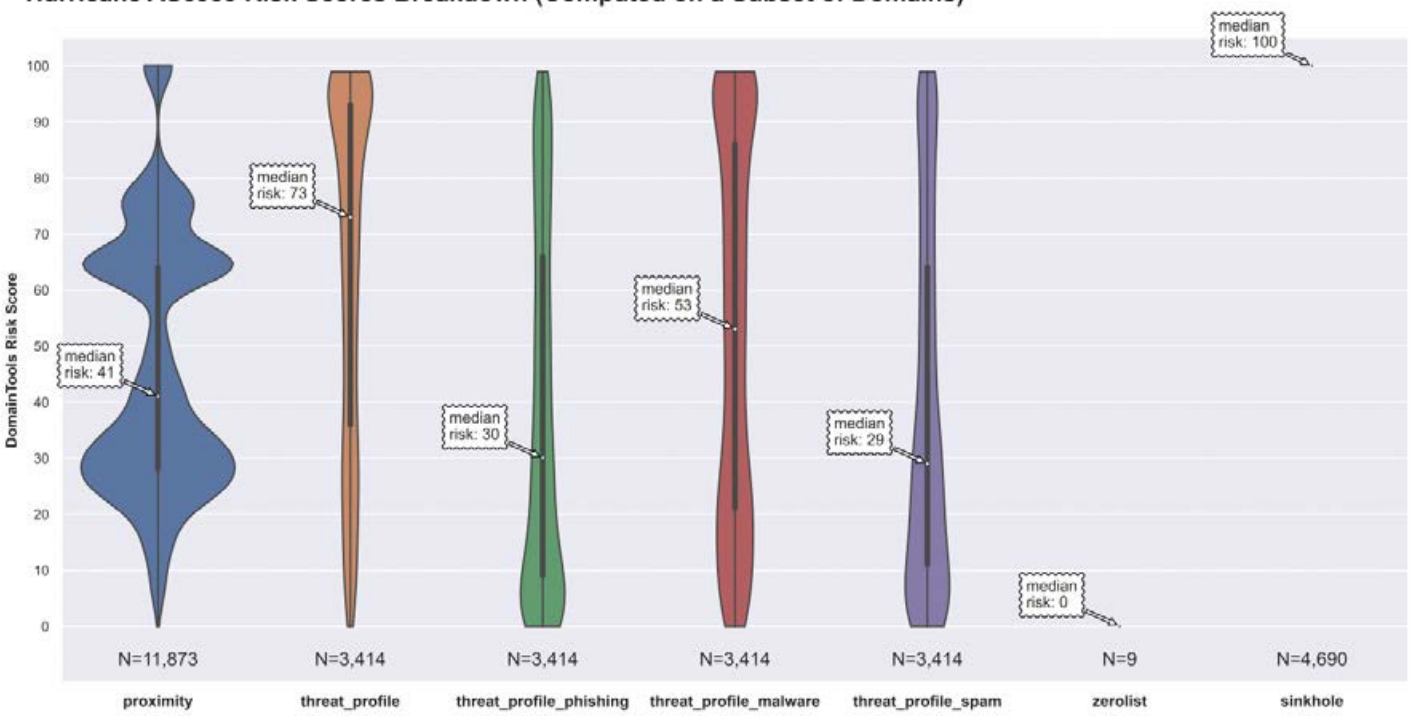
Huawei AS55990 Risk Scores Breakdown (Computed on a Subset of Domains)



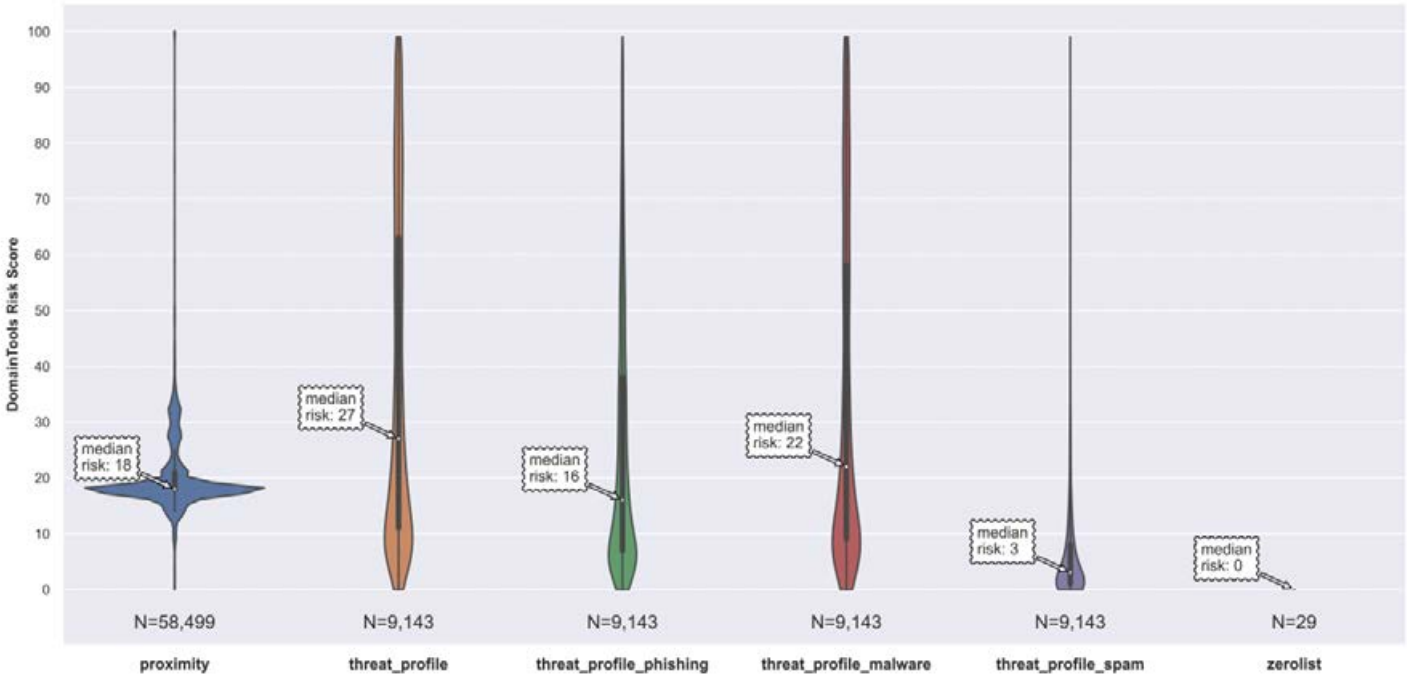
Huawei AS136907 Risk Scores Breakdown (Computed on a Subset of Domains)



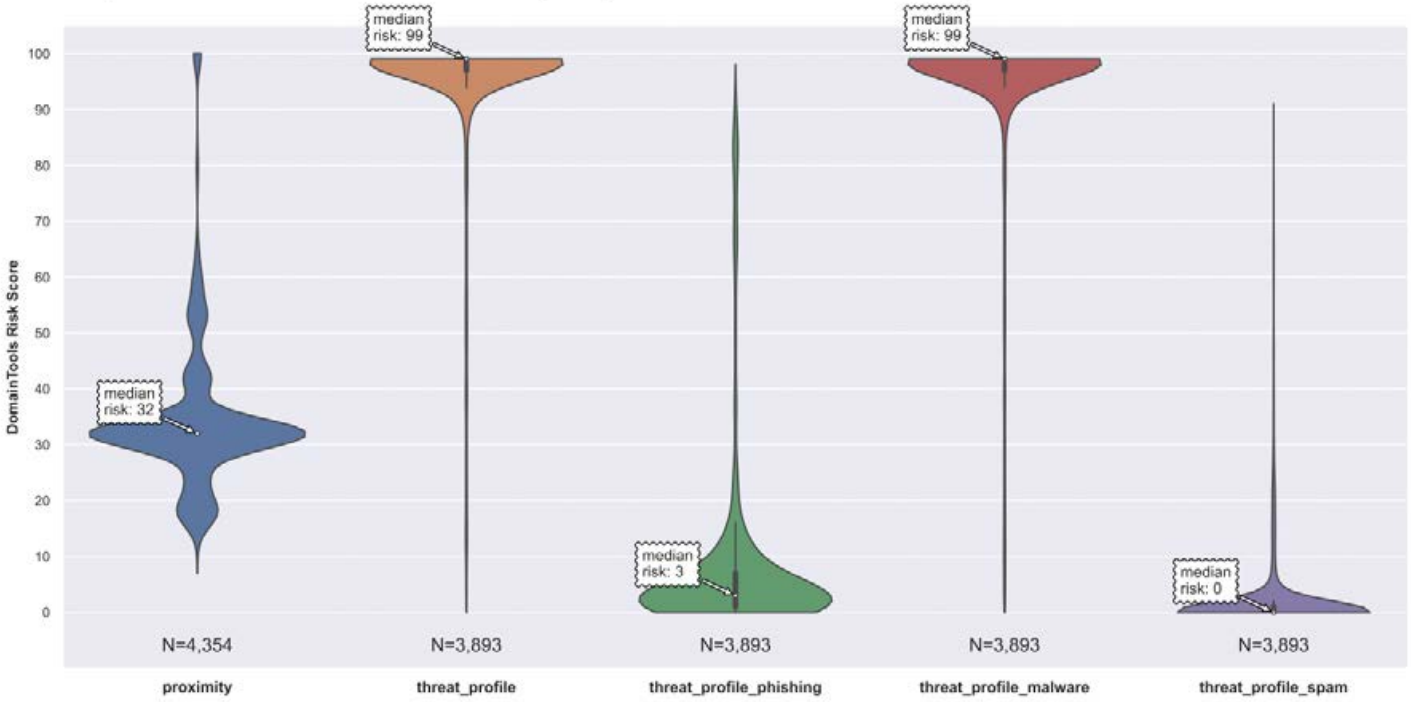
Hurricane AS6939 Risk Scores Breakdown (Computed on a Subset of Domains)



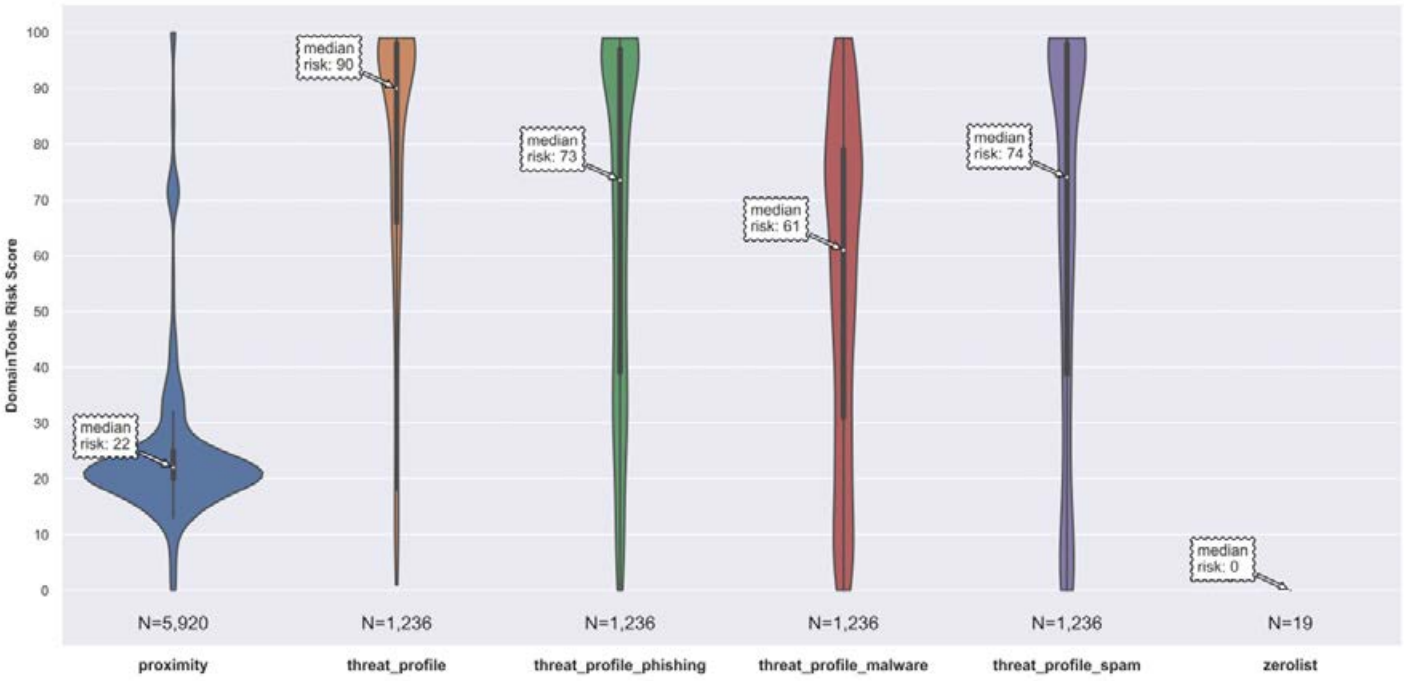
HZ Alibaba AS37963 Risk Scores Breakdown (Computed on a Subset of Domains)



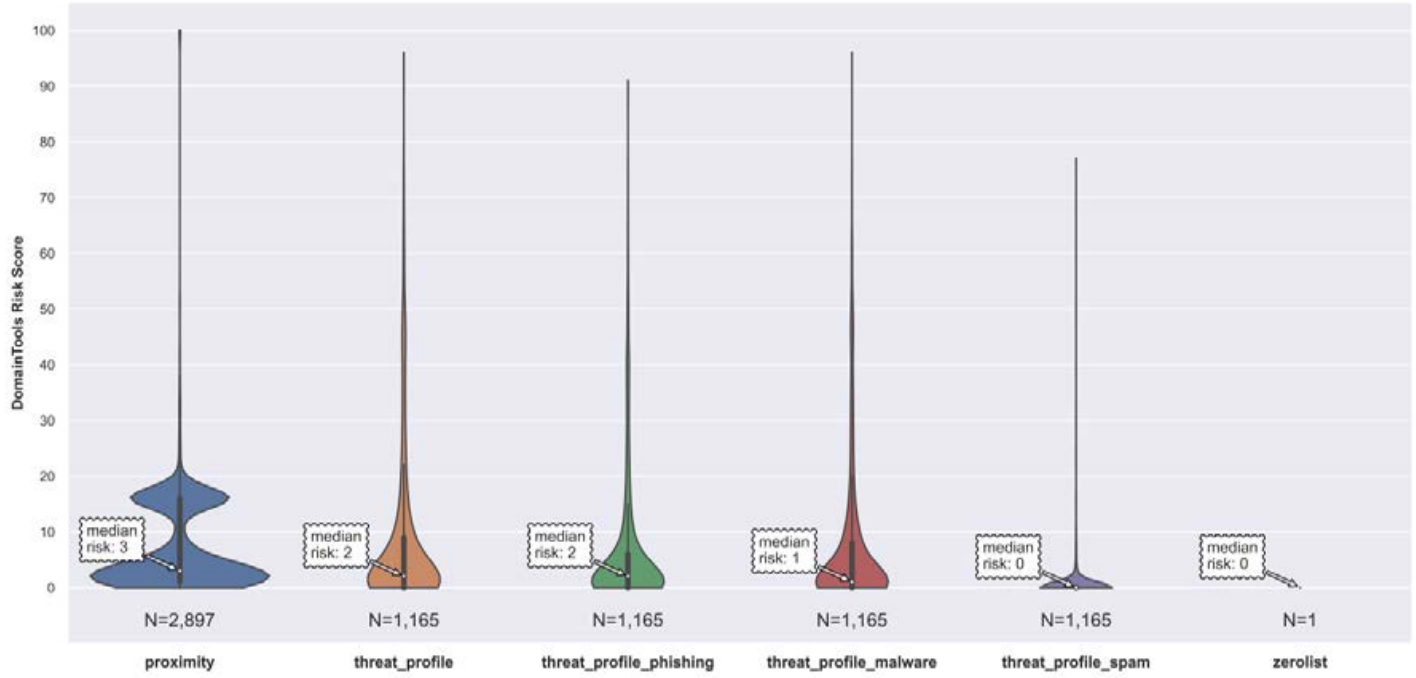
HZ Zhiyu AS59037 Risk Scores Breakdown (Computed on a Subset of Domains)



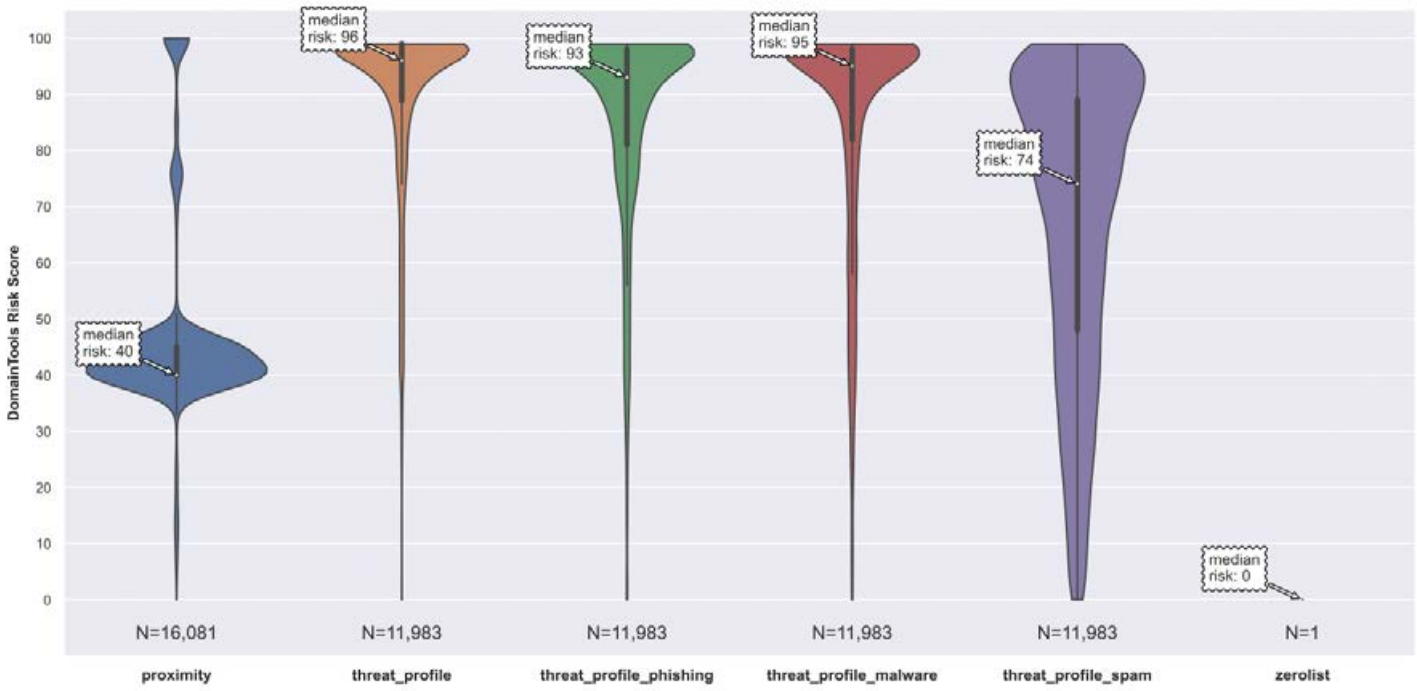
IDC Chinanet AS23724 Risk Scores Breakdown (Computed on a Subset of Domains)



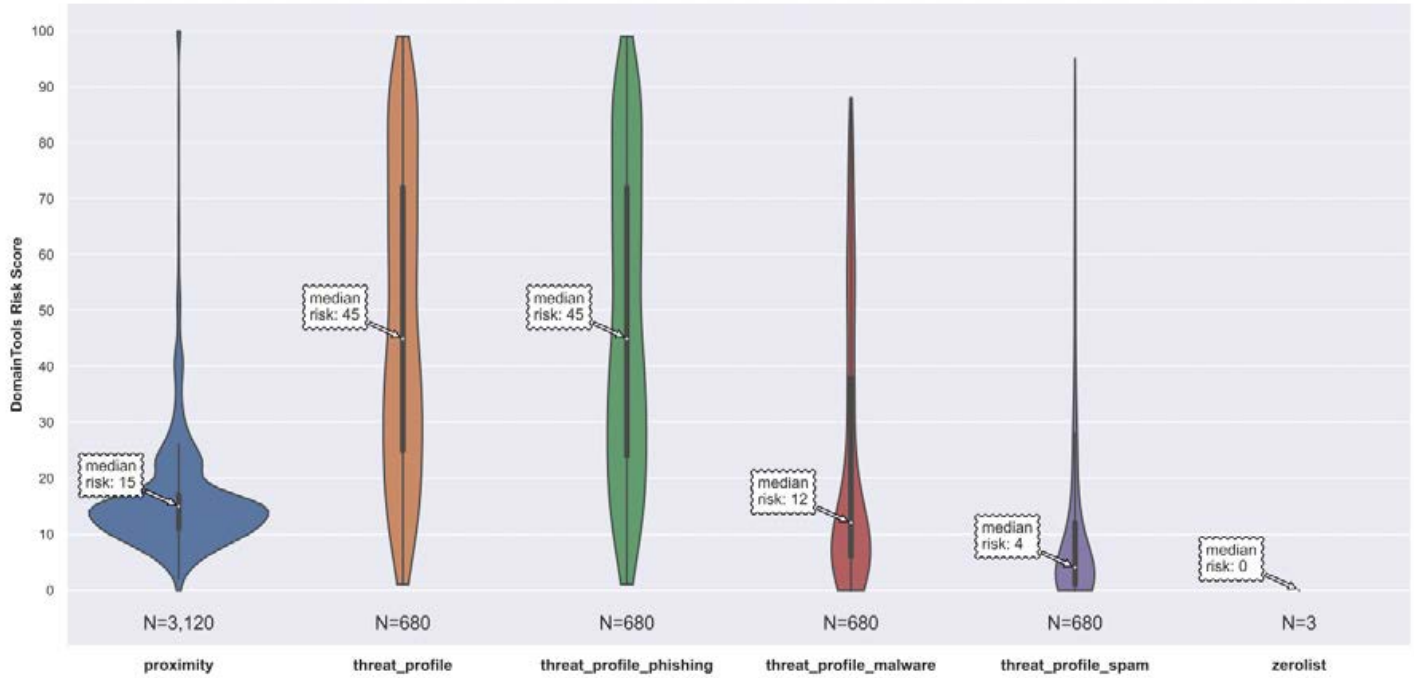
Infomaniak AS29222 Risk Scores Breakdown (Computed on a Subset of Domains)



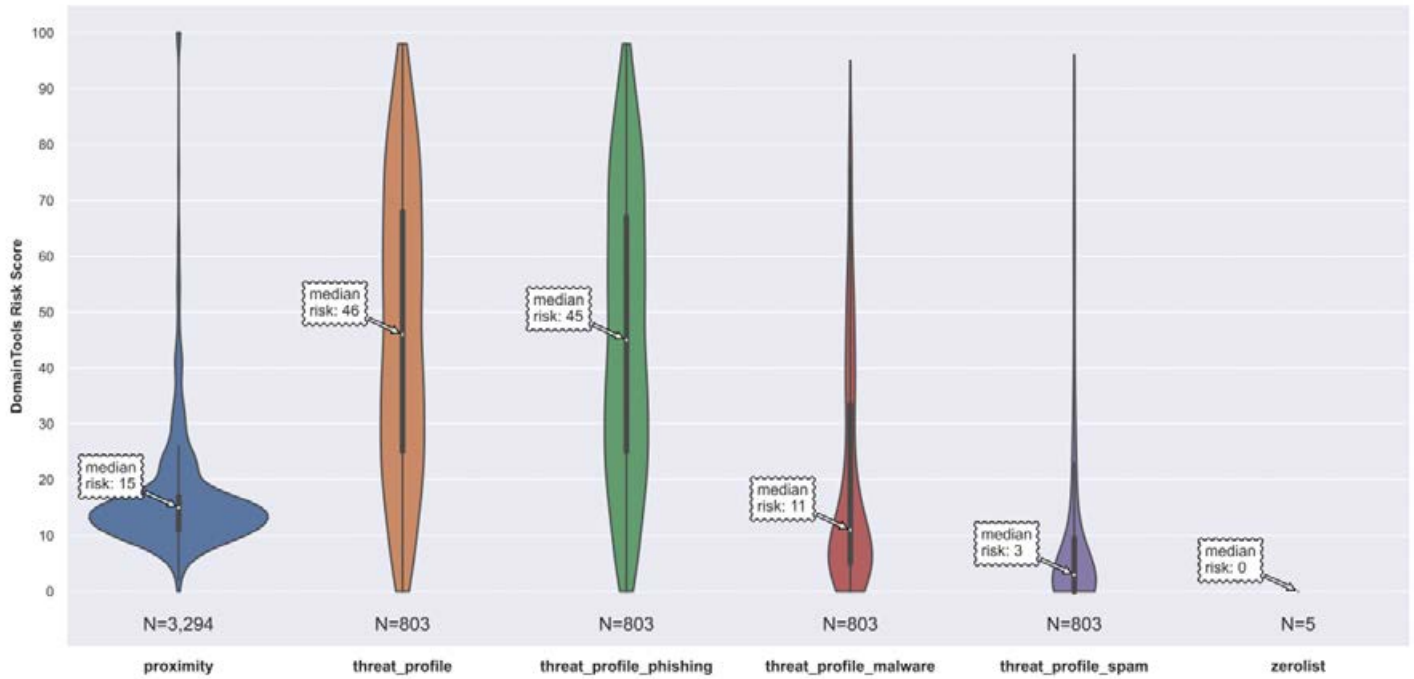
InMotion AS3842 Risk Scores Breakdown (Computed on a Subset of Domains)



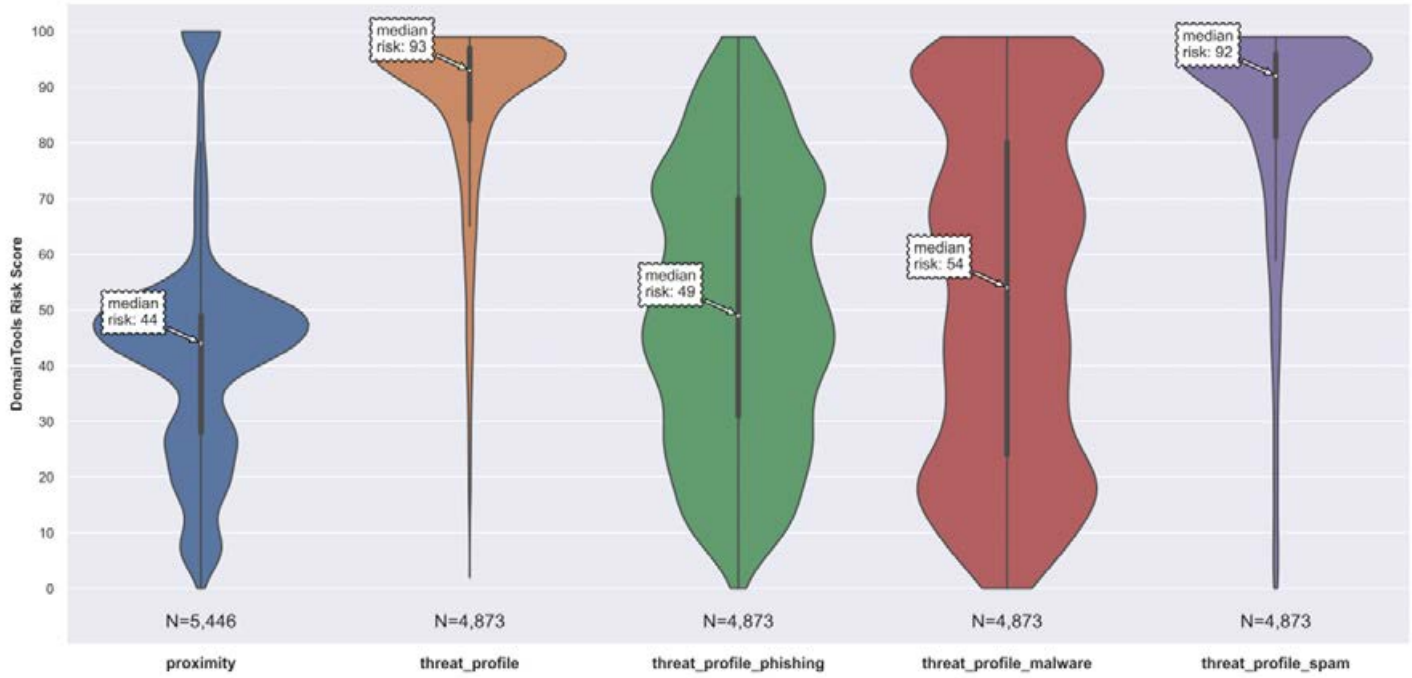
Inmotion AS22611 Risk Scores Breakdown (Computed on a Subset of Domains)



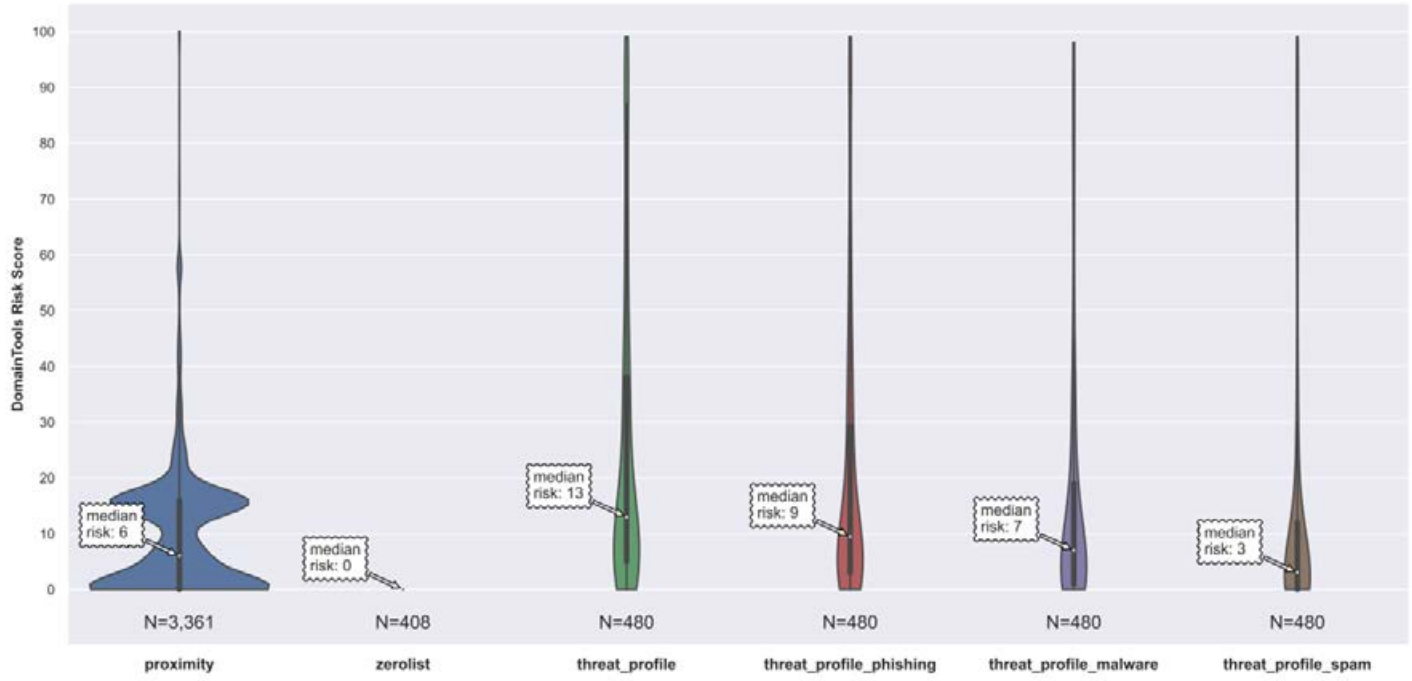
Inmotion AS54641 Risk Scores Breakdown (Computed on a Subset of Domains)



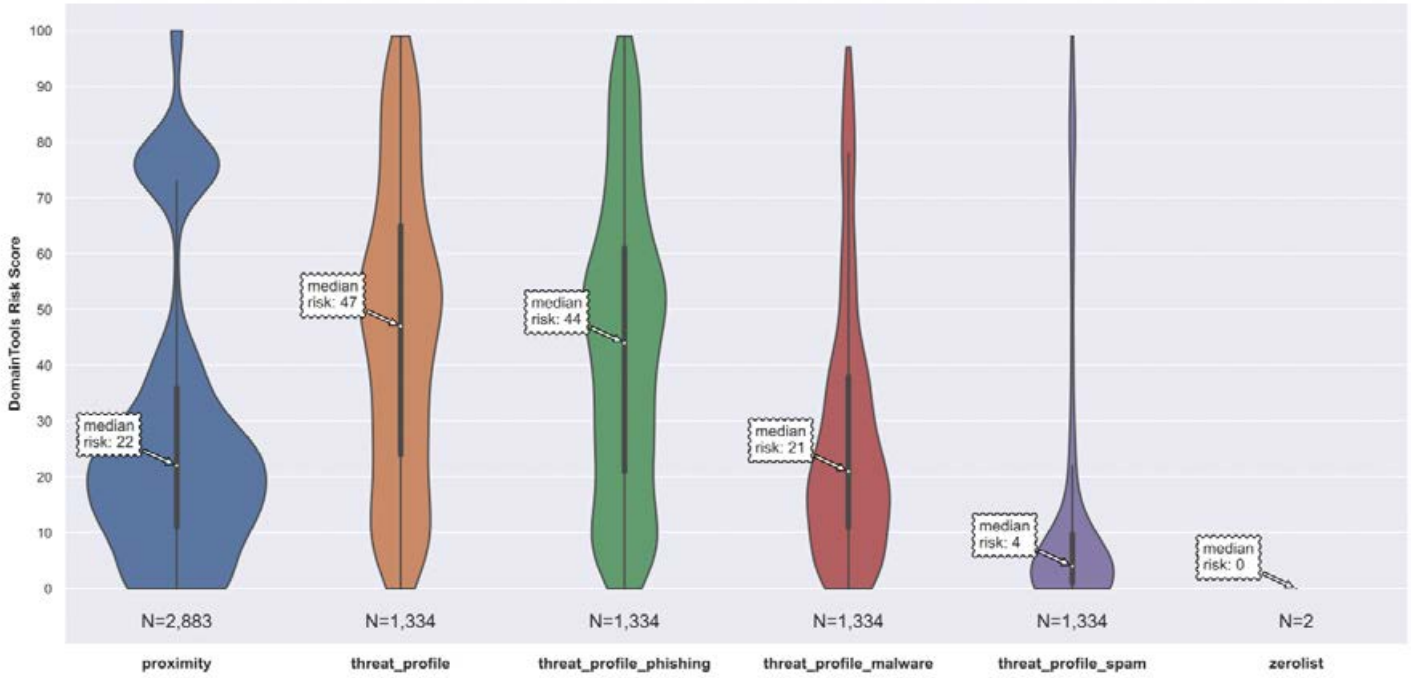
Intercontinental Internet AS398968 Risk Scores Breakdown (Computed on a Subset of Domains)



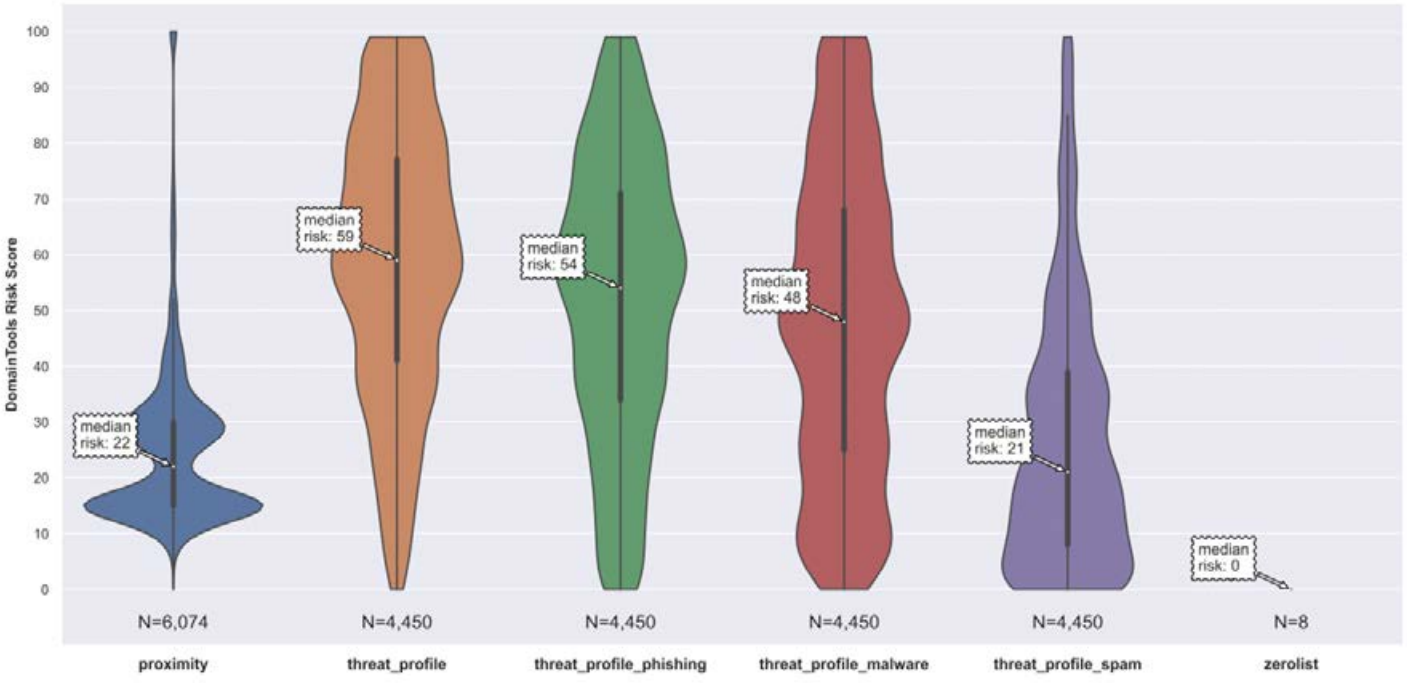
Internet Hostspace AS399674 Risk Scores Breakdown (Computed on a Subset of Domains)



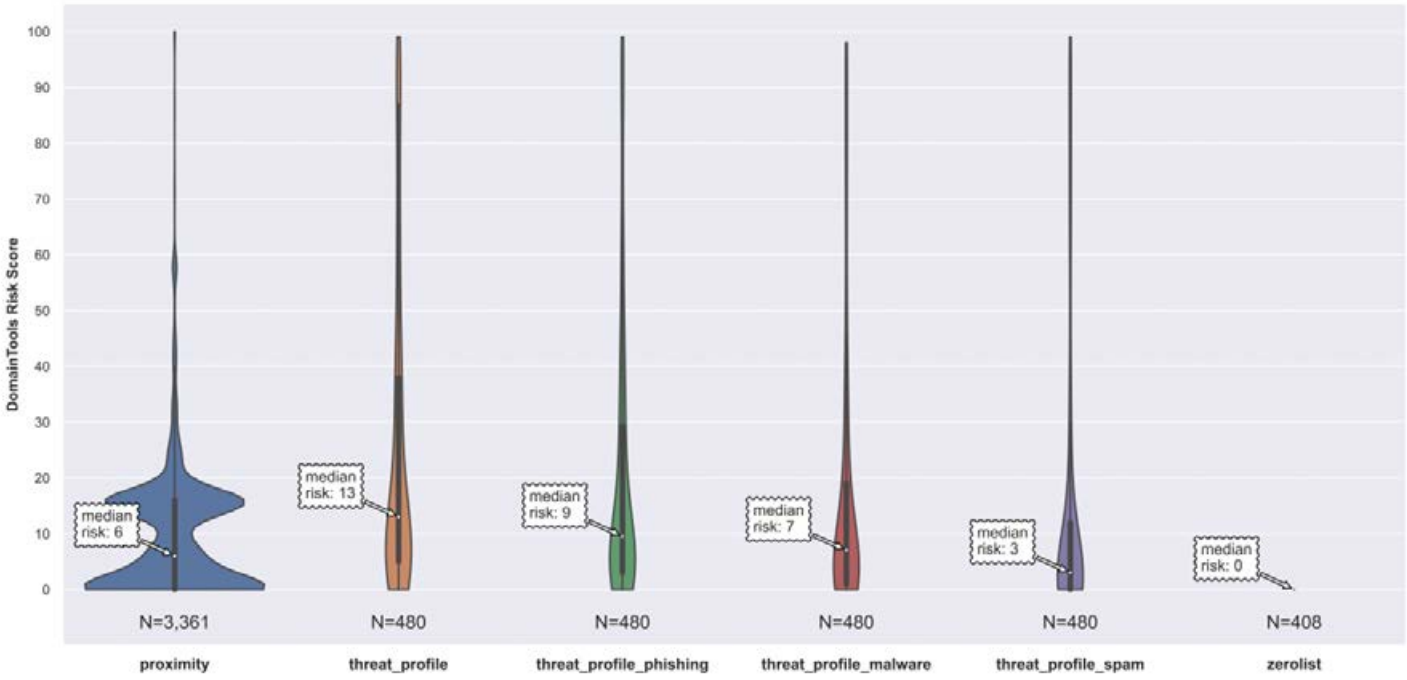
Interserver AS19318 Risk Scores Breakdown (Computed on a Subset of Domains)



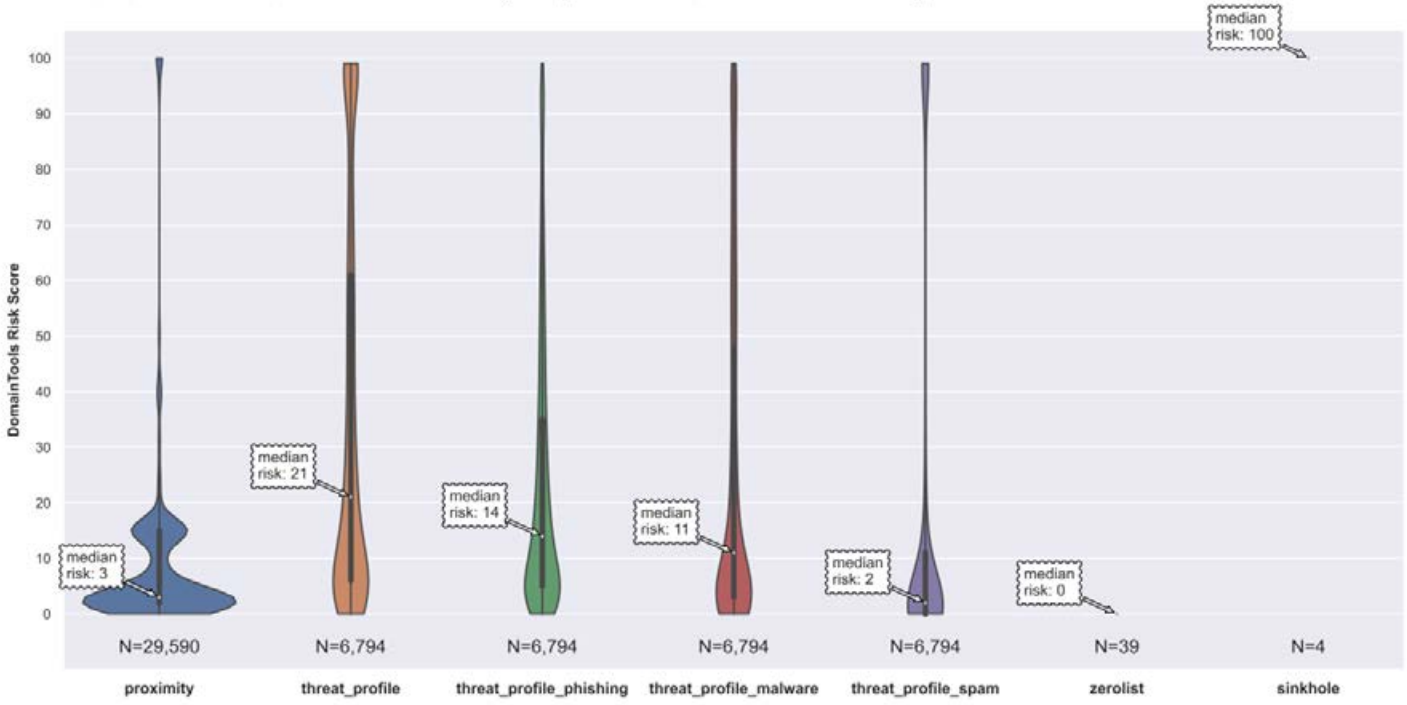
IO Flood AS53755 Risk Scores Breakdown (Computed on a Subset of Domains)



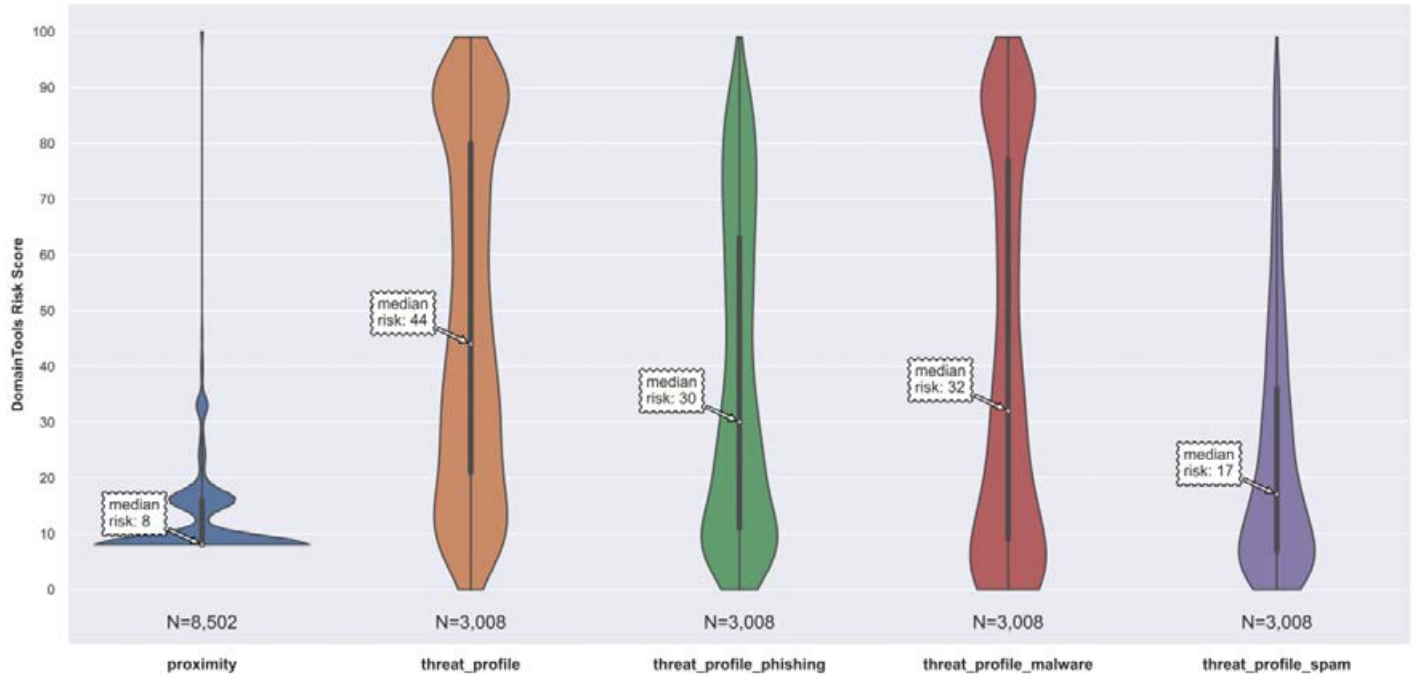
IO Mart AS20860 Risk Scores Breakdown (Computed on a Subset of Domains)



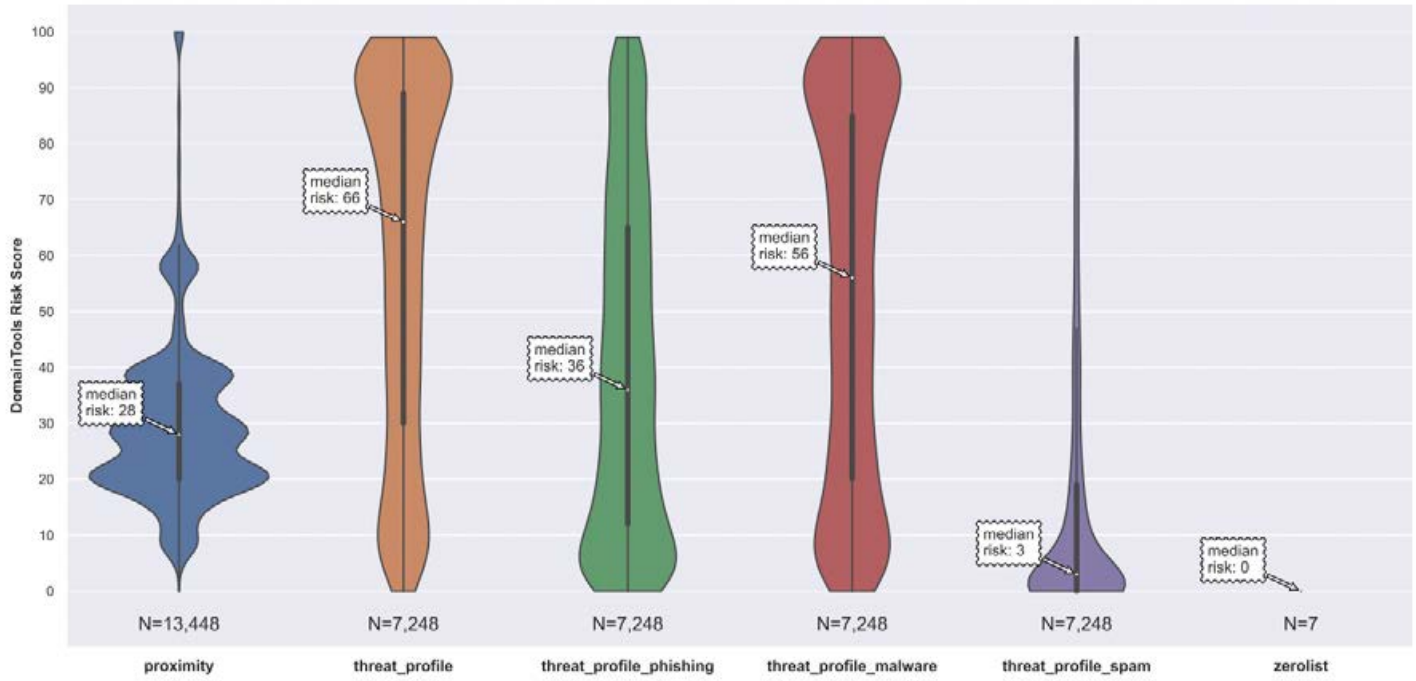
IONOS AS8560 Risk Scores Breakdown (Computed on a Subset of Domains)



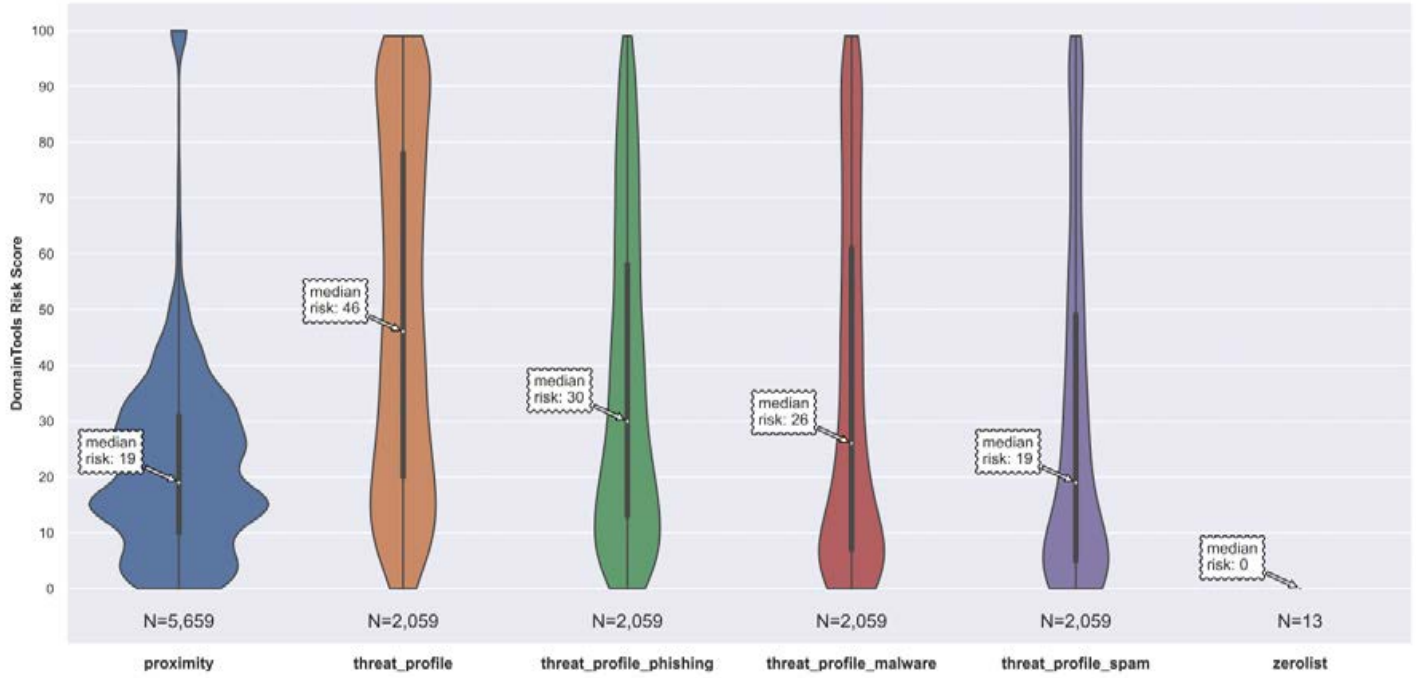
IP Volume AS58110 Risk Scores Breakdown (Computed on a Subset of Domains)



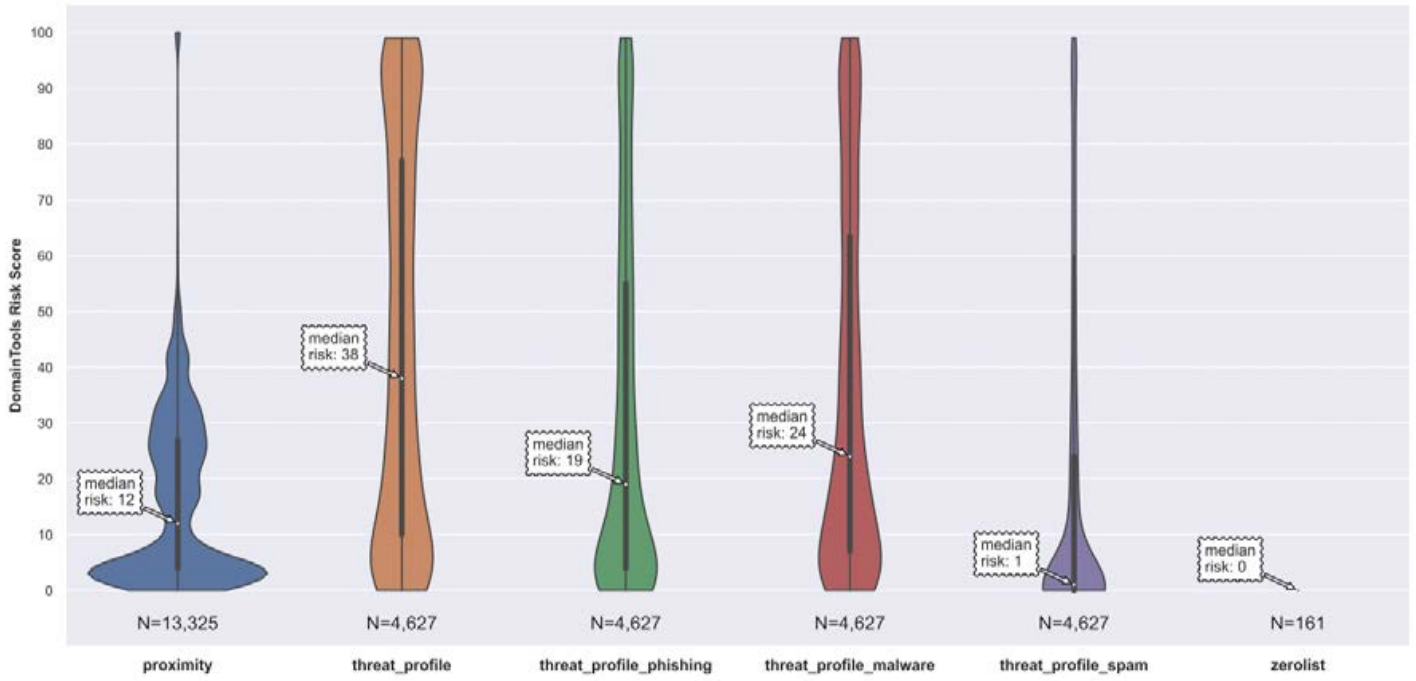
IT7 Networks AS25820 Risk Scores Breakdown (Computed on a Subset of Domains)



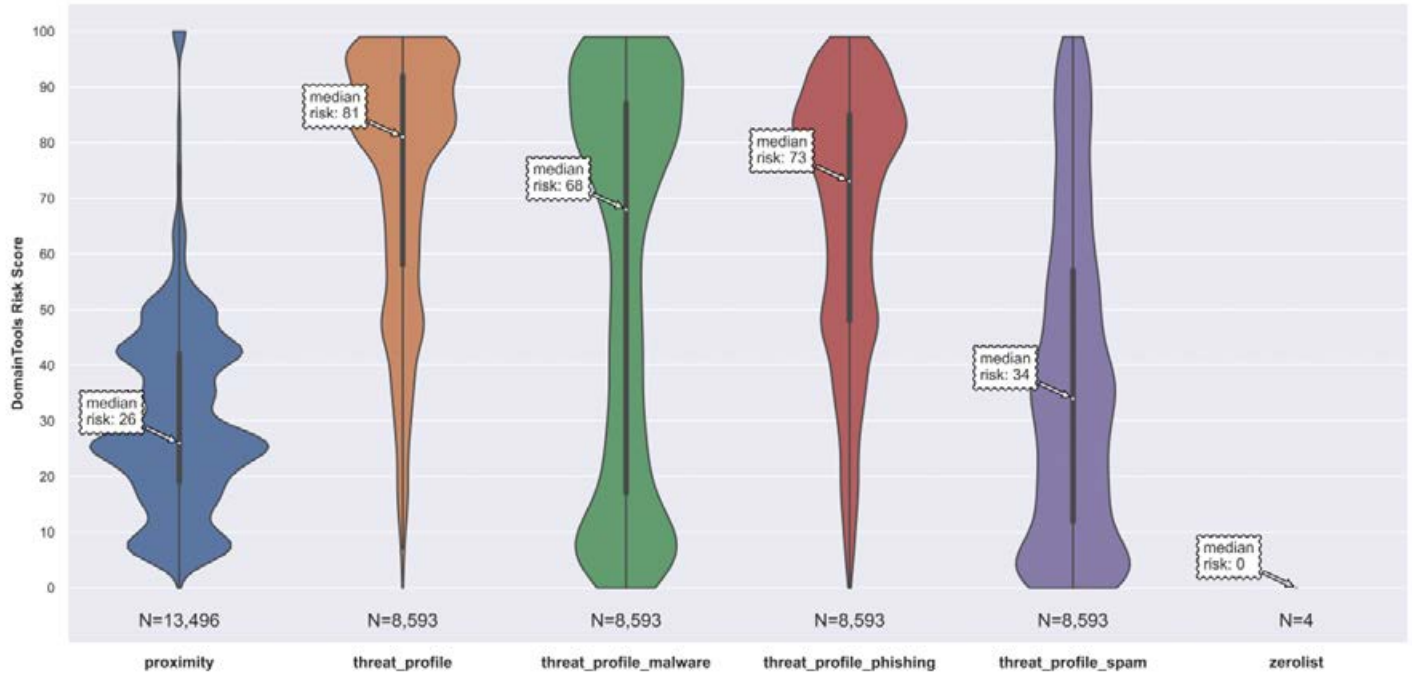
JSC IOT AS29182 Risk Scores Breakdown (Computed on a Subset of Domains)



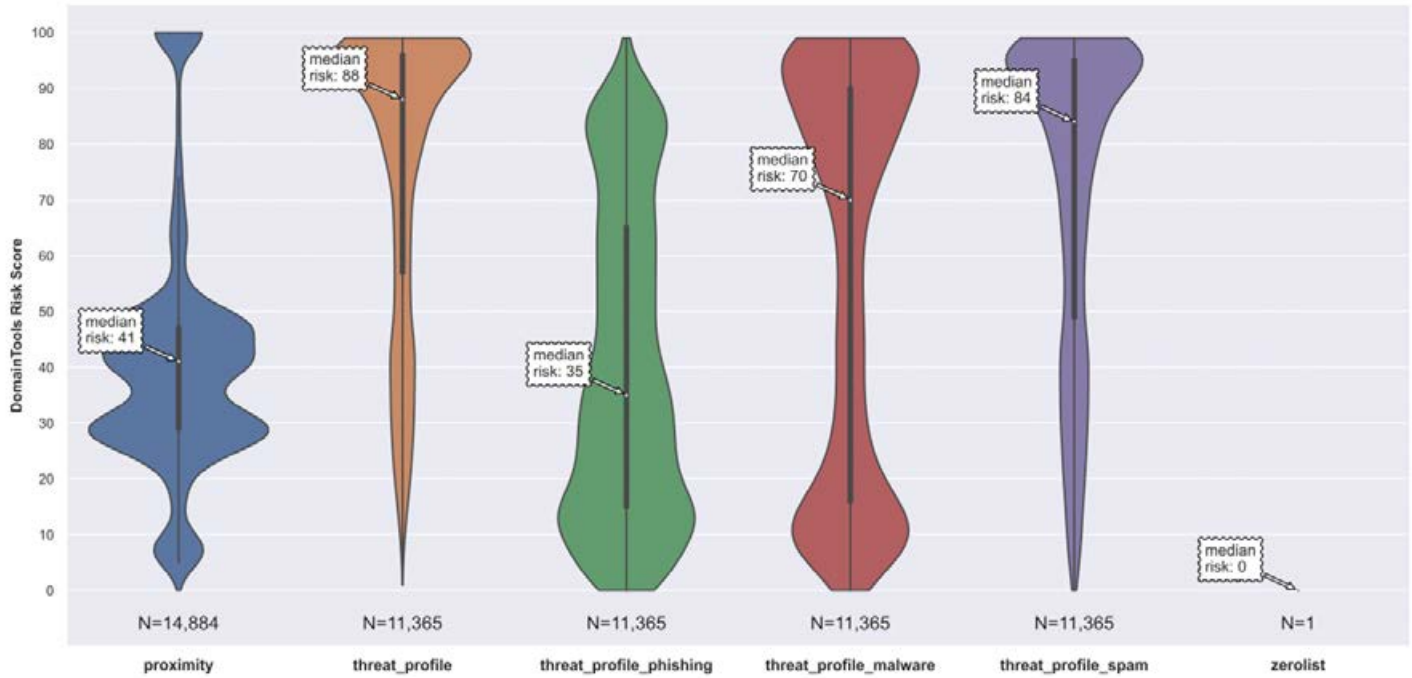
Korea Telecom AS4766 Risk Scores Breakdown (Computed on a Subset of Domains)



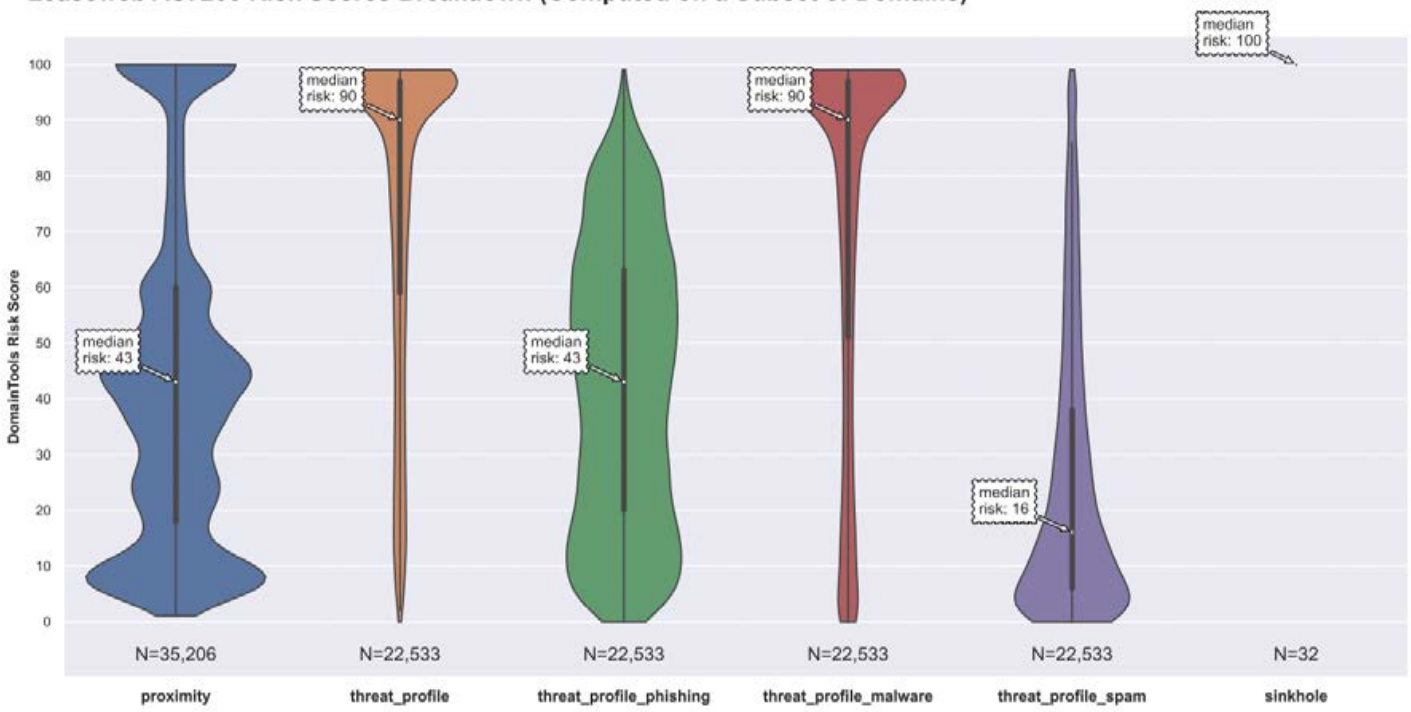
Krypt-AS35908 Risk Scores Breakdown (Computed on a Subset of Domains)



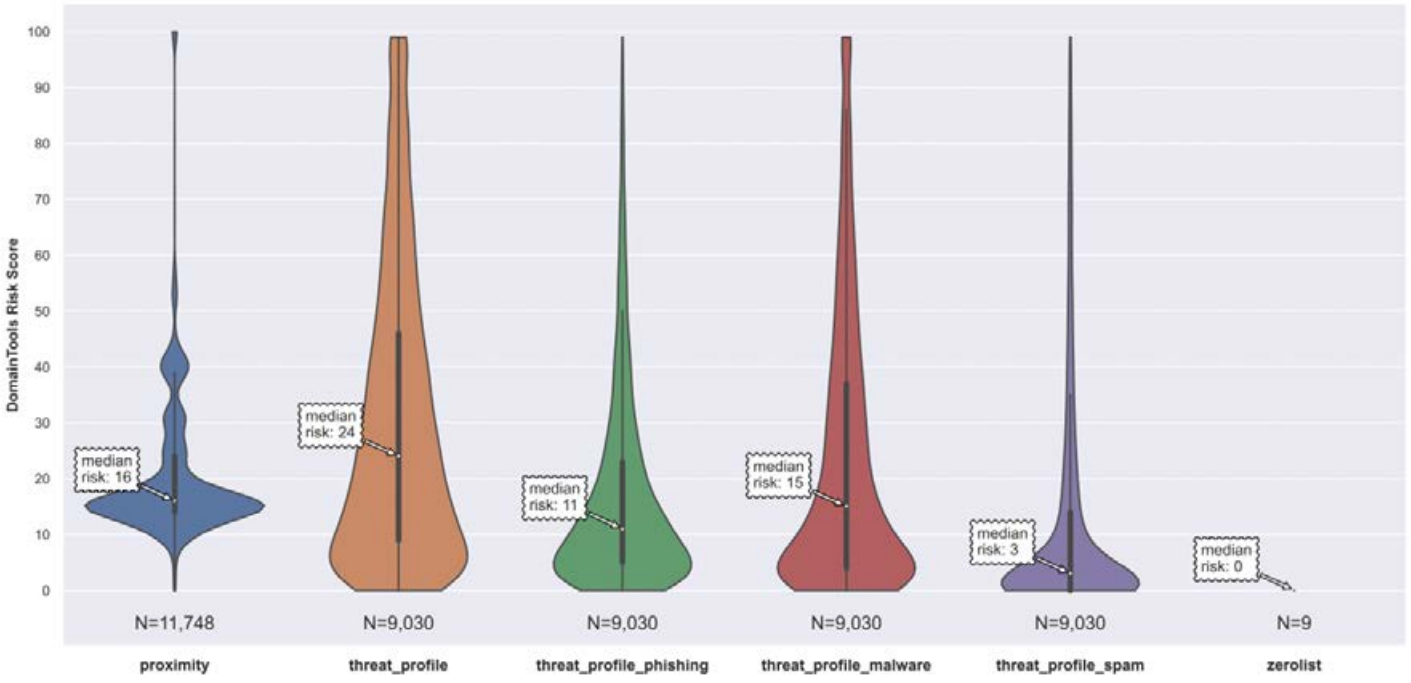
LayerHost AS46573 Risk Scores Breakdown (Computed on a Subset of Domains)



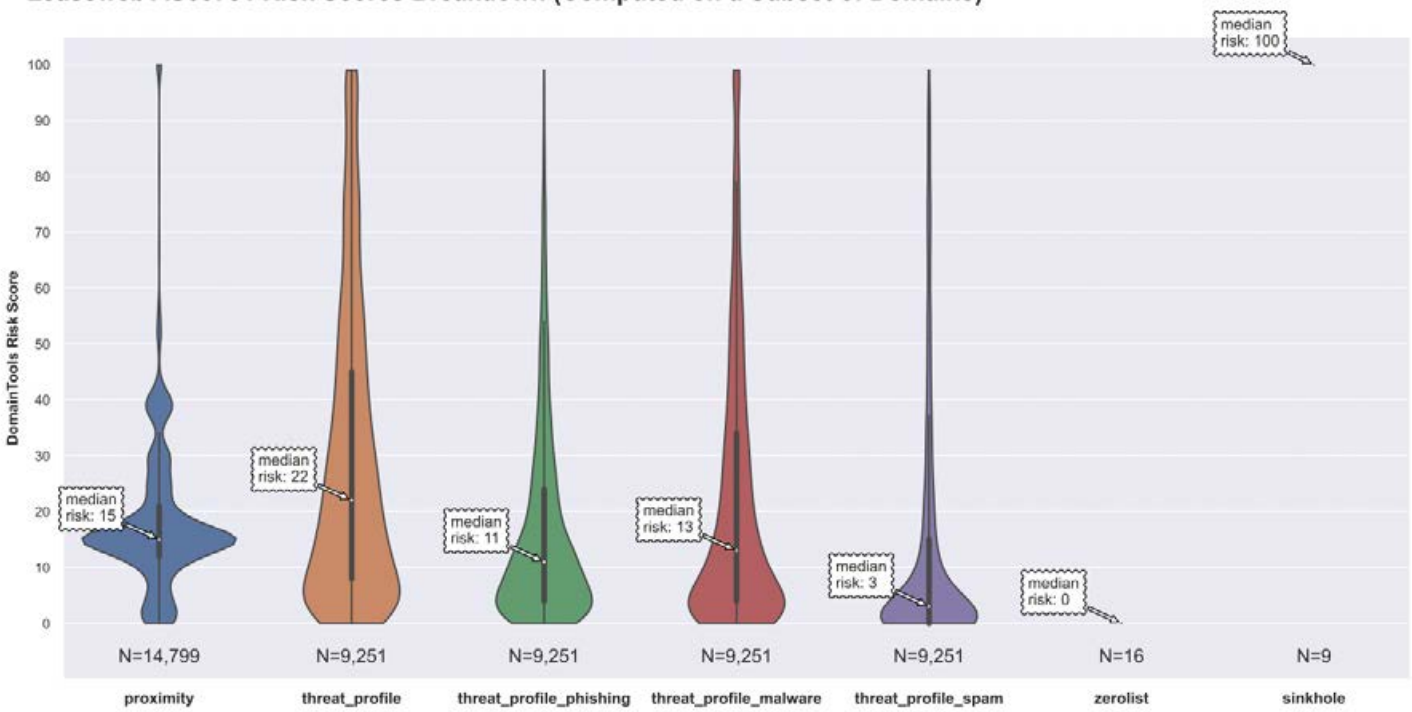
Leaseweb AS7203 Risk Scores Breakdown (Computed on a Subset of Domains)



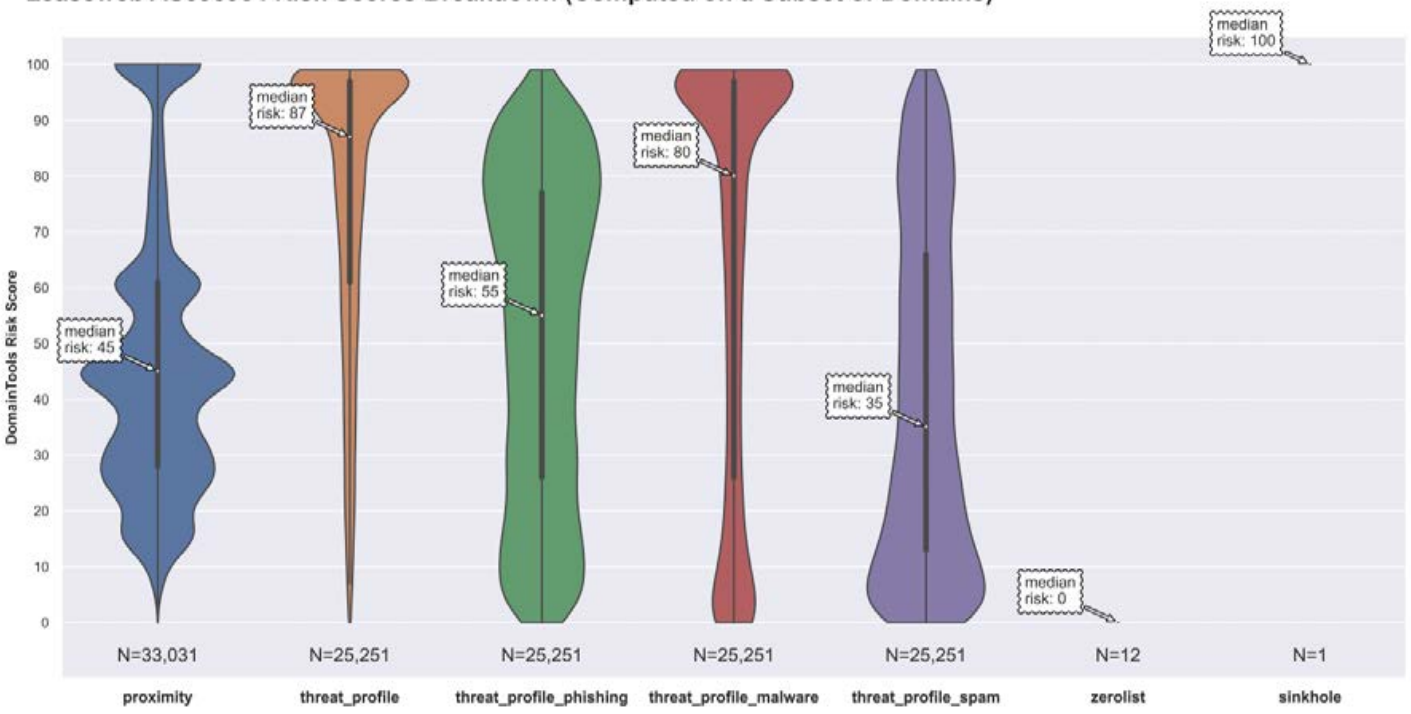
Leaseweb AS30633 Risk Scores Breakdown (Computed on a Subset of Domains)



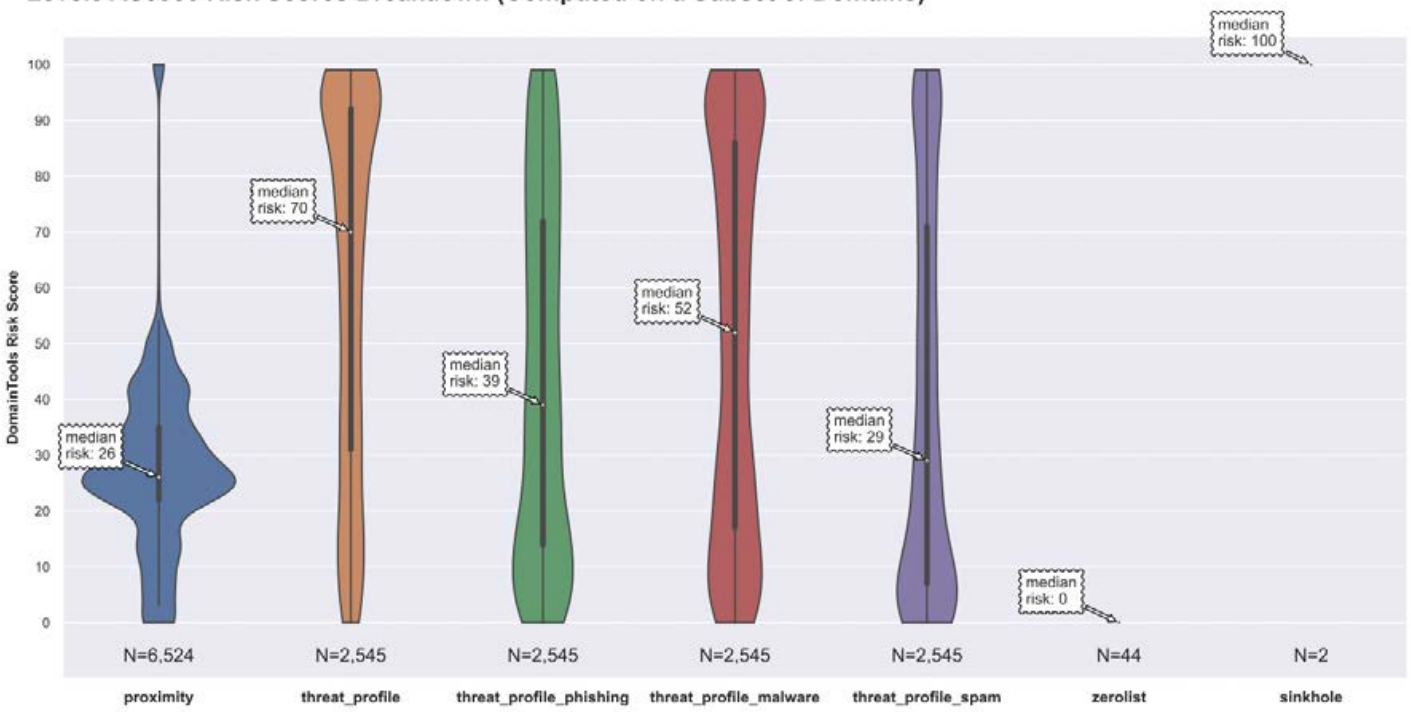
Leaseweb AS60781 Risk Scores Breakdown (Computed on a Subset of Domains)



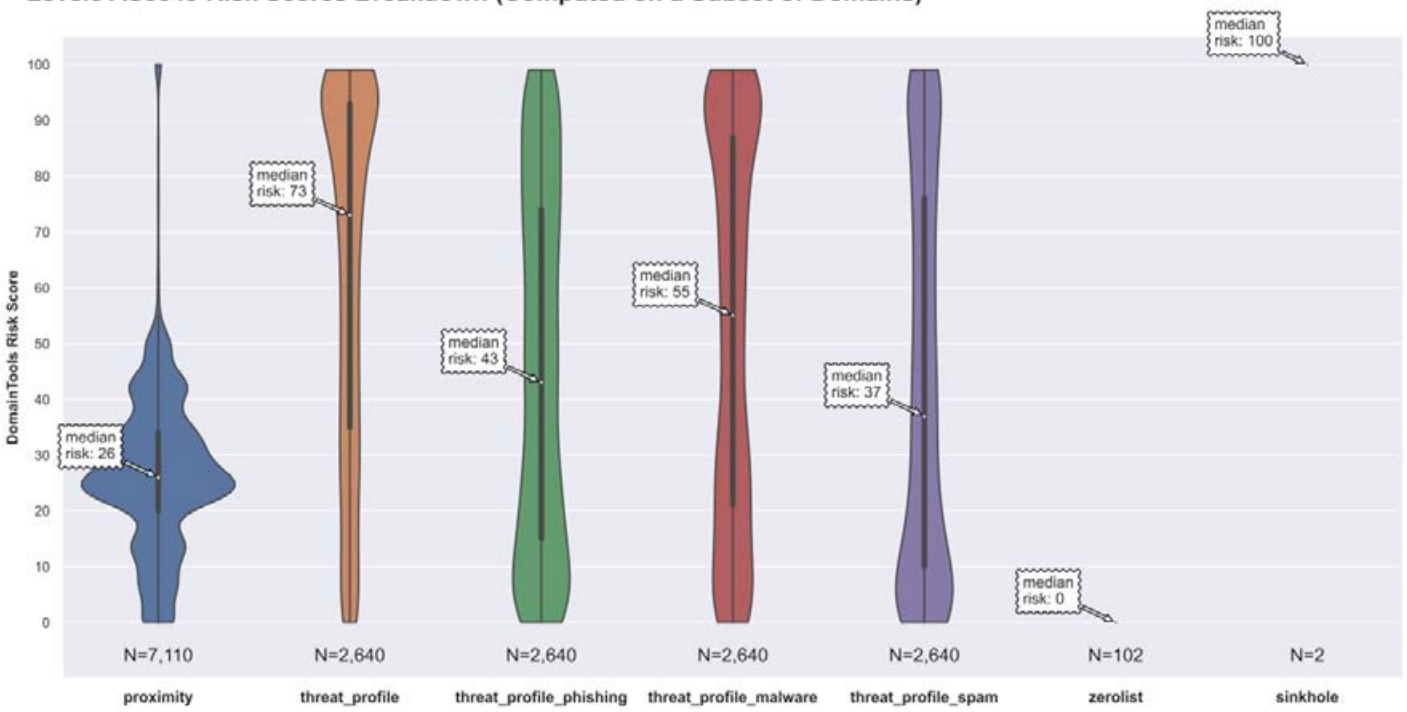
Leaseweb AS395954 Risk Scores Breakdown (Computed on a Subset of Domains)



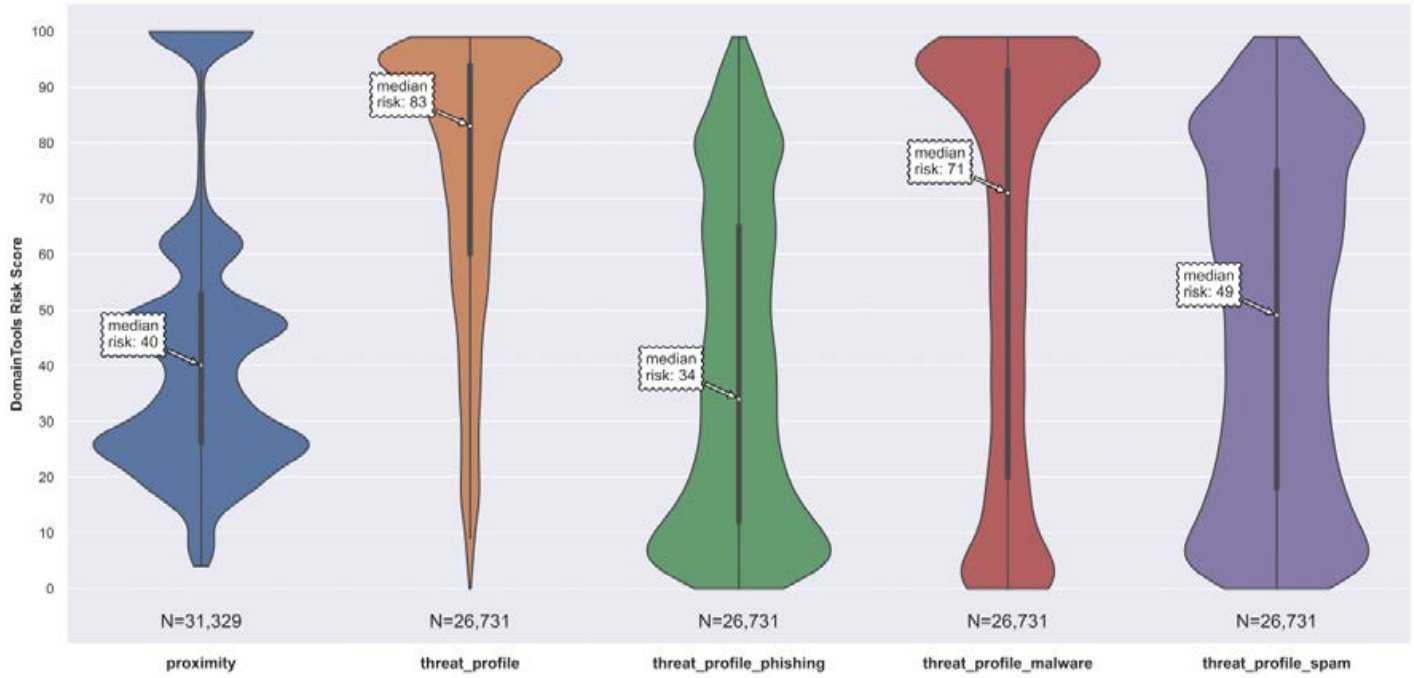
Level3 AS3356 Risk Scores Breakdown (Computed on a Subset of Domains)



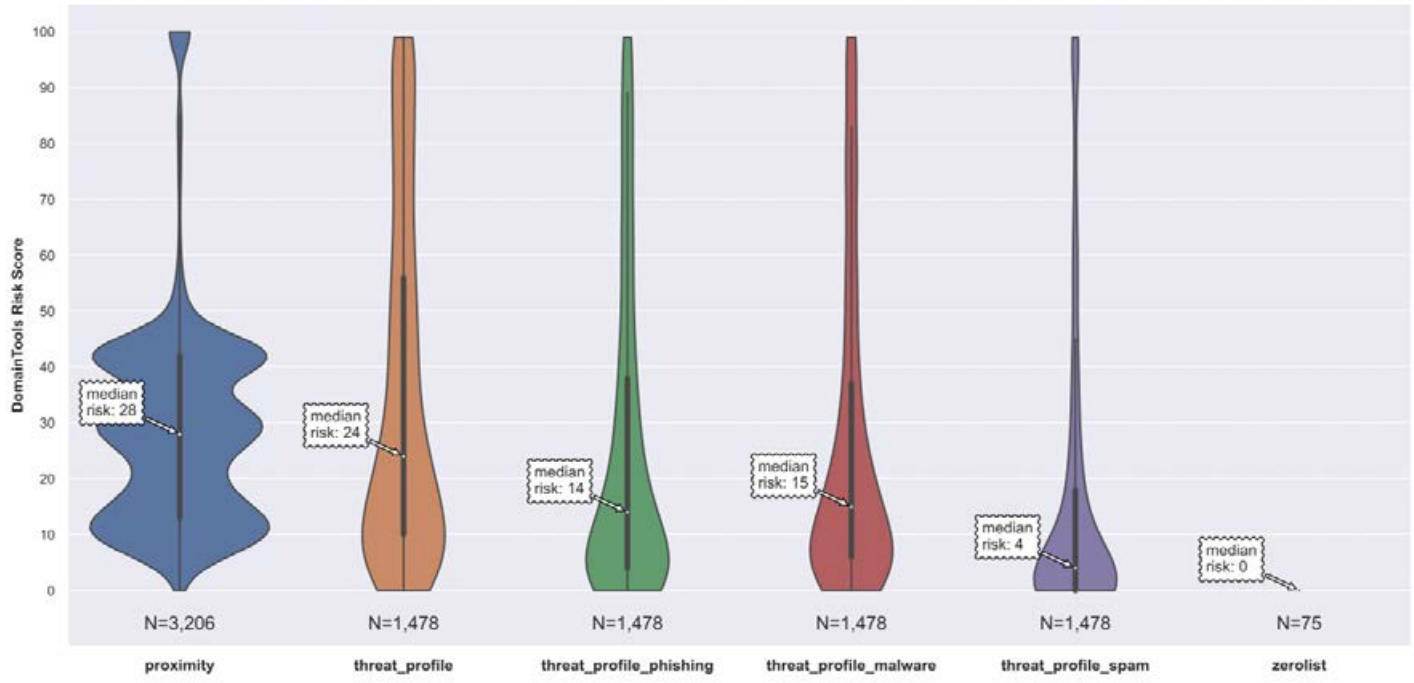
Level3 AS3549 Risk Scores Breakdown (Computed on a Subset of Domains)



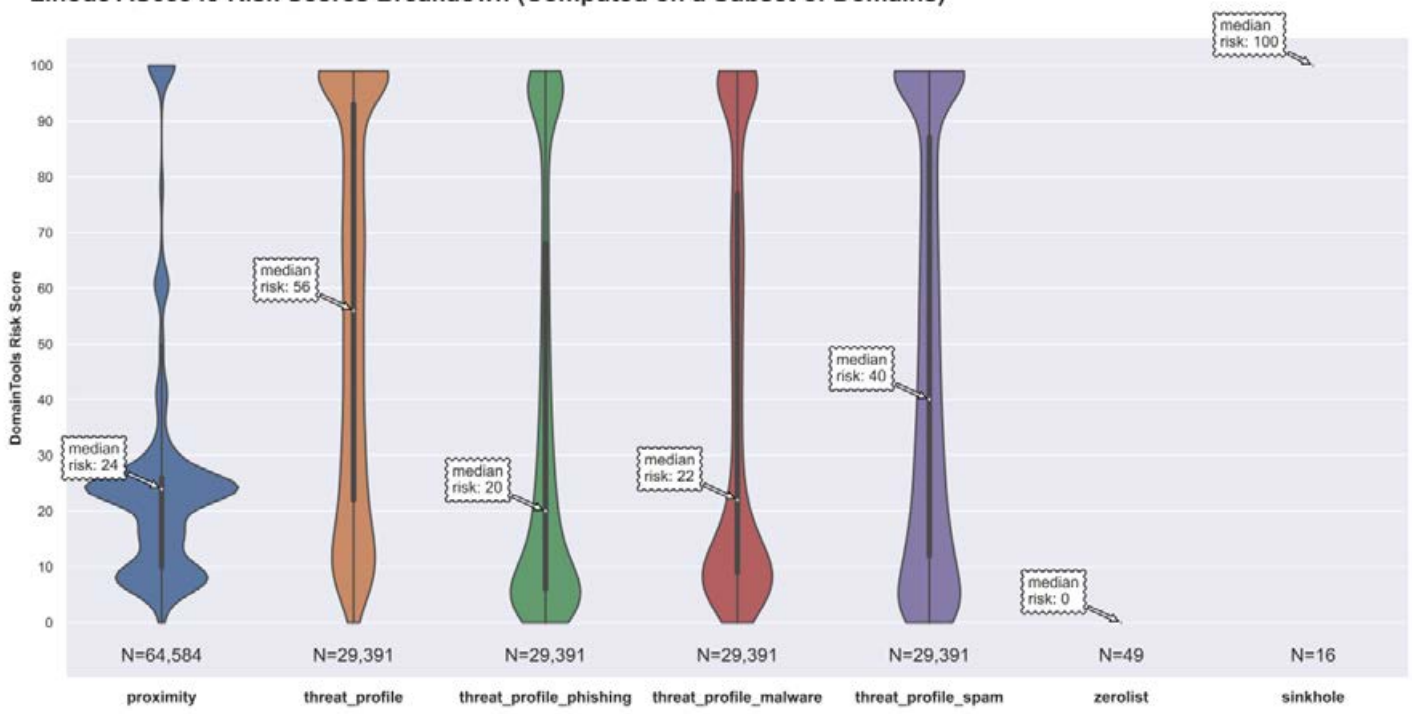
LG DACOM AS3786 Risk Scores Breakdown (Computed on a Subset of Domains)



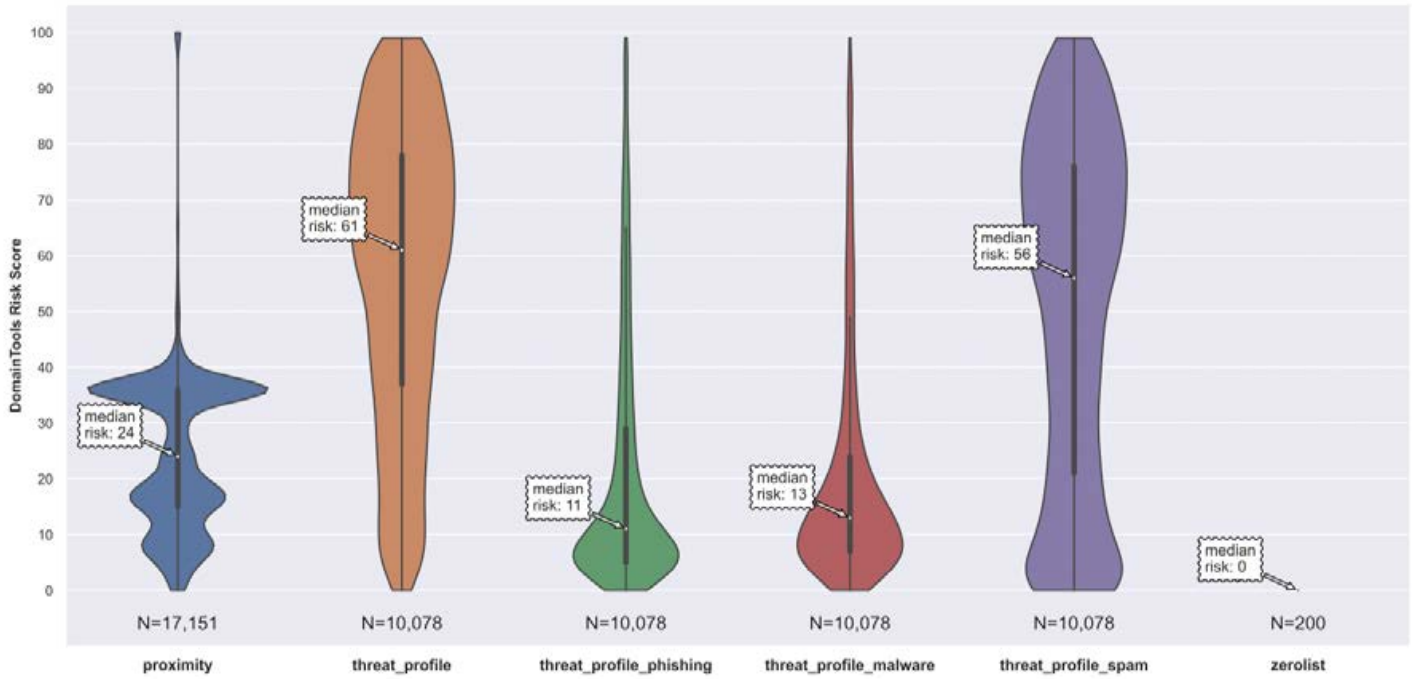
Limestone AS46475 Risk Scores Breakdown (Computed on a Subset of Domains)



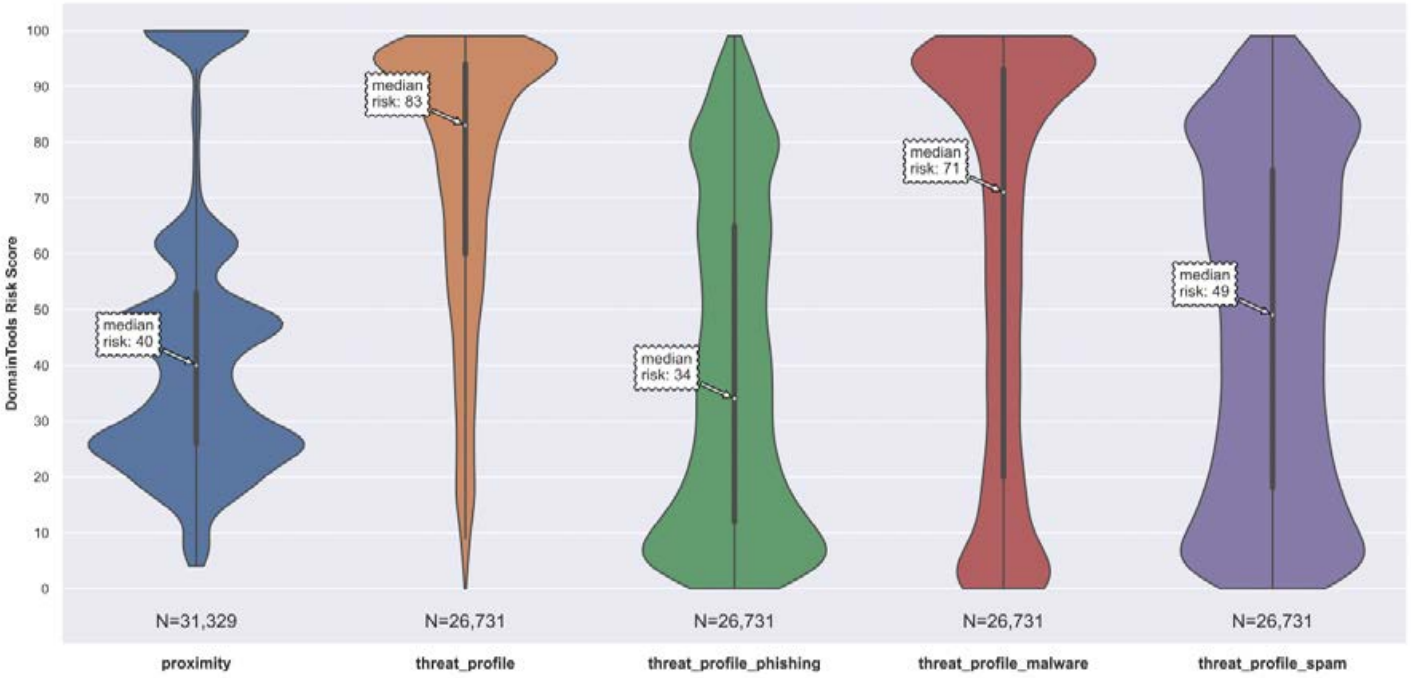
Linode AS63949 Risk Scores Breakdown (Computed on a Subset of Domains)



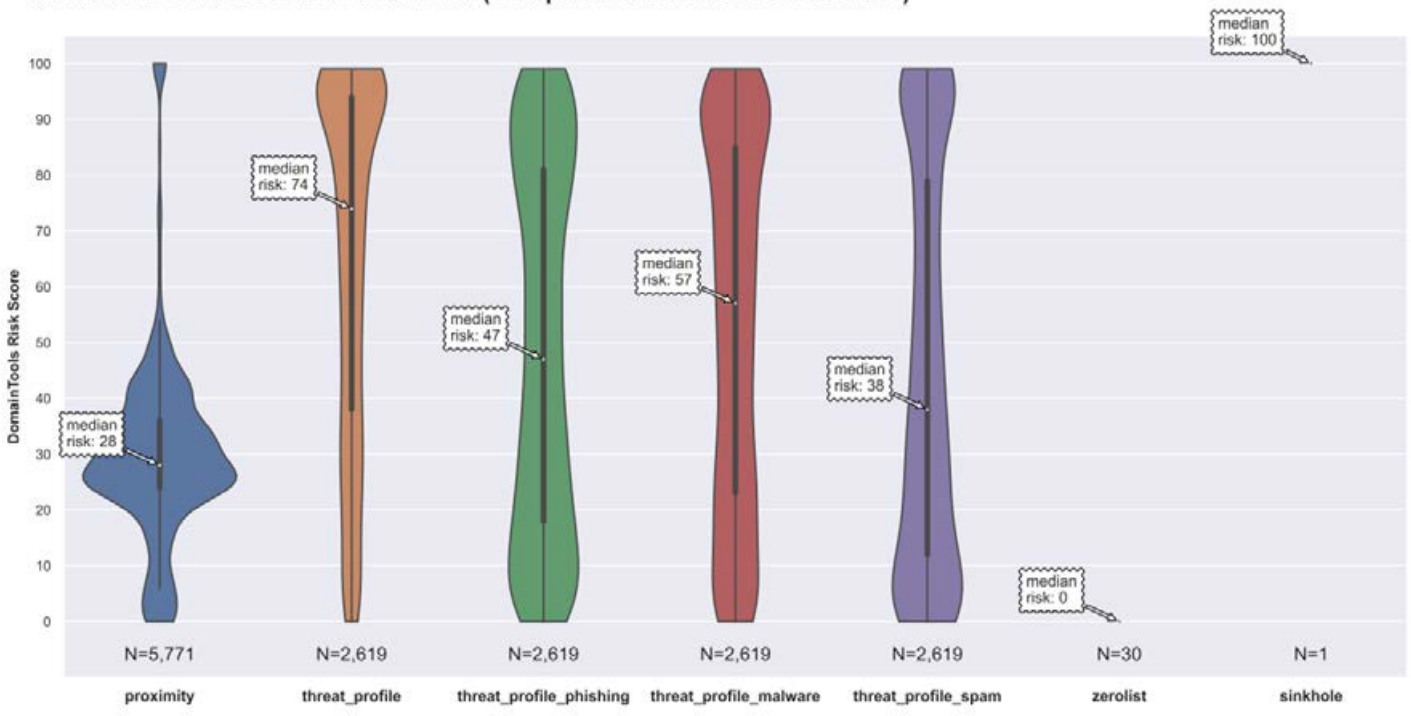
Liquid Web AS32244 Risk Scores Breakdown (Computed on a Subset of Domains)



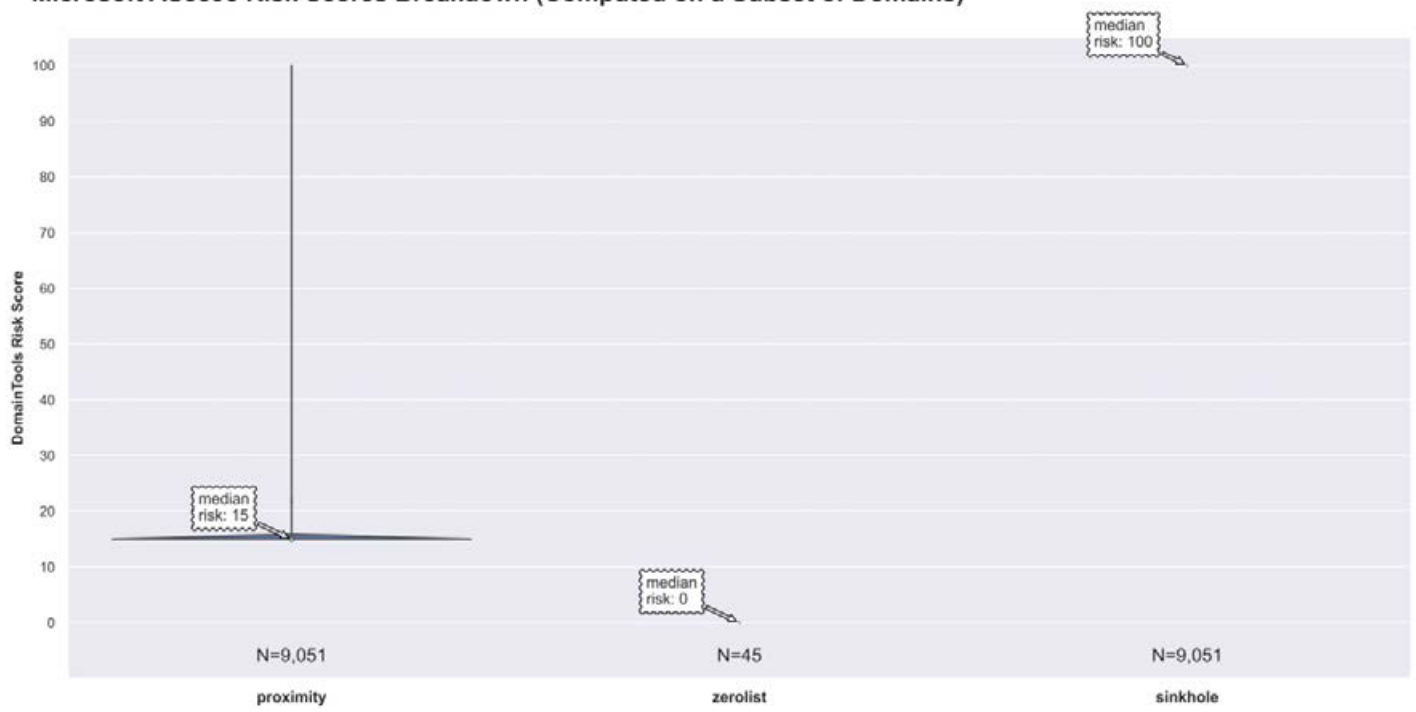
Luogelang AS135097 Risk Scores Breakdown (Computed on a Subset of Domains)



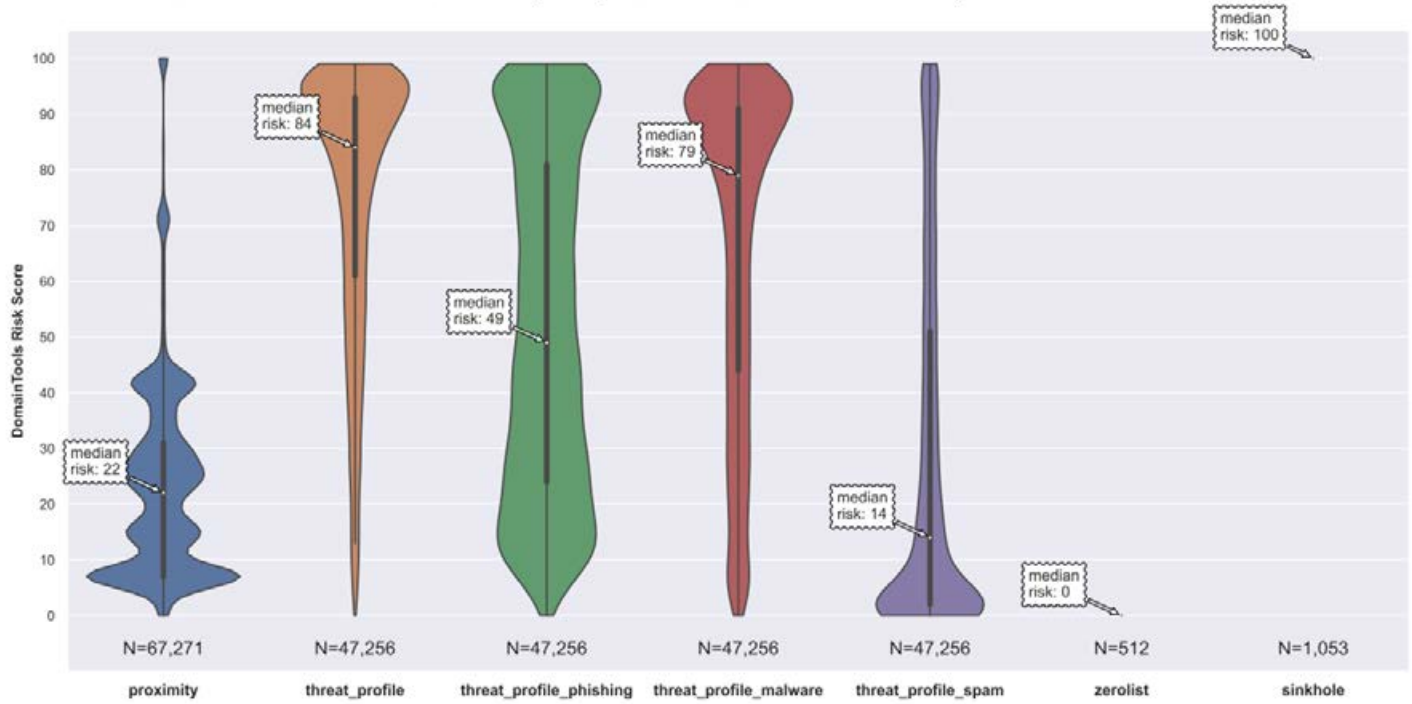
M247 AS9009 Risk Scores Breakdown (Computed on a Subset of Domains)



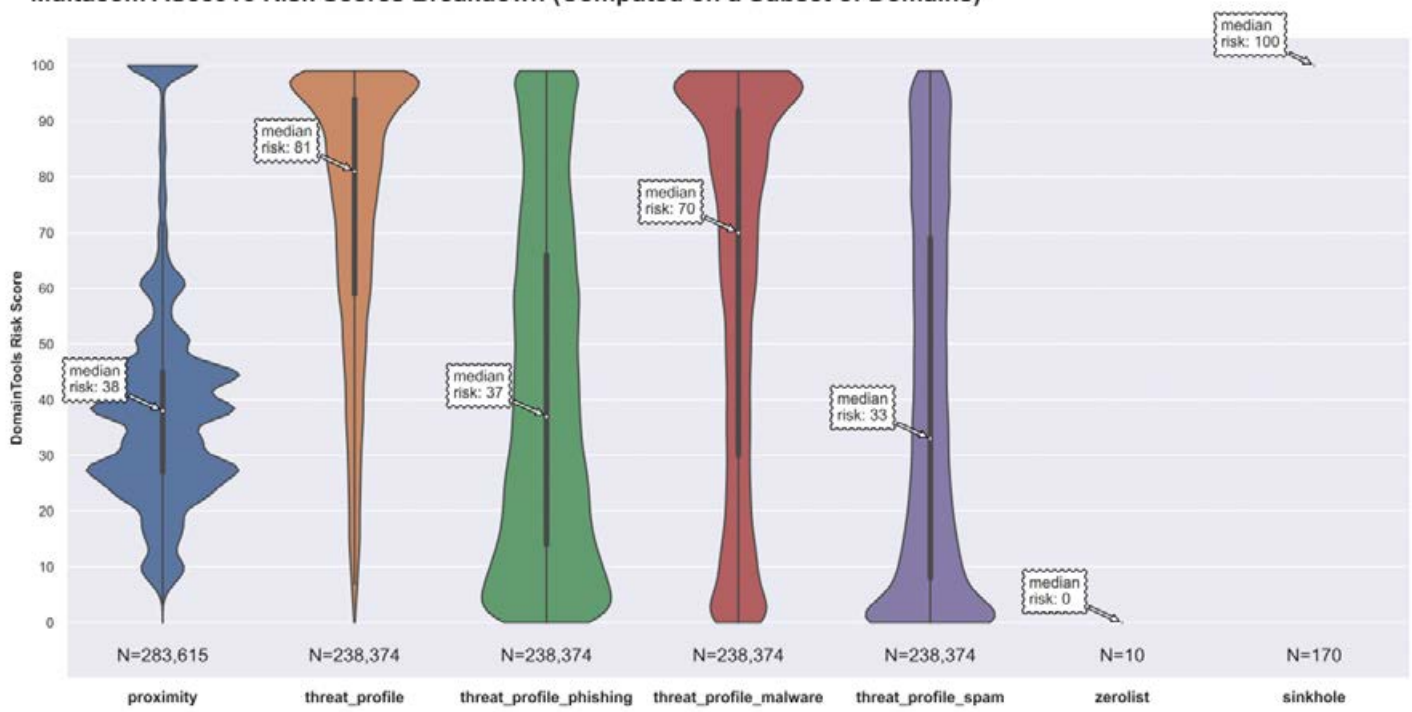
Microsoft AS3598 Risk Scores Breakdown (Computed on a Subset of Domains)



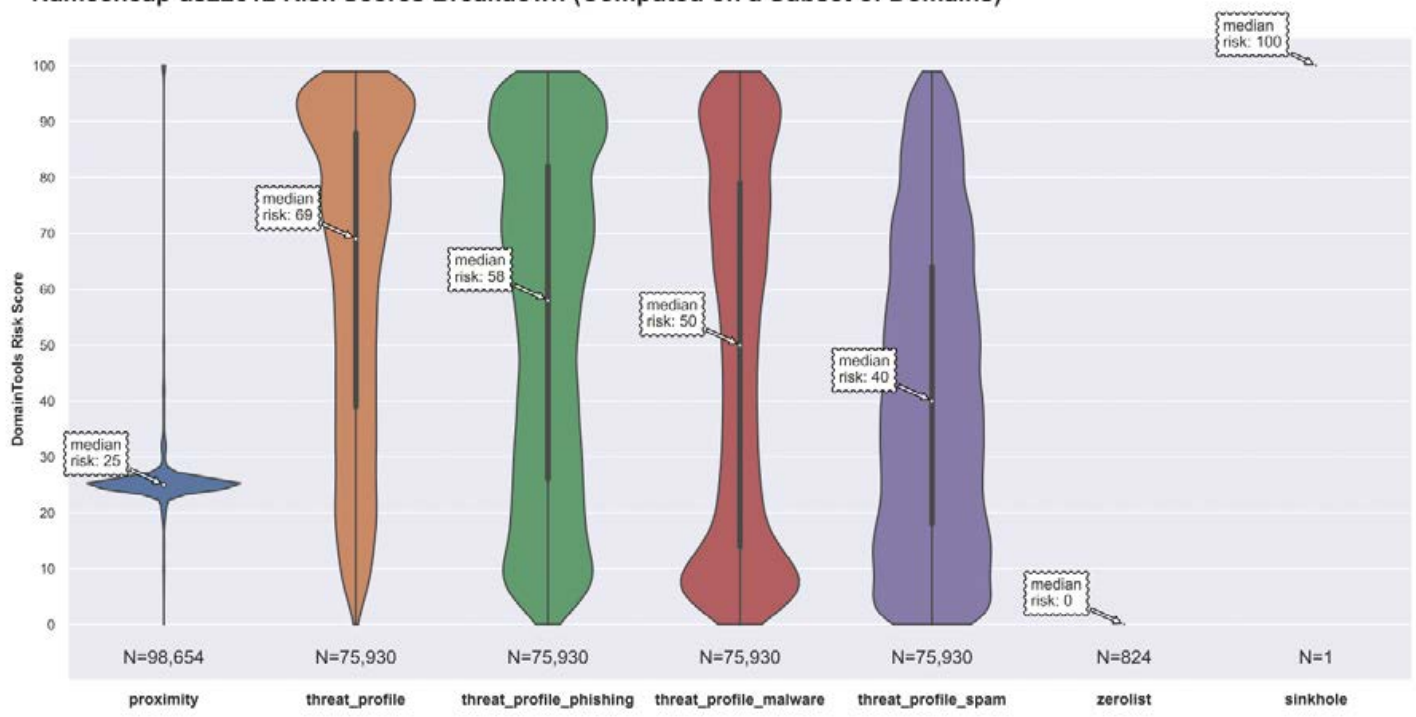
Microsoft AS8075 Risk Scores Breakdown (Computed on a Subset of Domains)



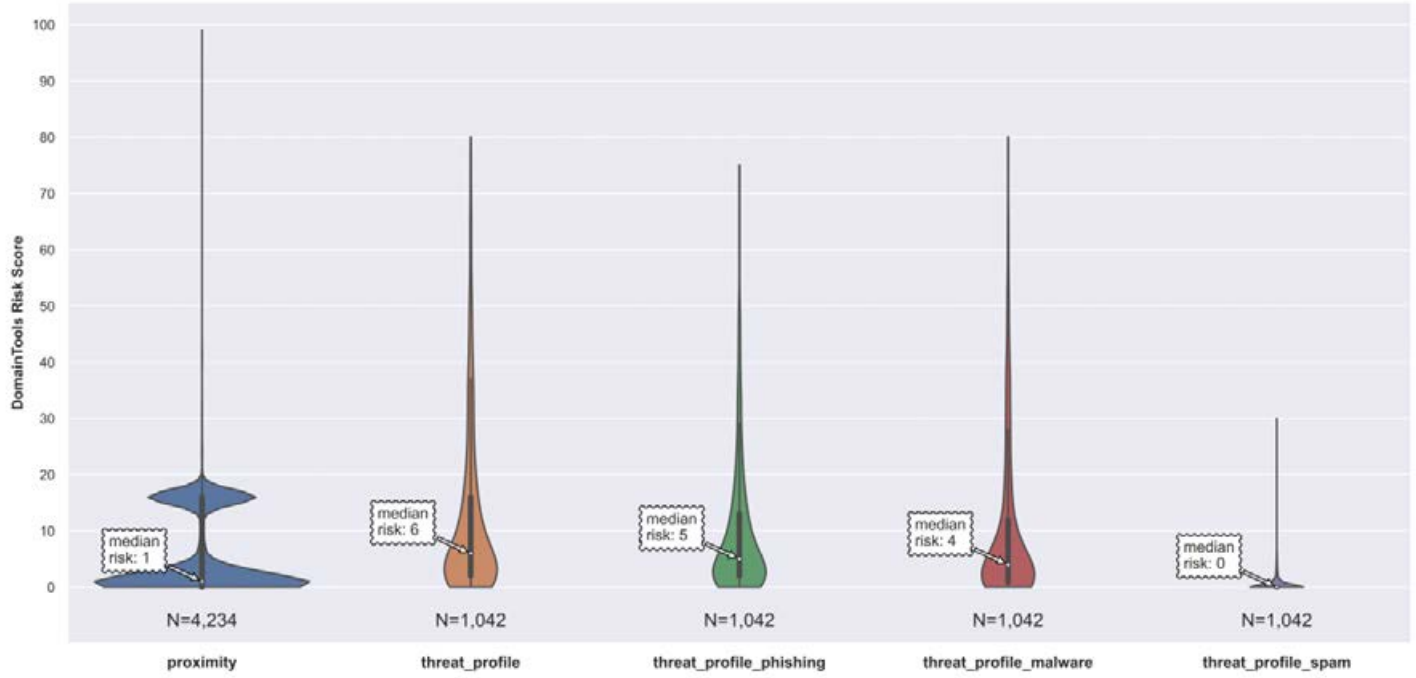
Multacom AS35916 Risk Scores Breakdown (Computed on a Subset of Domains)



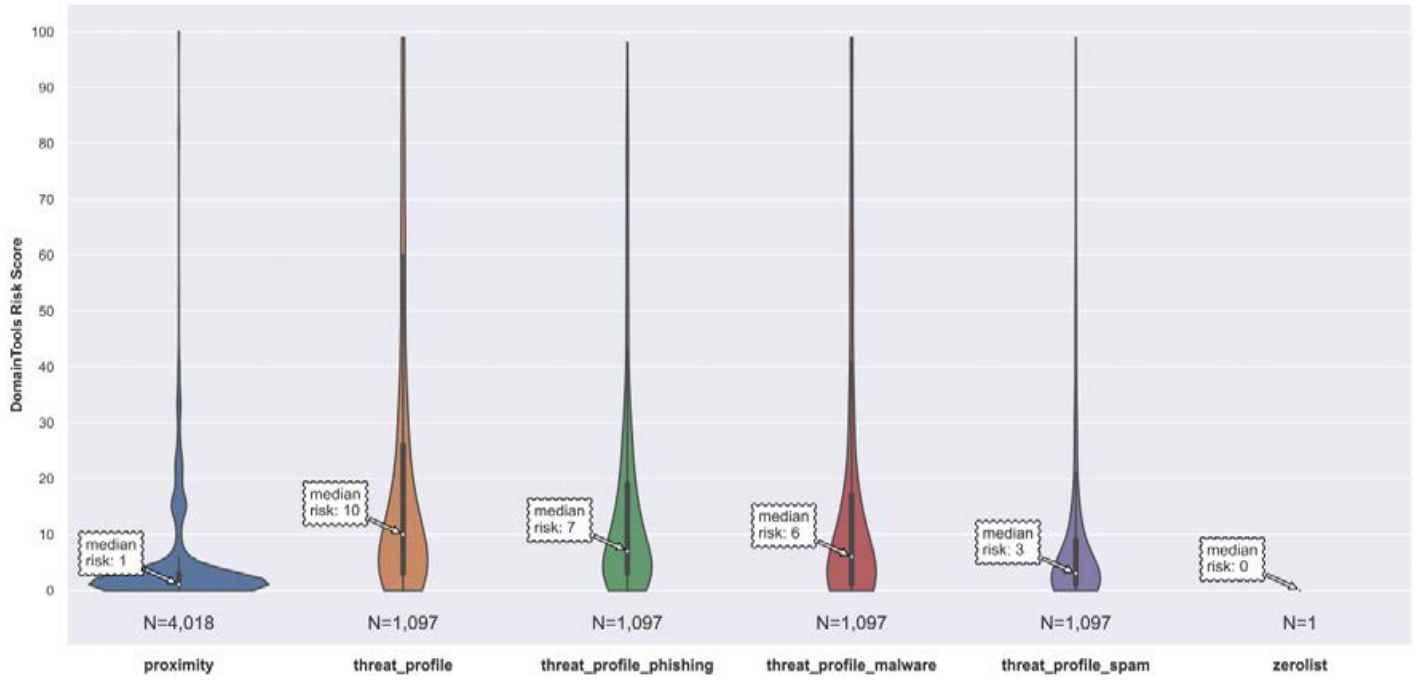
Namecheap-as22612 Risk Scores Breakdown (Computed on a Subset of Domains)



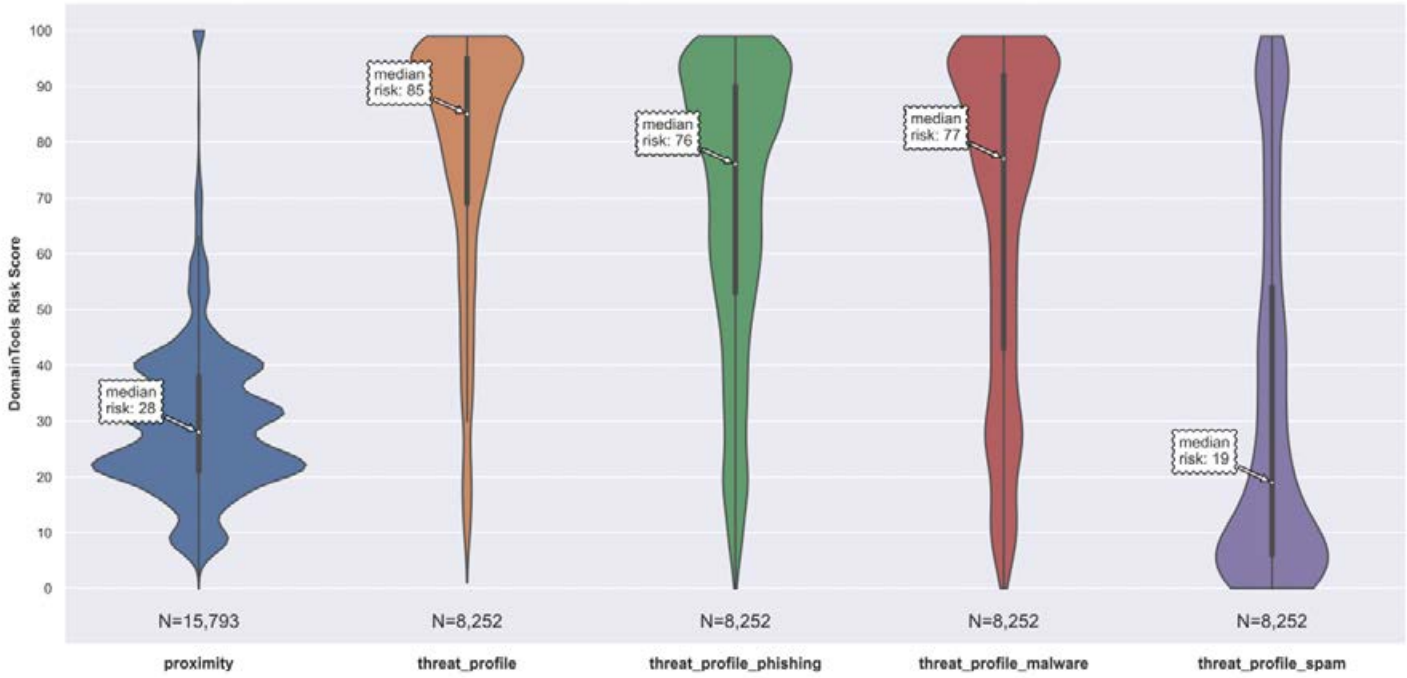
Namesco AS8622 Risk Scores Breakdown (Computed on a Subset of Domains)



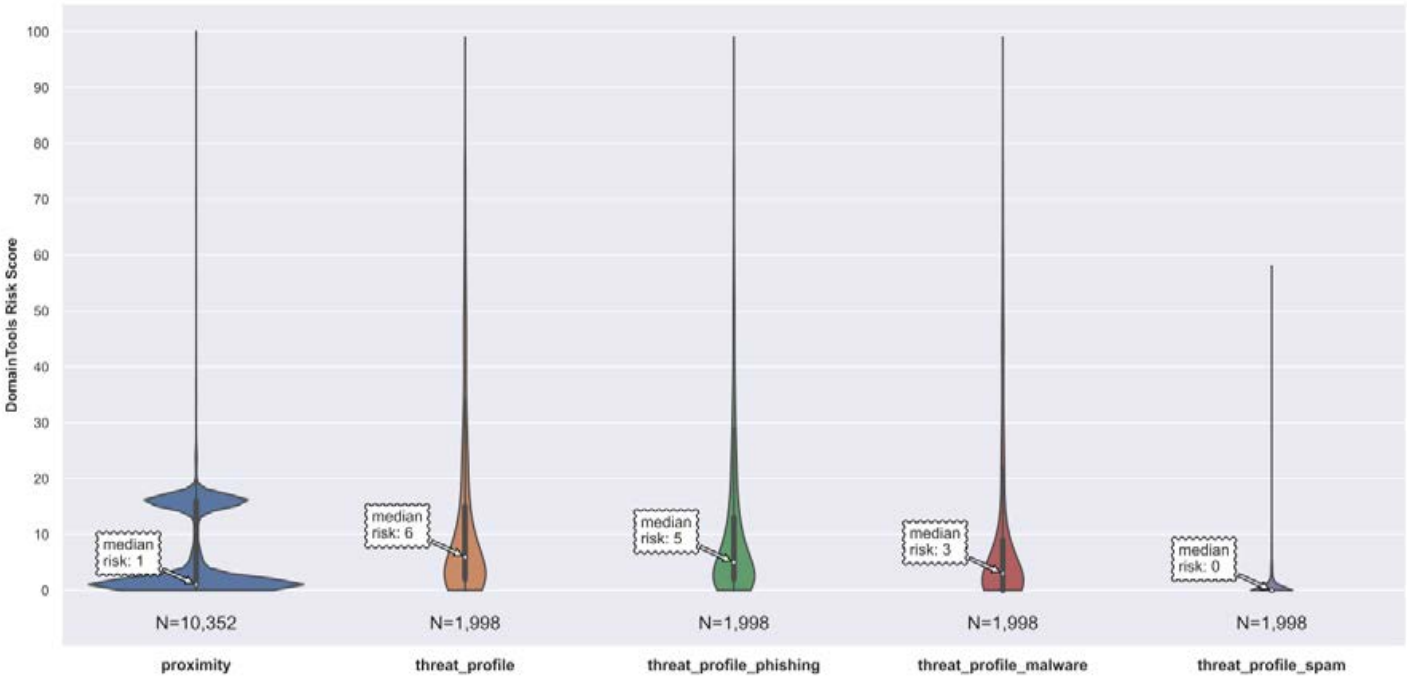
Netcup AS197540 Risk Scores Breakdown (Computed on a Subset of Domains)



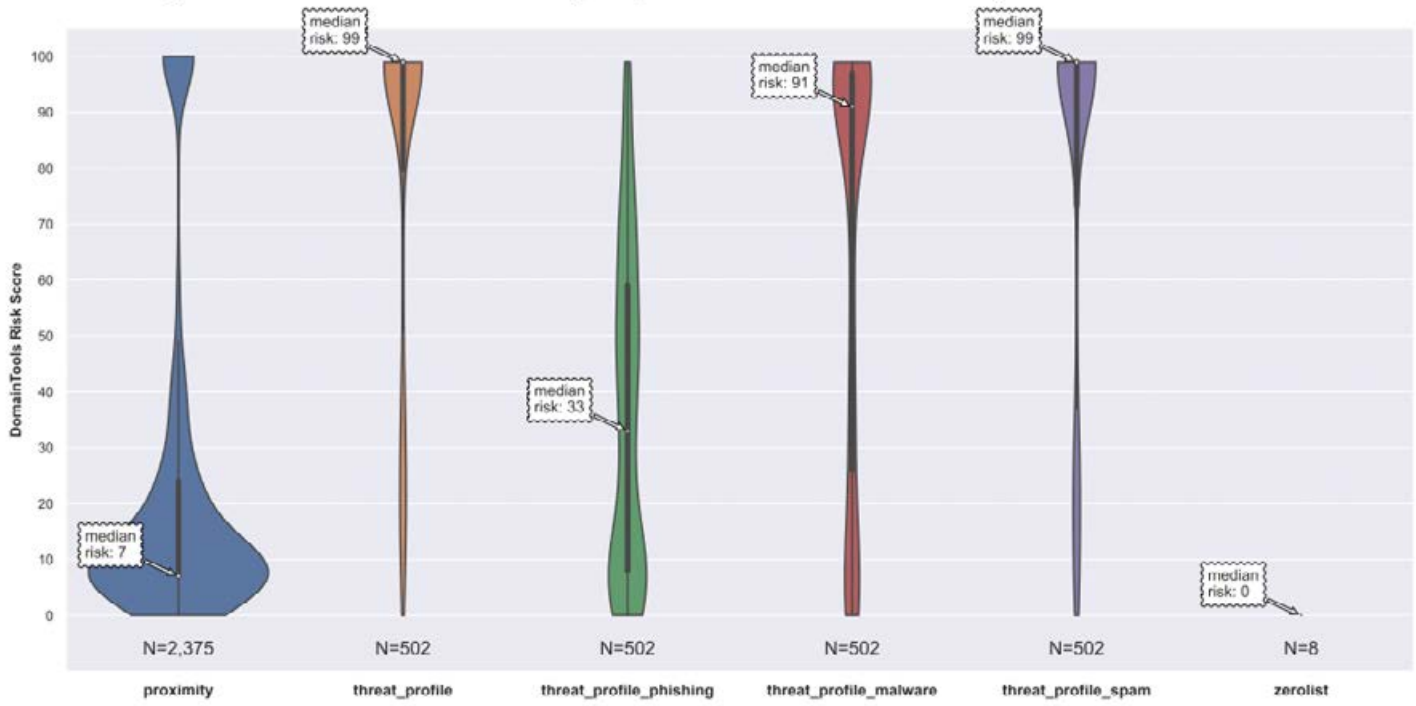
Netsec AS45753 Risk Scores Breakdown (Computed on a Subset of Domains)



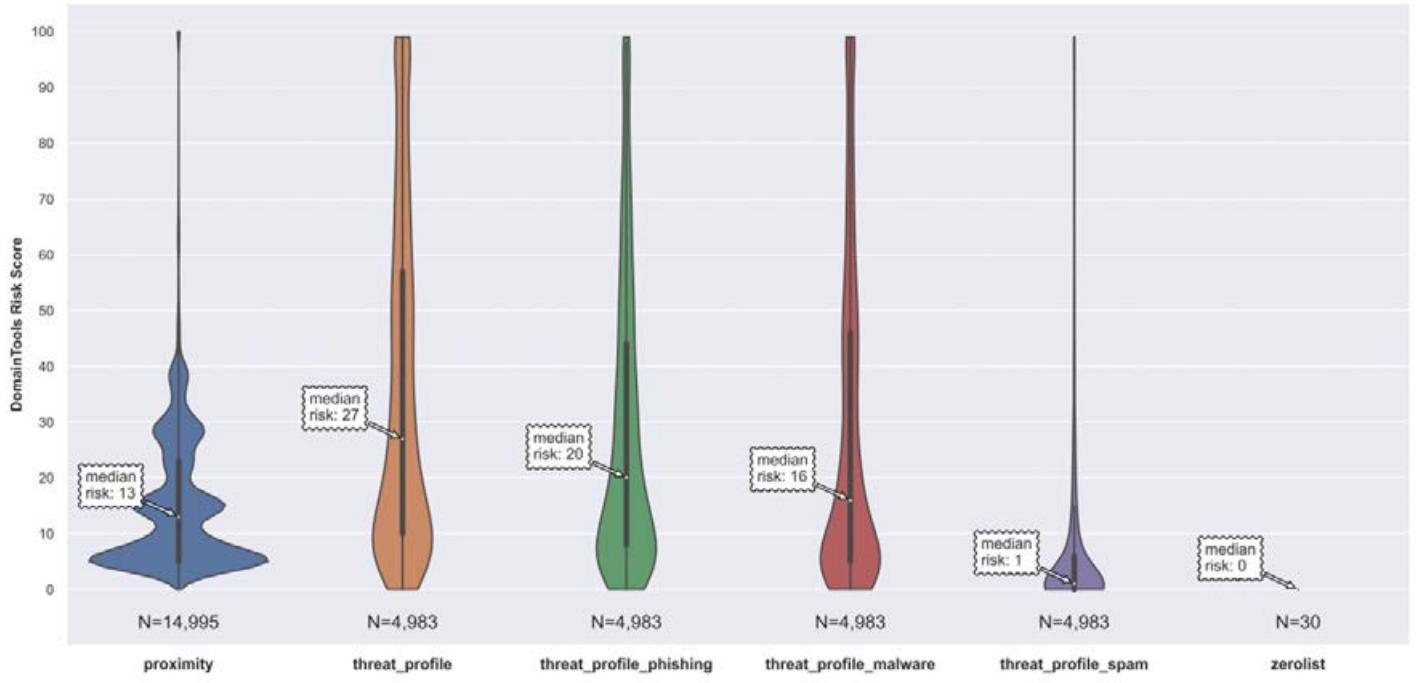
Neue Medien AS34788 Risk Scores Breakdown (Computed on a Subset of Domains)



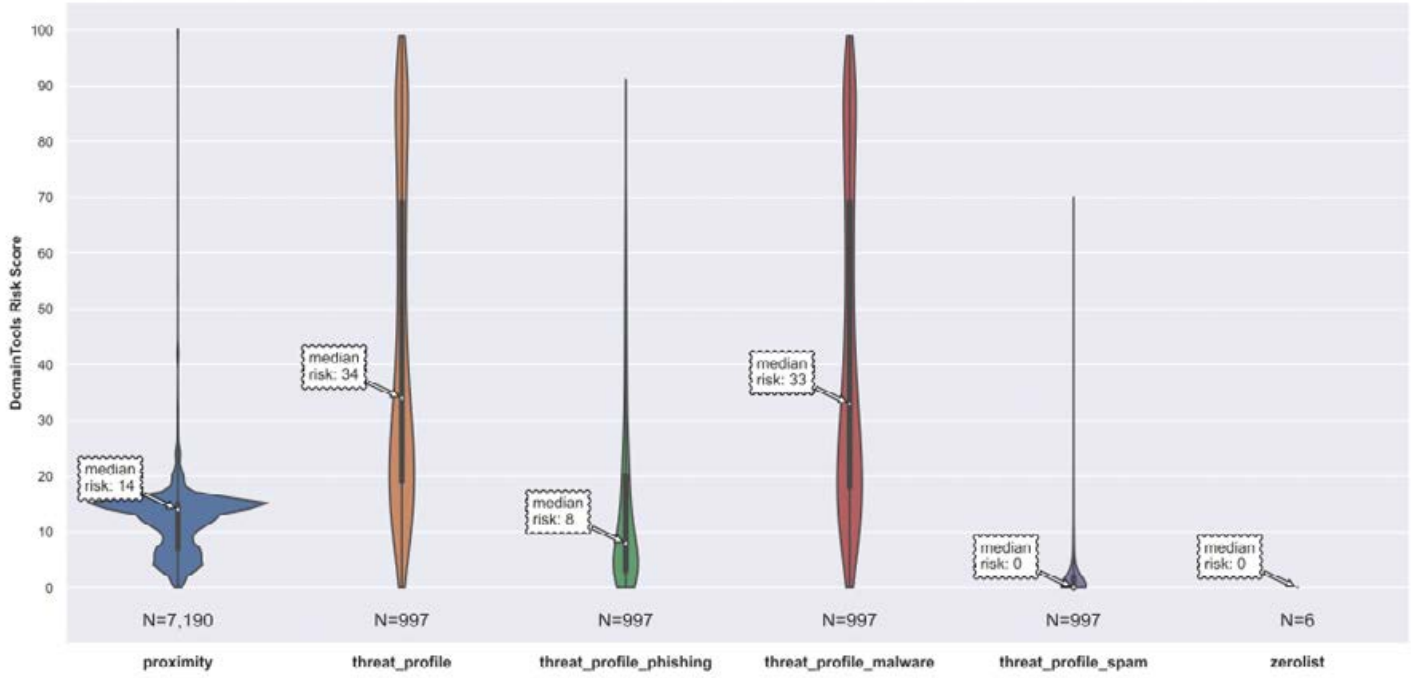
New Century AS9919 Risk Scores Breakdown (Computed on a Subset of Domains)



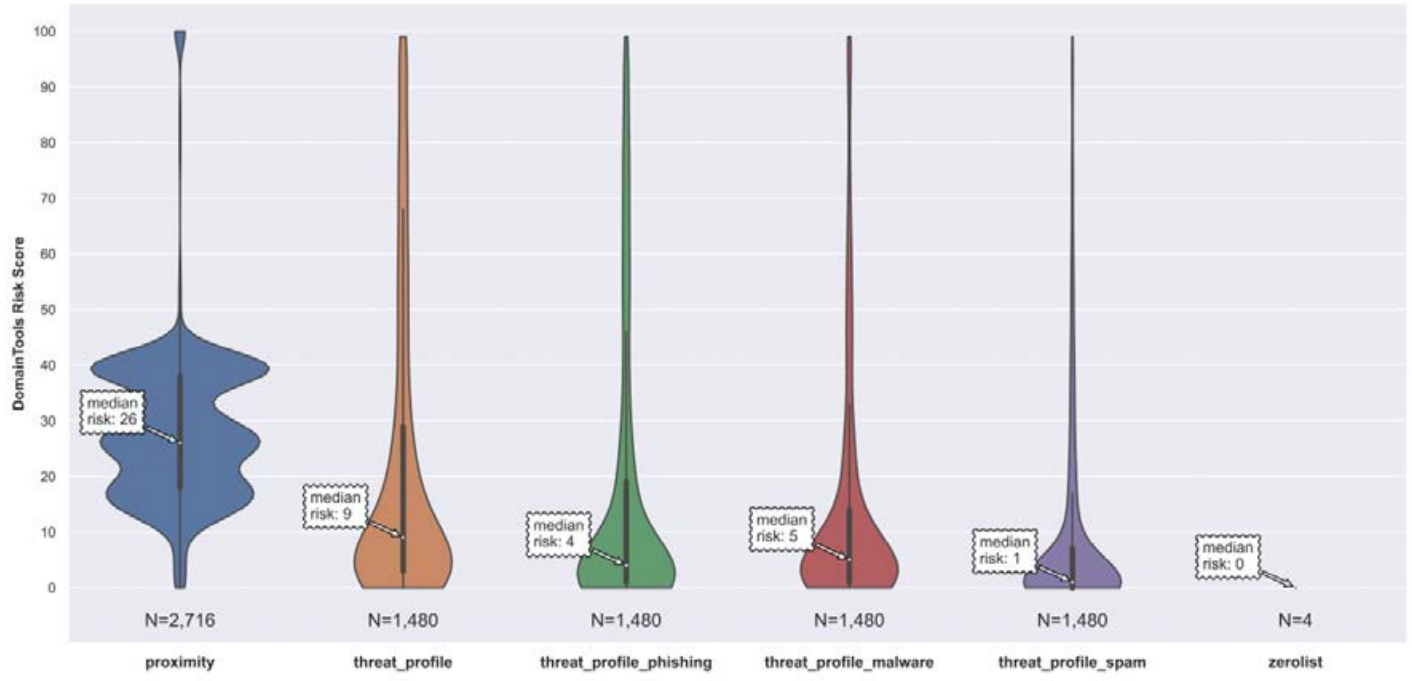
New Dream AS26347 Risk Scores Breakdown (Computed on a Subset of Domains)



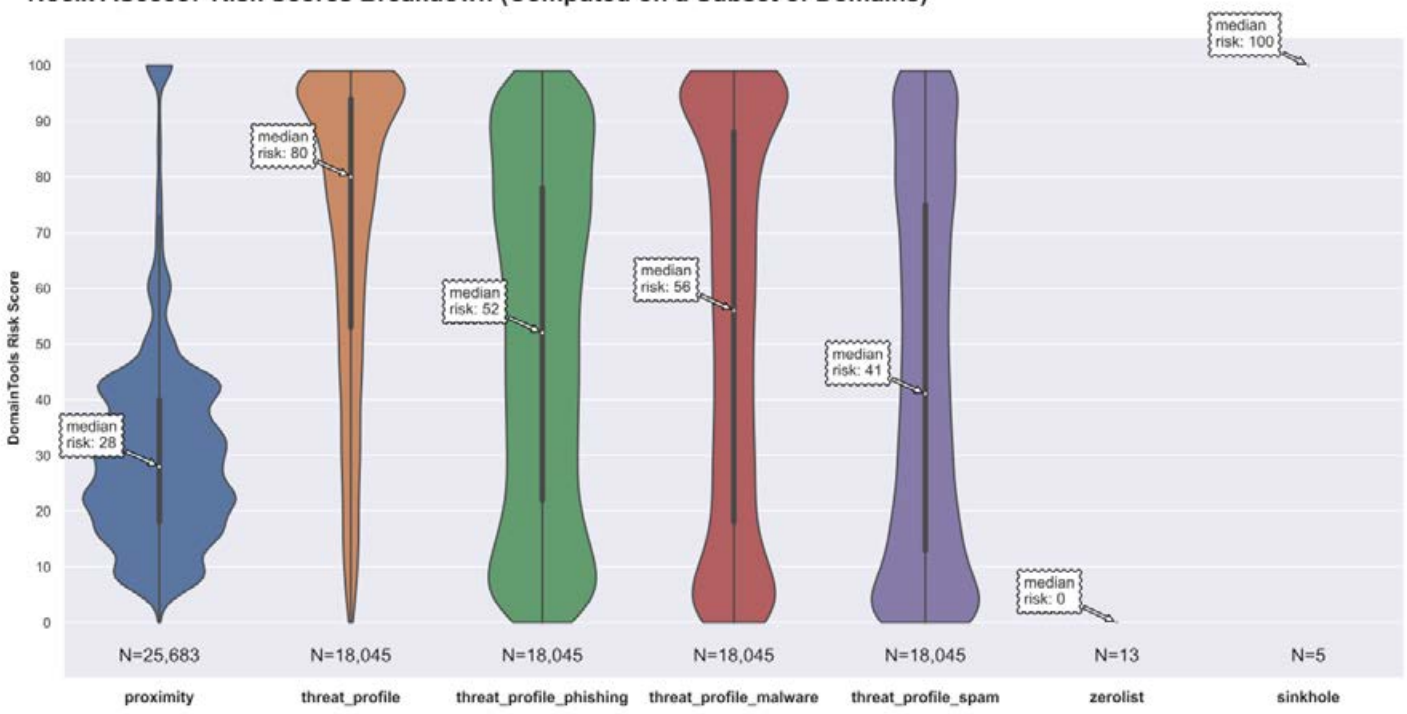
Newfold AS29873 Risk Scores Breakdown (Computed on a Subset of Domains)



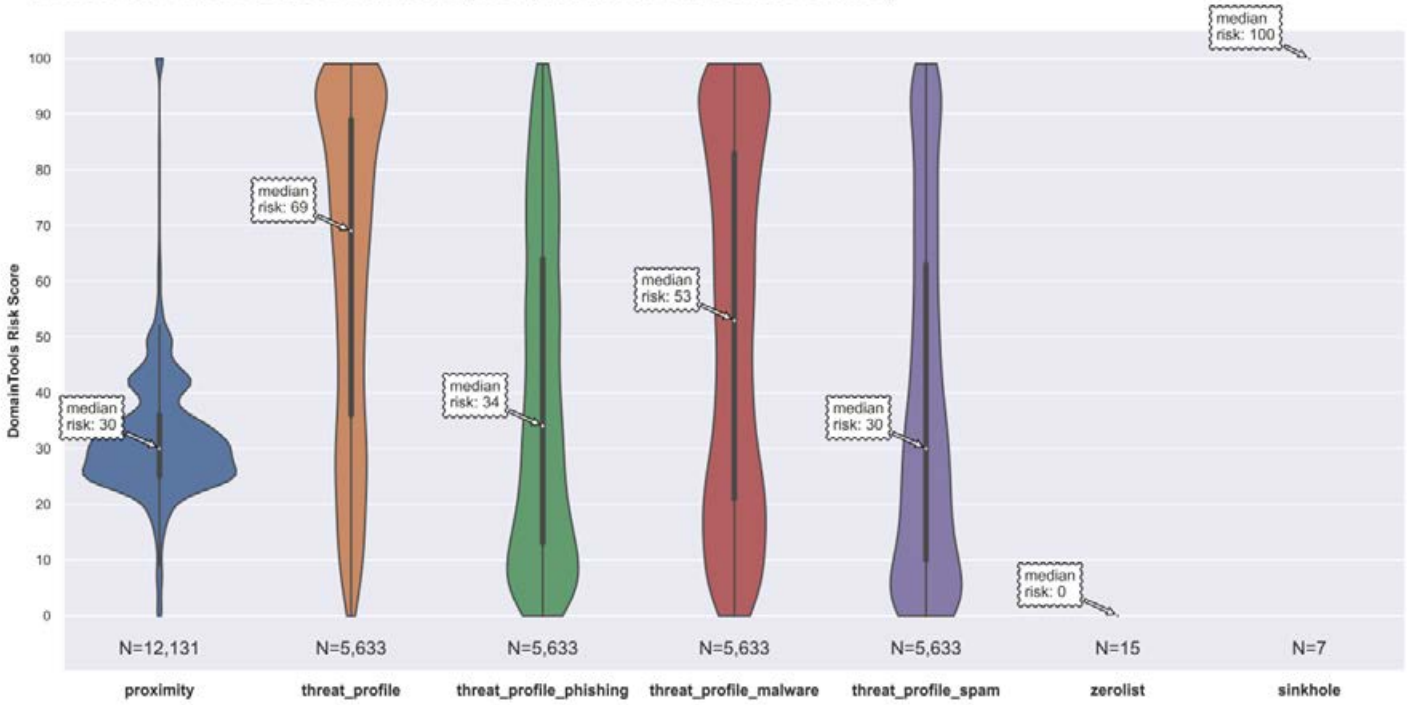
NForce AS43350 Risk Scores Breakdown (Computed on a Subset of Domains)



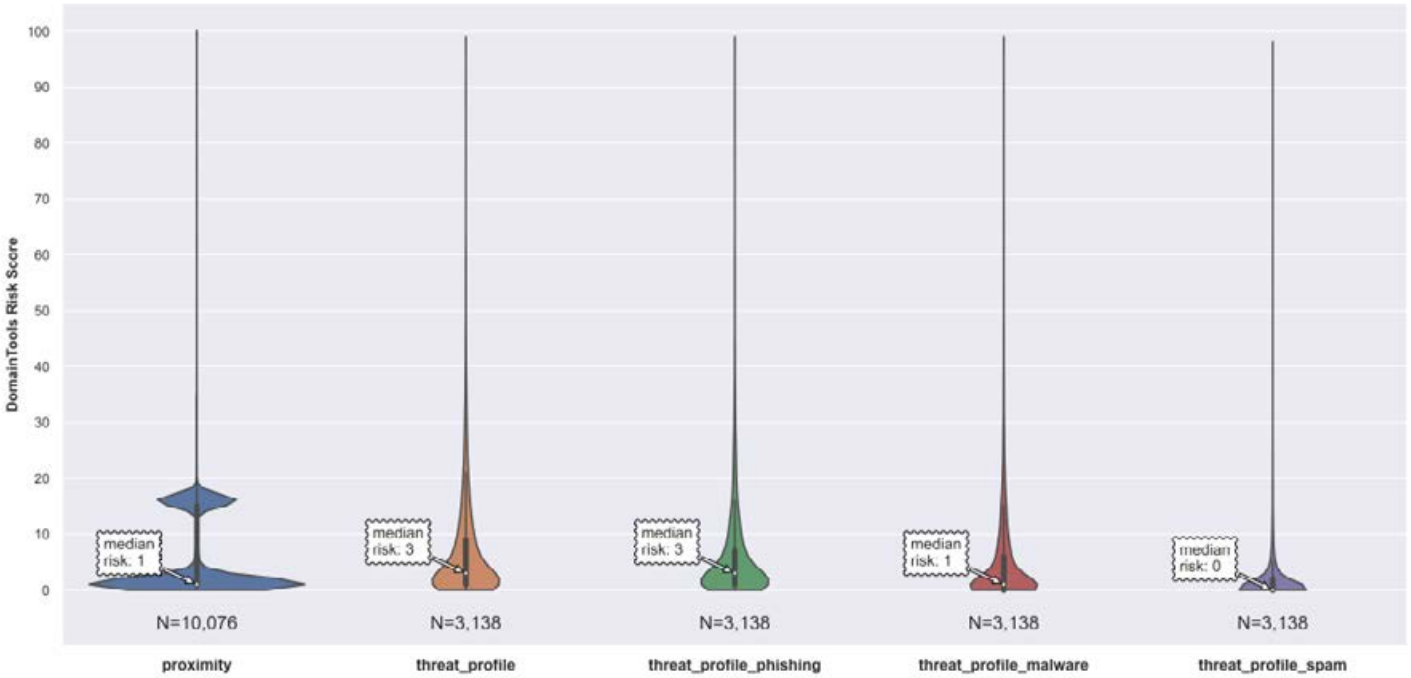
Nocix AS33387 Risk Scores Breakdown (Computed on a Subset of Domains)



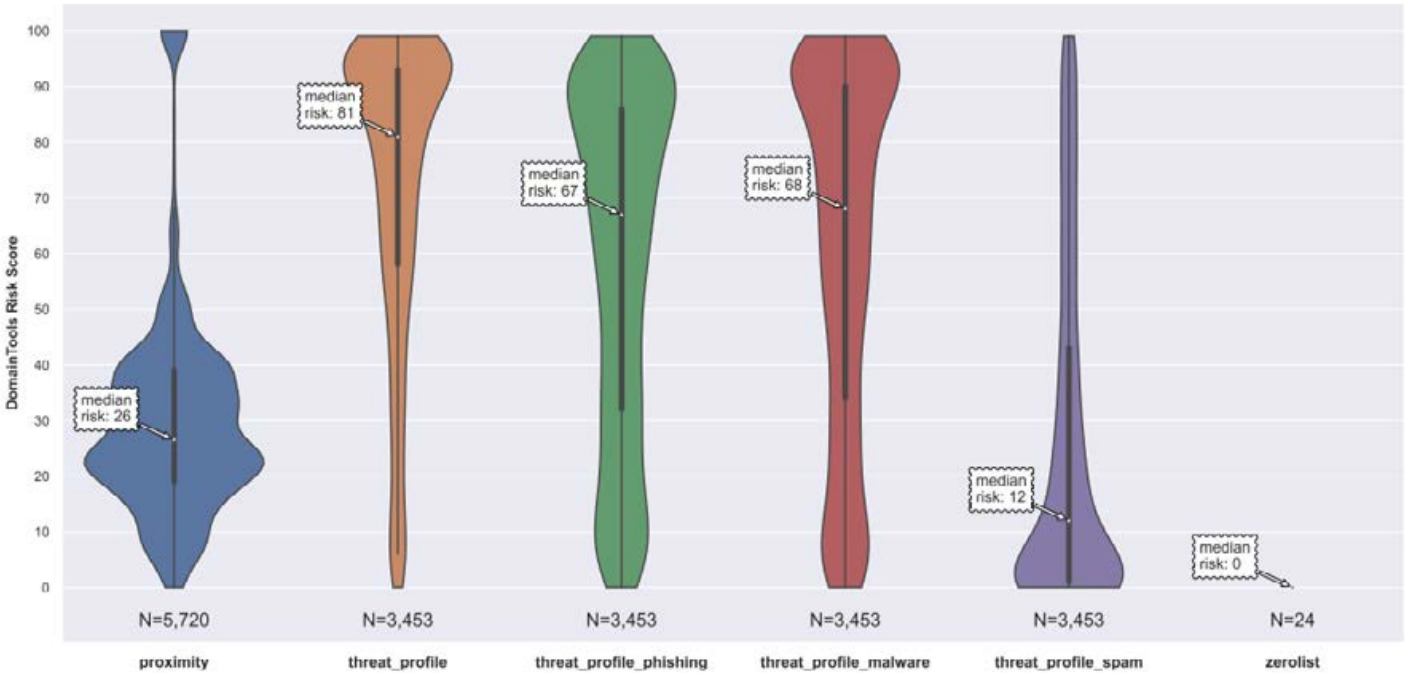
NTT AS2914 Risk Scores Breakdown (Computed on a Subset of Domains)



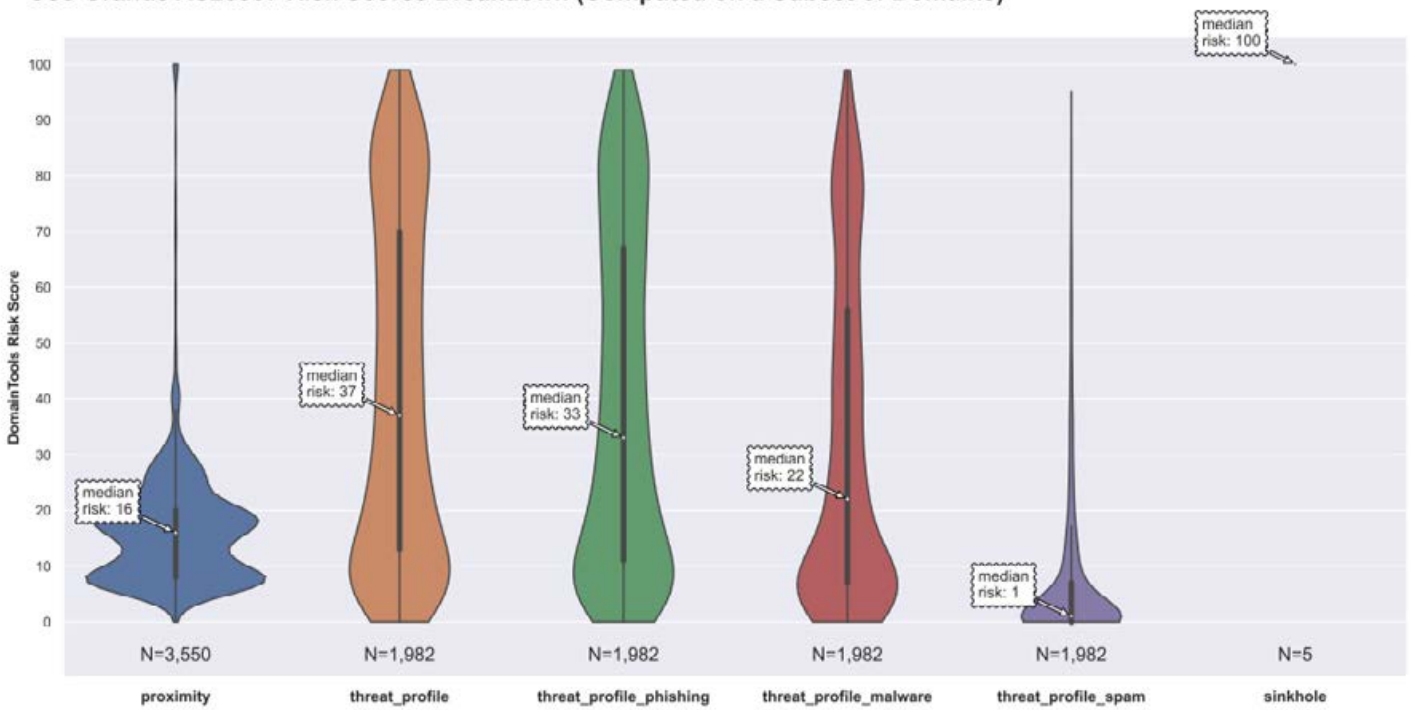
One AS51468 Risk Scores Breakdown (Computed on a Subset of Domains)



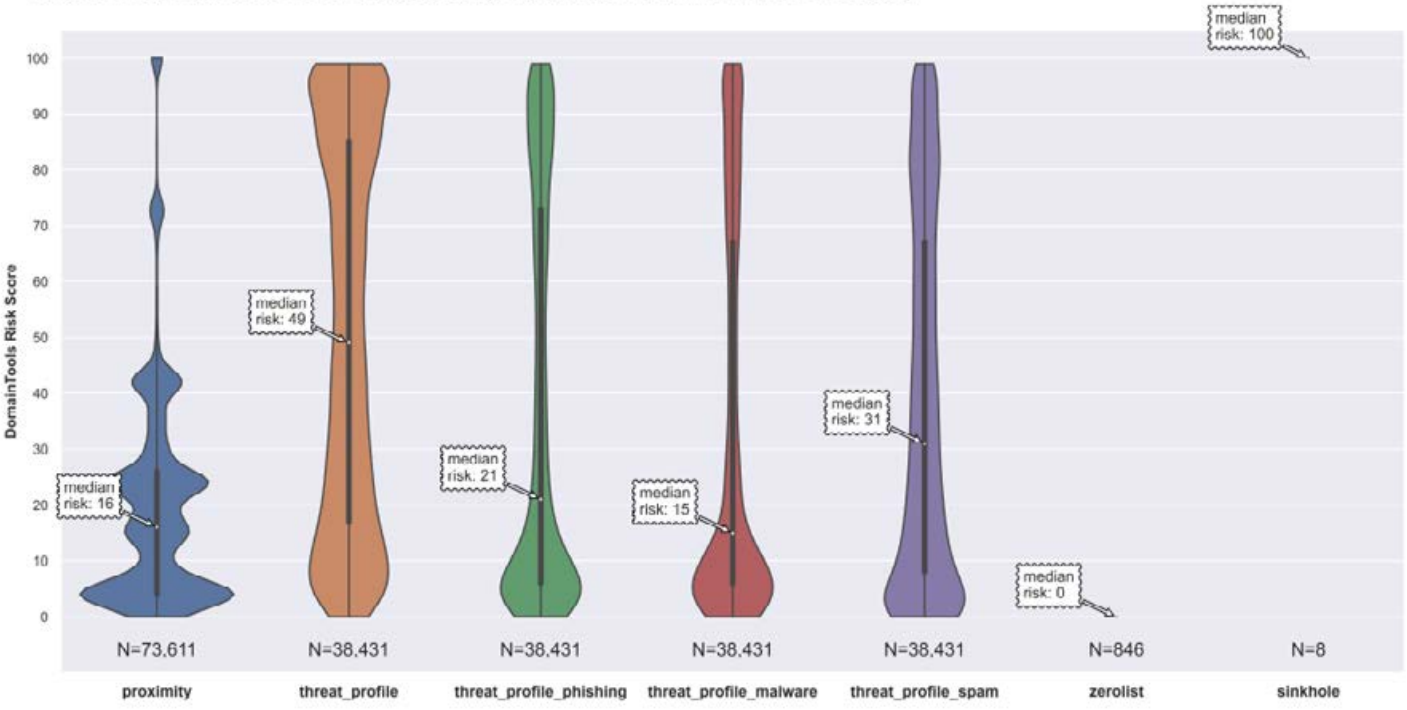
Oracle AS31898 Risk Scores Breakdown (Computed on a Subset of Domains)



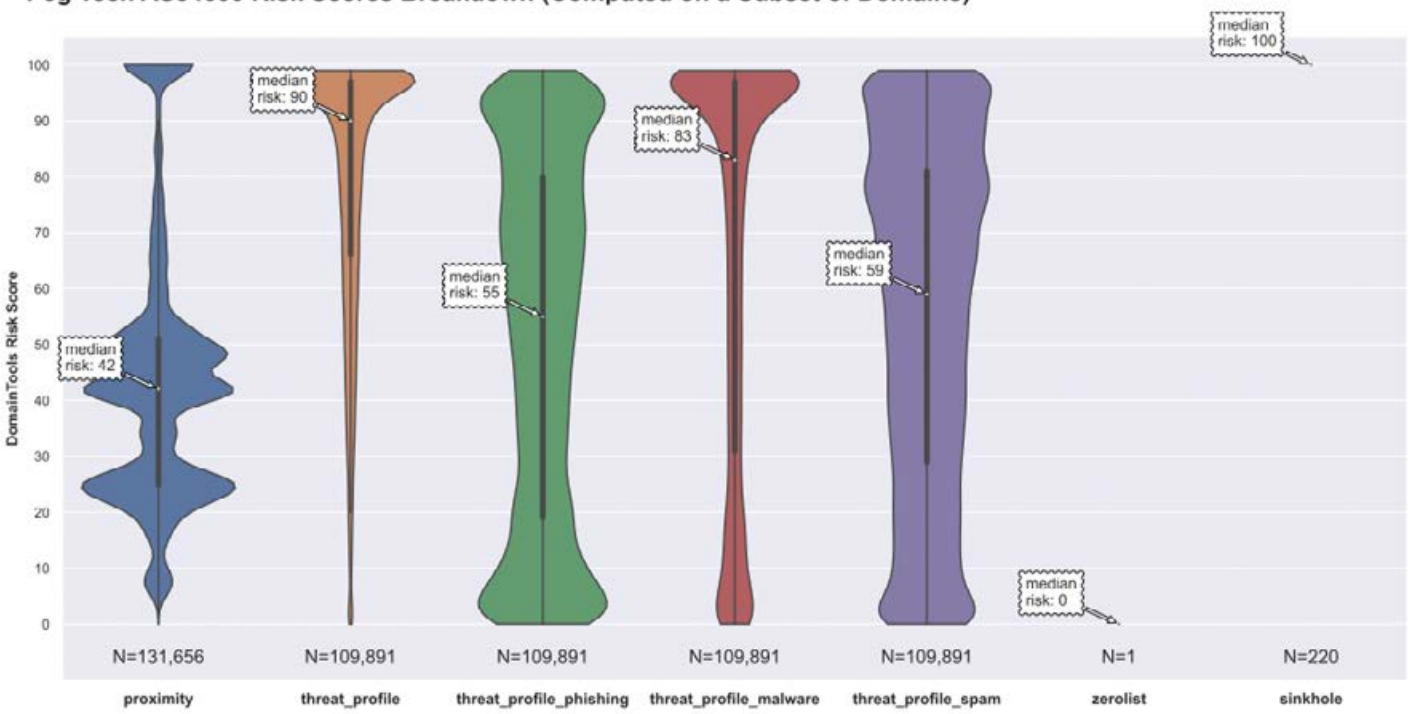
Oso Grande AS26337 Risk Scores Breakdown (Computed on a Subset of Domains)



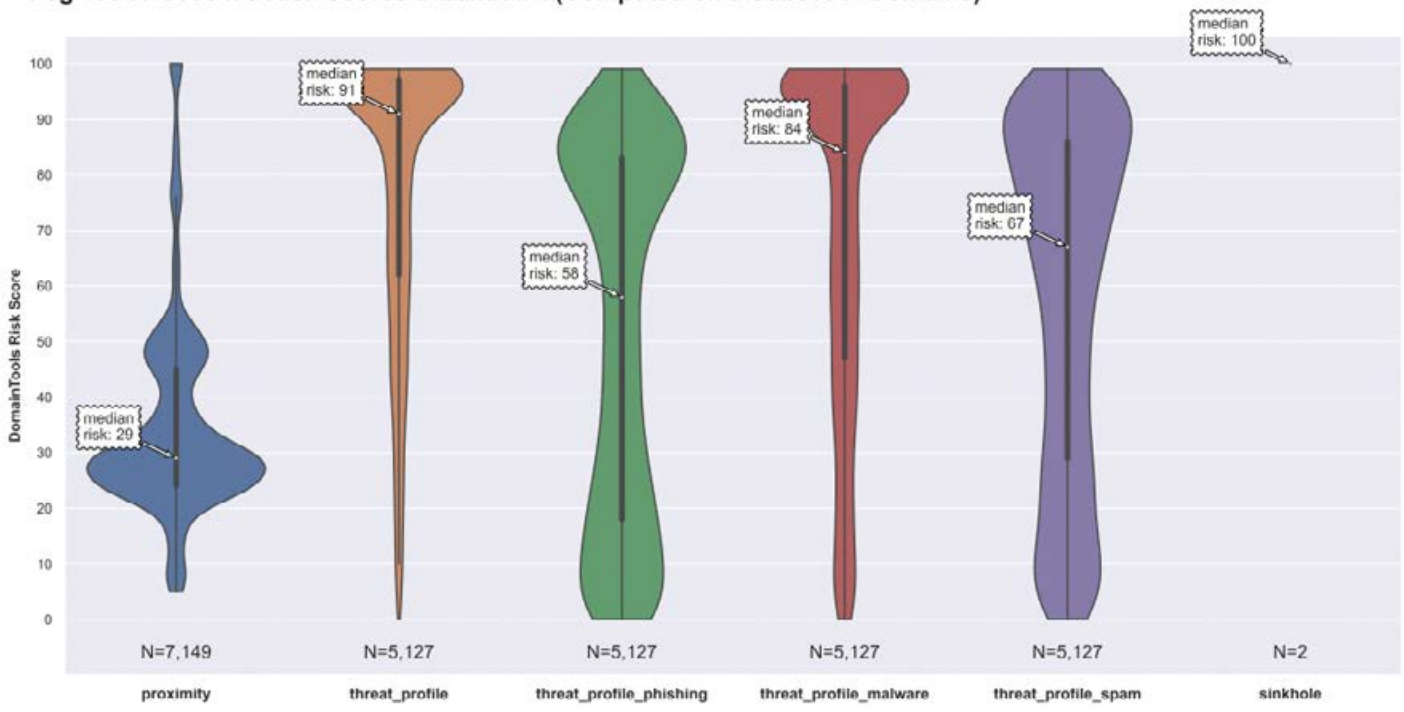
OVH AS16276 Risk Scores Breakdown (Computed on a Subset of Domains)



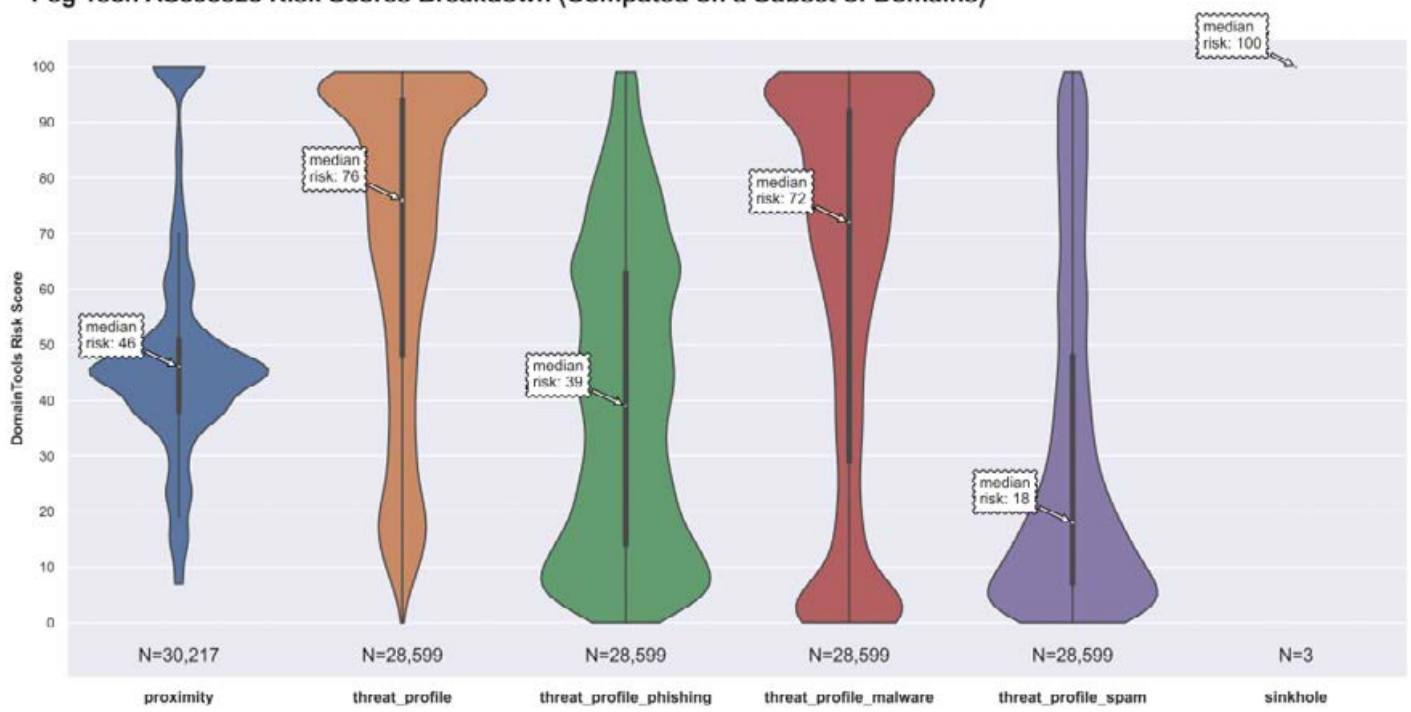
Peg Tech AS54600 Risk Scores Breakdown (Computed on a Subset of Domains)



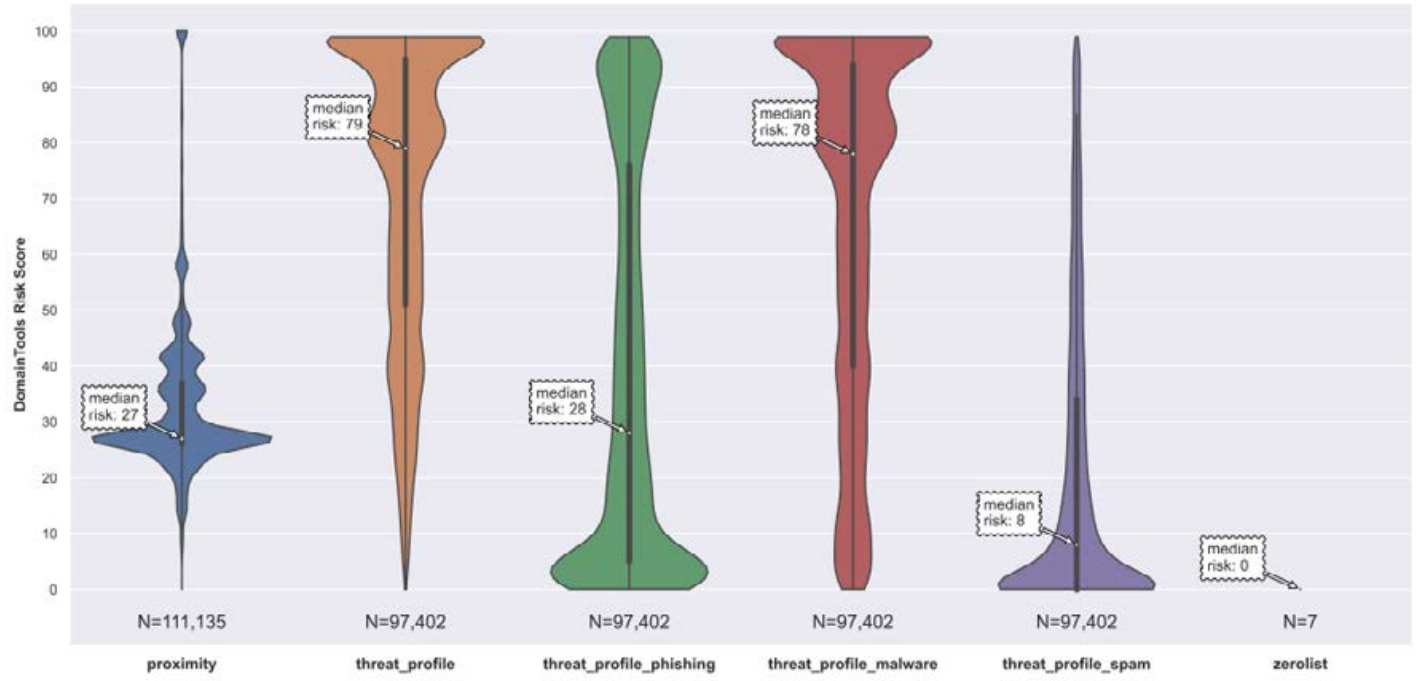
Peg Tech AS398478 Risk Scores Breakdown (Computed on a Subset of Domains)



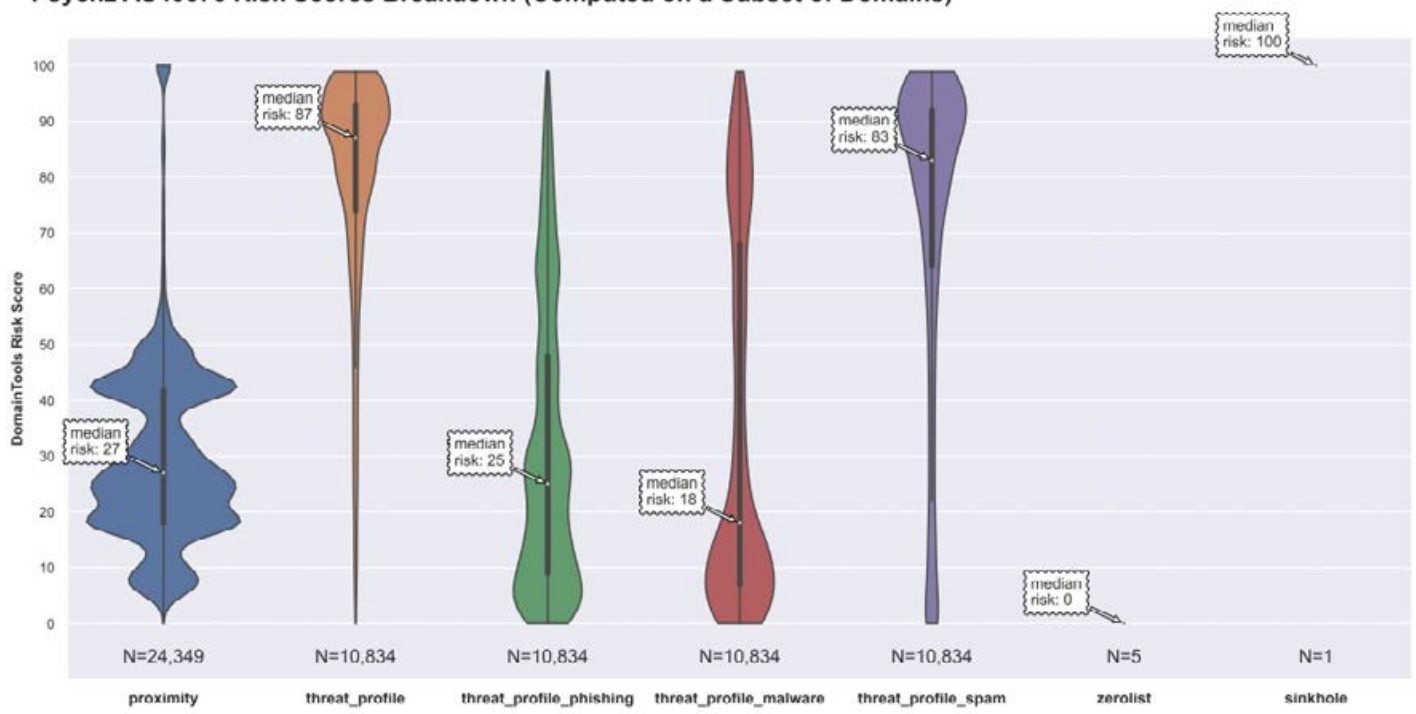
Peg Tech AS398823 Risk Scores Breakdown (Computed on a Subset of Domains)



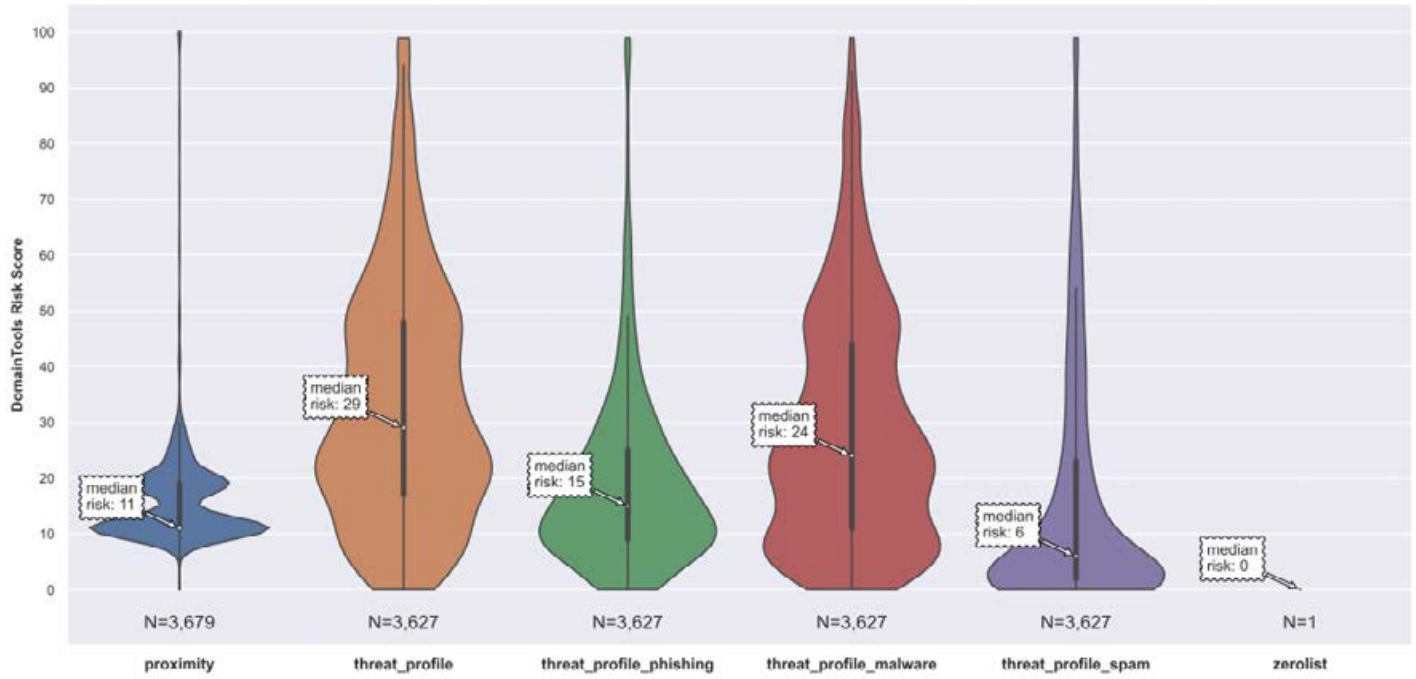
Power Line HK AS132839 Risk Scores Breakdown (Computed on a Subset of Domains)



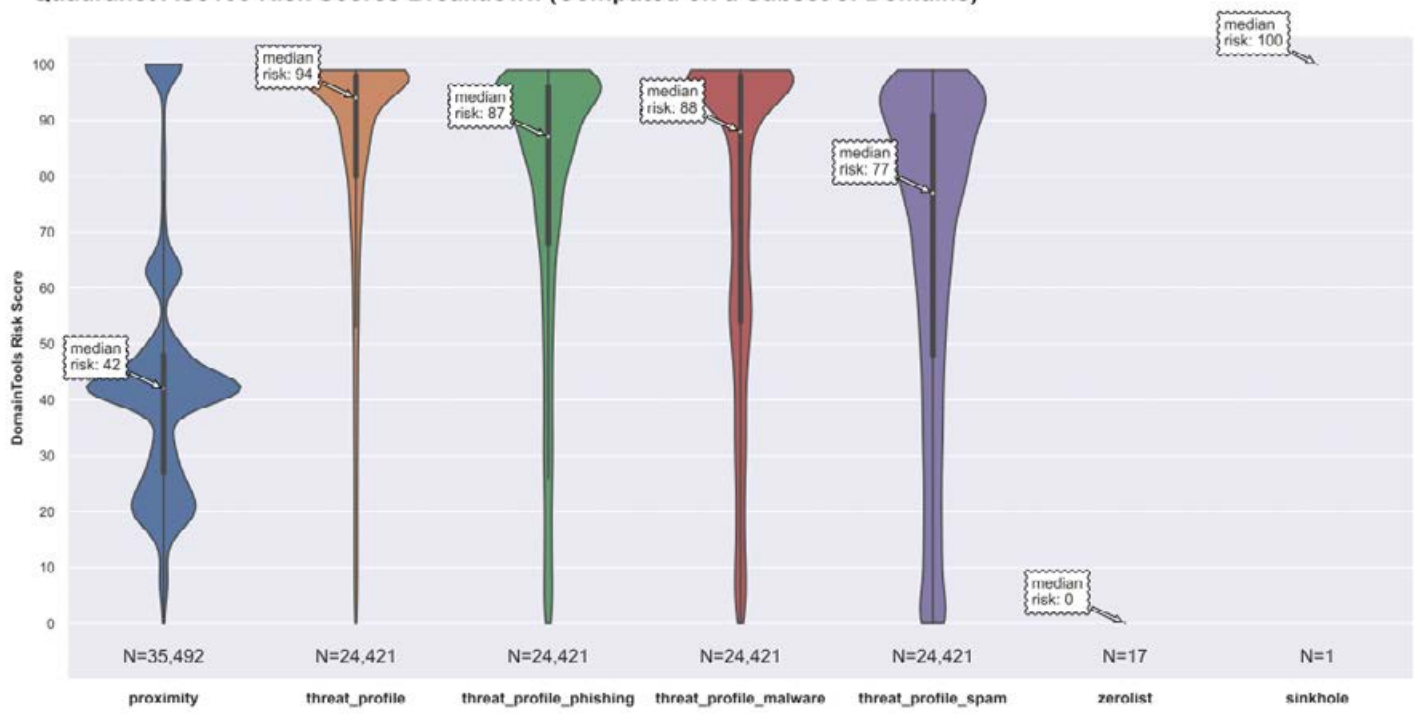
Psychz AS40676 Risk Scores Breakdown (Computed on a Subset of Domains)



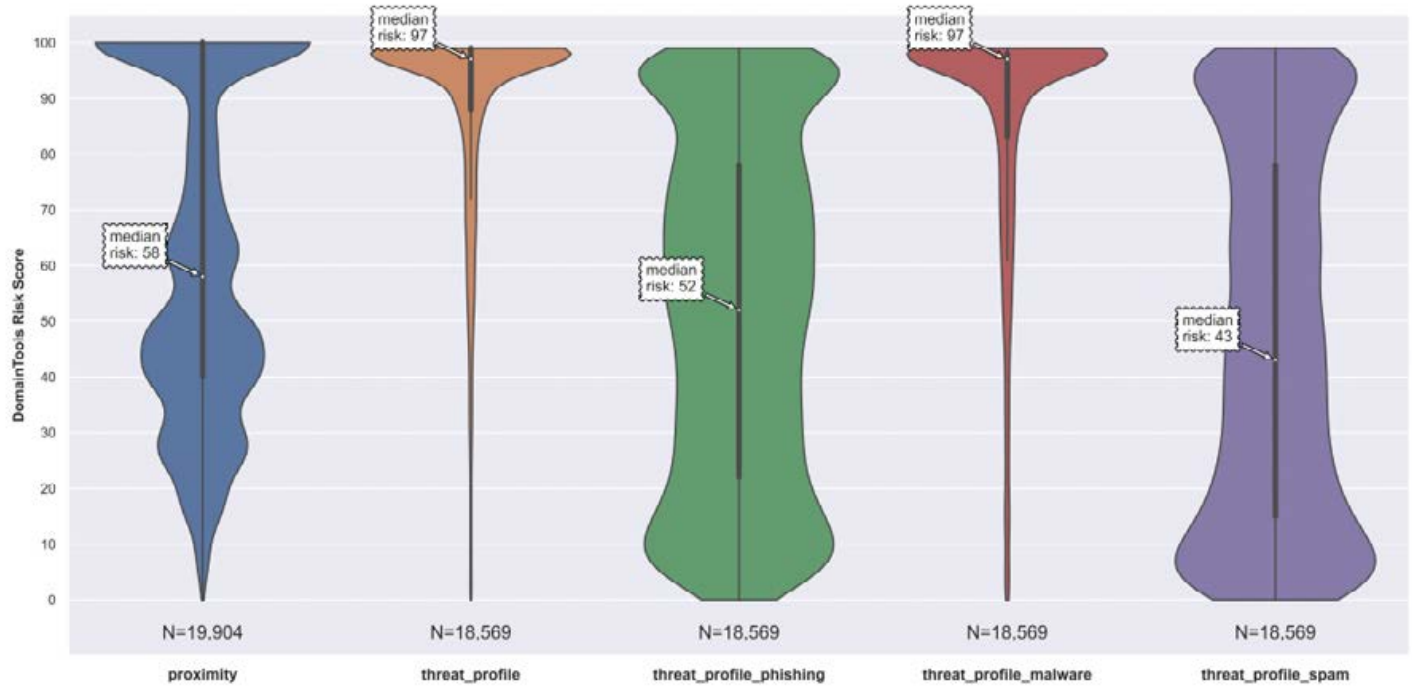
Private Layer AS51852 Risk Scores Breakdown (Computed on a Subset of Domains)



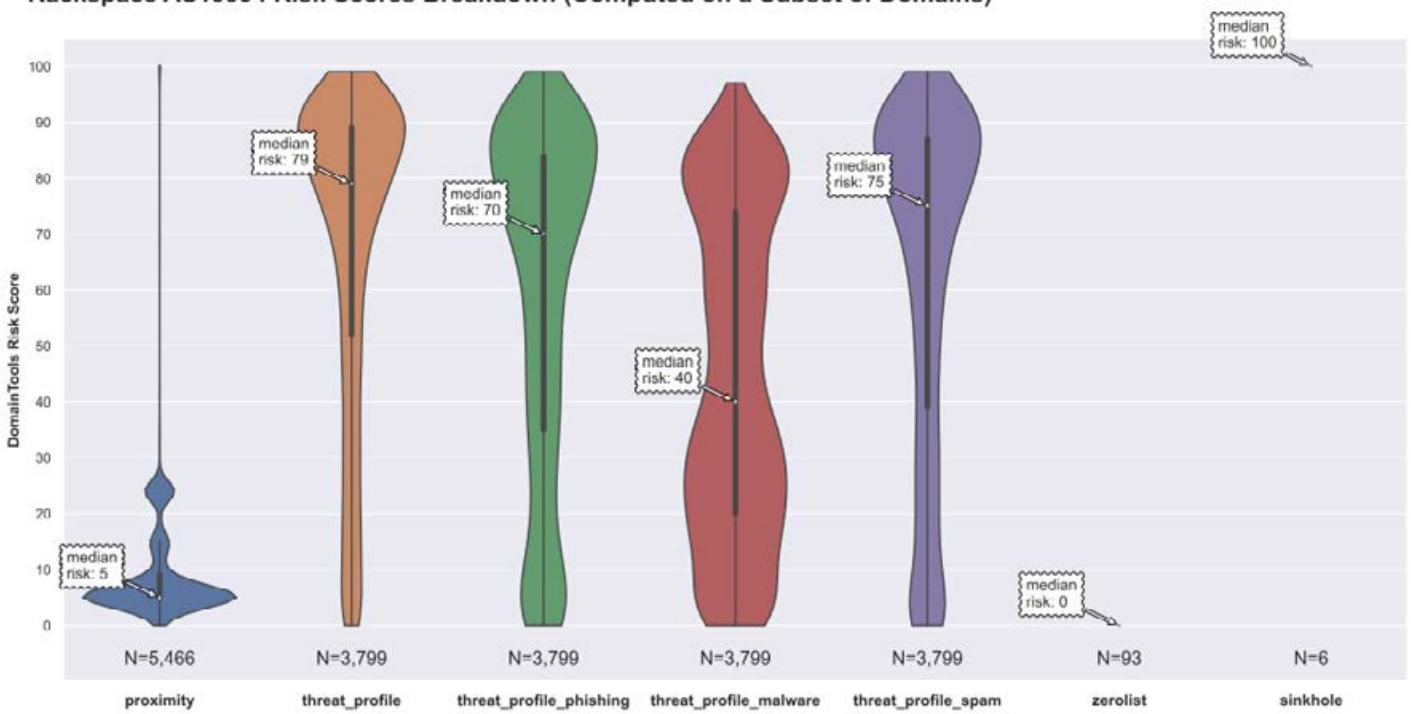
Quadranet AS8100 Risk Scores Breakdown (Computed on a Subset of Domains)



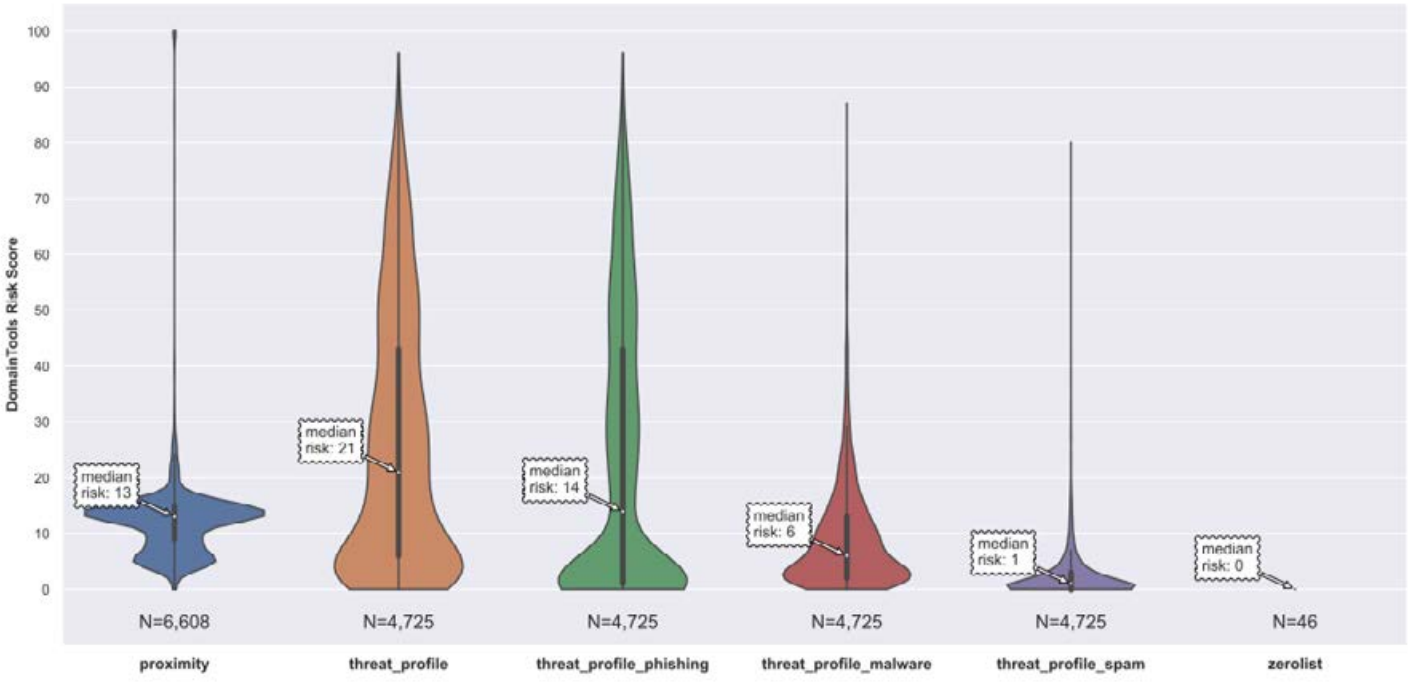
Quick Pick AS46261 Risk Scores Breakdown (Computed on a Subset of Domains)



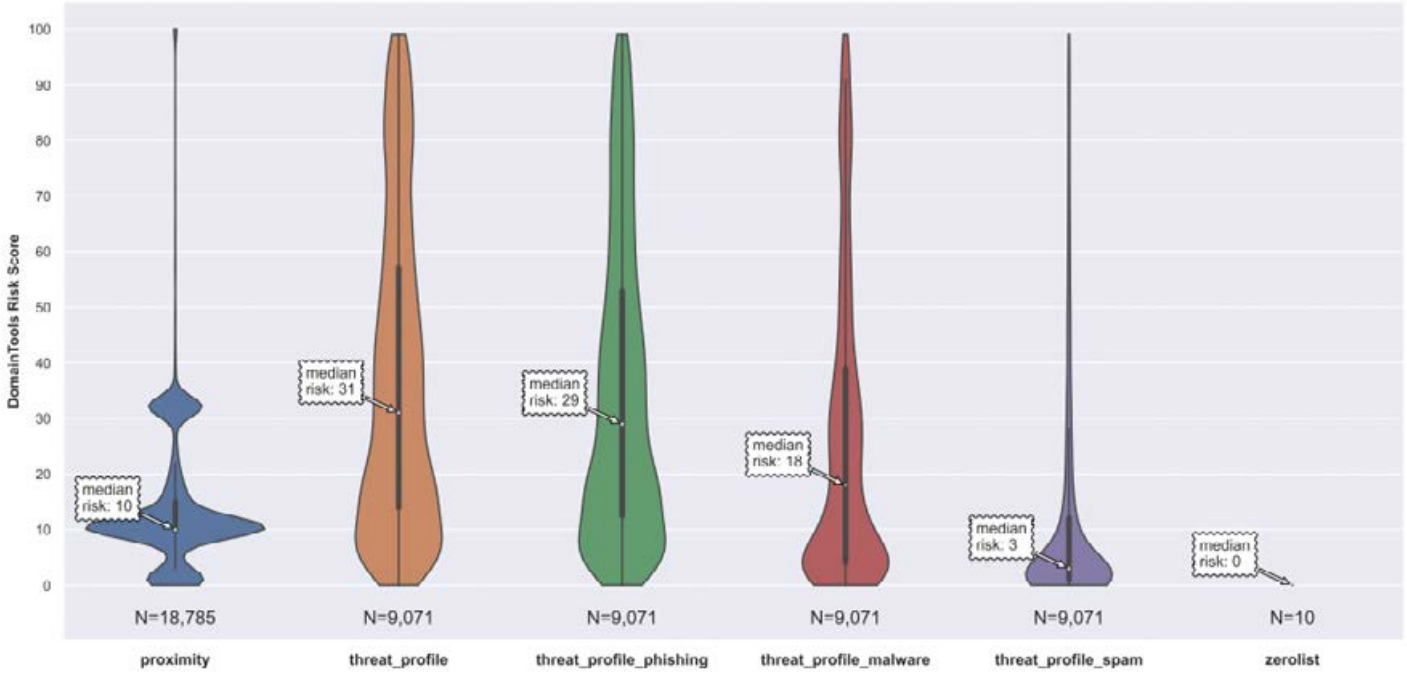
Rackspace AS19994 Risk Scores Breakdown (Computed on a Subset of Domains)



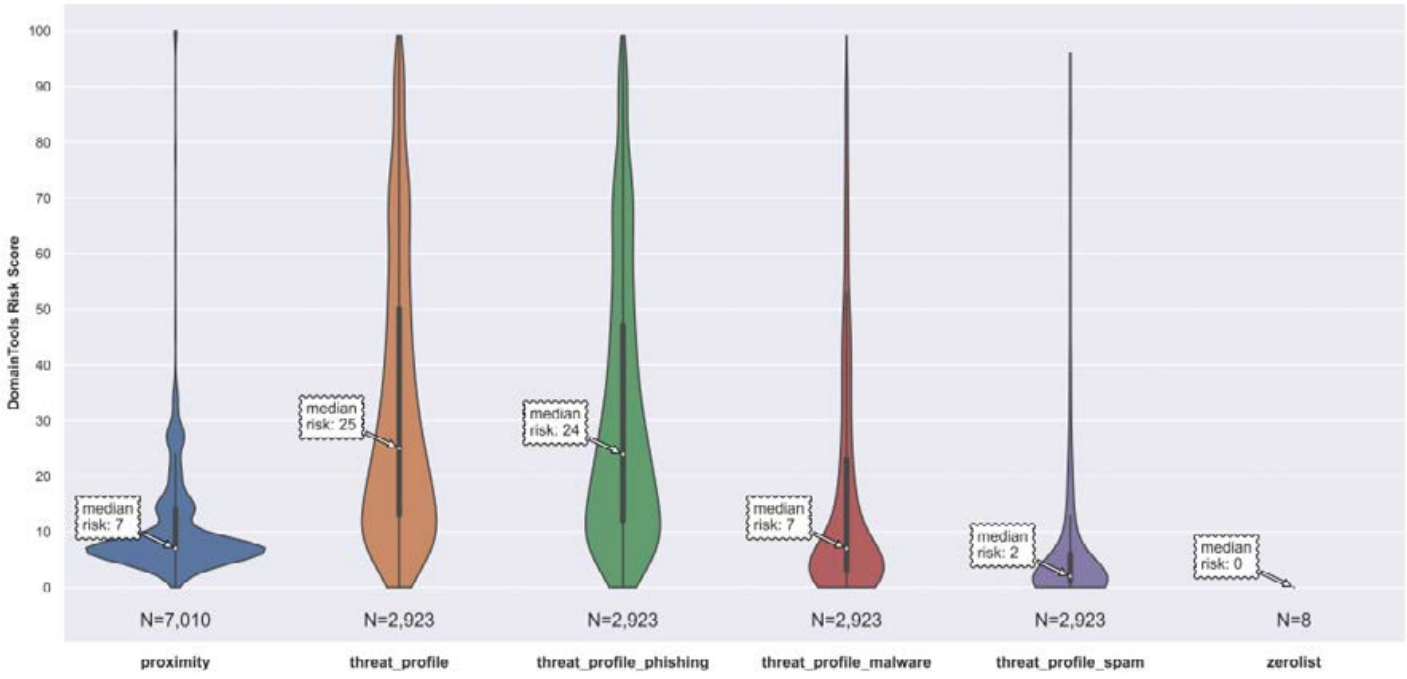
Rackspace AS33070 Risk Scores Breakdown (Computed on a Subset of Domains)



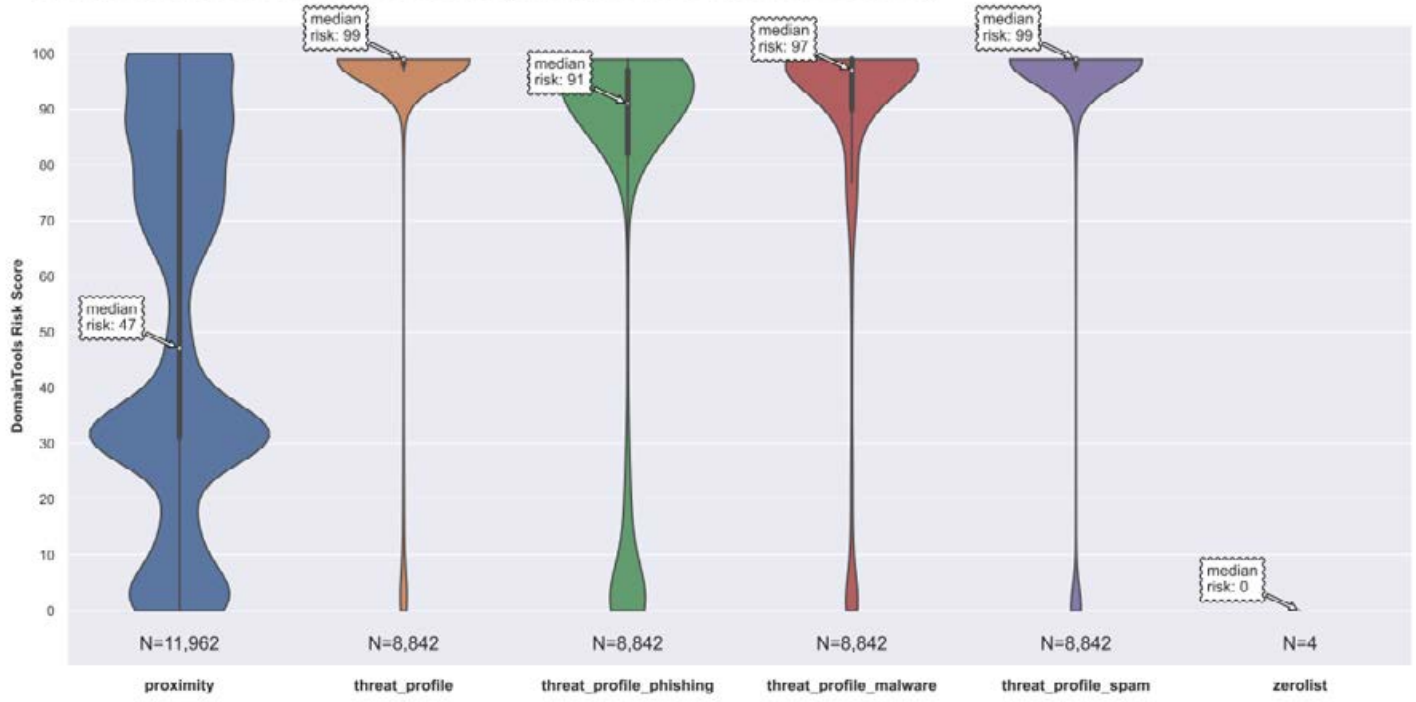
Regru AS197695 Risk Scores Breakdown (Computed on a Subset of Domains)



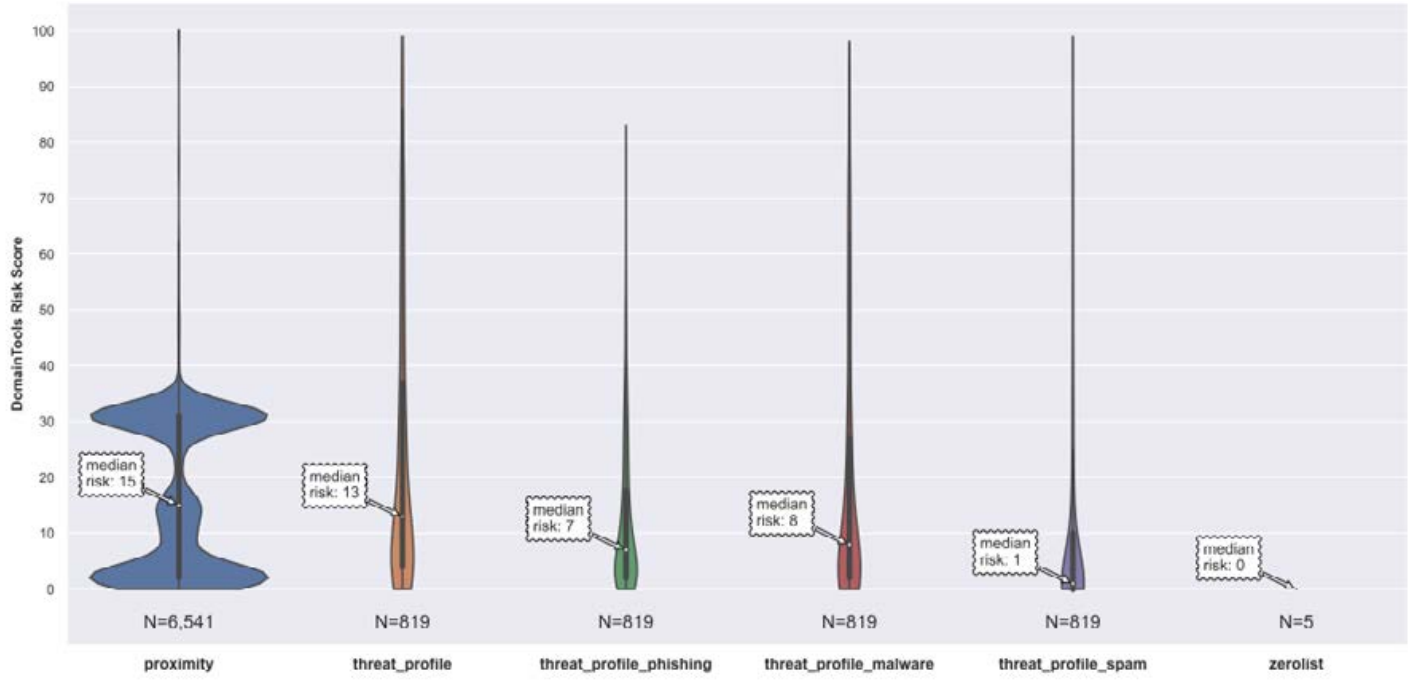
RU Center AS48287 Risk Scores Breakdown (Computed on a Subset of Domains)



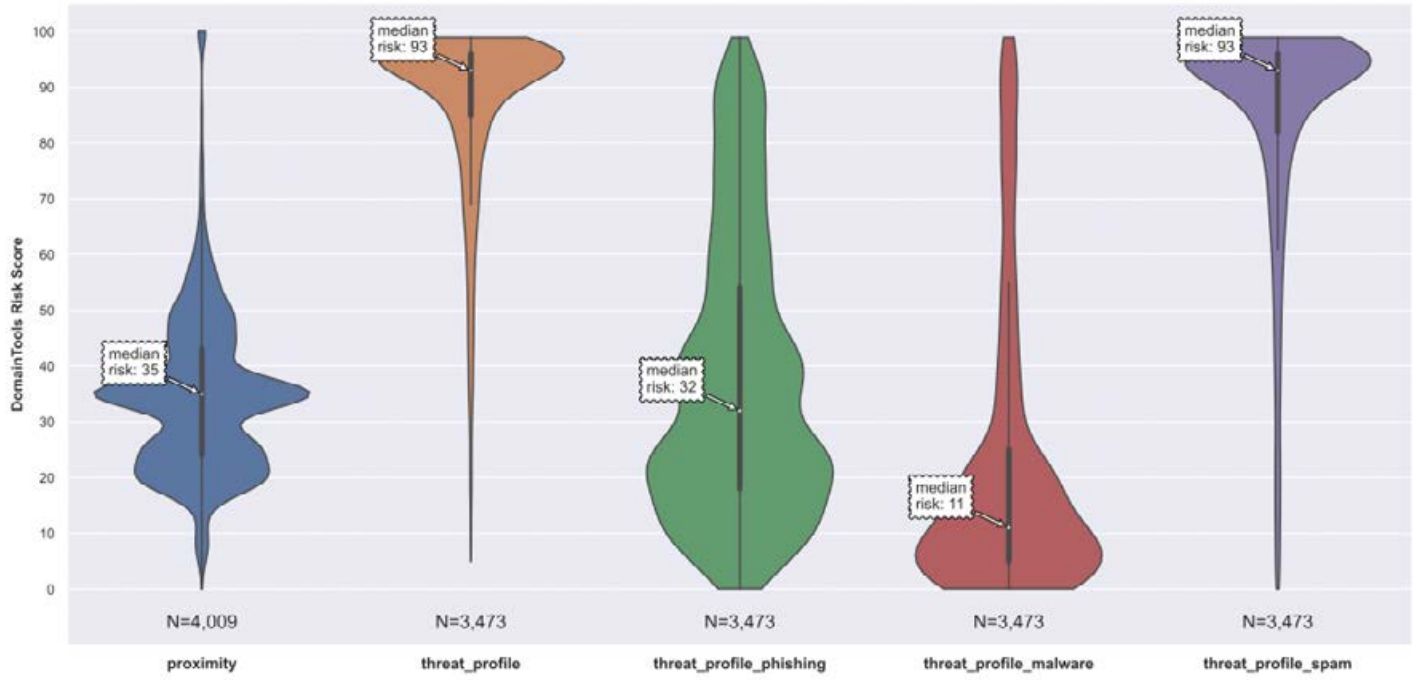
Sakura AS9370 Risk Scores Breakdown (Computed on a Subset of Domains)



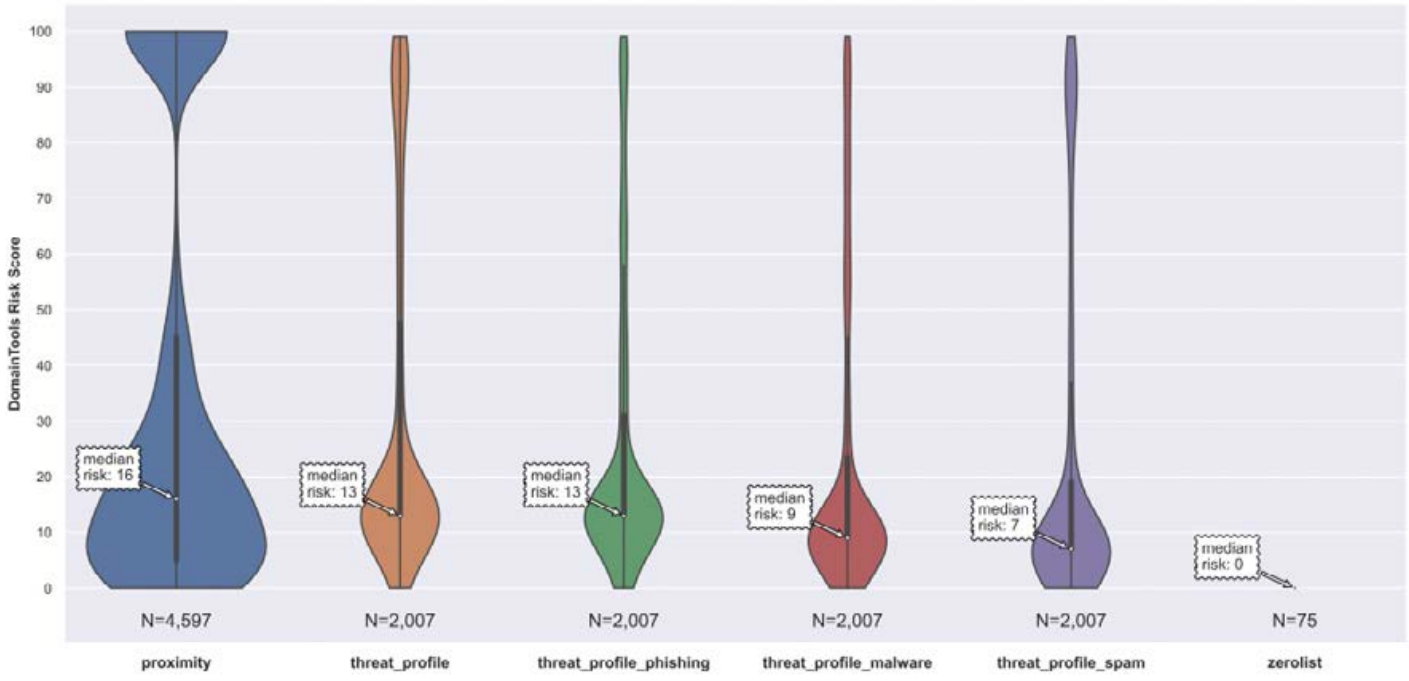
Sakura AS9371 Risk Scores Breakdown (Computed on a Subset of Domains)



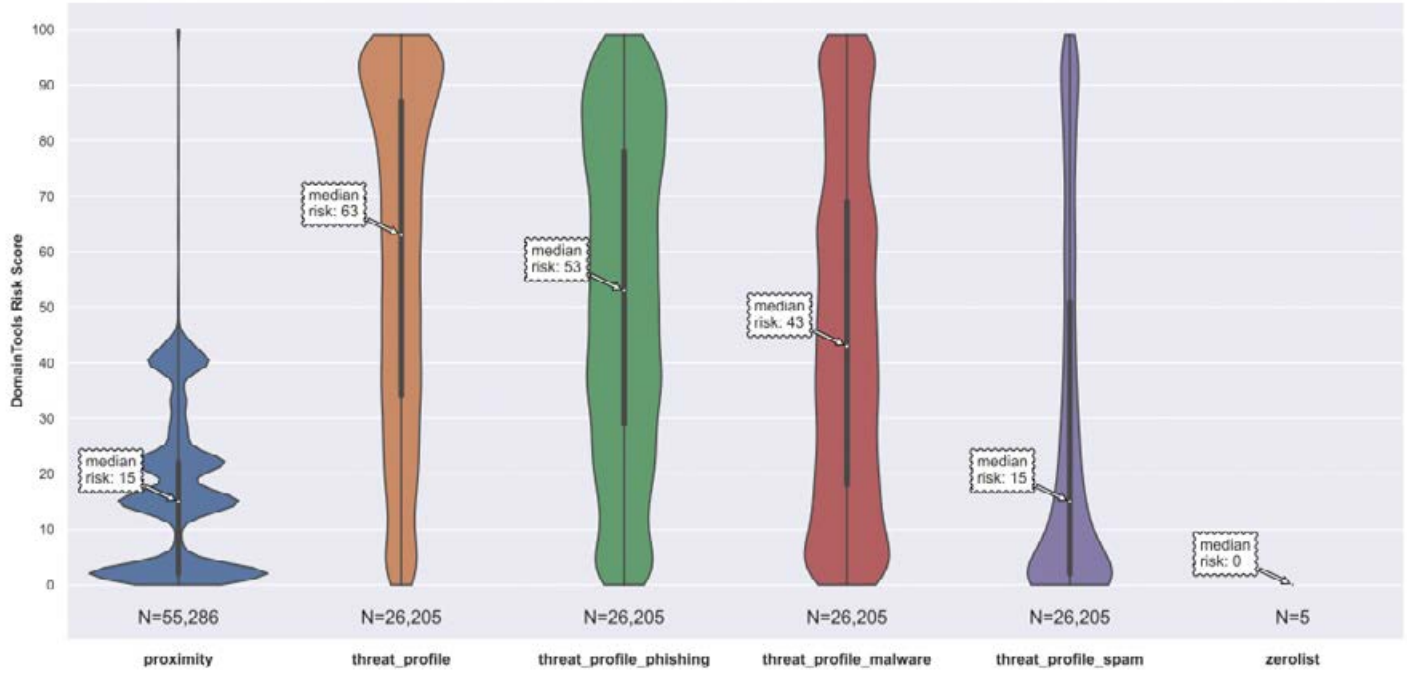
Sanren Data AS139330 Risk Scores Breakdown (Computed on a Subset of Domains)



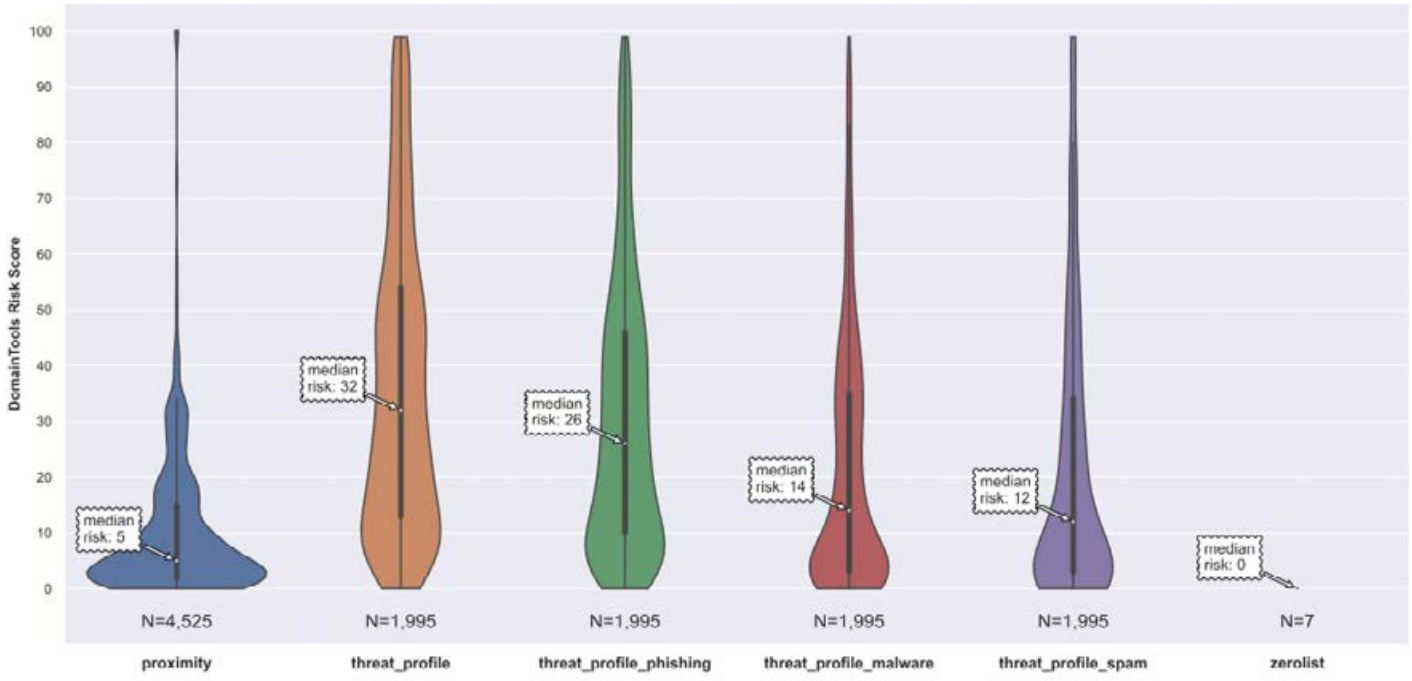
Scaleway AS12876 Risk Scores Breakdown (Computed on a Subset of Domains)



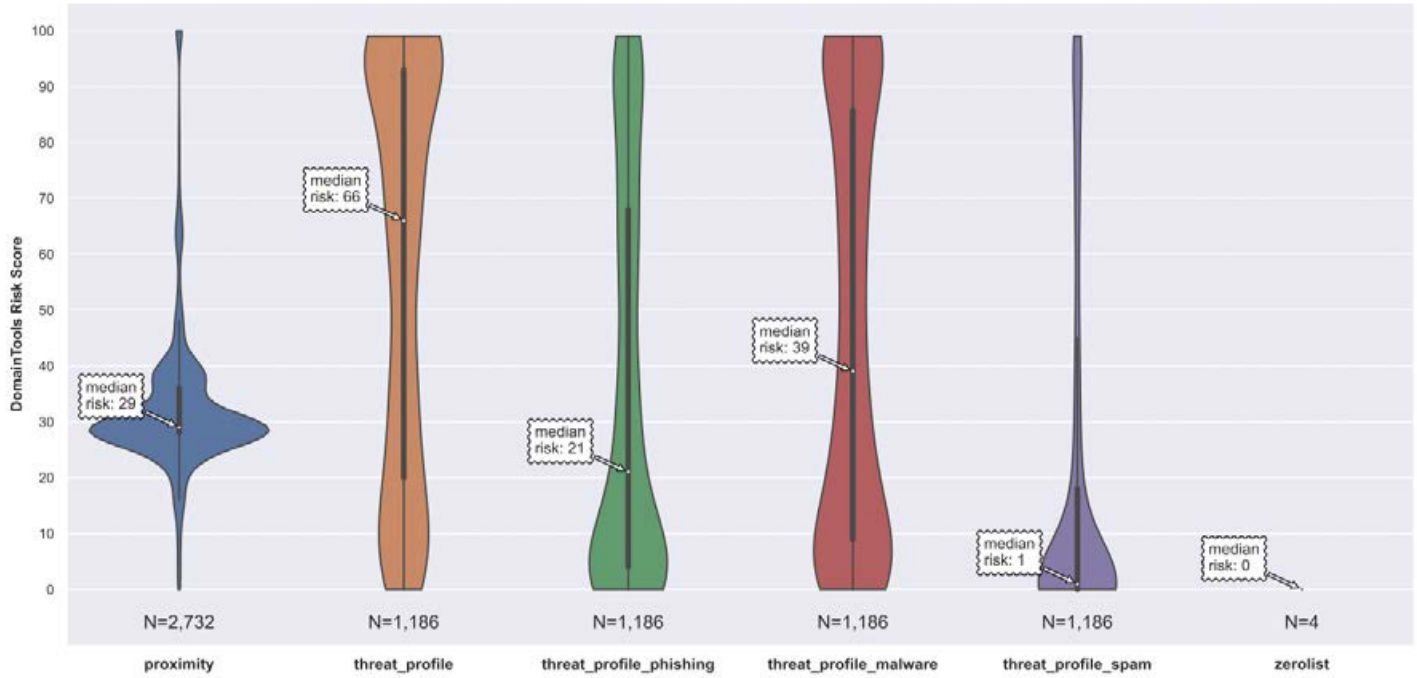
Sedo AS47846 Risk Scores Breakdown (Computed on a Subset of Domains)



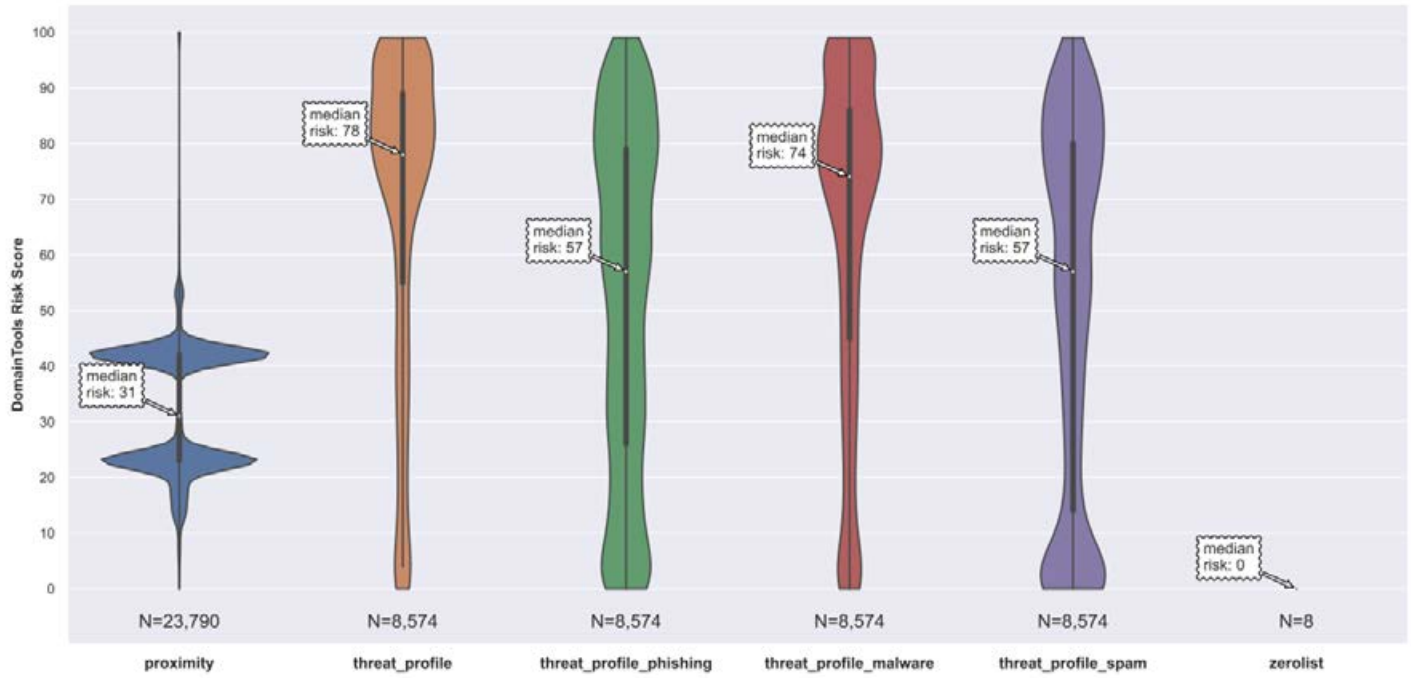
Selectel AS49505 Risk Scores Breakdown (Computed on a Subset of Domains)



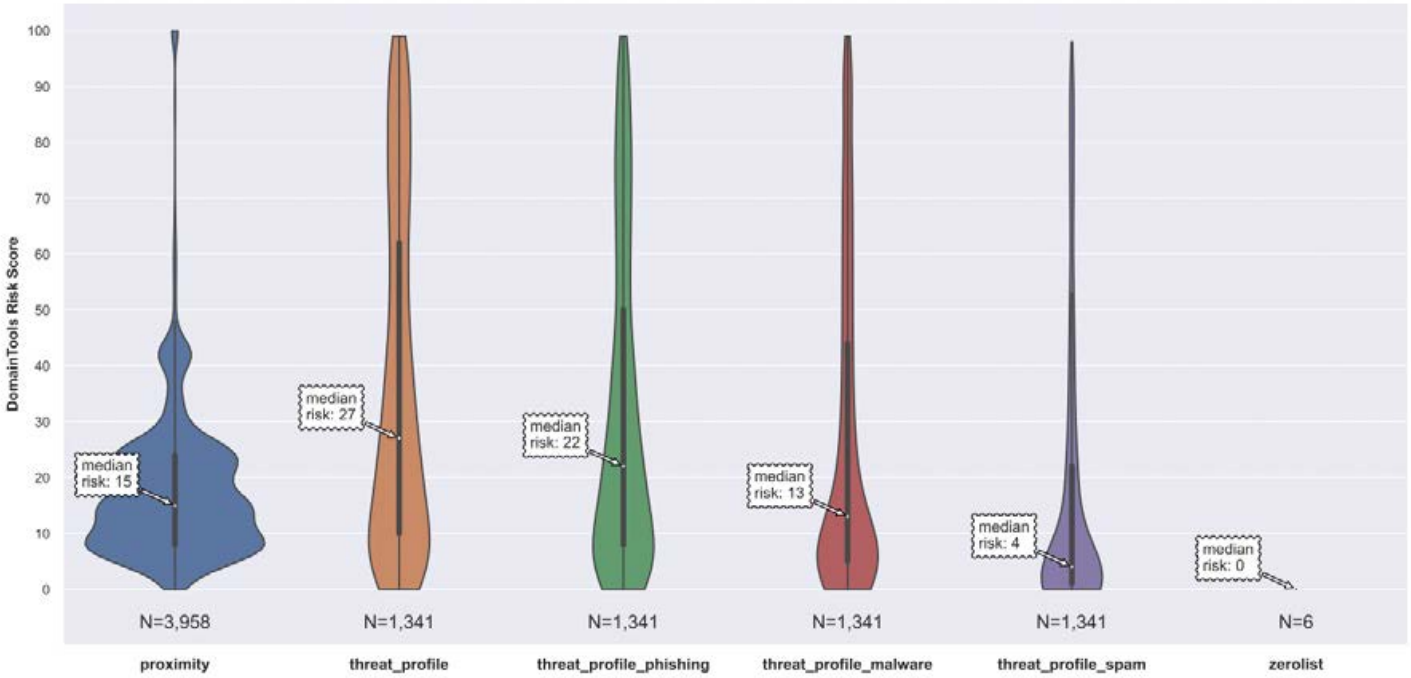
ServiHosting AS29119 Risk Scores Breakdown (Computed on a Subset of Domains)



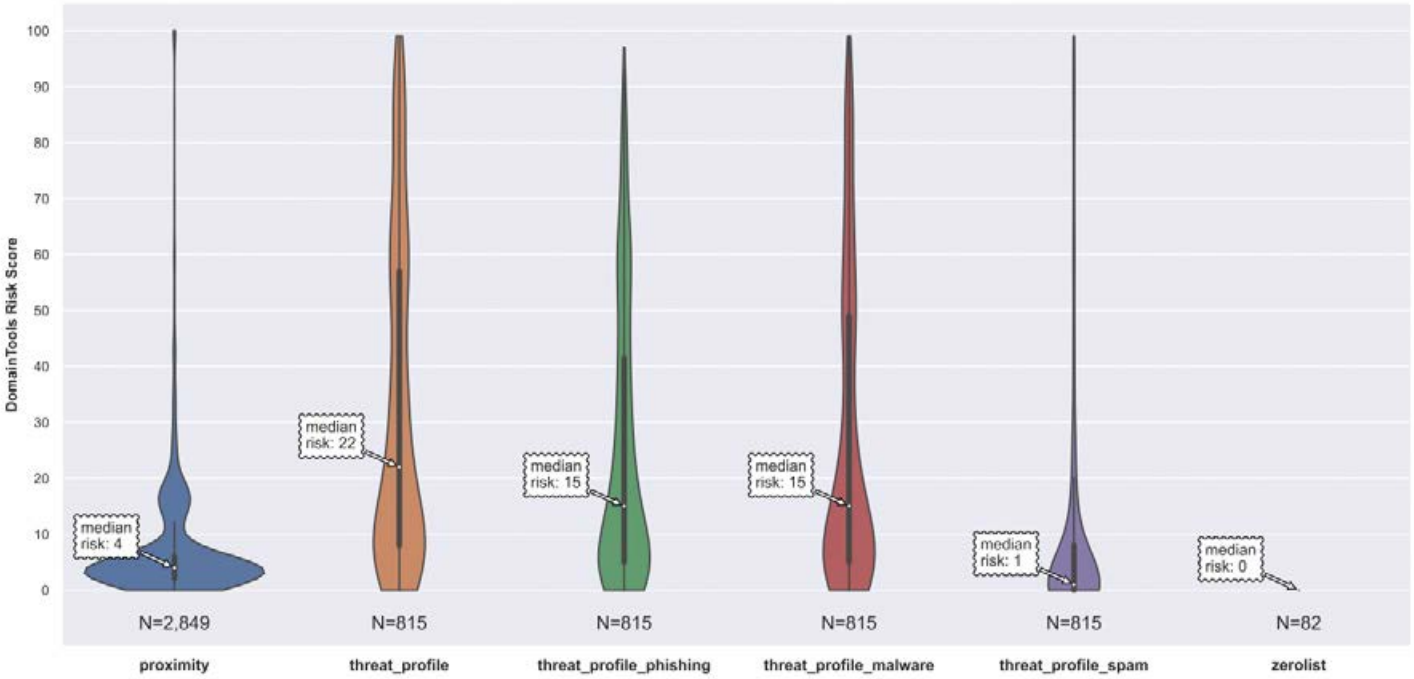
Sharktech AS46844 Risk Scores Breakdown (Computed on a Subset of Domains)



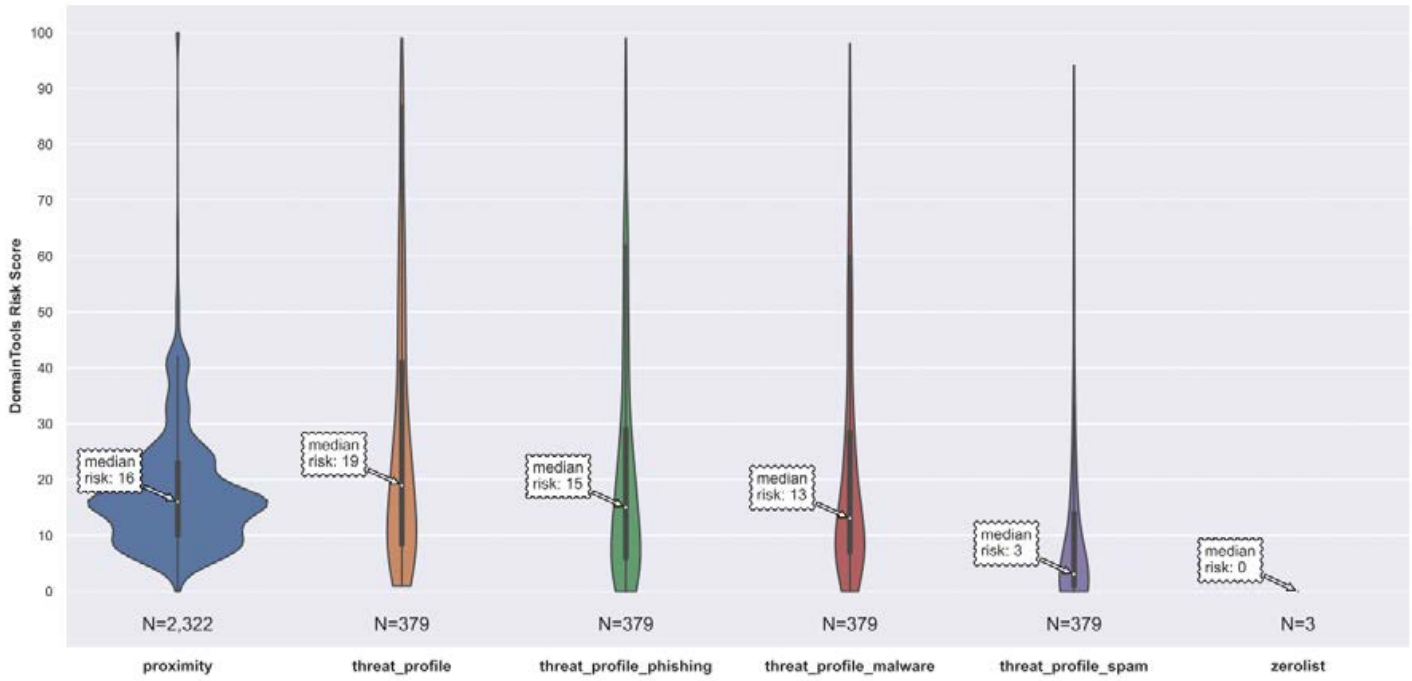
Singlehop AS32475 Risk Scores Breakdown (Computed on a Subset of Domains)



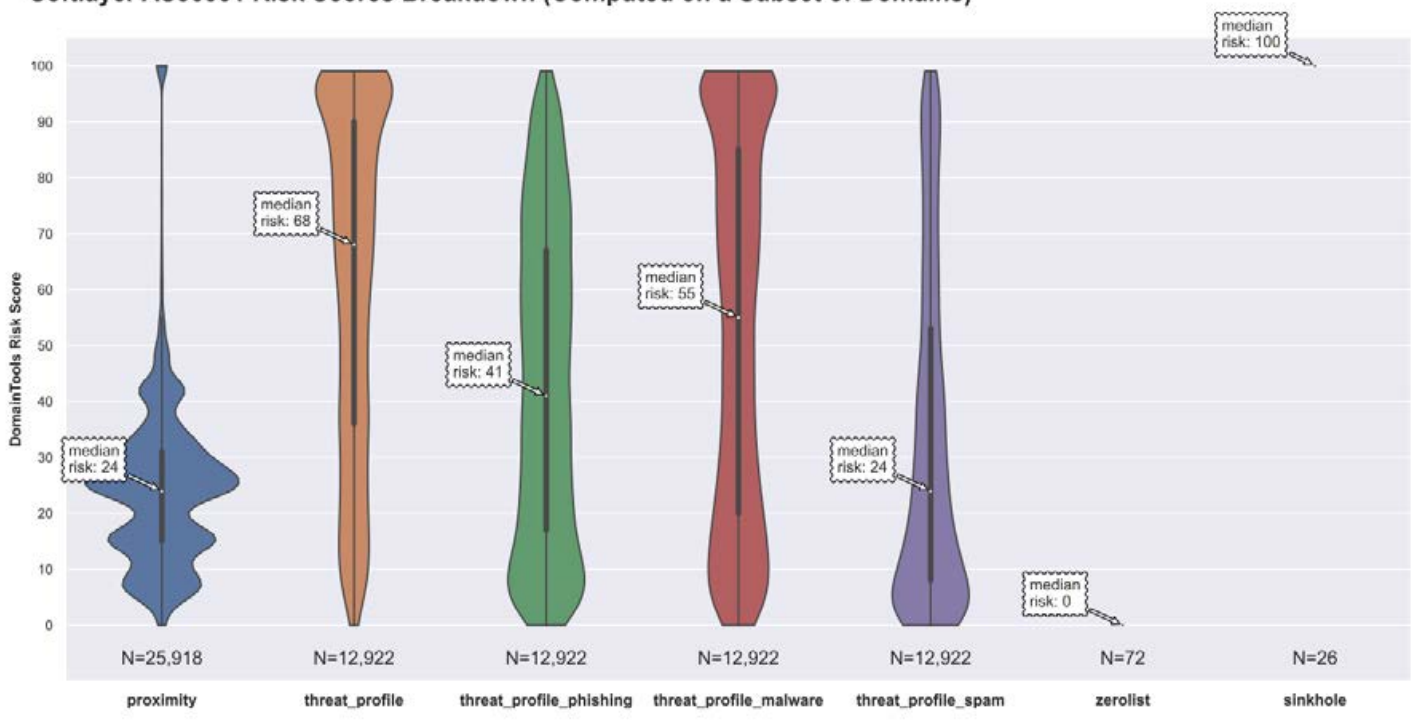
SK Broadband AS9318 Risk Scores Breakdown (Computed on a Subset of Domains)



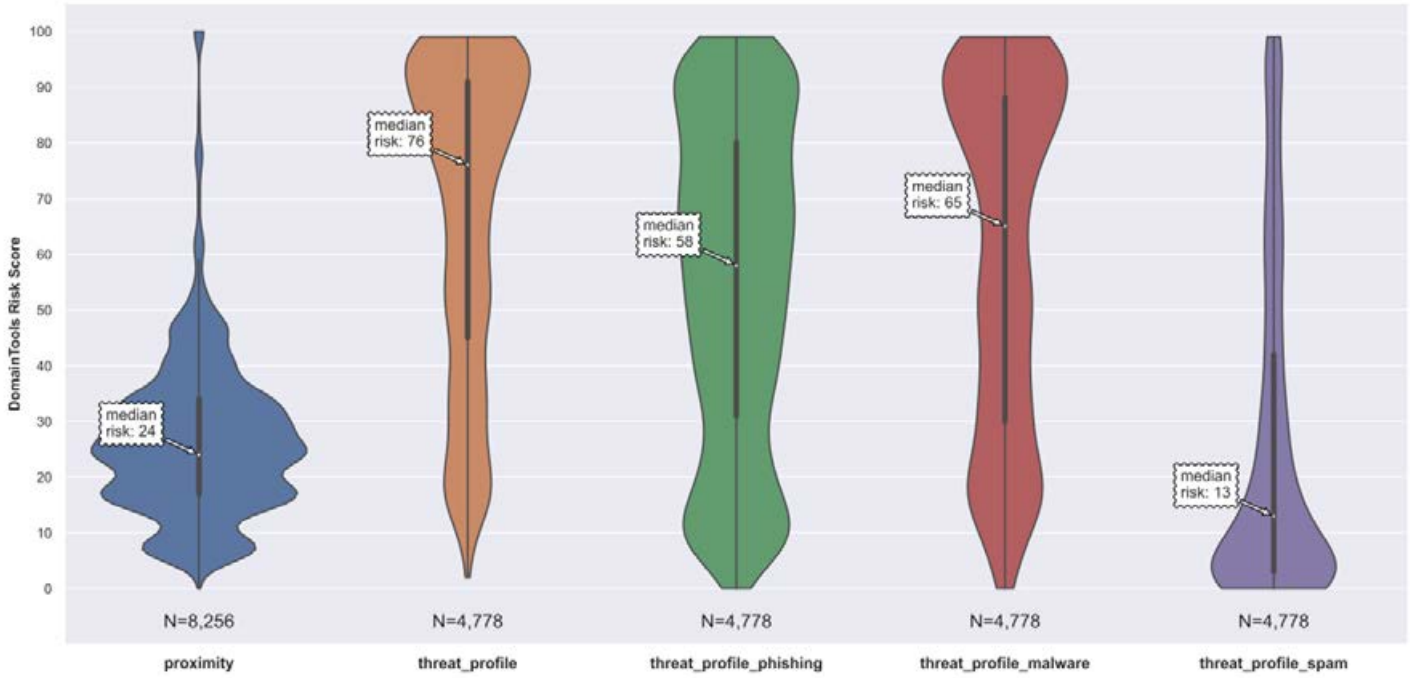
Small Orange AS62729 Risk Scores Breakdown (Computed on a Subset of Domains)



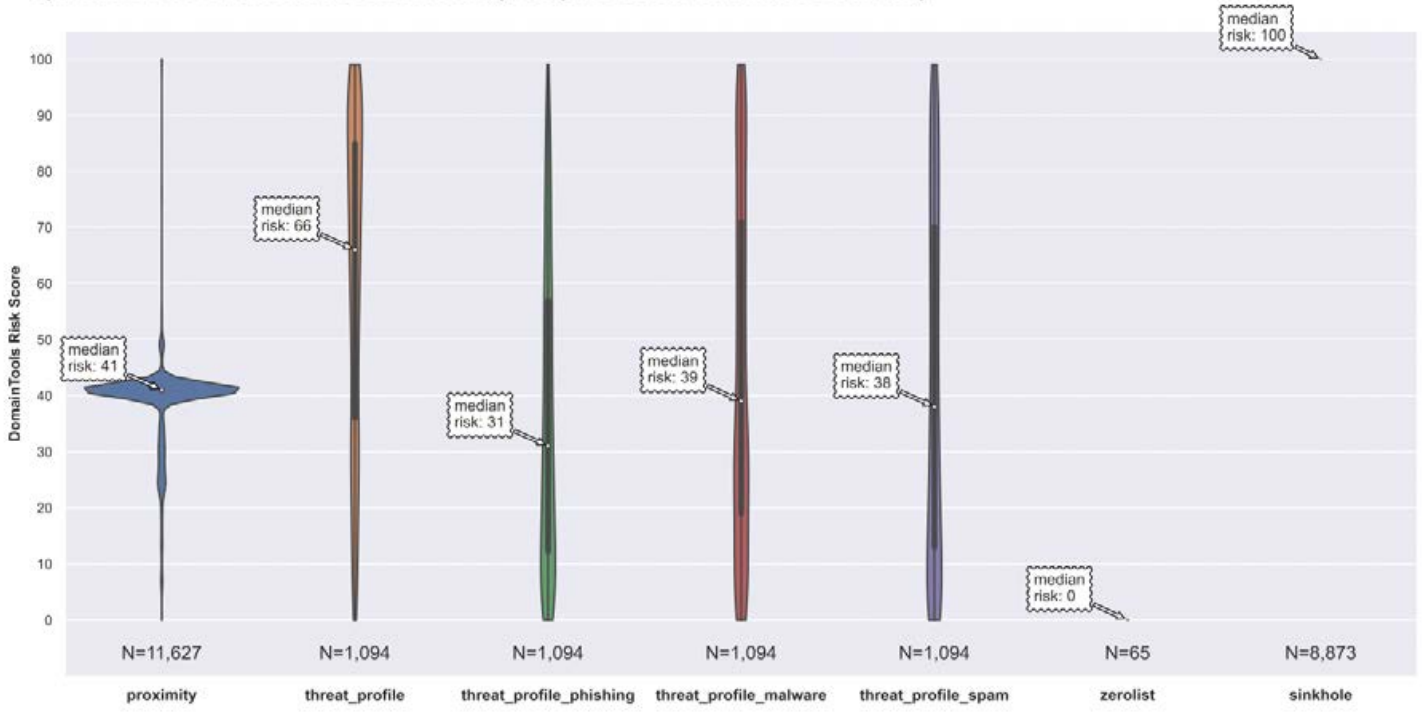
Softlayer AS36351 Risk Scores Breakdown (Computed on a Subset of Domains)



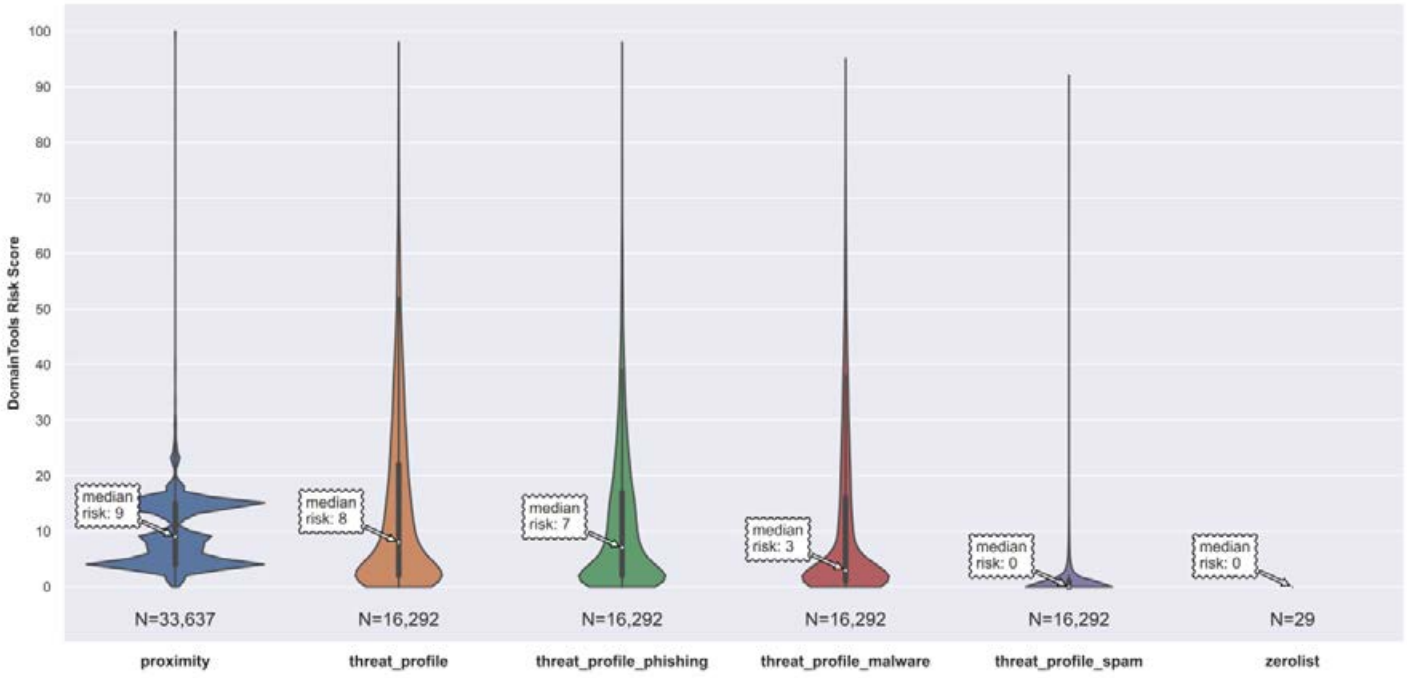
Sonder Cloud AS133199 Risk Scores Breakdown (Computed on a Subset of Domains)



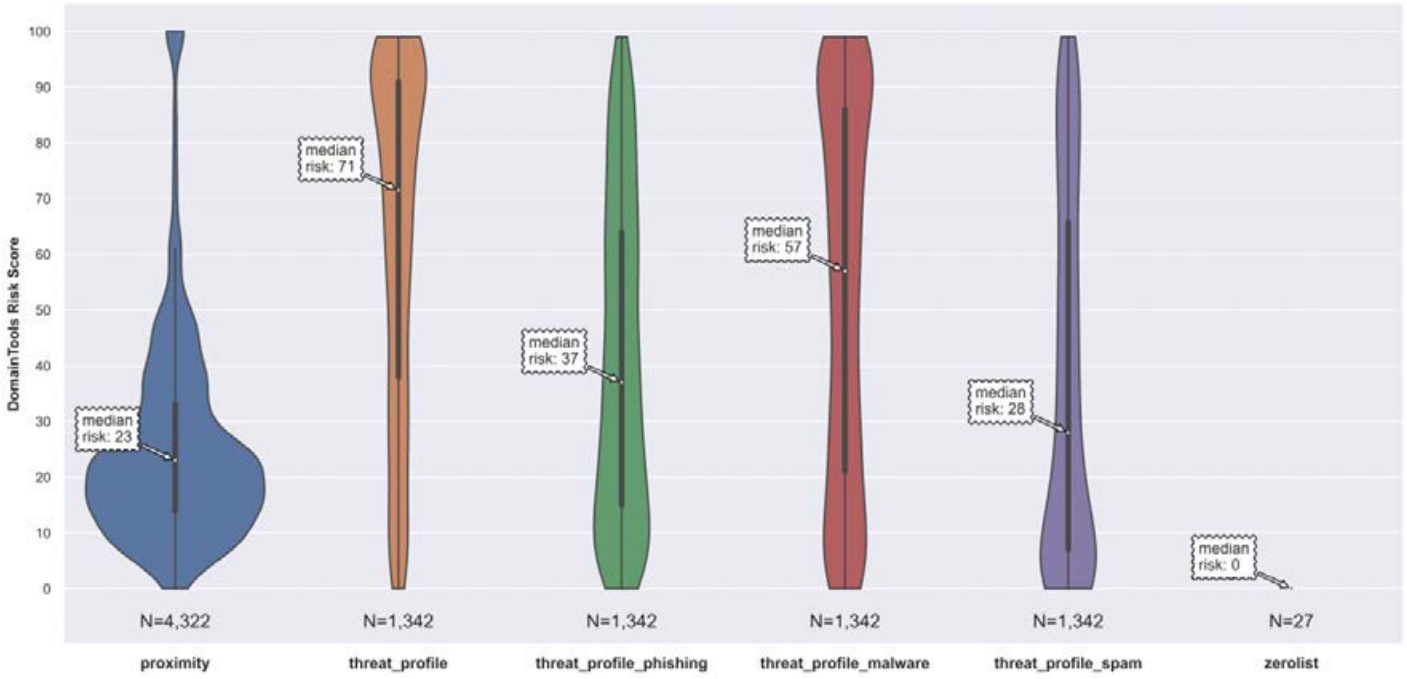
Sprint AS1239 Risk Scores Breakdown (Computed on a Subset of Domains)



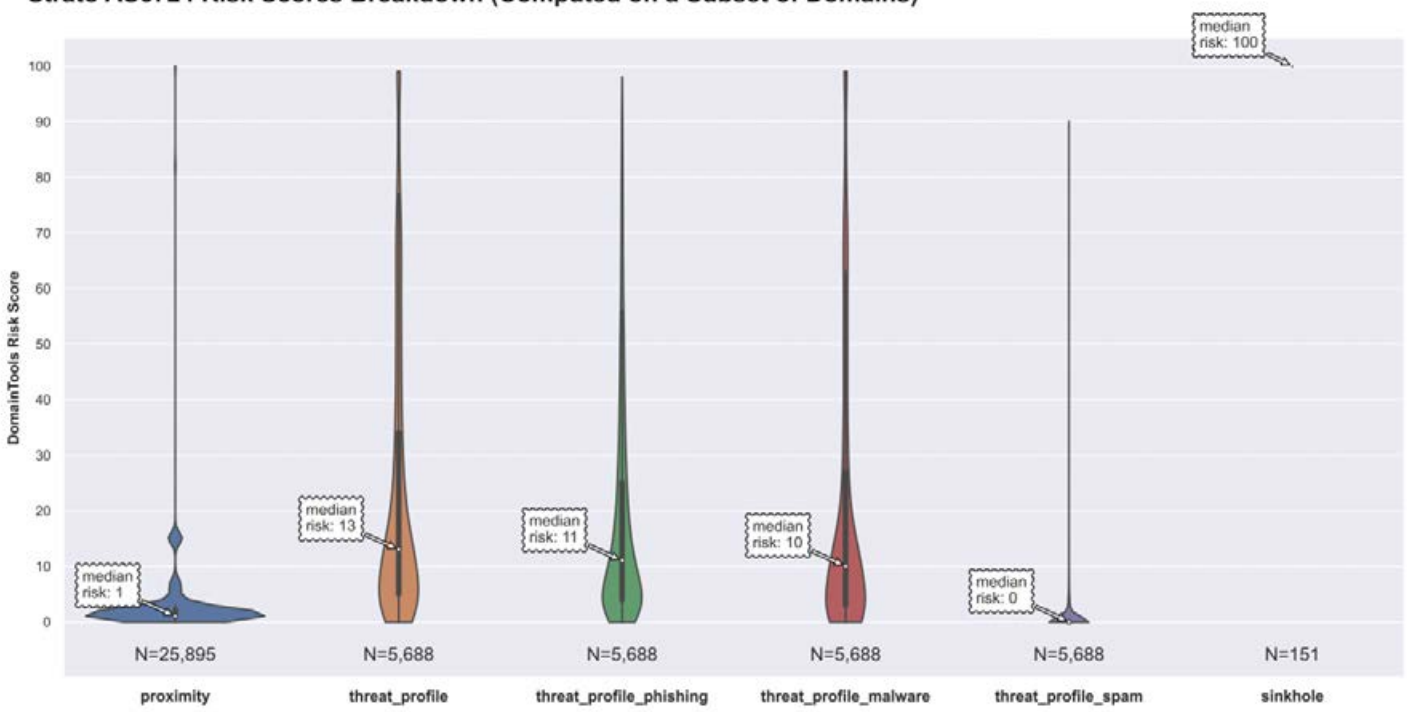
Squarespace AS53831 Risk Scores Breakdown (Computed on a Subset of Domains)



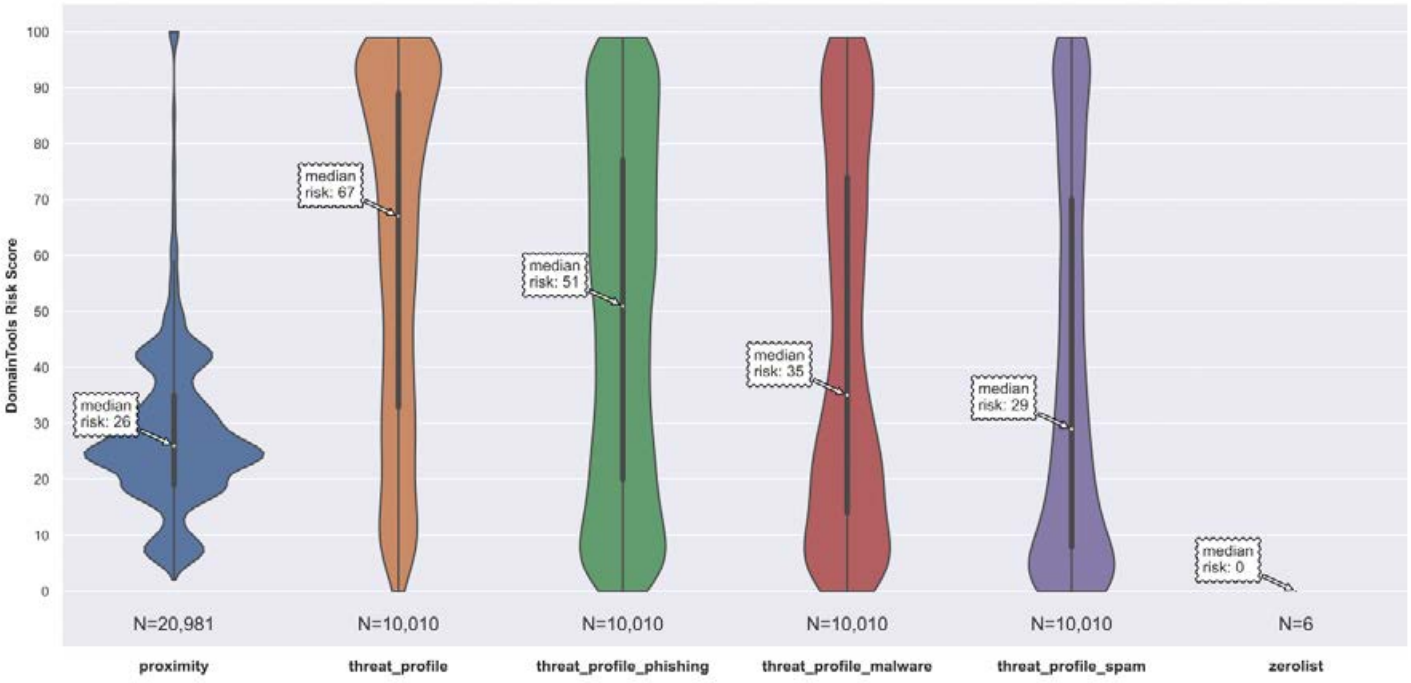
Stackpath AS33438 Risk Scores Breakdown (Computed on a Subset of Domains)



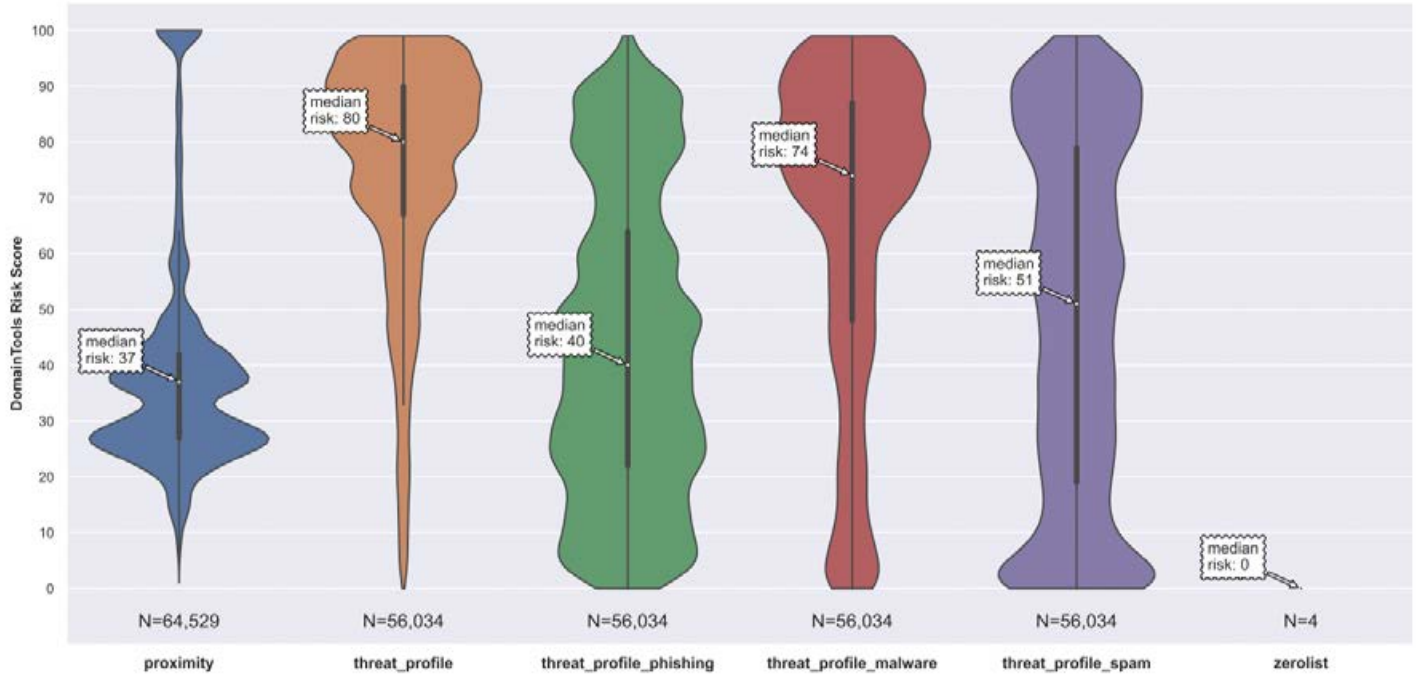
Strato AS6724 Risk Scores Breakdown (Computed on a Subset of Domains)



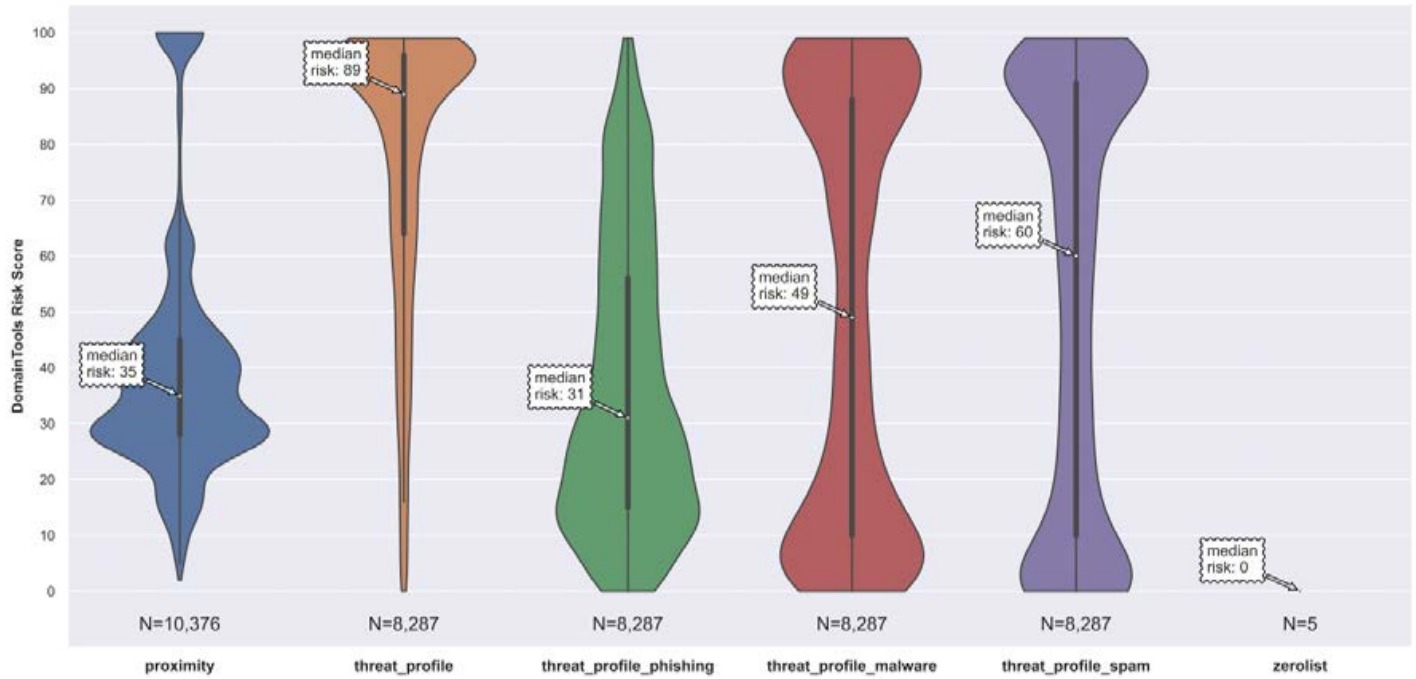
Sun Net AS38197 Risk Scores Breakdown (Computed on a Subset of Domains)



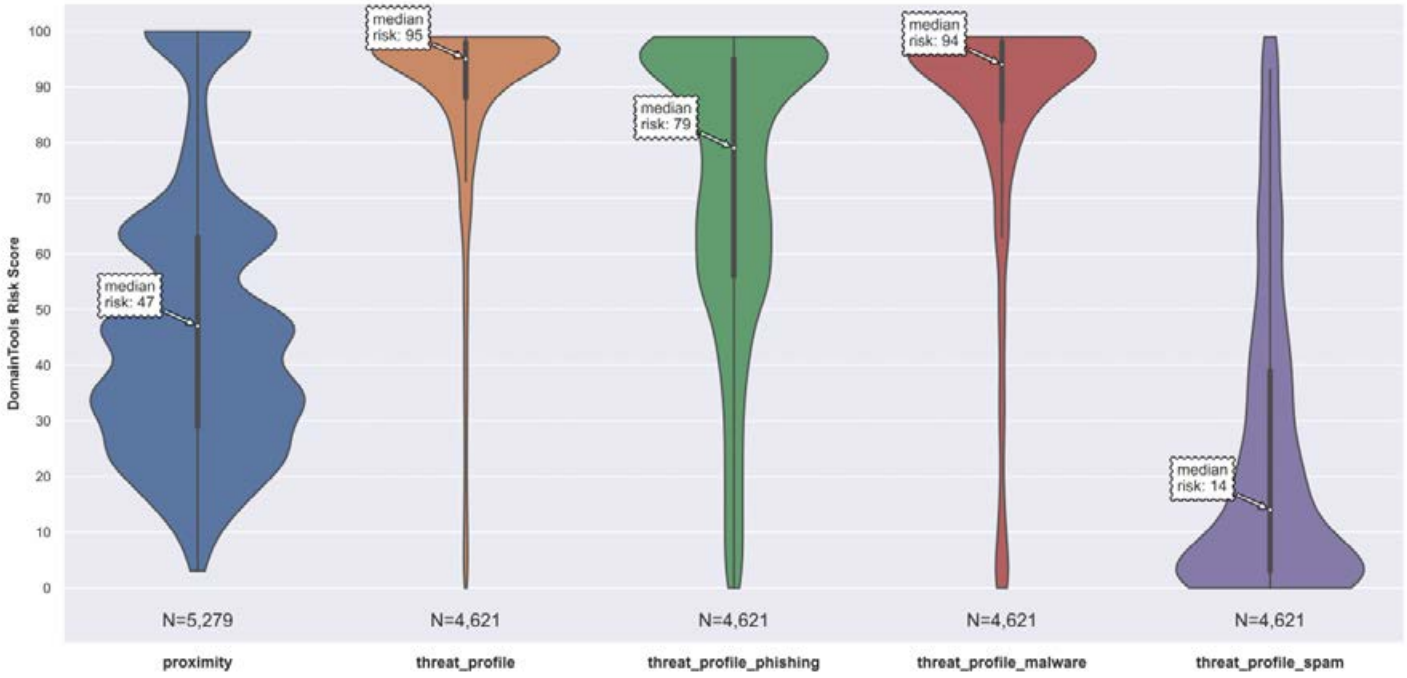
Sun Net AS136800 Risk Scores Breakdown (Computed on a Subset of Domains)



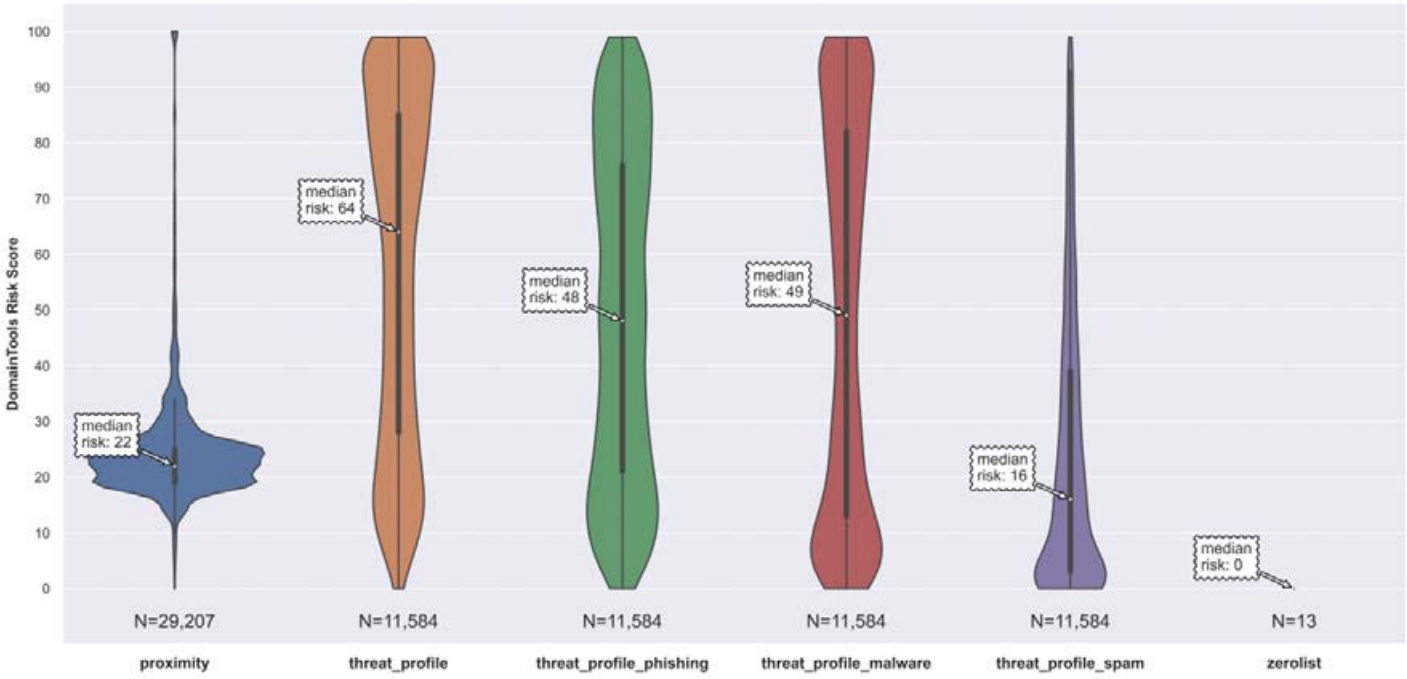
Take 2 Host AS20248 Risk Scores Breakdown (Computed on a Subset of Domains)



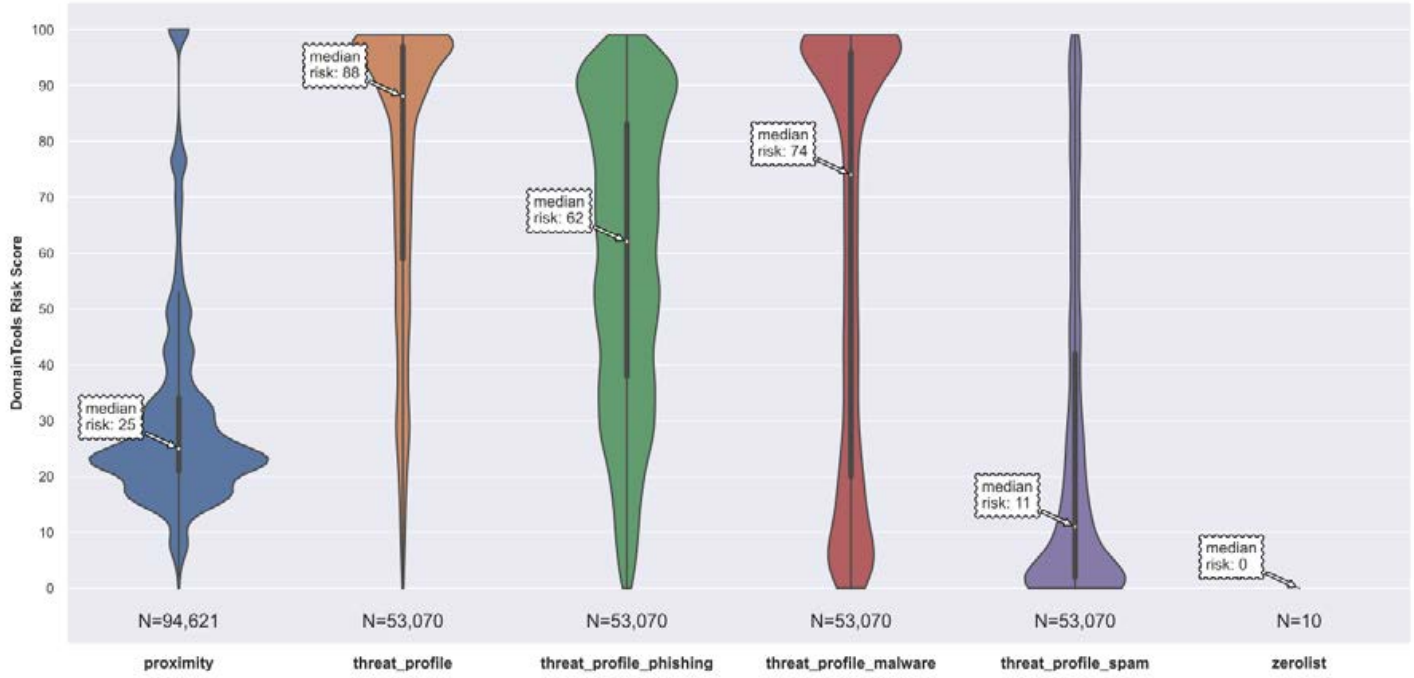
Tcloudnet AS399077 Risk Scores Breakdown (Computed on a Subset of Domains)



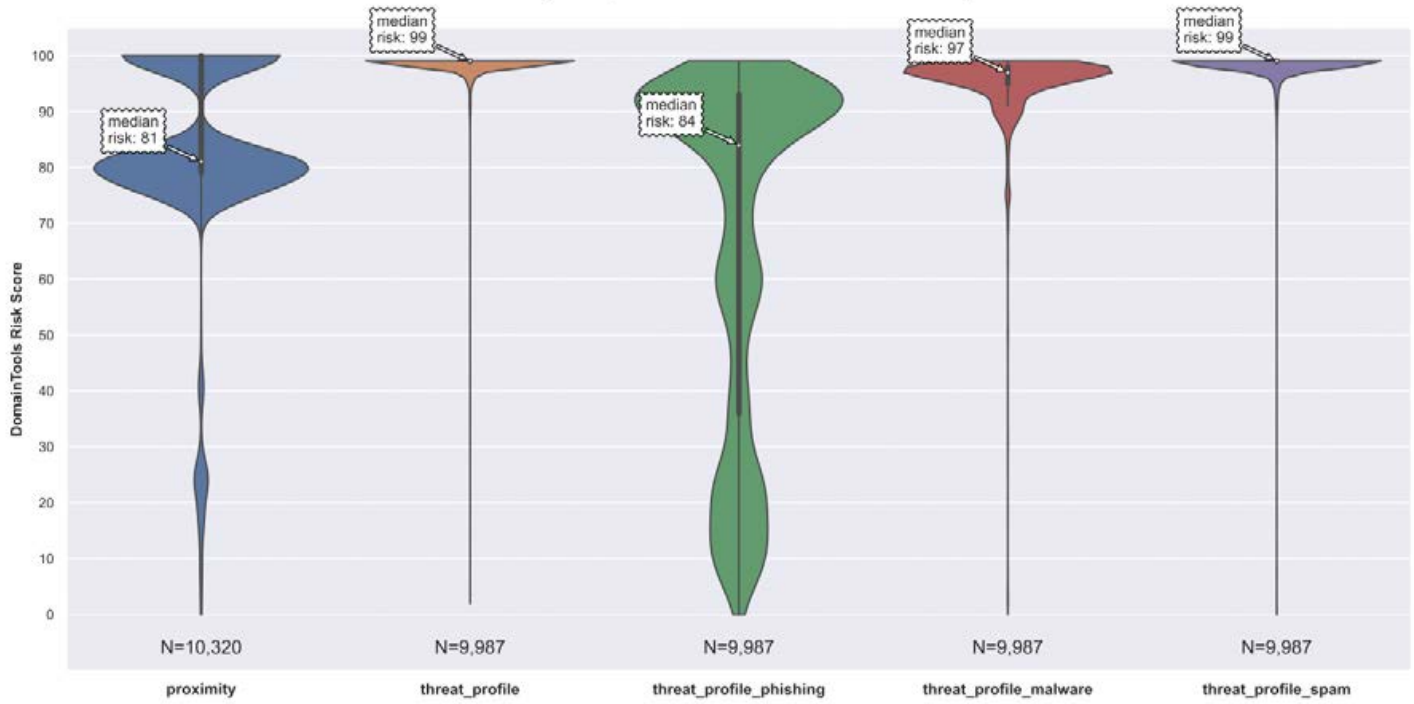
Tencent AS45090 Risk Scores Breakdown (Computed on a Subset of Domains)



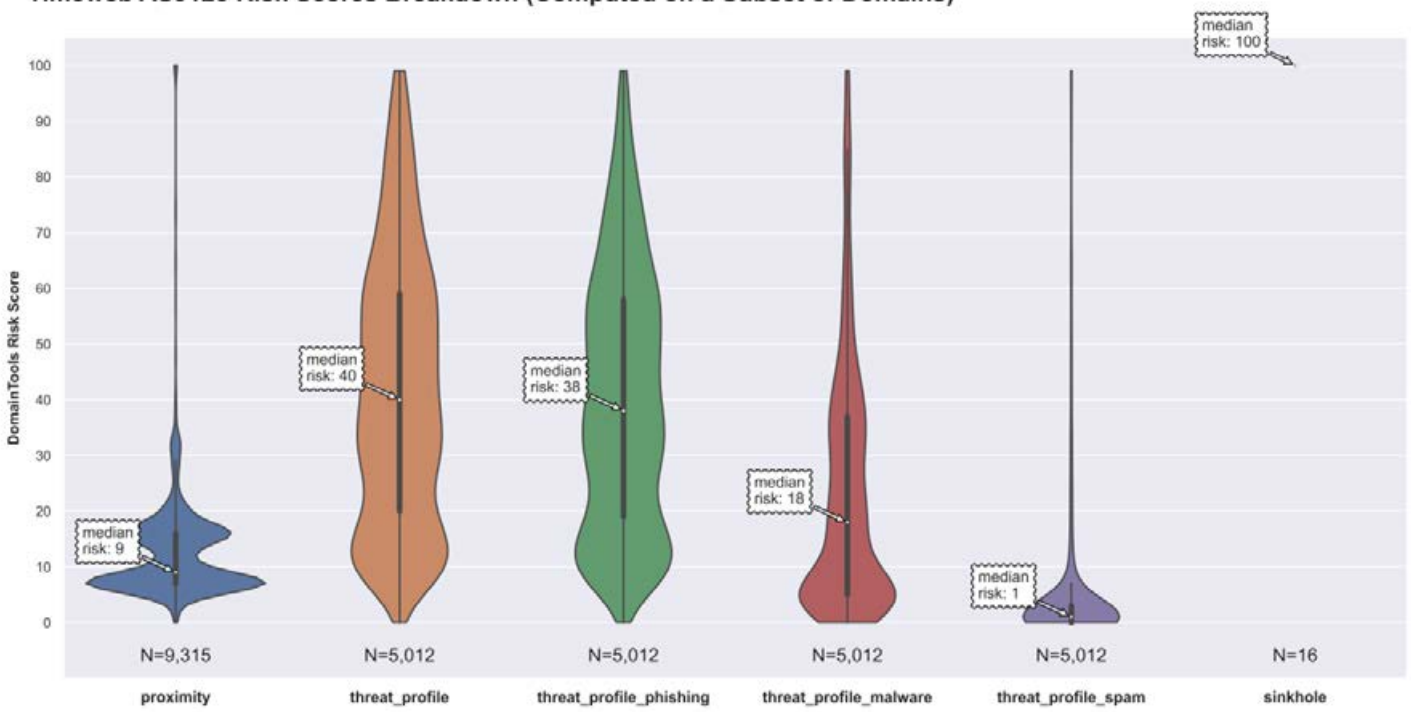
Tencent AS132203 Risk Scores Breakdown (Computed on a Subset of Domains)



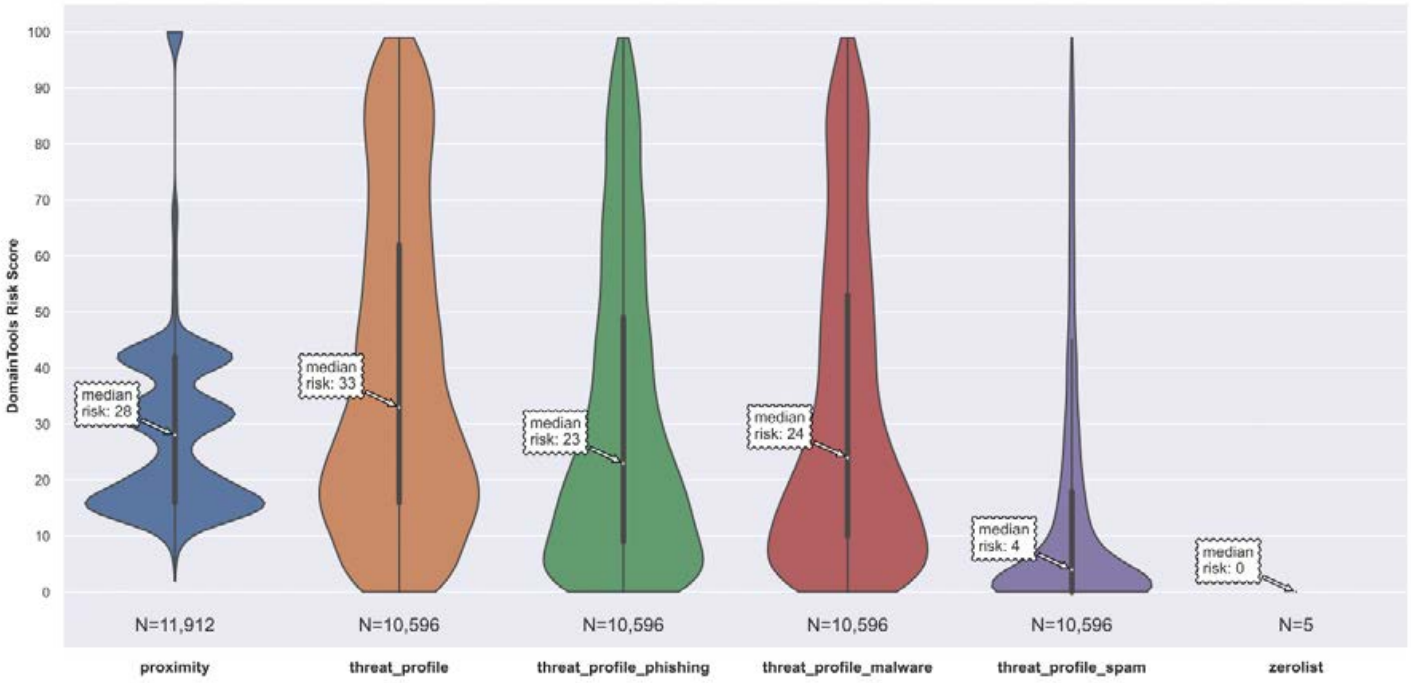
Tier Net AS397423 Risk Scores Breakdown (Computed on a Subset of Domains)



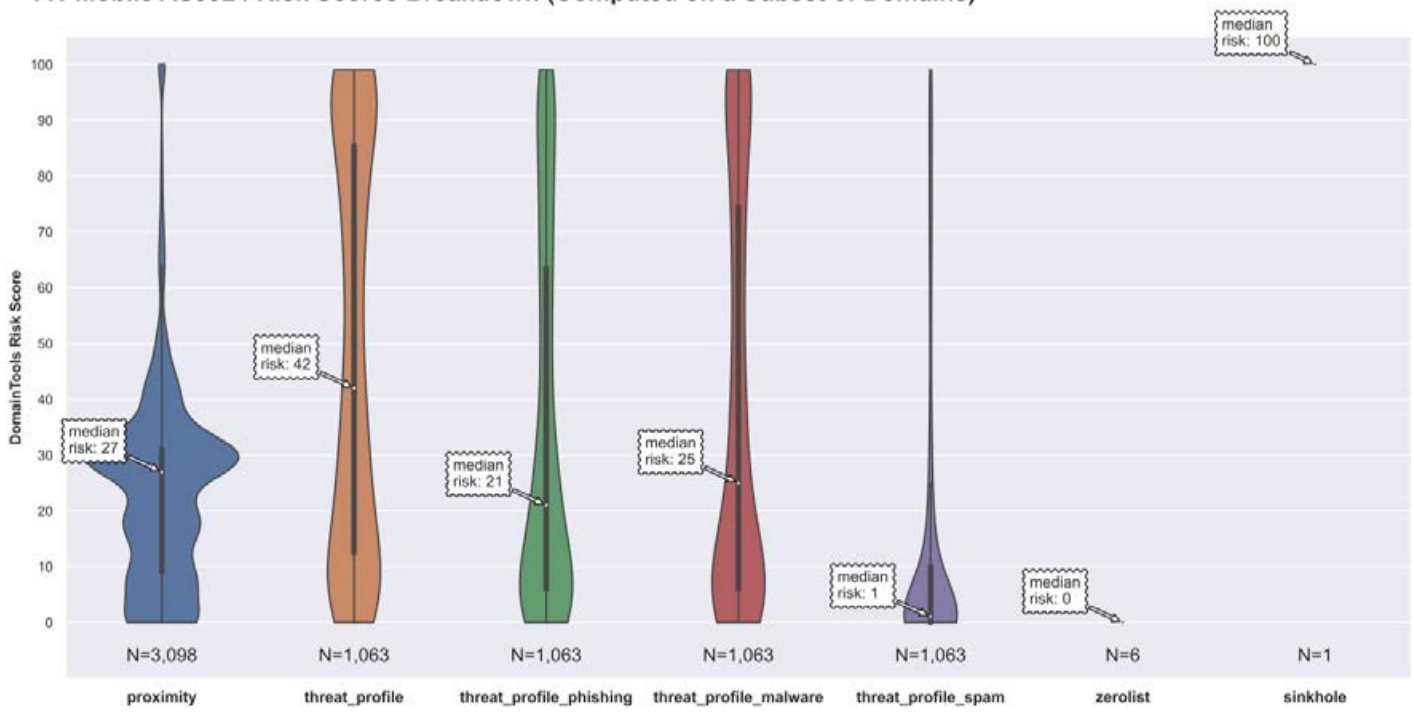
Timeweb AS9123 Risk Scores Breakdown (Computed on a Subset of Domains)



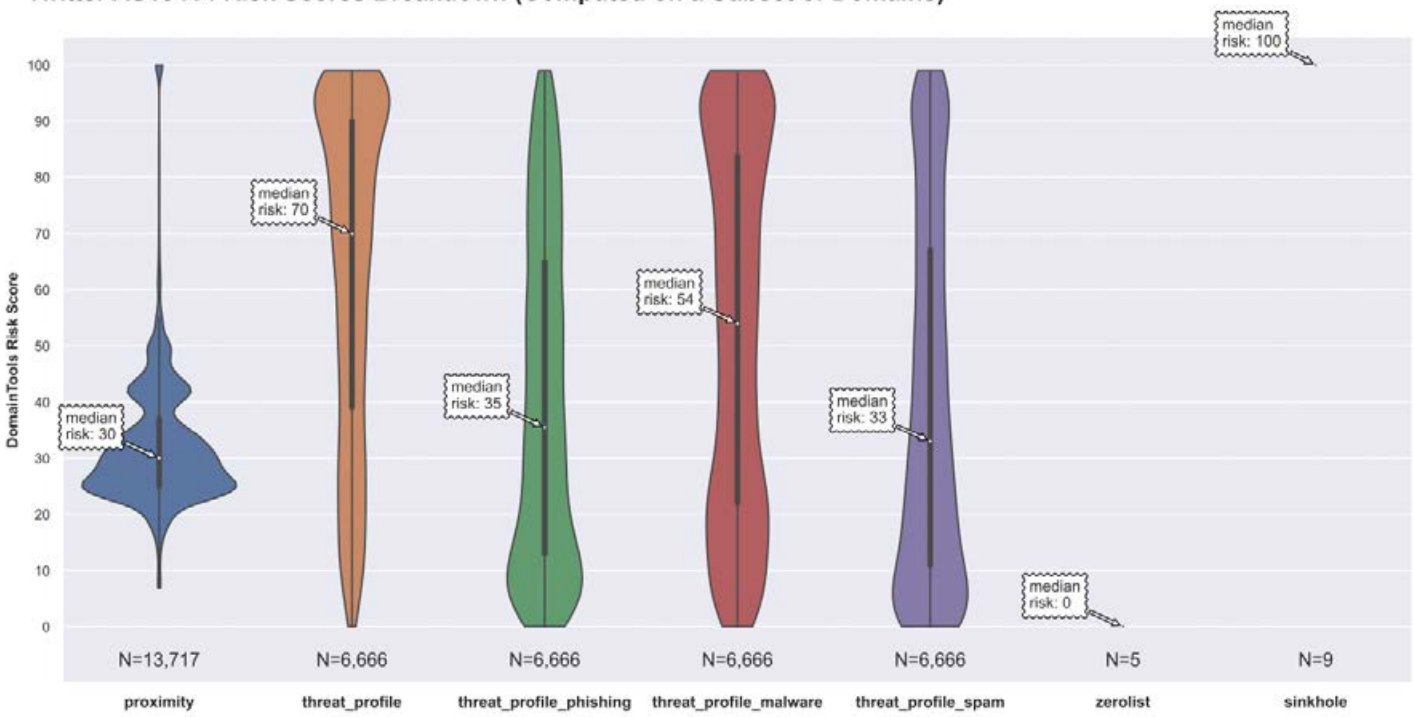
Trellian Pty AS133618 Risk Scores Breakdown (Computed on a Subset of Domains)



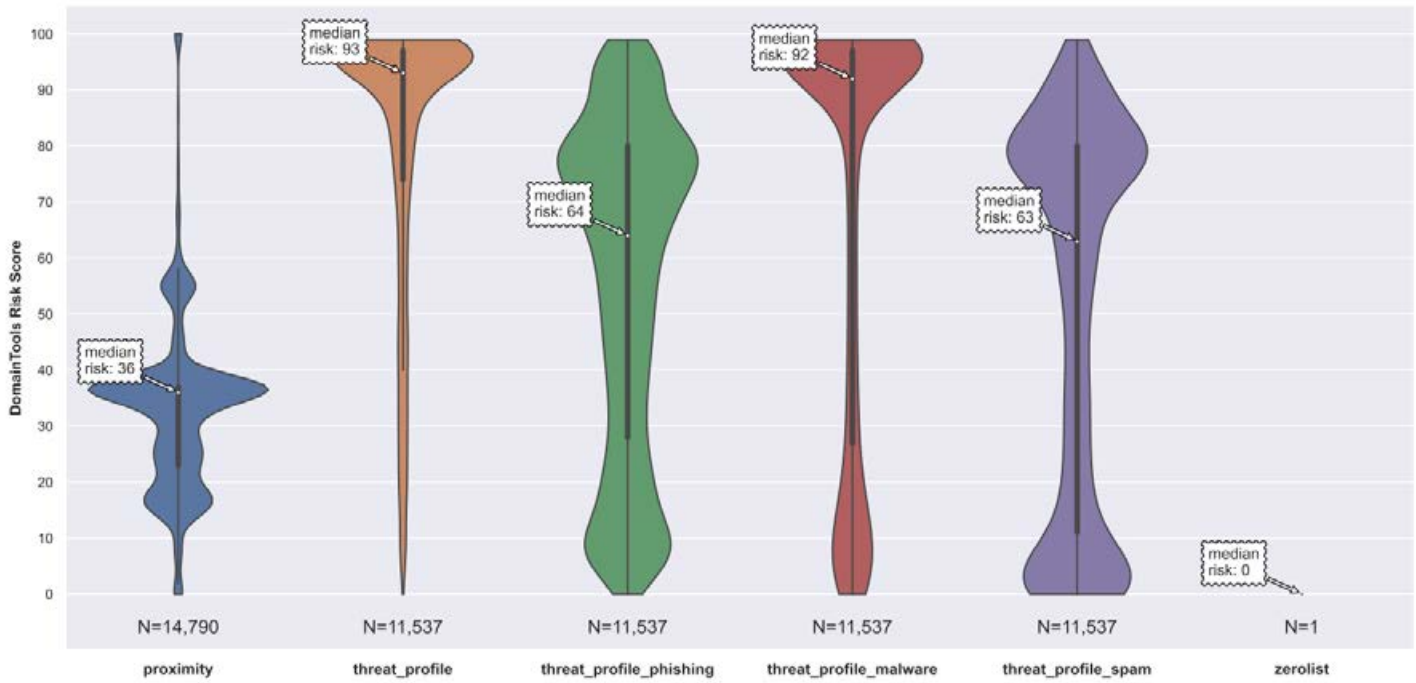
TW Mobile AS9924 Risk Scores Breakdown (Computed on a Subset of Domains)



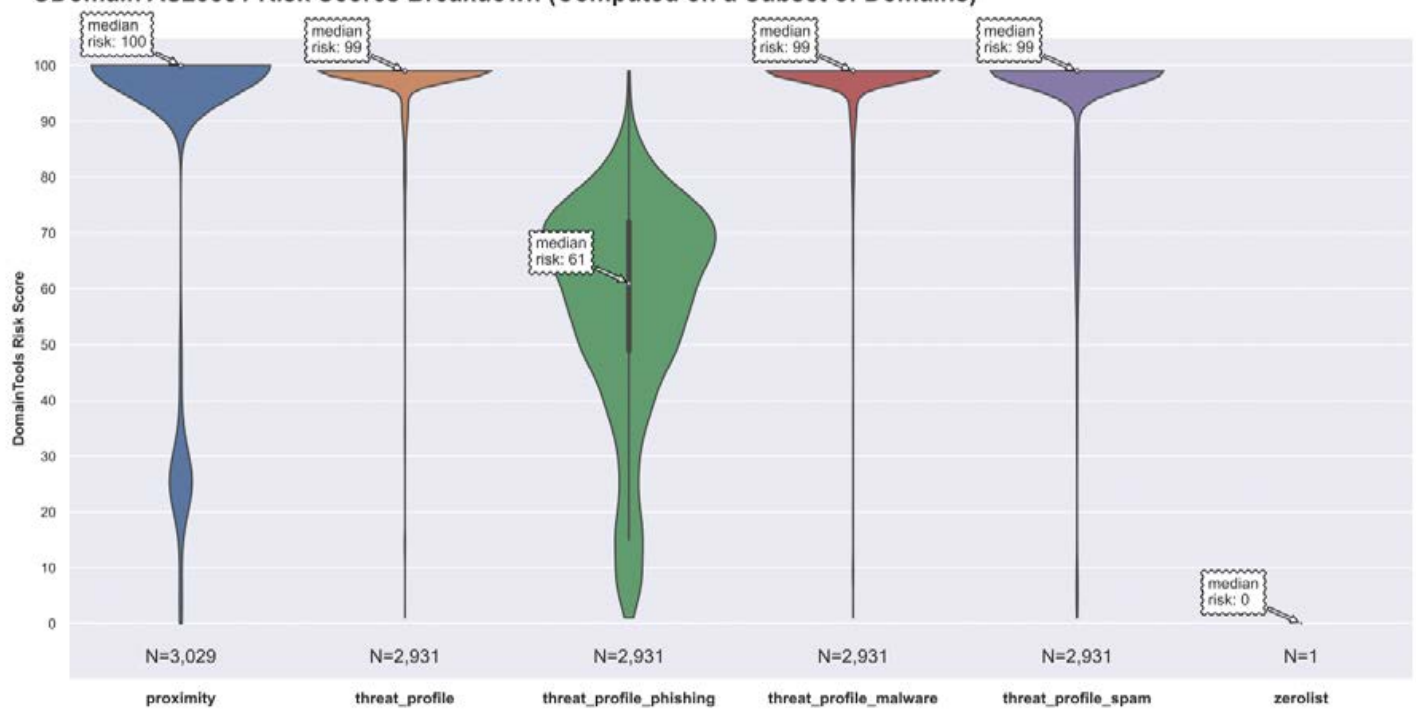
Twitter AS13414 Risk Scores Breakdown (Computed on a Subset of Domains)



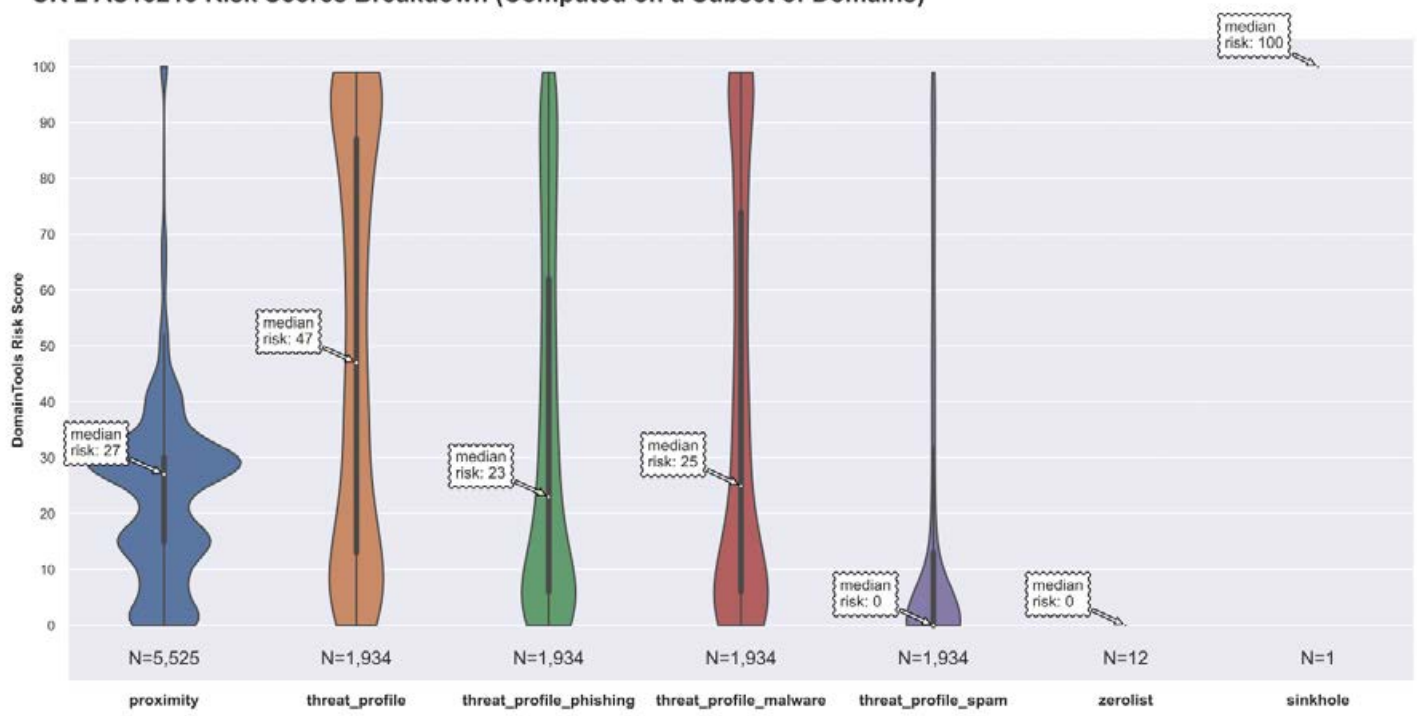
Ucloud AS135377 Risk Scores Breakdown (Computed on a Subset of Domains)



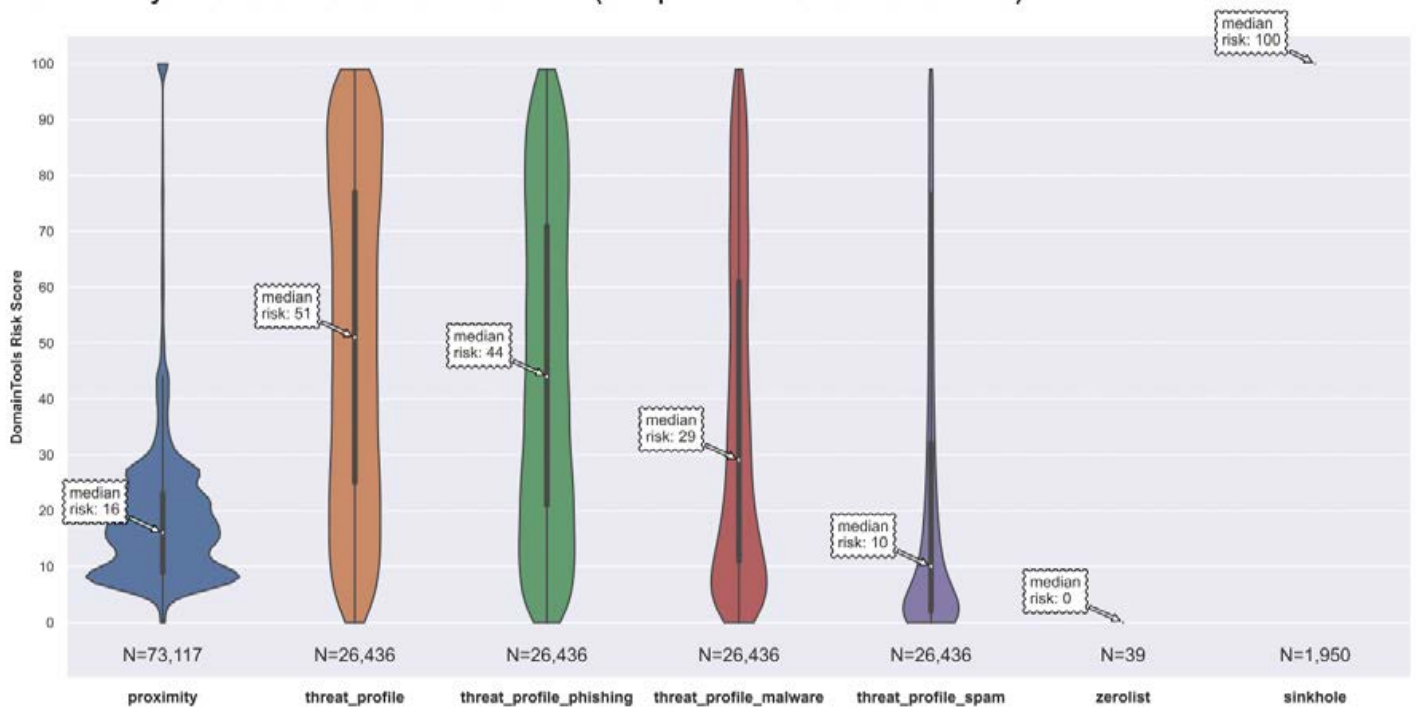
UDomain AS23881 Risk Scores Breakdown (Computed on a Subset of Domains)



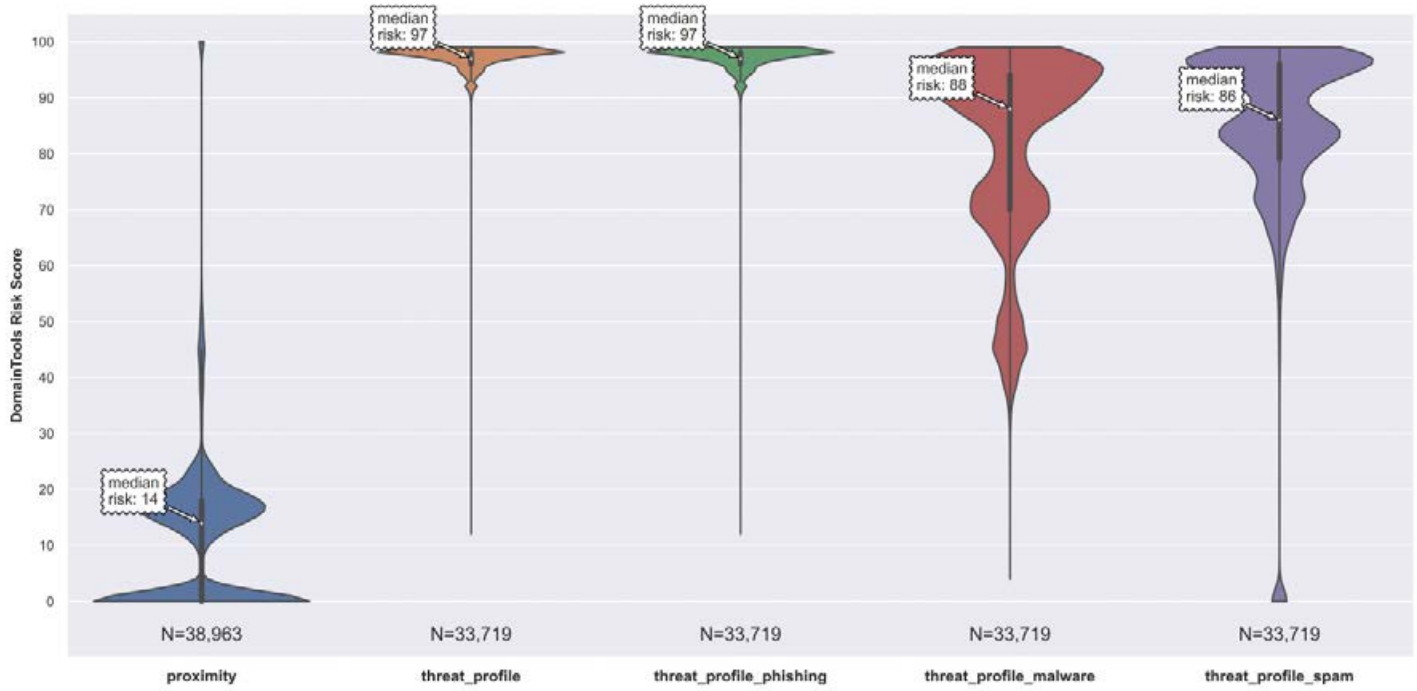
UK 2 AS13213 Risk Scores Breakdown (Computed on a Subset of Domains)



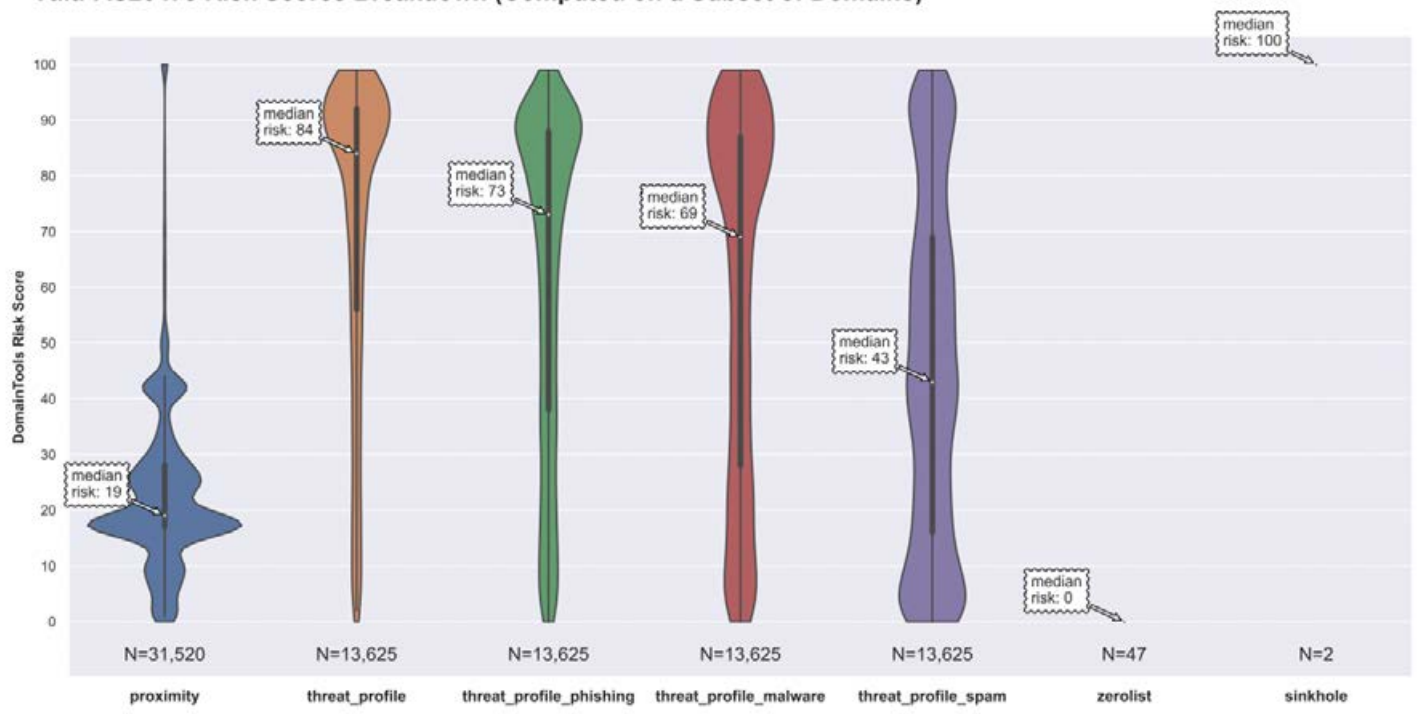
Unified Layer AS46606 Risk Scores Breakdown (Computed on a Subset of Domains)



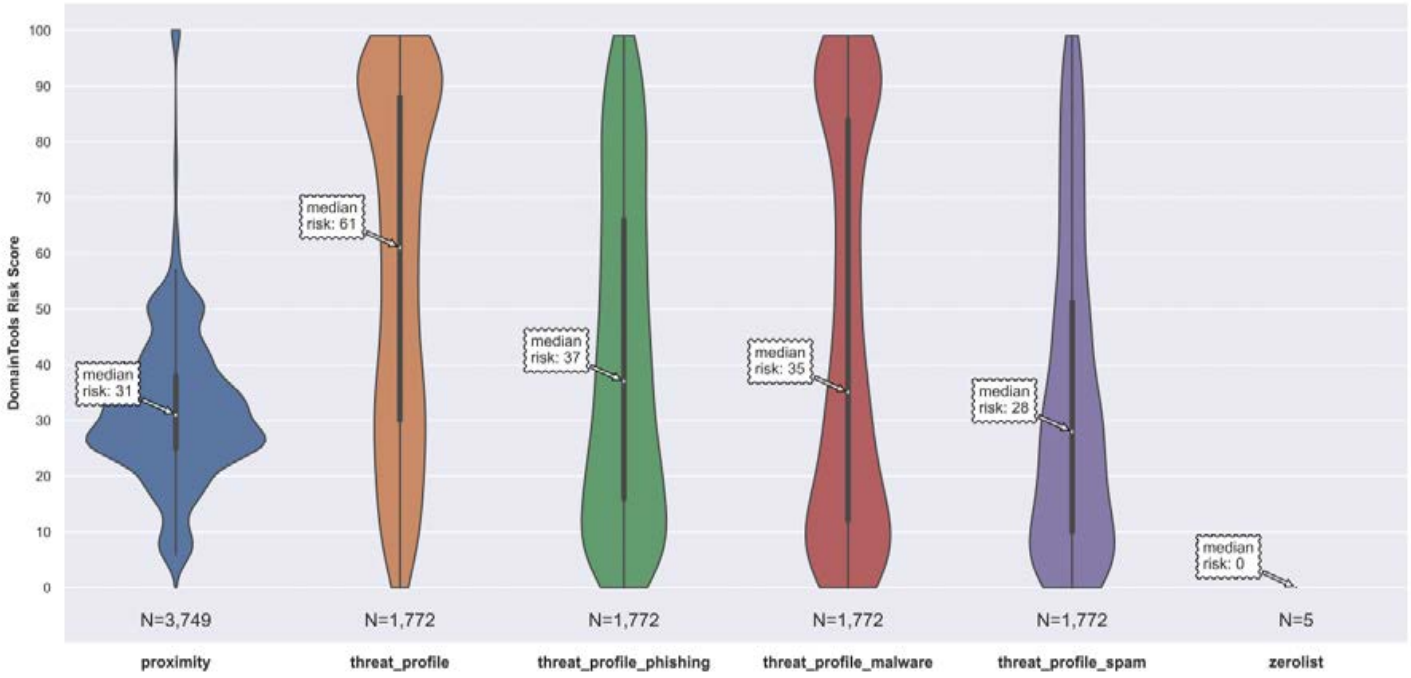
Verotel AS31624 Risk Scores Breakdown (Computed on a Subset of Domains)



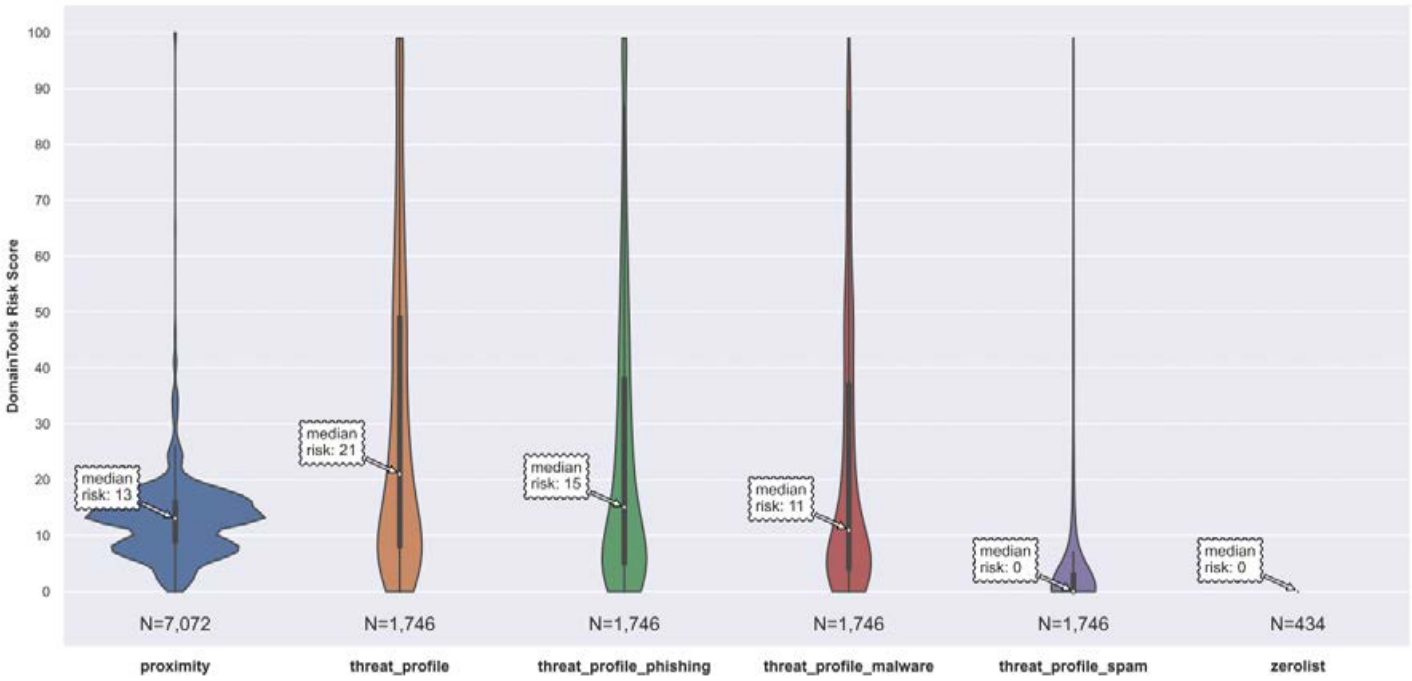
Vultr AS20473 Risk Scores Breakdown (Computed on a Subset of Domains)



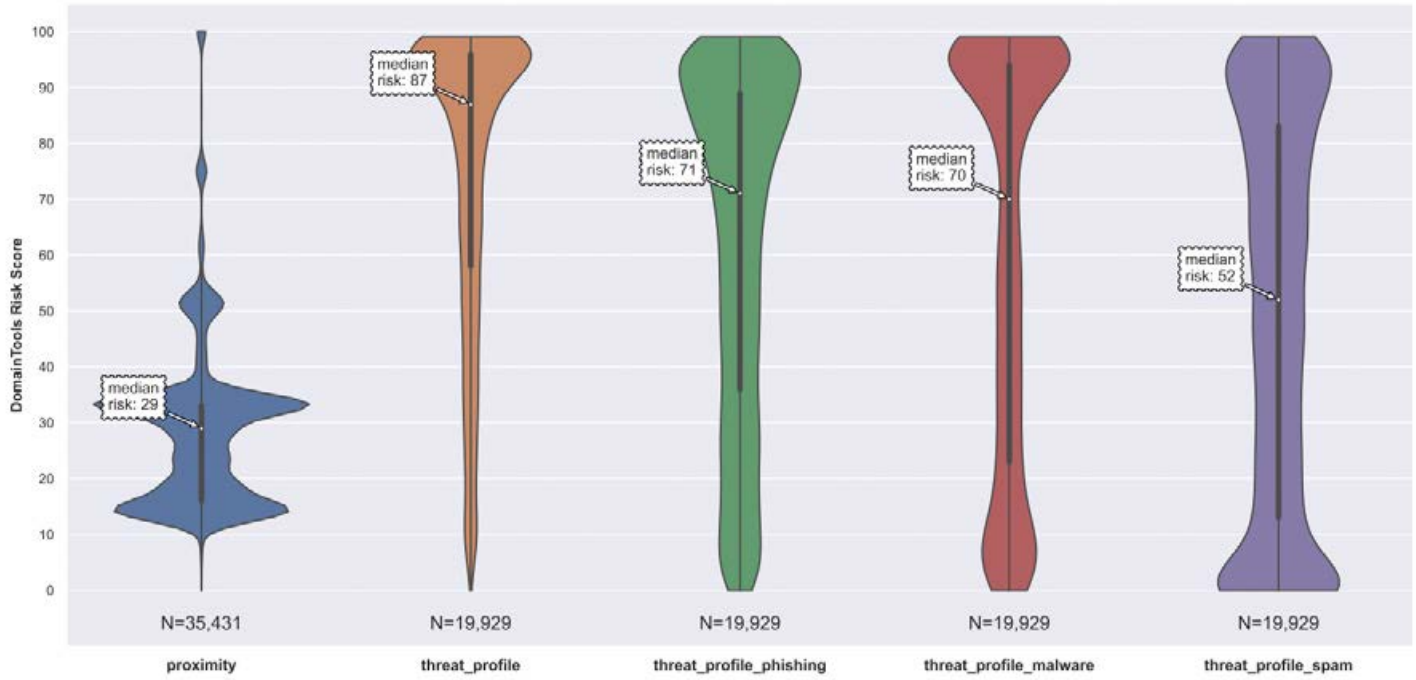
WebNX AS18450 Risk Scores Breakdown (Computed on a Subset of Domains)



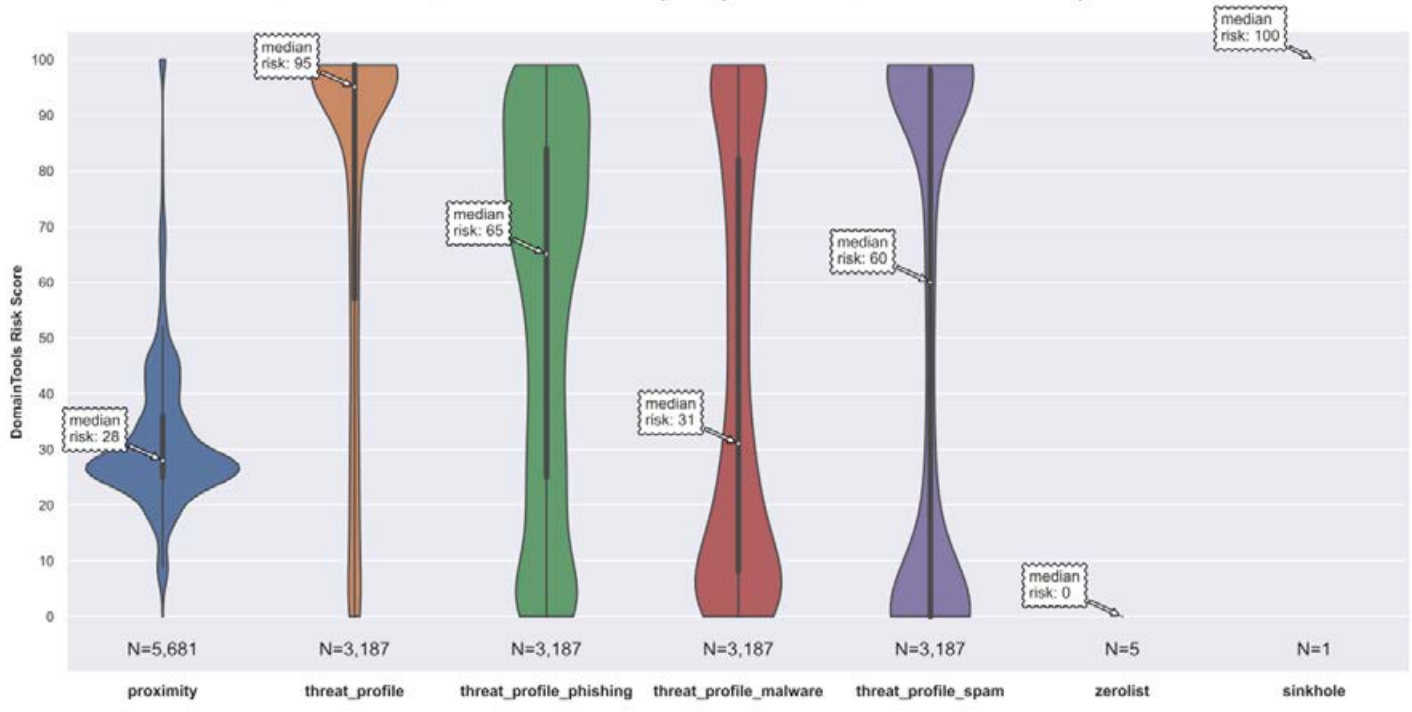
Weebly AS276417 Risk Scores Breakdown (Computed on a Subset of Domains)



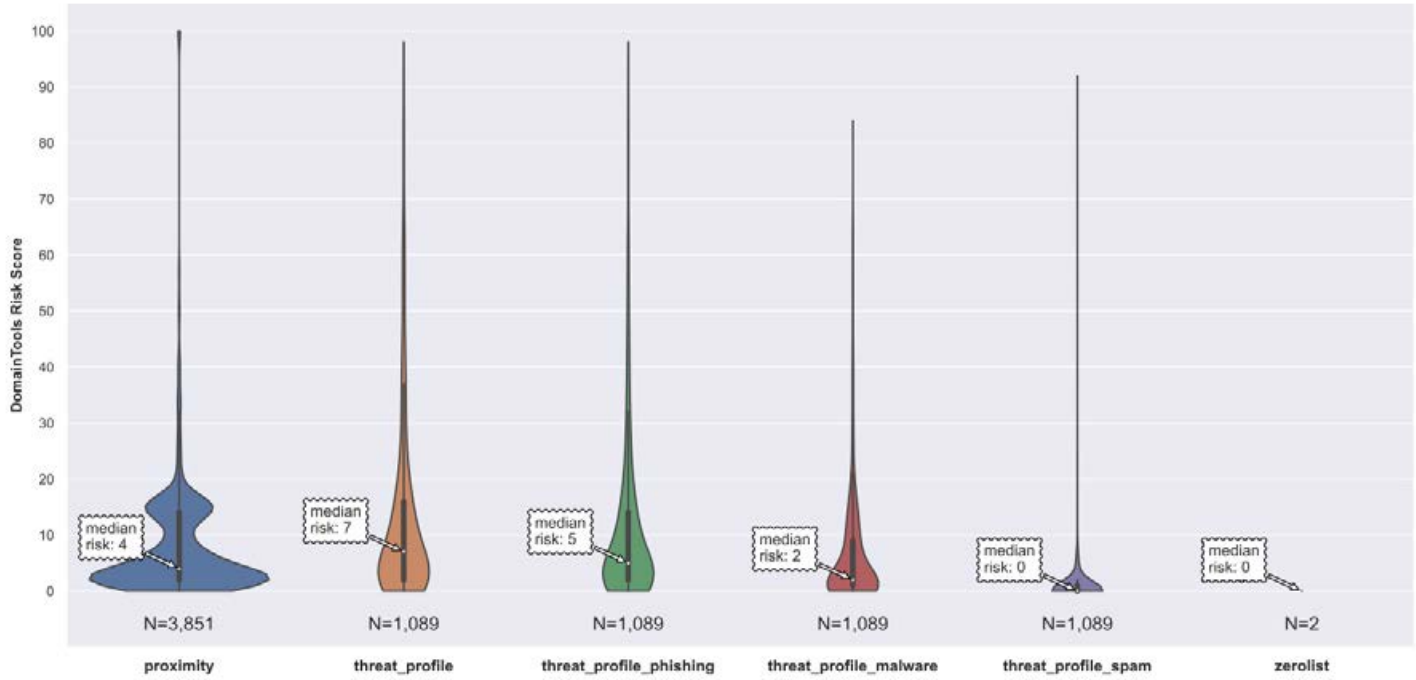
West 263 AS139021 Risk Scores Breakdown (Computed on a Subset of Domains)



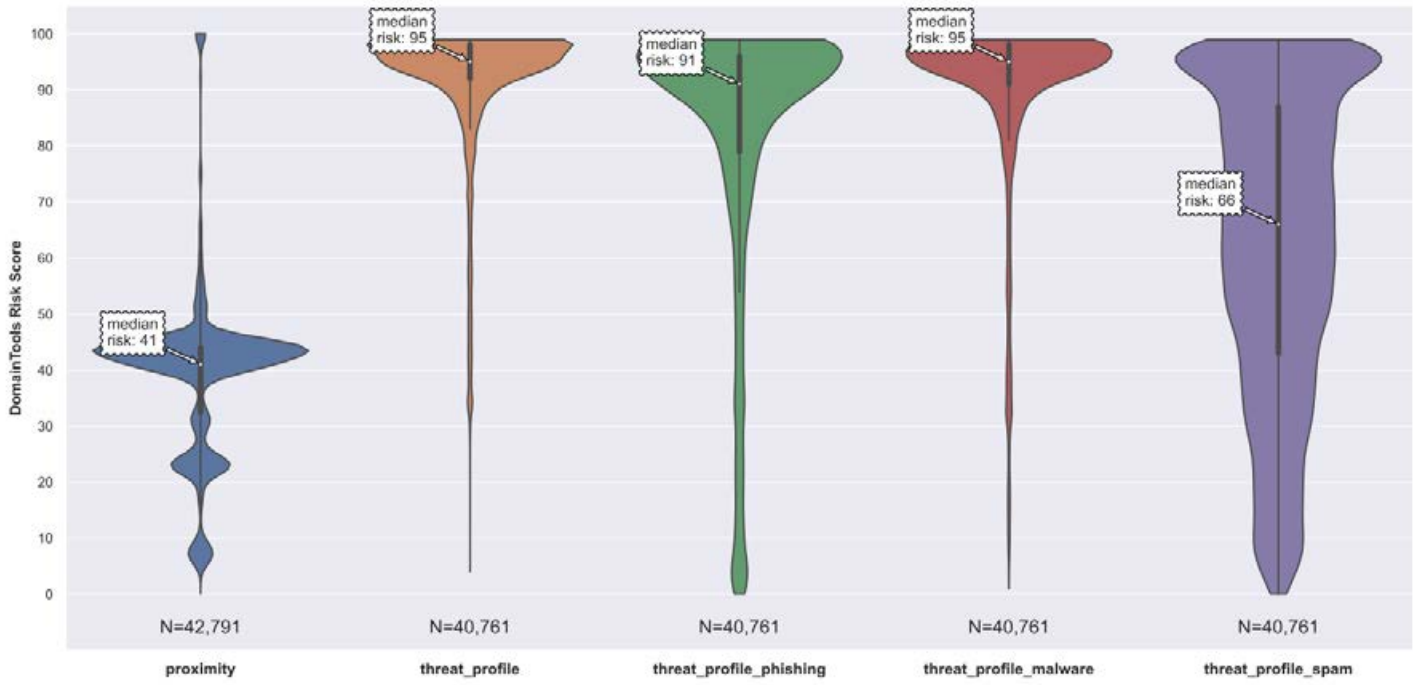
Wholesale Internet AS32097 Risk Scores Breakdown (Computed on a Subset of Domains)



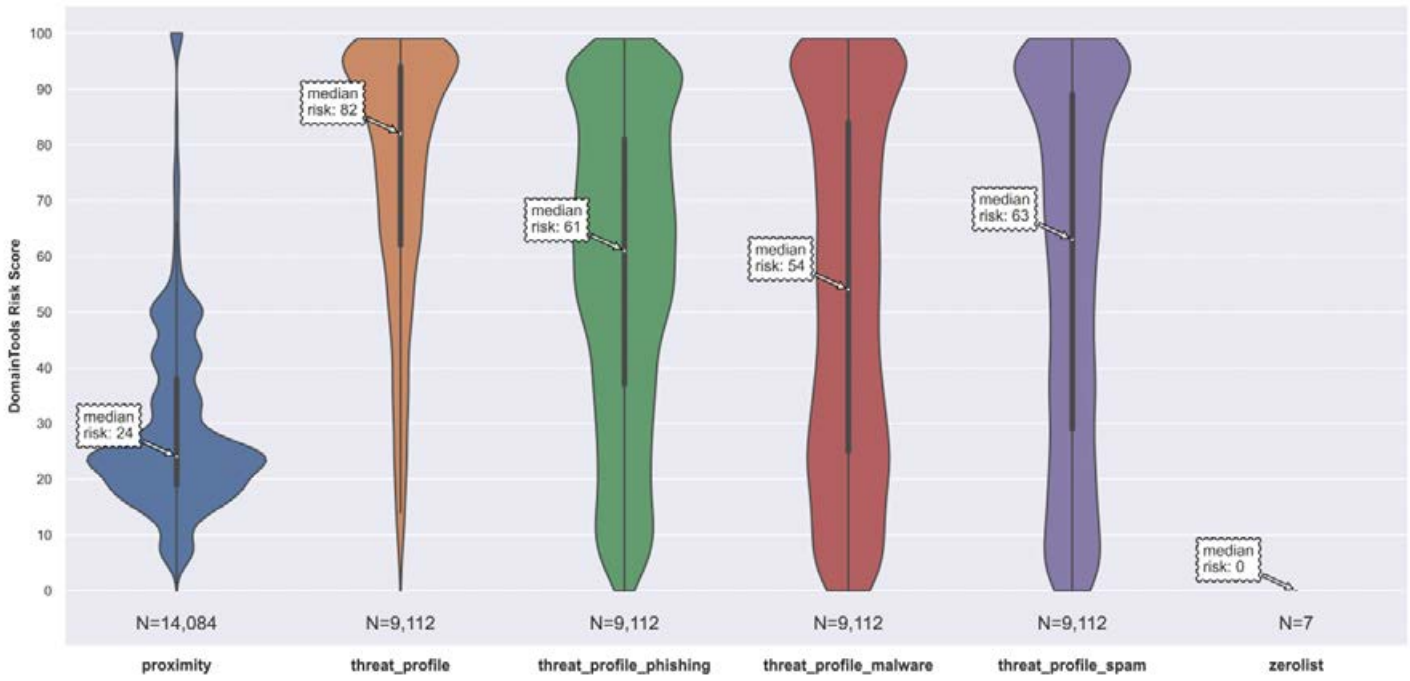
Xneelo AS37153 Risk Scores Breakdown (Computed on a Subset of Domains)



Yisu Cloud AS136970 Risk Scores Breakdown (Computed on a Subset of Domains)



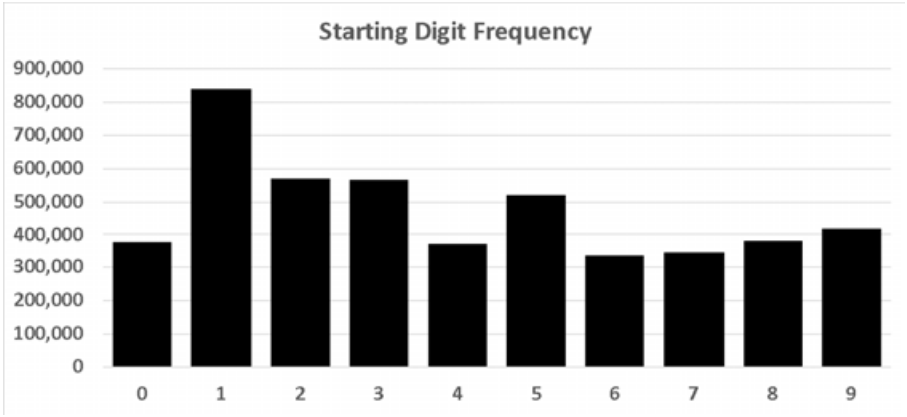
Zen Layer AS21859 Risk Scores Breakdown (Computed on a Subset of Domains)



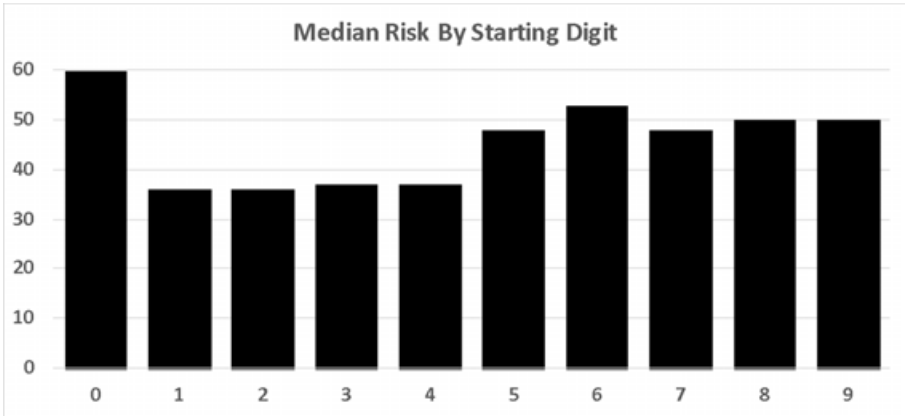
Part F. Other Quick Analyses

Does the Leading Digit Seem to Matter?

While our primary focus in the violin plots was on differences between ASNs and the contribution of the subscores. Since we had the data available, however, we also wanted to see if there were differences according to the leading first digit. Were the domains uniformly distributed across the ten possible values? Or do we see a disproportionate number of domains starting with "lucky" number 7? As it turns out, domains starting with the digit 1 are seen more often than other options:



And is there a difference in risk score for the different starting digits? Yes, for some (undetermined) reason ...



ANOVA

	sum_sq	df	F	PR(>F)
first_letter	8.289524e+07	9.0	7838.76317	0.0
Residual	5.571658e+09	4741818.0	NaN	NaN

first_letter	risk	count
0	60.0	379405
1	36.0	841119
2	36.0	569673
3	37.0	566699
4	37.0	372786
5	48.0	518792
6	53.0	337481
7	48.0	349411
8	50.0	384991
9	50.0	421471

Are There Differences in Median Risk Between ASN “Home Countries”?

We mapped ASNs to their home countries based on what was reported in WHOIS. We used that data for the little flags in the graphs. It's natural to wonder if there were differences between regions or countries when it comes to median risk scores. We checked.

Looking solely at countries where there were at least three ASNs included in the study, the broadest differences are apparent in Europe – the 14 German ASNs have exemplary low median risk scores, while the 3 NL ASNs are very high risk, at least when looking at domains beginning with a digit:

ASN Region/ "Home Country"	Studied ASNs	Median Risk	Count of Registrable Domains Starting With A Digit (These ASNs)
APNIC Region:			
HK	21	64	690,098
JP	3	31	46,983
TW	3	27	19,025
CN	18	26	388,367
KR	4	17	23,869
ARIN Region:			
US	79	48	2,815,071
RIPE Region:			
NL	3	96	56,433
FR	3	19	85,063
RU	7	16	52,155
UK	4	16	19,312
DE	14	9	270,367

Part G. Conclusions

Combine researcher interest in ASNs, unexpected data patterns (such as comparatively scant uptake of domains that begin with a digit) and a fondness for visualizing complex data, and you've got the nexus for this study. Let's conclude this report by summarizing just some of what we uncovered.

- Most importantly, we succeeded in demonstrating that yes, it is possible to compute risk scores for larger aggregates (such as computing risk scores on a per-autonomous system basis). Doing so required leveraging Farsight DNSDB data along with DomainTools risk scoring data as well as third party routing data, but this was a satisfying synergy to observe!
- Particularly exciting was the fact that we uncovered a wide range of risk scores on an ASN-by-ASN basis – we'd fully expected to see most ASNs clump in a "muddy puddle" of "middling" risk scores (except perhaps for some "known-sketchy" ASNs), but what we found was a much-better distributed spectrum of ASNs literally running from a median risk score of zero all the way up to a median risk score of 100. That range of scores bodes well for future ASN risk scoring work.
- While we could (and did) compute medians for each ASN, we were unsatisfied with just that table of results. Violin plots allowed us to visualize the distribution of risk scores for each ASN. If you're like us, you were probably surprised by some of the exotic patterns that emerged – these aren't some boring run-of-the-mill Gaussian distributions!
- While we'd expected to see some poorly-scoring ASNs, consistent with accepted "folklore" about various ASNs, we were surprised to see some expected-to-be-awful ASNs deliver mundane middle-of-the-road risk scores, and we were equally surprised to see some expected-to-be-good ASNs deliver poor risk scores.
- Given those surprising results, we felt compelled to dig in and go beyond just aggregate risk scores, digging into per-ASN subscores for proximity, phishing, malware, spam, zero-listing and sinkholes. Doing so resulted in further insights, including the fact that sinkholes, in particular, can be a bit of a "mixed bag:"
    - When a malicious domain is sinkholed by a researcher, a consumer machine will normally only attempt to contact that sinkhole if it is compromised. Thus, from a consumer-protection-centric point of view, it is appropriate to give the sinkhole domain a bad risk score – trying to reach that sinkhole is normally a sign that something's amiss on the consumer's system.
    - However, having a sinkhole hosted on an IP associated with an ASN is NOT an indicator that an ASN (as an aggregate) should be negatively scored. Sinkholes are important cyber security research infrastructure, do not cause any damage in and of themselves, and allowing one to be hosted on one of your IPs is a net positive (not a net negative) thing. ASNs with sinkholes using an IP the ASN routes should not be "punished" by sinkhole "100" scores in aggregations. In future research in this area, we may re-evaluate how we handle sinkhole scores.
- We'd initially planned to focus exclusively on computing risk scores per ASN, but there were some other "splits" that were just too obvious to overlook, such as the "home country" of the ASNs. We're glad we took the time to look at these splits, because it exposed substantive differences. While the APNIC region is often believed to be particularly strongly associated with "domains starting with digits," ARIN had the highest count, and RIPE countries had some of the best (and worst) risk scores based on data from our studied ASNs.
- As is often the case for our studies, we pushed on things in perhaps unexpected ways and discovered implicit (or explicit!) usage limits, and solutions that allow those who may seek to replicate or extend our work to "fly higher and go further" in the future. Just to mention a few of those:

- While DomainTools offers a Risk API endpoint, using Iris APIs may be more efficient due in part to its ability to take up to a hundred domains per charged query, returning risk data (as well as lots of other data).
- The Iris interface and APIs returns a status code and the "last known" risk score for domains that have ceased resolving, while the Risk API endpoint returns "not found." This is a subtle but important difference.
- Risk scores are computed and returned for "registrable domains." While many might normally look to the Public Suffix List when "stemming" FQDNs, the PSL is not perfectly aligned with what's registered. Unless you pay attention to those differences, you may "waste" queries re-asking what amounts to the same question, a question whose answer you may have already received.
- If you truly need extreme performance, processing downloaded risk score data into an MTBL-format file allows random access at rates sufficient to support over 2.5 million queries in 90 seconds using unoptimized code on a run-of-the-mill laptop. See "Efficiently Accessing a Moderately-Large Sorted and Uniquely-Keyed CSV File in Python3 with MTBL."

Truly, we hope we've piqued your curiosity, and perhaps inspired you to say, "Gosh, they should have checked <this> or I wonder about <that> – let me explore this a little more myself!"

For more information about access to DomainTools Iris, Risk Score, or Farsight DNSDB data, please feel free to contact us!

https://www.domaintools.com/contact

https://www.farsightsecurity.com/about-farsight-security/contacts/

Acknowledgements

Many at DomainTools provided help, insightful comments or other support in the preparation of this report, including (listed in alphabetical order by first name): Aaron Gee-Clough, Barry Rellis, Dan Nunes, Daniel Schwalbe, Jeremy Reed, Kali Fencl, Michael Klatt, Mike McCarthy, and Sean McNee. Many thanks to you all!

Any issues remaining are solely the responsibility of the author.

Appendices

Appendix A. The 174 Studied Autonomous Systems

ASN	Registrable Domains	Numeric Registrable Domains	Percent
alibaba-as45102	1,196,557	109,849	9.2
amazon-as14618	2,530,595	34,468	1.4
amazon-as16509	9,043,932	231,523	2.6
anchnet-as137443	50,705	10,232	20.2
aptum-as13768	453,640	7,536	1.7
att-as7018	175,543	2,197	1.3
baidu-as38365	54,240	2,238	4.1
baidu-as55967	44,249	2,452	5.5
bgpnet-global-as64050	684,083	117,560	17.2
bharti-as45609	2,808	2,296	81.8
chernyhov-as202984	16,584	2,965	17.9
china-mobile-as9808	126,326	5,886	4.7
china-mobile-as56046	45,180	7,372	16.3
china-mobile-as56048	76,291	2,165	2.8
china-telecom-as4812	69,225	3,324	4.8
china-unicom-as4808	146,102	6,386	4.4
china-unicom-as4837	243,367	18,252	7.5
chinanet-as4134	564,037	27,011	4.8
chinanet-as38283	82,792	4,538	5.5
clayer-as137951	765,210	43,319	5.7
cloud-inno-as134175	739,759	62,341	8.4
cloudflare-as13335	8,669,816	167,148	1.9
cloudie-as55933	180,580	23,456	13.0
cn-networks-ix-as4847	39,457	2,195	5.6
cnservers-as40065	683,816	95,566	14.0
cogent-as174	1,533,474	147,885	9.6
colo-crossing-as36352	204,172	6,205	3.0
comcast-as7922	208,484	2,578	1.2
confluence-as40034	723,141	35,952	5.0
contabo-as51167	610,144	5,949	1.0
ddos-guard-as57724	344,917	4,033	1.2
ddos-guard-as262254	36,413	14,387	39.5
dedipath-as35913	164,347	11,674	7.1
digitalocean-as14061	2,631,208	48,220	1.8
dreamscape-as38719	851,571	8,348	1.0

ASN	Registrable Domains	Numeric Registrable Domains	Percent
dtag-as3320	429,118	2,332	0.5
dxtl-as134548	1,057,038	91,329	8.6
egihosting-as18779	1,868,500	175,389	9.4
ehostict-as45382	11,527	2,469	21.4
enzu-as18978	78,603	5,887	7.5
eonix-as62904	351,775	16,184	4.6
esited-as22552	195,229	14,384	7.4
facebook-as32934	35,745	13,916	38.9
fastly-as54113	603,359	14,425	2.4
frantech-as53667	523,782	17,134	3.3
gandi-as29169	831,395	8,985	1.1
gigabit-hosting-as55720	67,597	15,340	22.7
gmo-int-as7506	1,570,492	28,471	1.8
godaddy-as26496	3,975,397	50,536	1.3
godaddy-as398101	1,340,184	17,128	1.3
google-as15169	7,230,458	280,187	3.9
google-as19527	1,103,591	11,317	1.0
google-as396982	2,336,595	50,816	2.2
gorilla-as53850	261,413	46,945	18.0
hetzner-as24940	3,709,118	45,299	1.2
hinet-as3462	157,224	13,538	8.6
hivelocity-as29802	261,726	4,793	1.8
hk-bb-as9269	54,931	18,561	33.8
hk-bb-as10103	49,455	18,413	37.2
hk-comm-as140227	177,305	57,538	32.5
hkbn-as17444	85,952	5,779	6.7
host-eu-as8972	1,034,200	9,129	0.9
host-eu-as20738	1,670,563	16,320	1.0
host-eu-as20773	3,323,627	35,654	1.1
host-eu-as21499	228,263	2,260	1.0
host-eu-as21501	1,950,398	22,427	1.1
host-eu-as34011	783,613	5,940	0.8
hostinger-as47583	1,810,976	14,789	0.8
hostus-as7489	32,309	3,335	10.3
huawei-as55990	69,987	4,038	5.8
huawei-as136907	69,436	7,048	10.2
hurricane-as6939	198,067	11,882	6.0
hz-alibaba-as37963	1,117,701	58,528	5.2
hz-zhiyu-as59037	10,144	4,354	42.9

ASN	Registrable Domains	Numeric Registrable Domains	Percent
idc-chinanet-as23724	144,679	5,935	4.1
infomaniak-as29222	323,038	2,898	0.9
inmotion-as3842	510,016	16,082	3.2
inmotion-as22611	282,864	3,123	1.1
inmotion-as54641	299,916	3,299	1.1
int-hostspace-as399674	280,986	19,991	7.1
intercontinental-as398968	88,997	5,445	6.1
interserver-as19318	272,474	2,464	0.9
io-flood-as53755	94,217	5,774	6.1
iomart-as20860	304,482	3,769	1.2
ionos-as8560	2,901,997	29,545	1.0
ip-volume-as58110	358,971	8,502	2.4
it7-networks-as25820	110,903	13,455	12.1
jsc-iot-as29182	211,739	5,649	2.7
kr-telecom-as4766	477,740	13,330	2.8
krypt-as35908	73,979	13,499	18.2
layerhost-as46573	248,096	14,885	6.0
leaseweb-as7203	296,344	35,206	11.9
leaseweb-as30633	582,382	11,757	2.0
leaseweb-as60781	920,299	14,750	1.6
leaseweb-as395954	411,894	33,042	8.0
level3-as3356	145,034	6,466	4.5
level3-as3549	202,211	7,192	3.6
lg-dacom-as3786	272,998	5,223	1.9
limestone-as46475	159,785	3,188	2.0
linode-as63949	1,471,421	64,515	4.4
liquid-web-as32244	825,133	17,183	2.1
luogelang-as135097	443,355	31,329	7.1
m247-as9009	164,467	5,772	3.5
microsoft-as3598	24,221	9,096	37.6
microsoft-as8075	1,124,110	67,781	6.0
multacom-as35916	3,221,106	283,327	8.8
namecheap-as22612	3,526,551	98,539	2.8
namesco-as8622	450,782	4,234	0.9
netcup-as197540	347,713	4,012	1.2
netsec-as45753	82,319	15,793	19.2
neue-medien-as34788	1,583,949	10,352	0.7
new-century-as9919	28,506	2,383	8.4
new-dream-as26347	1,084,945	15,024	1.4

ASN	Registrable Domains	Numeric Registrable Domains	Percent
newfold-as29873	1,259,523	7,196	0.6
nforce-as43350	117,000	2,720	2.3
nocix-as33387	1,056,800	25,695	2.4
ntt-as2914	79,693	12,141	15.2
one-as51468	1,239,945	10,076	0.8
oracle-as31898	205,074	5,743	2.8
oso-grande-as26337	458,308	3,550	0.8
ovh-as16276	5,286,442	72,659	1.4
peg-tech-as54600	1,227,219	131,657	10.7
peg-tech-as398478	72,922	7,149	9.8
peg-tech-as398823	295,529	30,217	10.2
power-line-hk-as132839	1,287,909	111,142	8.6
psychz-as40676	279,603	24,352	8.7
pvt-layer-as51852	436,643	3,680	0.8
quadranet-as8100	756,052	35,509	4.7
quick-pck-as46261	232,869	19,904	8.5
rackspace-as19994	175,545	5,559	3.2
rackspace-as33070	241,214	6,654	2.8
regru-as197695	1,100,170	18,656	1.7
ru-center-as48287	795,122	7,005	0.9
sakura-as9370	319,363	11,966	3.7
sakura-as9371	441,200	6,546	1.5
sanren-as139330	39,476	4,007	10.2
scaleway-as12876	319,309	3,419	1.1
sedo-as47846	1,748,279	55,254	3.2
selectel-as49505	264,925	4,532	1.7
servihosting-as29119	18,508	2,643	14.3
sharktech-as46844	147,671	23,784	16.1
singlehop-as32475	360,016	3,950	1.1
sk-broadband-as9318	140,526	2,847	2.0
small-orange-as62729	205,654	2,325	1.1
softlayer-as36351	1,008,481	25,985	2.6
sonder-cloud-as133199	53,763	8,256	15.4
sprint-as1239	68,789	11,692	17.0
squarespace-as53831	3,778,191	33,666	0.9
stackpath-as33438	225,972	4,345	1.9
strato-as6724	3,380,866	25,894	0.8
sun-net-as38197	145,910	20,987	14.4
sun-net-as136800	682,475	64,533	9.5

ASN	Registrable Domains	Numeric Registrable Domains	Percent
take-2-host-as20248	97,418	10,381	10.7
tcloudnet-as399077	55,793	5,279	9.5
tencent-as45090	433,342	29,213	6.7
tencent-as132203	788,781	94,631	12.0
tier-net-as397423	47,859	10,320	21.6
timeweb-as9123	423,468	9,315	2.2
trellian-pty-as133618	545,076	11,911	2.2
tw-mobile-as9924	22,109	3,104	14.0
twitter-as13414	31,545	13,722	43.5
ucloud-as135377	154,952	14,791	9.5
udomain-as23881	23,446	3,022	12.9
uk-2-as13213	594,304	5,537	0.9
unified-layer-as46606	6,208,378	73,156	1.2
verotel-as31624	1,096,972	38,963	3.6
vultr-as20473	743,085	31,512	4.2
webnx-as18450	69,770	3,751	5.4
weebly-as27647	757,349	7,506	1.0
west263-as139021	306,588	35,431	11.6
wholesale-int-as32097	87,307	5,686	6.5
xneelo-as37153	428,899	3,828	0.9
yisu-cloud-as136970	75,840	42,791	56.4
zen-layer-as21859	107,854	14,085	13.1
TOTAL	142,067,268	4,739,944	

Appendix B. The 611 Excluded ASNs (with <2,000 registered domains that begin with a digit, but at least 100 registered Domains

ASN	Registrable Domains	Numeric Registrable Domains	Percent
1-and-1-as8881	47,934	391	0.82
123-net-as12129	12,503	115	0.92
1337team-as51381	104	0	0.00
2-cloud-as39845	3,237	6	0.19
23m-as47447	62,698	494	0.79
24shells-as55081	24,786	433	1.75
3s-inf-as37671	1,783	26	1.46
ab-stract-as39287	4,439	73	1.64
abansys-as196713	24,607	86	0.35
abcle-as38661	8,831	210	2.38
abdilaziz-as211193	4,143	13	0.31
abelohost-as204196	3,726	67	1.80
acorus-as35280	257	5	1.95
active1-as197071	9,846	106	1.08
active-host-as51698	822	20	2.43
advania-as50613	7,807	82	1.05
advancedhosters-as39572	21,946	413	1.88
aeza-as210644	3,044	49	1.61
affinity-as3064	167,152	1,991	1.19
ais-as6130	7,837	80	1.02
akamai-as16625	12,815	62	0.48
akamai-as32787	70,042	552	0.79
akamai-intl-as20940	36,768	302	0.82
alexhost-as200019	8,156	245	3.00
altushost-as51430	12,837	197	1.53
alsycon-as208911	838	3	0.36
amarutu-as206264	18,588	269	1.45
amati-as42237	47,017	4	0.01
ancel-as6057	11,123	74	0.67
anonim-sikreti-as60446	117	1	0.85
another-corp-as32329	1,458	20	1.37
anycast-glbl-as58511	147	1	0.68
ap-telecom-as17709	1,871	45	2.41
aplikanusa-as4800	3,068	2	0.07
apple-as714	6,242	195	3.12

ASN	Registrable Domains	Numeric Registrable Domains	Percent
arelion-as1299	16,010	78	0.49
arteria-as2519	12,011	525	4.37
atom86-as8455	13,999	200	1.43
atria-convergence-as24309	2,396	13	0.54
att-as2386	3,755	17	0.45
att-as2686	1,093	26	2.38
att-as6389	9,875	63	0.64
att-as20057	513	10	1.95
awknet-as17048	352	12	3.41
axtel-as6503	3,163	24	0.76
b2-net-as55286	65,091	1,713	2.63
bahnhof-as8473	34,718	284	0.82
baxet-as51659	15,096	575	3.81
bekkoame-as4686	6,852	52	0.76
belcloud-as44901	7,734	133	1.72
bell-canada-as577	40,879	328	0.80
bell-canada-as855	3,159	38	1.20
best-idc-as59374	2,972	34	1.14
beyond-pl-as31229	65,358	534	0.82
bezeq-as8551	51,242	1,173	2.29
bgx-as55960	3,560	104	2.92
bharti-as9498	9,144	58	0.63
bharti-as24560	6,161	47	0.76
biznet-as17451	2,475	13	0.53
bl-networks-as399629	8,312	67	0.81
blacknight-as39122	121,623	696	0.57
bluevps-as62005	900	4	0.44
bouygues-as5410	10,964	122	1.11
bredband2-as45011	4,430	35	0.79
breezeline-as11776	2,368	51	2.15
bryan-barbolina-as213268	790	11	1.39
bsnl-as9829	4,533	39	0.86
bt-as2856	80,182	506	0.63
bt-as5400	1,927	3	0.16
bytes-as50321	352	7	1.99
c-spire-as11272	1,443	9	0.62
cable-one-as11492	3,986	35	0.88
cablevision-as6128	21,510	234	1.09
cantv-as8048	1,499	9	0.60

ASN	Registrable Domains	Numeric Registrable Domains	Percent
carinet-as10439	7,631	115	1.51
cdn-as13188	2,031	35	1.72
cegeka-as21286	1,514	1	0.07
centurylink-as209	81,486	848	1.04
centurylink-as3561	576,620	1,927	0.33
centurylink-as22561	2,013	24	1.19
charter-as7843	5,790	62	1.07
charter-as10796	30,668	326	1.06
charter-as12271	6,552	136	2.08
charter-as20115	52,974	440	0.83
charter-as33363	17,710	208	1.17
chief-telecom-as17408	6,016	366	6.08
china-mobile-as24444	2,928	113	3.86
china-mobile-as24445	2,116	98	4.63
china-mobile-as24547	1,087	39	3.59
china-mobile-as56040	10,829	722	6.67
china-mobile-as56041	12,305	1,816	14.76
china-mobile-as56047	1,155	47	4.07
china-telecom-as4811	20,073	1,013	5.05
china-telecom-as4835	61,876	1,942	3.14
china-telecom-as133774	18,738	827	4.41
china-telecom-as137691	2,800	65	2.32
china-tietong-as9394	3,299	405	12.28
china-unicom-as17621	12,378	609	4.92
china-unicom-as17622	2,285	118	5.16
china-unicom-as133119	23,427	818	3.49
china-unicom-as136958	9,257	507	5.48
china-unicom-as140979	210	7	3.33
chinanet-as134761	340	27	7.94
chinanet-as134768	2,038	129	6.33
chinanet-as134773	3,464	154	4.45
chinanet-gd-as58466	13,845	794	5.73
chinanet-jiangxi-as139201	976	71	7.27
chinanet-liaoning-as134762	10,660	686	6.44
chinanet-tj-as17638	1,909	75	3.93
citic-as4058	2,510	39	1.55
claro-as6400	1,547	16	1.03
claro-as28573	2,834	25	0.88
cloud-comp-as58519	7,977	396	4.96

ASN	Registrable Domains	Numeric Registrable Domains	Percent
cloudlite-as210200	1,047	19	1.81
clouvider-as62240	13,835	215	1.55
cn-r-and-e-as4538	8,395	456	5.43
cn-telecom-as4809	3,841	109	2.84
cn-telecom-hz-as58461	14,073	1,705	12.12
cn-unicom-gd-as134543	3,547	494	13.93
cn-unicom-sz-as17623	3,949	143	3.62
colo-america-as21769	3,796	45	1.19
colo-aus-as63956	45,318	374	0.83
colocenter-as58291	1,879	19	1.01
cologix-as40715	303	2	0.66
colt-as8220	57,563	410	0.71
columbus-networks-as23520	1,623	13	0.80
columbus-panama-as26426	152	4	2.63
comcast-as33491	24,244	275	1.13
comcor-as8732	4,326	95	2.20
comnet-bilgi-as61135	640	15	2.34
corespace-as54489	36,395	379	1.04
cox-as22773	100,808	913	0.91
craigslist-as22414	174	0	0.00
cyber-smart-as60118	1,396	13	0.93
cybercast-as27956	1,416	12	0.85
cybercon-as7393	65,226	662	1.01
cyberfort-as24958	8,794	46	0.52
cyrus-one-as62	9,802	57	0.58
data102-as33302	858	11	1.28
datacamp-as212238	786	16	2.04
datacamp-as60068	13,679	246	1.80
datacheap-as16262	3,897	73	1.87
dataclub-as52048	899	13	1.45
dataclub-as60567	498	1	0.20
datagroup-as3326	3,851	41	1.06
dataline-as49063	11,941	216	1.81
datapipe-as26228	3,301	30	0.91
datawagon-as27176	3,885	25	0.64
dc-star-as42160	7,298	57	0.78
dc74-as17216	4,443	156	3.51
defense-net-as55002	21,033	154	0.73
delis-llc-as211252	4,103	91	2.22

ASN	Registrable Domains	Numeric Registrable Domains	Percent
des-capital-as213035	11,538	204	1.77
dfn-as680	20,109	134	0.67
dianxintong-as17964	13,428	722	5.38
digital-fortress-as2044	864	5	0.58
digital-net-as12695	14,525	534	3.68
digital-space-as8607	2,721	31	1.14
digital-untd-as4780	4,707	131	2.78
dimension-data-as3741	18,483	216	1.17
diva-e-as44066	140,458	1,567	1.12
dlive-as10036	2,911	47	1.61
dod-as749	5,094	982	19.28
dongfong-as18046	453	10	2.21
dosarrest-as19324	42,566	674	1.58
dream-vps-as213038	3,018	35	1.16
dream-wave-as18068	2,234	222	9.94
dreamline-as9457	3,553	77	2.17
earthlink-iq-as50710	222	16	7.21
eastern-telecom-as9658	1,523	36	2.36
edgecast-as15133	23,801	342	1.44
edgenap-as61414	1,347	8	0.59
edgoo-as47787	314	11	3.50
ee-ltd-as12576	2,530	96	3.79
enes-koken-as209371	1,986	16	0.81
enet-as10297	8,649	91	1.05
eonix-as49532	708	5	0.71
equinix-jp-as17941	24,248	1,028	4.24
estruxture-as10929	24,716	250	1.01
eunetworks-as13237	43,072	639	1.48
exabytes-as38532	30,262	322	1.06
fairpoint-as13977	3,580	34	0.95
faster-as24641	67,597	367	0.54
fcd-as209813	3,604	48	1.33
fiber-grid-as37518	29,309	76	0.26
fiberhub-as53340	4,475	66	1.47
fiberlight-as13876	308	18	5.84
fibrenoire-as22652	19,319	142	0.74
filanco-as29076	16,885	307	1.82
first-light-as13536	1,780	11	0.62
flexential-as13649	83,939	1,190	1.42

ASN	Registrable Domains	Numeric Registrable Domains	Percent
flokinet-as200651	4,641	61	1.31
flyservers-as209588	1,201	4	0.33
fnk-as43317	7,644	110	1.44
fnx-as60503	335	5	1.49
fop-hornostay-as212913	4,561	64	1.40
forcepoint-as44444	154	1	0.65
fortressitx-as25653	7,523	67	0.89
fpt-telecom-as18403	74,298	663	0.89
free-sas-as12322	67,267	771	1.15
freedom-internet-as206238	2,243	29	1.29
freenet-as5430	10,641	100	0.94
freenet-as31148	698	14	2.01
frontier-as5650	28,222	322	1.14
fusion-as17184	4,170	47	1.13
g-core-as199524	4,178	142	3.40
g-core-202422	16,869	320	1.90
gamma-as31655	11,675	150	1.28
garant-park-as47196	4,362	70	1.60
gateway-as132827	2,252	126	5.60
gds-as45079	363	5	1.38
geekyworks-as203999	114	3	2.63
ghostnet-as12586	16,411	133	0.81
gkg-as18710	19,945	313	1.57
glesys-as42708	54,981	397	0.72
global-layer-as49453	5,559	87	1.57
globacom-as37148	133	6	4.51
globalconnect-as12552	59,575	489	0.82
globe-telecom-as4775	1,643	50	3.04
globe-telecom-as132199	439	9	2.05
globotech-as36666	72,003	761	1.06
gmo-z-as59349	1,039	29	2.79
godaddy-as30083	70,084	1,097	1.57
google-as16591	5,697	71	1.25
gransy-as60592	38,031	1,719	4.52
grupoice-as11830	1,924	34	1.77
gsl-networks-as137409	15,739	15	0.10
gtt-as3257	104,450	1,139	1.09
gtt-as5669	2,999	24	0.80
ha-vel-as15935	1,148	11	0.96

ASN	Registrable Domains	Numeric Registrable Domains	Percent
hathway-as17488	1,491	10	0.67
hcn-as7562	131	3	2.29
heficed-as61317	27,975	647	2.31
hgc-as18116	401	16	3.99
hgc-global-as9304	9,115	225	2.47
hien-quy-as140803	11,509	120	1.04
high-tech-as197765	274	2	0.73
hitron-as9311	2,087	141	6.76
hkbn-as2706	1,209	27	2.23
hkbn-as9381	23,028	491	2.13
hkt-as4515	11,094	347	3.13
host-eu-as29066	59,015	1,195	2.02
host-eu-as29486	27,284	226	0.83
host-eu-as34088	9,301	500	5.38
host-eu-as39783	4,800	26	0.54
host-eu-as44273	653	12	1.84
host-lincoln-as58040	3,830	36	0.94
host-sailor-as60117	6,722	94	1.40
hostdime-as33182	263,213	1,813	0.69
hosthatch-as63473	6,915	438	6.33
hosting-sol-as14576	13,524	252	1.86
hosting-technology-as207651	16,353	275	1.68
hostinger-as204915	18,009	169	0.94
hostkey-as395839	3,004	6	0.20
hostkey-as57043	5,317	41	0.77
hostmysite-as20021	39,214	389	0.99
hostroyale-as203020	1,391	35	2.52
hostwinds-as45290	71,174	1,312	1.84
hot-net-as12849	983	18	1.83
hp-as71	2,948	135	4.58
hugeservers-as25780	177	1	0.56
i3d-net-as49544	23,249	258	1.11
i4hk-as58779	2,939	70	2.38
idc-frontier-as4694	128,289	1,726	1.35
ielo-liazo-as29075	1,282	12	0.94
ifx-as18747	31,077	141	0.45
ihor-as35196	11,973	408	3.41
iij-as2497	17,338	279	1.61
ikoula-as21409	38,804	322	0.83

ASN	Registrable Domains	Numeric Registrable Domains	Percent
incapsula-as19551	146,013	1,863	1.28
indosat-as4761	691	5	0.72
inetcom-as35598	417	7	1.68
inf-sis-as61272	8,287	128	1.54
init7-as13030	7,891	80	1.01
inter-connects-as46805	3,928	29	0.74
inter-connects-as57858	1,231	19	1.54
inter-connects-as60485	337	1	0.30
inter-connects-as63119	1,975	9	0.46
interkvm-as25198	780	22	2.82
internap-as10912	6,379	185	2.90
internap-as13789	9,592	269	2.80
internap-as29791	10,349	237	2.29
internap-jp-as24295	562	32	5.69
internet-cz-as24806	204,430	1,792	0.88
internet-it-as200313	5,418	252	4.65
ip-projects-as48314	50,052	456	0.91
ip-volume-as202425	3,253	60	1.84
iptp-as41095	642	15	2.34
iran-telecom-as58224	5,727	71	1.24
isp-pro-as35366	12,650	178	1.41
istanbuldc-as59447	30,046	307	1.02
itl-as21100	26,586	368	1.38
itproximus-as203061	174	1	0.57
iweb-as32613	141,920	1,232	0.87
ix-reach-as4455	21,982	76	0.35
ix-reach-as43531	21,911	77	0.35
jastel-as45629	2,033	32	1.57
jcom-as9824	617	9	1.46
jingdong-as131486	10,884	407	3.74
joes-datacenter-as19969	8,712	126	1.45
jp-morgan-as7743	956	3	0.31
kaan-girgin-as211376	1,763	15	0.85
kar-tel-as21299	608	14	2.30
kcom-as206509	13,435	93	0.69
kddi-as2516	41,963	836	1.99
kddi-as9597	61,434	576	0.94
keyweb-as31103	72,823	680	0.93
kgix-as53861	8,294	44	0.53

ASN	Registrable Domains	Numeric Registrable Domains	Percent
kingsoft-as59019	3,444	215	6.24
kingsoft-cloud-as137280	3,001	175	5.83
kinx-as9957	2,432	46	1.89
koos-as18042	1,109	36	3.25
kpn-as286	102,442	1,117	1.09
kpn-as1136	51,602	463	0.90
krypt-as4213	9,243	443	4.79
kyivstar-as15895	1,470	34	2.31
lanset-as16578	2,148	45	2.09
layerbridge-as3280	345	2	0.58
las-vegas-as26277	14,036	157	1.12
leapswitch-as132335	18,016	117	0.65
leaseweb-as28753	51,426	698	1.36
leaseweb-as38930	2,213	38	1.72
leaseweb-as59253	110,146	1,357	1.23
leaseweb-as133752	21,646	444	2.05
leaseweb-as394380	9,195	43	0.47
lg-hellovision-as38091	2,166	45	2.08
liberty-global-as6830	46,945	300	0.64
lightower-as46887	5,947	61	1.03
limelight-as22822	503	11	2.19
limelight-as38622	166	7	4.22
linkedin-as14413	885	0	0.00
linknet-fastnet-as23700	224	3	1.34
liquid-as30844	1,708	7	0.41
liquid-web-as36444	49,293	465	0.94
mail-ru-as47764	4,683	95	2.03
managed-net-as19366	1,425	6	0.42
manitu-as34240	52,692	392	0.74
media-land-as206728	159	3	1.89
mediacom-as30036	4,854	52	1.07
medianet-as200736	335	6	1.79
mega-cable-as13999	1,203	12	1.00
megafon-as31133	4,207	112	2.66
megapath-as4565	2,228	18	0.81
megapath-as18566	5,507	68	1.23
melbikomas-as8849	253	66	26.09
melbikomas-as56630	4,597	544	11.83
meo-as3243	7,505	60	0.80

ASN	Registrable Domains	Numeric Registrable Domains	Percent
mercedes-benz-as31399	1,024	101	9.86
micron21-as38880	12,326	100	0.81
miran-as41722	3,699	56	1.51
mivocloud-as39798	4,221	53	1.26
mod-mission-as39855	581	9	1.55
moldtelecom-as8926	580	9	1.55
mtn-sa-as16637	27,521	273	0.99
mts-as8359	7,321	116	1.58
my-tech-bz-as52449	921	0	0.00
myloc-as24961	156,113	1,382	0.89
natl-info-res-as17841	1,555	23	1.48
naver-as135354	72,330	1,192	1.65
net-by-net-as12714	3,809	72	1.89
net-sol-as47674	2,692	59	2.19
netactuate-as36236	13,128	155	1.18
netcologne-as8422	26,982	189	0.70
netdna-as54104	222	5	2.25
netia-as12741	11,625	74	0.64
netinternet-as51559	55,042	455	0.83
netminders-as7040	4,059	46	1.13
netnod-as8674	764	5	0.65
netvigator-as4760	9,302	212	2.28
netzbetrieb-as201011	4,937	53	1.07
neustar-as19905	66,132	415	0.63
new-media-as38001	9,254	134	1.45
nice-it-as49447	2,619	6	0.23
ningxia-west-as135629	2,809	67	2.39
noack-hosting-as30893	325	0	0.00
novatel-as41313	432	7	1.62
nss-as16814	26,822	103	0.38
ntt-as2514	92,614	1,206	1.30
ntt-as4713	99,965	1,114	1.11
ntt-as9293	6,034	115	1.91
ntx-tech-as50113	3,552	239	6.73
nuvera-as7385	10,675	119	1.11
ods-jsc-as45538	29,287	264	0.90
offshore-racks-as52469	1,725	15	0.87
oleksandr-siedinkin-as56485	17,464	210	1.20
ooredoo-tunisie-as37693	2,966	42	1.42

ASN	Registrable Domains	Numeric Registrable Domains	Percent
oracle-as7160	7,322	138	1.88
orange-as3215	121,670	883	0.73
orange-as5511	791	15	1.90
orange-as5617	16,701	138	0.83
orange-as9050	6,488	58	0.89
orange-es-as12479	11,128	254	2.28
orion-net-as41564	9,163	32	0.35
pair-as7859	95,894	1,103	1.15
pakistan-telecom-as17557	1,080	44	4.07
partner-comm-as12400	36,648	441	1.20
paypal-as17012	241	2	0.83
pccw-as3491	7,814	367	4.70
pengelola-as132647	837	10	1.19
petersburg-internet-network-as34665	7,637	189	2.47
petersburg-internet-network-as44050	569	17	2.99
ph-long-dist-as9299	3,584	64	1.79
phoenix-nap-as12189	985	7	0.71
phoenix-nap-as60558	6,437	94	1.46
pinvds-as41909	427	2	0.47
pkt-exch-as58065	10,242	15	0.15
pkt-host-as54825	32,299	451	1.40
plus-server-as61157	134,361	947	0.70
powercomm-as17858	2,958	158	5.34
pptech-as48090	353	3	0.85
primus-as6407	1,904	15	0.79
privex-as210083	306	6	1.96
proper-support-as51490	101	0	0.00
proximus-as5432	26,536	202	0.76
qinghai-as140061	671	22	3.28
quantil-as54994	3,048	301	9.88
rack-sphere-as39782	182	1	0.55
rackco-as36529	5,286	68	1.29
rackspace-as27357	90,224	835	0.93
rackspace-ltd-as15395	94,406	827	0.88
rascom-as20764	163	2	1.23
rcs-rds-as8708	21,092	174	0.82
reflect-as29789	12,423	116	0.93
regru-as49352	11,516	130	1.13
reliance-as18101	2,002	6	0.30

ASN	Registrable Domains	Numeric Registrable Domains	Percent
retn-as9002	7,608	49	0.64
retn-ltd-as9002	7,826	59	0.75
rices-as48693	162	1	0.62
rogers-as812	24,527	242	0.99
root-sa-as5577	3,595	53	1.47
rostelecom-as12389	29,140	753	2.58
rostelecom-as42610	4,840	127	2.62
routelabel-as198203	19,002	191	1.01
saudi-telecom-as25019	2,285	41	1.79
saudi-telecom-as39891	805	22	2.73
savecom-as9676	1,479	53	3.58
seed-net-as4780	4,673	130	2.78
sejong-as9848	19,106	335	1.75
sendgrid-as11377	35,400	277	0.78
serbia-bb-as31042	19,278	92	0.48
server-central-as23352	256,576	1,900	0.74
serverastra-as56322	722	7	0.97
servereasy-as60798	13,926	88	0.63
serverion-as399471	2,632	200	7.60
serverius-as50673	143,102	1,683	1.18
servers-as7979	45,076	784	1.74
serverstack-as46652	2,683	24	0.89
sfr-as15557	20,268	197	0.97
shanghai-mobile-as24400	2,345	146	6.23
shaw-as6327	25,332	293	1.16
shinjiru-as45839	22,699	266	1.17
shock-hosting-as395092	13,141	211	1.61
sia-nano-as52173	714	8	1.12
siamdata-as56309	12,773	195	1.53
singtel-as7473	346	4	1.16
skb-as64425	4,720	128	2.71
sky-as5607	4,261	96	2.25
skylink-as44592	5,081	46	0.91
snthostings-as140947	1,410	36	2.55
softbank-as4725	7,153	93	1.30
softbank-as17676	8,094	332	4.10
softqloud-as208006	4,018	58	1.44
sol-corp-as57910	155,584	988	0.64
sony-as18182	8,887	284	3.20

ASN	Registrable Domains	Numeric Registrable Domains	Percent
south-internet-as57416	1,381	33	2.39
spacenet-as5539	49,470	373	0.75
spirit-as2711	2,644	10	0.38
stackpath-as20446	49,909	640	1.28
stealth-as8002	2,586	20	0.77
sucuri-as30148	190,569	1,730	0.91
suddenlink-as19108	10,077	109	1.08
sungard-as7381	10,336	162	1.57
superhosting-as201200	103,581	1,024	0.99
superonline-as34984	61,096	353	0.58
swisscom-as3303	78,522	503	0.64
sz-blue-cloud-as58593	4,945	170	3.44
talktalk-as12708	1,433	14	0.98
tamatiya-as50360	518	8	1.54
tata-as4755	12,193	64	0.52
tata-as6453	3,909	25	0.64
tata-as17762	1,806	10	0.55
tata-as45820	3,774	32	0.85
tattelecom-as28840	898	20	2.23
tdc-as3292	42,286	425	1.01
tds-telecom-as4181	5,885	55	0.93
team-internet-as61969	177,510	1,648	0.93
teas-as8452	4,099	94	2.29
tel-algeria-as36947	2,030	85	4.19
tel-arg-as22927	1,022	18	1.76
tel-indo-as7713	8,586	32	0.37
tele-asia-as133398	2,029	98	4.83
tele2-as1257	54,855	351	0.64
telecentro-as27747	839	19	2.26
telecity-as15830	37,981	300	0.79
telecom-italia-as3269	34,869	283	0.81
telecom-italia-as6762	515	3	0.58
telecom-macau-as4609	1,747	36	2.06
telefonica-as12956	323	6	1.86
telefonica-as18881	12,831	90	0.70
telefonica-es-as3352	53,193	271	0.51
telehouse-as57344	781	5	0.64
telekom-srbija-as8400	4,137	30	0.73
telemar-as7738	2,474	16	0.65

ASN	Registrable Domains	Numeric Registrable Domains	Percent
telenor-as2119	18,082	143	0.79
telia-as1299	16,010	78	0.49
telia-as3301	38,208	221	0.58
telkom-as5713	7,006	48	0.69
telkom-as37457	3,686	35	0.95
telmex-as8151	11,759	125	1.06
telstra-as1221	28,326	243	0.86
telstra-as4637	2,961	58	1.96
telus-as852	22,028	164	0.74
terrahost-as56655	44,659	735	1.65
the-hut-as197651	941	5	0.53
tierpoint-as7349	7,936	87	1.10
tierpoint-as36024	11,006	601	5.46
time-as9930	2,915	35	1.20
tisp-limited-as63888	340	25	7.35
tm-net-as4788	8,686	76	0.87
tmobile-as21928	911	24	2.63
tokai-as10010	10,052	149	1.48
tot-as23969	2,115	24	1.13
tpg-telecom-as7545	19,933	163	0.82
trabia-as43289	3,821	50	1.31
transtelco-as32098	857	5	0.58
transtelecom-as20485	1,892	43	2.27
trunkoz-as58641	3,684	21	0.57
ttnet-as47331	9,545	70	0.73
tun-telecom-as2609	72,608	571	0.79
turk-telekom-as9121	19,692	148	0.75
turknet-as12735	3,231	16	0.50
tw-gateway-as9505	149	4	2.68
tw-infra-as18049	453	18	3.97
tzulo-as11878	6,664	108	1.62
uab-cherry-as16125	45,865	736	1.60
uk-dedi-as42831	65,536	985	1.50
ucsd-as7377	1,064	133	12.50
ucsd-sdsc-as195	339	5	1.47
ufanet-as24955	1,480	35	2.36
ufinet-as52468	551	6	1.09
ukr-r-and-e-as3255	1,553	14	0.90
ukr-telecom-as6849	4,418	45	1.02

ASN	Registrable Domains	Numeric Registrable Domains	Percent
uninet-as8151	11,997	123	1.03
union-bb-as24164	246	15	6.10
unitas-as1828	820	5	0.61
united-network-as39134	54,565	1,001	1.83
uniti-as13760	1,341	6	0.45
vee-time-as17809	221	10	4.52
verdina-as201133	7,871	1,035	13.15
verizon-as6167	1,303	80	6.14
verizon-as701	117,388	1,550	1.32
verizon-as702	33,868	150	0.44
verizon-as703	957	3	0.31
verizon-as2828	16,088	180	1.12
viasat-as7155	233	4	1.72
videotron-as5769	16,266	110	0.68
vidolu-as327813	4,871	39	0.80
viettel-as7552	41,736	489	1.17
viettel-as24086	717	7	0.98
vimpelcom-as3216	11,026	164	1.49
vimpelcom-as8371	467	11	2.36
vimpelcom-as8402	6,994	137	1.96
virgin-as5089	38,307	400	1.04
virt-sys-as30860	10,761	303	2.82
vkontakte-as47541	112	0	0.00
vnpt-as45899	91,976	1,248	1.36
vocus-as4826	19,662	132	0.67
vodafone-as1273	7,420	85	1.15
vodafone-as3209	99,685	732	0.73
vodafone-as8386	258	14	5.43
vodafone-as12302	5,288	39	0.74
vodafone-as15897	491	6	1.22
vodafone-as15924	12,315	104	0.84
vodafone-as31334	11,640	138	1.19
vodafone-as33915	51,594	452	0.88
vodafone-as38266	154	16	10.39
vodafone-as55410	1,241	8	0.64
volico-as33724	5,972	42	0.70
voxility-as3223	24,677	232	0.94
vpsville-as59504	4,596	1,260	27.42
wave-bb-as11404	6,127	222	3.62

ASN	Registrable Domains	Numeric Registrable Domains	Percent
we-dare-as20495	12,518	160	1.28
web-elite-as19383	1,225	10	0.82
web-werks-as33480	3,547	17	0.48
web-werks-as133296	52,772	293	0.56
webair-as20264	145	0	0.00
webair-as27257	57,809	968	1.67
webzilla-as35415	20,733	748	3.61
wenzhou-as134771	9,426	1,495	15.86
wholesail-as20055	2,591	23	0.89
wholesale-as45671	62,018	1,580	2.55
wide-open-west-as16724	1,407	6	0.43
wikimedia-as14907	1,272	30	2.36
wildcard-as34119	100,945	1,355	1.34
wind-as1267	26,108	147	0.56
windstream-as7029	36,304	426	1.17
wnet-telecom-as1820	1,308	27	2.06
worldstream-bv-as49981	66,404	1,111	1.67
wow-as12083	9,980	203	2.03
wowrack-as23033	10,762	129	1.20
wz-comm-as40824	12,707	195	1.53
xs4all-as3265	53,040	526	0.99
xsserver-gmbh-as207959	541	5	0.92
xtom-as9312	533	57	10.69
yahoo-as26101	184,577	1,944	1.05
yandex-as13238	2,857	72	2.52
yandex-cloud-as200350	40,939	652	1.59
yisp-as58073	1,743	93	5.34
you-as18207	818	6	0.73
zayo-as6461	22,200	239	1.08
zayo-as8218	10,003	62	0.62

TOTAL	12,246,452	154,592	

Appendix C. Filter List

```
$ cat stuff-to-filter.txt
001www\.com
12hp\.at
12hp\.ch
12hp\.de
1337\.pictures
2ix\.at
2ix\.ch
3utilities\.com
4lima\.at
4lima\.ch
4lima\.de
64-b\.it
advisor\.ws
affinitylottery\.org\.uk
africa\.com
amazonaws\.com
amscompute\.com
and\.mom
anthropology\.museum
appchizi\.com
applinzi\.com
appspaceusercontent\.com
aquila\.it
art\.pl
arvo\.network
at-band-camp\.net
ath\.cx
authgear-staging\.com
authgearapps\.com
awdev\.ca
backplaneapp\.io
bar1\.net
barsy\.bg
barsy\.in
barsy\.net
barsy\.online
barsyonline\.co\.uk
barsyonline\.com
be\.gy
belau\.pw
better-than\.tv
bhz\.br
biz\.my
blogdns\.com
blogdns\.net
blogdns\.org
blogsite\.org
bmoattachments\.org
boldlygoingnowhere\.org
bounceme\.net
bplaced\.com
bplaced\.de
bplaced\.net
br\.com
broke-it\.net
bsb\.br
bss\.design
builtwithdark\.com
c66\.me
cafjs\.com
camdvr\.org
casacam\.net
catholic\.edu\.au
cc\.fl\.us
cdn77\.org
cechire\.com
channelsdvr\.net
cherkasy\.ua
ciscofreak\.com
clan\.rip
cloudera\.site
cloudjiffy\.net
cloudns\.asia
cloudns\.biz
cloudns\.cc
cloudns\.eu
cloudns\.in
cloudns\.info
cloudns\.org
cloudns\.pro
cloudns\.pw
cloudns\.us
cn\.com
cn\.eu\.org
co\.business
co\.education
co\.events
co\.financial
co\.na
co\.network
co\.place
co\.technology
com\.la
com\.ss
coop\.ar
cryptonomic\.net
curitiba\.br
cx\.ua
cya\.gg
damnserver\.com
dappnode\.io
```

```
datadetect\.com
dattolocal\.com
dattolocal\.net
dd-dns\.de
ddns\.me
ddns\.net
ddnsfree\.com
ddnsgeek\.com
ddnsking\.com
ddnss\.de
ddnss\.org
de\.com
de\.cool
debian\.net
dedyn\.io
deno-staging\.dev
deno\.dev
dev\.br
digitaloceanspaces\.com
diskstation\.eu
diskstation\.me
diskstation\.org
ditchyourip\.com
dnsalias\.com
dnsalias\.net
dnsalias\.org
dnsdojo\.com
dnsdojo\.net
dnsdojo\.org
dnsfor\.me
dnshome\.de
dnsiskinky\.com
does-it\.net
doesntexist\.com
doesntexist\.org
dontexist\.com
dontexist\.net
doomdns\.com
doomdns\.org
dopaas\.com
dray-dns\.de
drayddns\.com
draydns\.de
dreamhosters\.com
drud\.io
drud\.us
dscloud\.biz
dscloud\.me
dscloud\.mobi
dsmynas\.com
dsmynas\.net
dsmynas\.org
duckdns\.org
dvrcam\.info
dvrdns\.org
dyn-ip24\.de
dyn-o-saur\.com
dyn-vpn\.de
dynalias\.com
dynalias\.net
dynalias\.org
dynathome\.net
dyndns-at-home\.com
dyndns-at-work\.com
dyndns-free\.com
dyndns-home\.com
dyndns-ip\.com
dyndns-mail\.com
dyndns-office\.com
dyndns-remote\.com
dyndns-server\.com
dyndns-web\.com
dyndns-wiki\.com
dyndns-work\.com
dyndns1\.de
dyndns\.biz
dyndns\.info
dyndns\.org
dyndns\.tv
dyndns\.ws
dynns\.com
dynu\.net
dynv6\.net
dynvpn\.de
e164\.arpa
eating-organic\.net
ed\.pw
edu\.eu\.org
edu\.it
edu\.ye
eero\.online
elasticbeanstalk\.com
en-root\.fr
encoway\.cloud
endofinternet\.net
endofinternet\.org
endoftheinternet\.org
enscaled\.us
eu\.com
evennode\.com
familyds\.com
familyds\.net
familyds\.org
fastly\.net
```

```
fastlylb\.net
faststacks\.net
fbxos\.fr
feste-ip\.net
fh-muenster\.io
firebaseapp\.com
firewall-gateway\.com
firewall-gateway\.de
floripa\.br
fnwk\.site
folionetwork\.site
for-more\.biz
for-our\.info
for\.men
for\.mom
for\.one
for\.sale
forgot\.his\.name
forumz\.info
framer\.app
freeboxos\.fr
freeddns\.org
freemyip\.com
from-ak\.com
from-al\.com
from-az\.net
from-ca\.com
from-dc\.com
from-fl\.com
from-ga\.com
from-il\.com
from-in\.com
from-ks\.com
from-ky\.com
from-la\.net
from-md\.com
from-me\.com
from-me\.org
from-mi\.com
from-ms\.com
from-mt\.com
from-nc\.com
from-nh\.com
from-nj\.com
from-nm\.com
from-ny\.net
from-oh\.com
from-ok\.com
from-pa\.com
from-pr\.com
from-sc\.com
from-sd\.com
from-tn\.com
from-tx\.com
from-ut\.com
from-va\.com
from-vt\.com
from-wa\.com
from-wi\.com
from-wv\.com
from-wy\.com
frusky\.de
fuettertdasnetz\.de
game-server\.cc
gdynia\.pl
geekgalaxy\.com
getmyip\.com
gets-it\.net
giize\.com
githubusercontent\.com
gitlab\.io
gleeze\.com
glitch\.me
gmail\.co\.com
goip\.de
golffan\.us
gotdns\.ch
gotdns\.com
gotdns\.org
goupile\.fr
gov\.fj
groks-the\.info
gwiddle\.co\.uk
half\.host
ham-radio-op\.net
hicam\.net
hk\.com
hobby-site\.com
home-webserver\.de
homedns\.org
homeftp\.net
homeftp\.org
homeip\.net
homelinux\.com
homelinux\.net
homelinux\.org
homeoffice\.gov\.uk
homeunix\.com
homeunix\.net
homeunix\.org
hopto\.me
hopto\.org
hotelwithflight\.com
hu\.com
```

```
i234\.me
iamallama\.com
ignorelist\.com
ik-server\.com
impertrix\.com
in-addr\.arpa
in-the-band\.net
instantcloud\.cn
ip6\.arpa
is-a-anarchist\.com
is-a-chef\.com
is-a-chef\.net
is-a-geek\.com
is-a-geek\.net
is-a-geek\.org
is-a-hunter\.com
is-a-lawyer\.com
is-a-liberal\.com
is-a-musician\.com
is-a-painter\.com
is-a-photographer\.com
is-a-player\.com
is-a-rockstar\.com
is-a-student\.com
is-a-teacher\.com
is-a-therapist\.com
is-an-actress\.com
is-an-artist\.com
is-an-engineer\.com
is-by\.us
is-certified\.com
is-leet\.com
is-lost\.org
is-slick\.com
is-very-sweet\.org
is-with-theband\.com
isa-geek\.com
isa-geek\.net
issmarterthanyou\.com
istmein\.de
jelastic\.cloud
jelastic\.com
jele\.host
jotelulu\.cloud
jp\.net
kicks-ass\.net
kicks-ass\.org
knowsitall\.info
kommune\.no
kozow\.com
kyiv\.ua
l-o-g-i-n\.de
lib.ms\.us
likes-pie\.com
likescandy\.com
lima-city\.at
lima-city\.ch
lima-city\.de
lima-city\.rocks
lima\.zone
linodeusercontent\.com
lmpm\.com
log\.br
loginto\.me
lolipop\.io
loseyourip\.com
lublin\.pl
magentosite\.cloud
mayfirst\.org
mcdir\.ru
mex\.com
mil\.ph
mine\.nu
mircloud\.us
mo-siemens\.io
my-firewall\.org
my-gateway\.de
my-router\.de
my-wan\.de
myasustor\.com
myddns\.rocks
mydissent\.net
mydobiss\.com
mydrobo\.com
myds\.me
myfirewall\.org
myfritz\.net
myftp\.biz
myftp\.org
myhome-server\.de
myiphost\.com
mypep\.link
myphotos\.cc
mypi\.co
mypsx\.net
myqnapcloud\.com
mysecuritycamera\.com
mysecuritycamera\.net
myspreadshop\.at
myspreadshop\.be
myspreadshop\.ca
myspreadshop\.ch
myspreadshop\.co\.uk
myspreadshop\.com
```

myspreadshop\.de
myspreadshop\.dk
myspreadshop\.es
myspreadshop\.fi
myspreadshop\.fr
myspreadshop\.ie
myspreadshop\.it
myspreadshop\.net
myspreadshop\.nl
myspreadshop\.no
myspreadshop\.pl
myspreadshop\.se
myvnc\.com
mywire\.org
n4t\.co
nabu\.casa
name\.ng
neat-url\.com
nerdpol\.ovh
net\.eu.org
net\.mz
netlify\.app
newyork\.museum
ngo\.ph
ngrok\.io
nhs\.uk
nid\.io
nl\.eu\.org
no-ip\.biz
no-ip\.ca
no-ip\.info
no-ip\.net
no-ip\.org
nodeart\.io
noho\.st
nohost\.me
noip\.me
noip\.us
noop\.app
now\.sh
nsupdate\.info
ntdll\.top
office-on-the\.net
official\.academy
omg\.lol
on-rio\.io
onavstack\.net
ong\.br
onthewifi\.com
ooguy\.com
or\.pw
org\.ph
outsystemscloud\.com
ownip\.net
oxa\.cloud
pagefrontapp\.com
pagexl\.com
panel\.gg
pantheonsite\.io
pixolino\.com
platform\.sh
platformsh\.site
platter-app\.com
platterp\.us
plesk\.page
poa\.br
podzone\.net
point2this\.com
pointto\.us
police\.uk
ponpes\.id
pony\.club
prequalifyme\.today
primetel\.cloud
pstmn\.io
quickconnect\.to
quipelements\.com
radio\.am
radio\.fm
raffleentry\.org\.uk
read-books\.org
reclaim\.cloud
redirectme\.net
regruhosting\.ru
remotewd\.com
render\.com
repair\.men
repl\.co
reservd\.com
reserve-online\.com
reserve-online\.net
ribeirao\.br
rio\.br
ripe\.net
roma\.museum
rs\.gov\.br
ru\.com
ru\.net
sa\.com
sandcats\.io
scaleforce\.net
schulserver\.de
scienceandindustry\.museum
securitytactics\.com

```
seg\.br
seidat\.net
selfip\.biz
selfip\.com
selfip\.net
selfip\.org
sells-it\.net
sendgrid\.com
senseering\.net
servebbs\.com
servebbs\.net
servebbs\.org
servebeer\.com
serveexchange\.com
serveftp\.com
serveftp\.net
serveftp\.org
servegame\.com
servegame\.org
servehalflife\.com
servehttp\.com
servehumour\.com
serveirc\.com
serveminecraft\.net
servemp3\.com
servepics\.com
servequake\.com
shoemakers\.com\.ph
shop\.th
shopware\.store
shw\.io
siiites\.com
skygearapp\.com
spdns\.de
spdns\.eu
spdns\.org
square7\.ch
square7\.de
square7\.net
stackhero-network\.com
stufftoread\.com
syncloud\.it
syno-ds\.de
synology-ds\.de
synology\.me
sytes\.net
tabitorder\.co\.il
taifun-dns\.de
tcp4\.me
teaches-yoga\.com
tec\.br
temp-dns\.com
tempurl\.host
that\.win
theworkpc\.com
thingdustdata\.com
thruhere\.net
tickets\.io
topology4\.dyndns\.atlas\.ripe\.net
traeumtgerade\.de
trafficplex\.cloud
trendhosting\.cloud
tst\.site
tuxfamily\.org
uk\.com
under\.one
unispace\.io
unusualperson\.com
upli\.io
us\.com
vapor\.cloud
vaporcloud\.io
vbrplsbx\.io
vercel\.app
vercel\.dev
virtual-user\.de
vp4\.me
vpndns\.net
vpnplus\.to
wallonie\.museum
we\.bs
we\.tc
web\.app
webhop\.biz
webhop\.info
webhop\.me
webhop\.net
webhop\.org
webredirect\.org
webspace\.rocks
weeklylottery\.org\.uk
wellbeingzone\.eu
woltlab-demo\.com
workisboring\.com
worse-than\.tv
wpmudev\.host
writesthisblog\.com
x443\.pw
xnbay\.com
yahoo\.com\.ph
ynh\.fr
zapto\.org
```

Appendix D. 2nd-level-dom-large script

```
$ cat 2nd-level-dom-large.pl
#!/usr/bin/perl
use strict;
use warnings;
use IO::Socket::SSL::PublicSuffix;

my $pslfile = '/usr/local/share/public_suffix_list.dat';
my $ps = IO::Socket::SSL::PublicSuffix->from_file($pslfile);

while (my $line = <STDIN>) {
        chomp($line);
        my $root_domain = $ps->public_suffix($line,1);
        printf( "%s\n", $root_domain );
}
```

Appendix E. Sample Python3 Code to Make a Multi-ASN Violin Plot

```
$ cat read_csv.py
#!/usr/local/bin/python3
""" produce sample violin plot for multiple ASNs """

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib.offsetbox import OffsetImage
from matplotlib.offsetbox import AnnotationBbox


### set various configuration variables
datafile='all-data.csv'
first_title="Per-ASN Risk Scores (Computed on a Subset of Domains)\n"
second_title="Selected Assortment of ASNs (page 1 of 8)\n"
mytitle=first_title+second_title
myplot='violin-plot.pdf'
flags_datafile='asn-to-country-code.csv'

### variables we're reading in: asn,domain,risk
### assumes those column names are present in the data file, otherwise
### define them here
df = pd.read_csv(datafile)
asns_seen = df.asn.unique()
asn_count = asns_seen.size

### uncomment the following to see all rows
# pd.set_option('display.max_rows', None)
# print(df)

### we want to break our data into groups by asn, and then sort by
### the media risk score of each of the groups. That requires adding
### that median risk score to the data frame
df['median_risk'] = df.groupby('asn')['risk'].transform('median')

### sort the data frame by the median risk score per group, and
### if there are any ties, break the ties by alphabetizing the asn name
mydf = df.sort_values(['median_risk', 'asn'], ascending = [True, True])

### make the obs numbers ascending across the newly reordered groups
mydf = mydf.reset_index(drop=True)

### if you want to print all rows for checking, uncomment the next two lines
# pd.set_option('display.max_rows', None)
# print(mydf)

### format the output for legal format (14x8.5") with half inch margins
### on the top and sides, and an inch on the bottom
plt.figure(figsize=(13, 7.0))
```

```
### use the default seaborn theme (gray background, white grid lines)
### https://seaborn.pydata.org/generated/seaborn.set_theme.html
sns.set_theme()

### I quote: 'Many people find the moderated hues of the default
### "deep" palette to be aesthetically pleasing', other options at
### https://seaborn.pydata.org/tutorial/color_palettes.html
sns.color_palette("deep")

### set_context tweaks a bunch of font parameters
### https://seaborn.pydata.org/generated/seaborn.set_context.html
### note that "talk" appears prone to making asn labels overlap
sns.set_context("paper")

### finally ready to do the violinplot!
### https://seaborn.pydata.org/generated/seaborn.violinplot.html
###
### cut = 0 ensures the violins won't extend beyond the observed data
### scale options? None of the options are perfect/all have limitations.
ax = sns.violinplot(data=mydf, x='asn', y='risk', cut=0, scale='width')

### we need the grouped median risk and asn names for annotating purposes
### note that we MUST use sort=False or the annotation won't line up
### with the graphs!
median_df = mydf.groupby('asn',sort=False)['risk'].median()
count_df  = mydf.groupby('asn',sort=False)['risk'].count()
asn_df    = mydf.groupby('asn',sort=False).groups.keys()

### we'll convert the asn names and median risk scores to a dictionary
res = dict(zip(asn_df, median_df))

### now we'll convert the asn names and counts to a dictionary, too
res2 = dict(zip(asn_df, count_df))

### to verify the content of the dictionaries, uncomment the following
# print(str(res))
# print(str(res2))

### loop over the selected asns, annotating each group with the median risk
### score (we need the "i" variable to track which violin we're working on)
i=0
for k,v in res.items():
    ax.annotate('median\nrisk: '+str(int(v)), size=10,
        xy=(i,v), xycoords='data',
        xytext=(-60,+12), textcoords='offset points',
        bbox=dict(boxstyle='roundtooth, pad=.5', facecolor='white',\
            edgecolor='black'),
        arrowprops=dict(arrowstyle='simple', facecolor='white', \
            edgecolor='black'))
    i=i+1
```

```
### and now handle adding the counts below each plot
i2=0
for k2,v2 in res2.items():
    ### add commas between thousands, and a prefix N=
    centered_final=f'N={v2:,}'

    ### now write the count, centered below each violin
    ### we surround it with a hidden "box" set in background grey color
    ax.annotate(f'{centered_final:^s}',size=7,\
        xy=(i2,-3.5), xycoords='data',\
        horizontalalignment='center',\
        bbox=dict(boxstyle='roundtooth, pad=.2',\
        facecolor='#EAEAF2', edgecolor='#EAEAF2'))
    i2=i2+1

### put up a title
plt.suptitle(mytitle, size=16, fontweight="bold", y=.97, x=.03,
    horizontalalignment='left', verticalalignment='top')

### want to maximize use of the plot area (margins are "bad" :-))
plt.autoscale(enable=False, axis='both', tight=True)

### this right hand title block is purely a matter of full disclosure
### and could be omitted if desired.
plt.title("Process: (1) For each ASN, get the IPv4 prefixes it originates  "+\
"(2) Lookup prefixes in DNSDB (timefencing to past 30 days)\n"+\
"(3) Condense to unique base domains  (4) Keep domains matching ^[0-9].*$"+\
"  (5) Get Domaintools risk scores  (6) Graph.\n", \
size=8, horizontalalignment='left',verticalalignment='bottom')

### handle the horizontal axis (no axis label required)
plt.xlabel('')
plt.xticks(fontsize=10, weight='bold')

### handle the vertical axis
### the y is set to 104 to ensure we get the 100 label intact
plt.ylabel('DomainTools Risk Score', fontweight="bold")
plt.yticks(np.arange(-10, 104, step=10))

### maximize usable space by asking for tight margins
plt.subplots_adjust(left=0.055, right=0.98, top=0.90, bottom=0.08)

### now going to annotate the plots with the provider's home country's flag
### variables we're reading in: provider_asn,country_code
flags_df = pd.read_csv(flags_datafile)

### we'll convert the provider_asn and country code to a dictionary
res3 = flags_df.set_index('provider_asn').to_dict()['country_code']

### Now actually add the little flags to the bottom of each violin plot
i3=0
for k,v in res.items():
```

```
    my_cc = res3.get(k)
    ### handle possibility that we missed mapping an ASN to relevant flag
    if my_cc:
        pass
    else:
        print(f'country code for {k} is unknown.\nupdate {flags_datafile}')
    ### construct filepath
    my_argument = "flags/"+str(my_cc)+".png"
    ### read the image
    ### https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.imread.html
    arr_image = plt.imread(my_argument, format='png')
    ### scale it to an appropriate size (my std US flag is 74x39 pixels)
    ### beware pixelation effects!
    imagebox = OffsetImage(arr_image, zoom=0.3)
    ### the follow "bounding box" is in "magic background" grey
    ab = AnnotationBbox(imagebox, (i3, -7), pad=0,
        bboxprops=dict(edgecolor='#EAEAF2'))
    ### we're cramming a random blob onto the plot so we'll use "add_artist"
    ### see https://matplotlib.org/3.5.0/tutorials/intermediate/artists.html
    ax.add_artist(ab)
    i3=i3+1

# This final obscure code lays a white blob over the bottom-most Y axis label
# because I don't want it to say "-10" on the graph (yes, this is a hack)
ax.annotate('   ', size=10,
        xy=(0,-7), xycoords='data',
        xytext=(30,-15), textcoords=('figure points','offset points'),
        bbox=dict(boxstyle='round',pad=.6,facecolor='white',edgecolor='white'))

### https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.draw.html
plt.draw()

### save the output to a plot file
### myplot has a pdf extension, so it will write a PDF format file by default
plt.savefig(myplot,dpi=300)
```

Appendix F. Python3 Data Reformatting Code to Set Up the Decomposed (“Per-Risk Subscore”) Single-ASN Violin Plot Data

```
$ cat read_jsonl.py
#!/usr/local/bin/python3
""" read and reformat the risk scores for per-risk plotting """

import ijson

### need to get the risk score output
datafile='amazon-as16509.jsonl'
### we'll use the ASN info to tag our output as belong to this ASN
my_asn='amazon-as16509'

with open(datafile, encoding="utf8") as f:

    ### For choice of backend, see https://pypi.org/project/ijson/#id3
    mybackend = ijson.get_backend('python')

    ### https://pypi.org/project/ijson/#id1 says:
    ### "The multiple_values option (defaults to False) controls whether
    ### multiple top-level values are supported. JSON content should
    ### contain a single top-level value (see the JSON Grammar). However
    ### there are plenty of JSON files out in the wild that contain
    ### multiple top-level values, often separated by newlines. By default
    ### ijson will fail to process these with a parse error: trailing
    ### garbage error unless multiple_values=True is specified."
    ###
    ### Welcome to JSON Lines format, my friends :-)
    ###
    parser = ijson.items(f, 'response', multiple_values=True)

    for objects in parser:
        mydomain=objects['domain']
        myrisk_score=objects['risk_score']
        mycomps=objects['components']

        ### we're NOT writing headers here, but if we were, we'd add
        ### my_asn,mydomain,type,risk

        ### mycomps may have multiple values, so we iterate over them here
        for myobject in mycomps:
            print(my_asn,          ",",
                mydomain,          ",",
                myobject['name'], ",",
                myobject['risk_score'])
```

Appendix G. Sample Python3 Code to Make the Decomposed ("Per-Risk Subscore") Single-ASN Violin Plots

```
$ cat make_per_as_plot.py
#!/usr/local/bin/python3
""" make a per-ASN violin plot breaking out the various risk components """

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

datafile='all-risks-output.csv'
first_title="Amazon-as16509 Risk Scores Breakdown (Computed on a Subset of
Domains)\n"
mytitle=first_title
myplot='bar-plot.pdf'

### variables we're reading in: my_asn,mydomain,type,risk
### assume those column names are present in the data file, otherwise
### define them here
df = pd.read_csv(datafile)

### if you want to print all rows for checking, uncomment the next two lines
# pd.set_option('display.max_rows', None)
# print(df)

### format the output for legal format (14x8.5") with half inch margins
### on the top and sides, and an inch on the bottom
plt.figure(figsize=(13, 7.0))

### use the default seaborn theme (gray background, white grid lines)
### https://seaborn.pydata.org/generated/seaborn.set_theme.html
sns.set_theme()

### I quote: 'Many people find the moderated hues of the default
### "deep" palette to be aesthetically pleasing', other options at
### https://seaborn.pydata.org/tutorial/color_palettes.html
sns.color_palette("deep")

### set_context tweaks a bunch of font parameters
### note that "talk" appears prone to making asn labels overlap
sns.set_context("paper")

### finally ready to do the violinplot!
### https://seaborn.pydata.org/generated/seaborn.violinplot.html
###
### cut = 0 ensures the violins won't extend beyond the observed data
### scale options? None of the options are perfect/all have limitations.
### NOTE THAT THIS SCALE OPTION IS DIFFERENT FROM THE OTHER VIOLIN PLOT
ax = sns.violinplot(data=df, x='type', y='risk', cut=0, scale='count')
```

```
### we need the grouped median risk and asn names for annotating purposes
### note that we MUST use sort=False or the annotation won't line up
### with the graphs!
median_df = df.groupby('type',sort=False)['risk'].median()
count_df  = df.groupby('type',sort=False)['risk'].count()
type_df   = df.groupby('type',sort=False).groups.keys()

### we'll convert the asn names and median risk scores to a dictionary
res = dict(zip(type_df, median_df))

### now we'll convert the asn names and counts to a dictionary
res2 = dict(zip(type_df, count_df))

### to verify the content of the dictionaries, uncomment the following
# print(str(res))
# print(str(res2))

### loop over the types, annotating each group with the median risk score
### we need the "i" variable to track which violin we're working on
i=0
for k,v in res.items():
    ax.annotate('median\nrisk: '+str(int(v)), size=10,
       xy=(i,v), xycoords='data',
       xytext=(-60,+12), textcoords='offset points',
       bbox=dict(boxstyle='roundtooth, pad=.5', facecolor='white',
           edgecolor='black'),
       arrowprops=dict(arrowstyle='simple', facecolor='white',
           edgecolor='black'))
    i=i+1

### and now handle adding the counts below each plot
i2=0
for k2,v2 in res2.items():
    ### add commas between thousands, and a prefix N=
    centered_final=f'N={v2:,}'

    ### now write the count, centered below each violin
    ### we surround it with a hidden "box" set in background grey color
    ax.annotate(f'{centered_final:^s}',size=12,\
        xy=(i2,-7), xycoords='data',\
        horizontalalignment='center',\
        bbox=dict(boxstyle='roundtooth, pad=.2',\
        facecolor='#EAEAF2', edgecolor='#EAEAF2'))
    i2=i2+1

### put up a title
plt.suptitle(mytitle, size=16, fontweight="bold", y=.97, x=.03,
     horizontalalignment='left', verticalalignment='top')

plt.autoscale(enable=False, axis='both', tight=True)
```

```
### handle the horizontal axis (no axis label required)
plt.xlabel('')
plt.xticks(fontsize=10, weight='bold')

### handle the vertical axis
### the y is set to 104 to ensure we get the 100 label intact
plt.ylabel('DomainTools Risk Score', fontweight="bold")
plt.yticks(np.arange(-10, 104, step=10))

### maximize usable space by asking for tight margins
plt.subplots_adjust(left=0.055, right=0.98, top=0.90, bottom=0.08)

### This final obscure code lays a white blob over the bottom-most Y axis label
### because I don't want it to say "-10" on the graph (yes, this is a hack)
ax.annotate('   ', size=10,
        xy=(0,-7), xycoords='data',
        xytext=(30,-15), textcoords=('figure points','offset points'),
        bbox=dict(boxstyle='round',pad=.6,facecolor='white',edgecolor='white'))

### https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.draw.html
plt.draw()

### save the output to a plot file
### myplot has a pdf extension, so it will write a PDF format file by default
plt.savefig(myplot,dpi=300)
```

Eric Johnson, *Cliffs of Dover* (1984)
(Grammy Award for Best Rock Instrumental Performance (1992))

https://www.youtube.com/watch?v=vyntQ-WZUI8
[657,480 views as of October 18, 2022]