DomainTools Investigations

YEAR IN REVIEW

2024

The sheer volume of newly observed domains in 2024 was over 106 million—approximately 289,000 daily—creating a significant challenge for security teams. Rapid identification and evaluation are critical. This report provides actionable insights by examining a large sampling of worldwide publicly reported malicious domains and the global scale of all newly observed domains of that year. We showcase various analytical techniques, including:

- Domain Attribute Analysis
- Website Title Analysis
- Risk Scoring
- Domain Generation Algorithm (DGA) Detection
- Keyword Likeness Assessment
- New Top-Level Domain (TLD) Analysis
- IDN Homoglyphs / Topic Likeness Distance Analysis

These techniques can reveal valuable domain behavior insights. This collection of adaptable approaches aims to support security researchers, threat intelligence analysts, brand protection teams, and incident responders in collaboratively understanding domain intelligence.

Domain Intelligence for Security Teams

Domain intelligence data is a powerful tool for security teams and organizations to understand and manage domain-related risks for mitigating spam and phishing attempts, informing incident response (IR) efforts and detection engineering, monitoring for brand infringements, and providing real-time information for threat intelligence and mitigations.

DomainTools prides itself on being a leading provider of domain intelligence and provides a suite of services to monitor and investigate a comprehensive range of domain-related data.

Retrospectives identify trends and anomalous behaviors and help isolate models of behavior that can be later used to identify suspicious activities and domains before they're used for malicious purposes.

To identify potential threat trends, this report compares 106 million newly observed domains from 2024 against a reference set of 395,000 known malicious domains. These malicious domains encompass infrastructure utilized by both nation-state sponsored Advanced Persistent Threat (APT) groups and cybercrime operations. Their uses are diverse and harmful, including hosting websites designed for malware delivery and credential harvesting, serving as Command and Control (C2) servers to manage compromised systems, functioning as relay and obfuscation networks to hide malicious activity, operating as part of botnets for large-scale attacks, and facilitating phishing campaigns to deceive users.

Overview of the Methods Employed:

1. Domain Attribute Analysis (Registration and Resolution Details):

Methods: Analyzing IP addresses, ISPs, registrars, nameservers, SSL issuers, and combinations thereof.

Purpose: Identify patterns and correlations between these attributes and malicious activity and reveal common hosting and registration practices used by threat actors. This helps establish proximity risk associations and identify high-risk providers.

2. Website Title Analysis:

Methods: Examining the titles of websites associated with domains.

Purpose: Identify content themes and keywords indicative of malicious intent, such as those related to phishing, scams, or malware distribution.

3. Risk Score Assessments:

Methods: Assigning risk scores based on various domain attributes and behaviors.

Purpose: Quantify the likelihood of a newly registered domain being malicious; enabling prioritization of domains for further investigation and threat mitigation.

4. DGA Detection (Entropy, Length, Standard Deviations):

Methods: Statistical analysis of domain name characteristics (entropy, length, standard deviations) to identify Domain Generation Algorithms (DGAs).

Purpose: Uncover domains generated by automated systems used by malware to evade detection, revealing communication channels used by botnets and other threats.

5. Keyword and Topic Analysis:

Methods: Search for keywords and analyze topics within domain names and associated content.

Purpose: Identify domains related to specific malicious activities (malware delivery, credential harvesting, scams) and emerging threat trends.

6. New TLD Analysis:

Methods: Focus on newly registered Top-Level Domains (TLDs).

Purpose: Identify emerging threat vectors and understand how threat actors utilize new TLDs in their campaigns.

7. IDN Homoglyphs / Topic Likeness Distance Analysis:

Methods: Measure the similarity of domain names to those of high-profile media events or brands.

Purpose: Identify domains used for typosquatting, phishing, and other deceptive tactics that exploit public interest in current events.

Newly Observed Domains and Threat Indicator Domain Counts in 2024

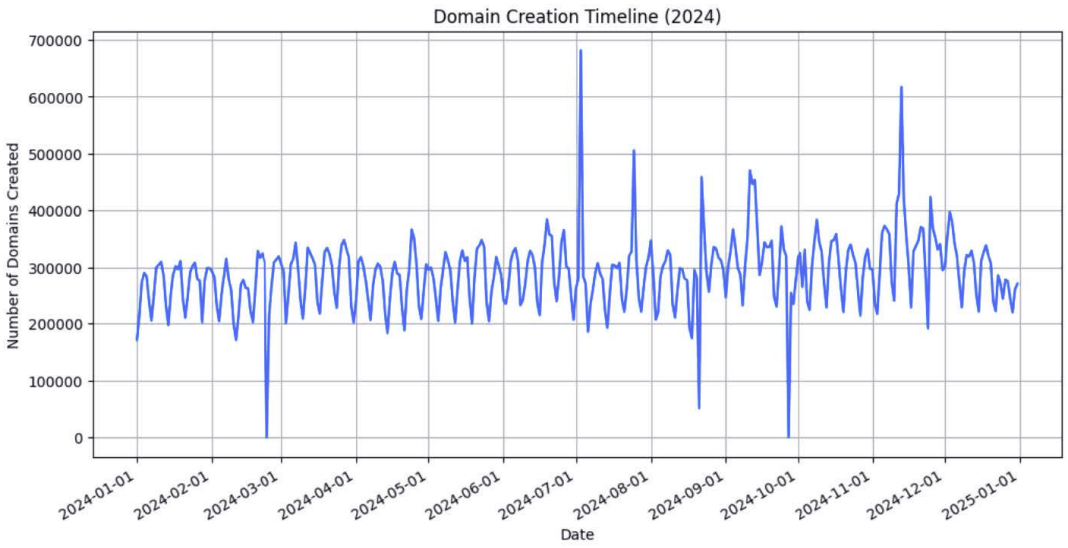
Analysis of domain intelligence data for 2024 reveals a clear upward trend in both Newly Observed Domains (NODs) and Threat Indicator Domain counts. Histograms representing the daily distribution of these metrics illustrate a consistent increase in volume throughout the year. Notably, both datasets exhibit significant spikes, ranging from 1.5 to 2.5 times the average daily count, concentrated primarily within the second half of 2024. These pronounced fluctuations, visualized below, warrant further investigation, and subsequent sections of this report will delve into techniques for analyzing and understanding these anomalous surges in domain activity.

significant spikes, ranging from

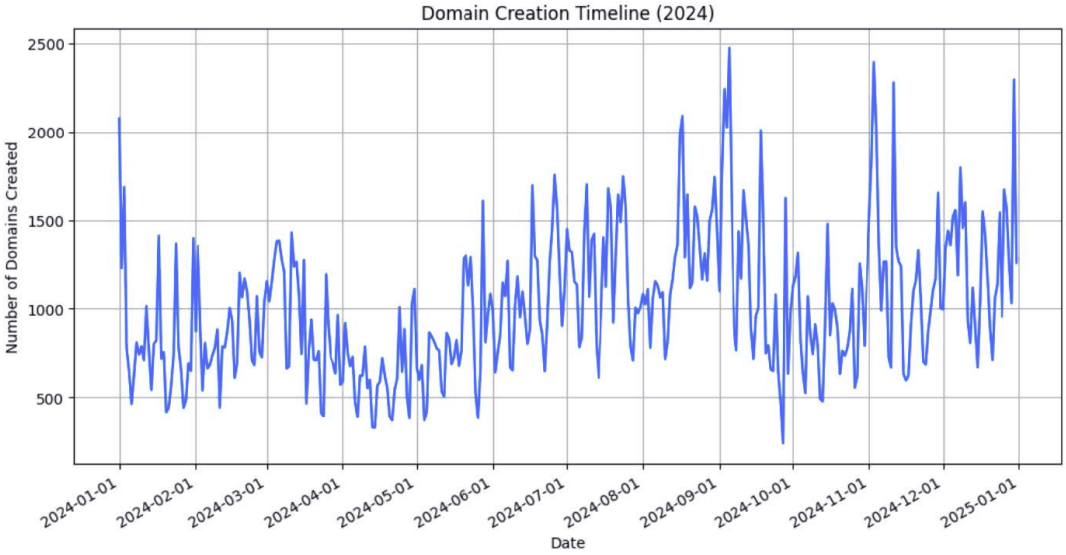
1.5 to **2.5**

times the average daily count

106 Million NODs Daily Counts over 2024



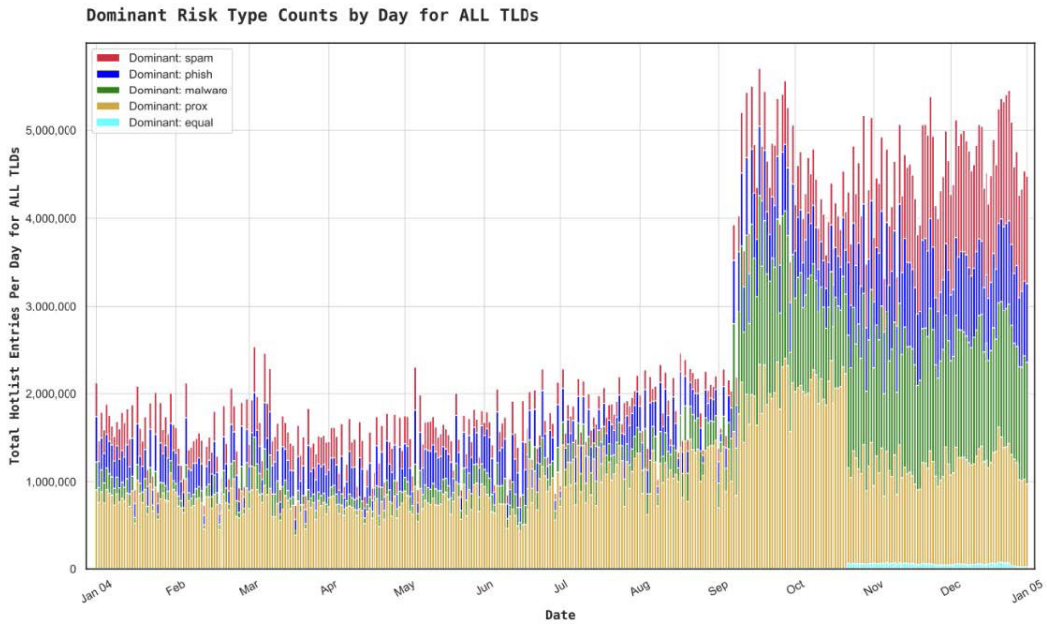
380 Thousand Threat Indicator Domain Daily Creation Counts Over 2024



DomainTools routinely improves its data visibility and risk scoring methodologies. In October 2024, a new category “equal” was introduced along with additional improvements to existing risk scoring categories. These improvements are reflected in the diagram below. The “equal” category consists of domains that scored ”equally badly” on all four risk subscores, an infrequent occurrence.

Overall Risk Counts by Subscore Category:

(”Malware” vs. ”Phishing” vs. ”Spam” vs. ”Proximity”)



Examining Newly Observed Public Threat Indicator Domains in 2024

Analyzing the hosting and registration information of publicly reported malicious domains reveals recurring patterns, such as common Registrars, ISPs, Name Servers and SSL Issuers. These focal points allow threat researchers to establish proximity risk associations. For instance, a high percentage of malicious domains hosted on a specific ISP or using a particular nameserver may indicate an elevated risk. When new domains exhibit these same ISP and nameserver combinations, they warrant further scrutiny due to their proximity to known high-risk environments.

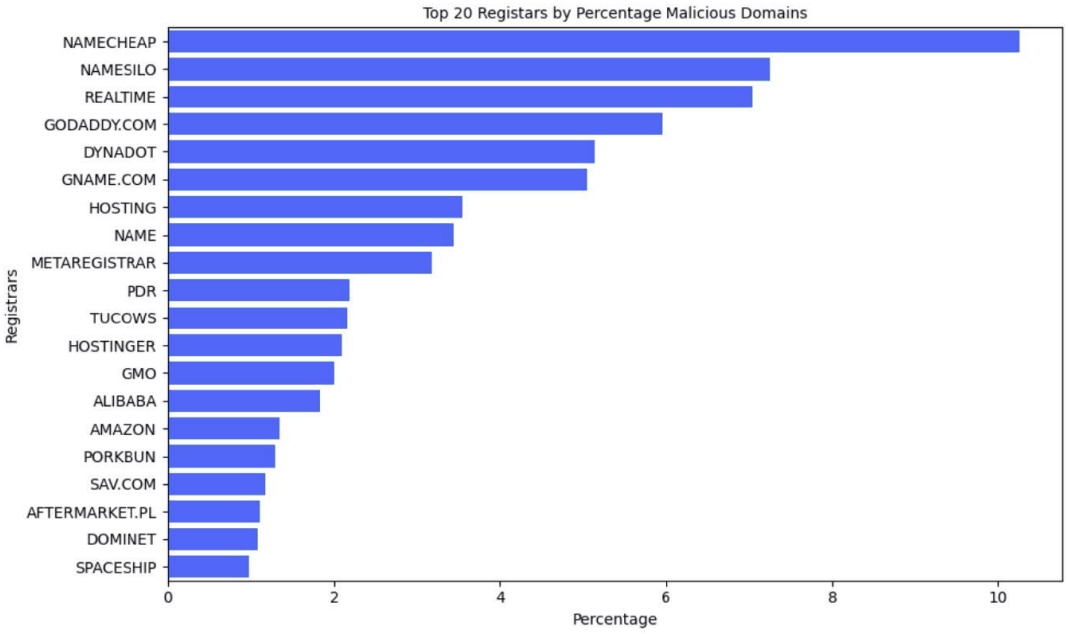
To understand the operational characteristics of malicious domains, security researchers analyze various webpage attributes that reveal domain usage and intent. These attributes, including MX records, website response codes, redirect values, SSL certificate information, and site analytics/tracking codes, provide insights into how threat actors are leveraging domains for malicious activities. Observing these elements allows security researchers to identify patterns in attack infrastructure, understand the scale and sophistication of malicious campaigns, and develop more effective detection strategies. The following table details the prevalence of these attributes within the analyzed set of malicious domains.

Attribute	Domain Count
Total Indicator Domains	394,889
MX Record	146,301
Website Response	310,182
SSL Info	335,293
Redirect Value	28,786
Google Analytics	16,148
GTM Codes	8,863
GA4 Codes	29,307
Facebook Codes	3,968
Baidu Codes	9,393
Matomo Codes	205
Hotja Codes	1,168

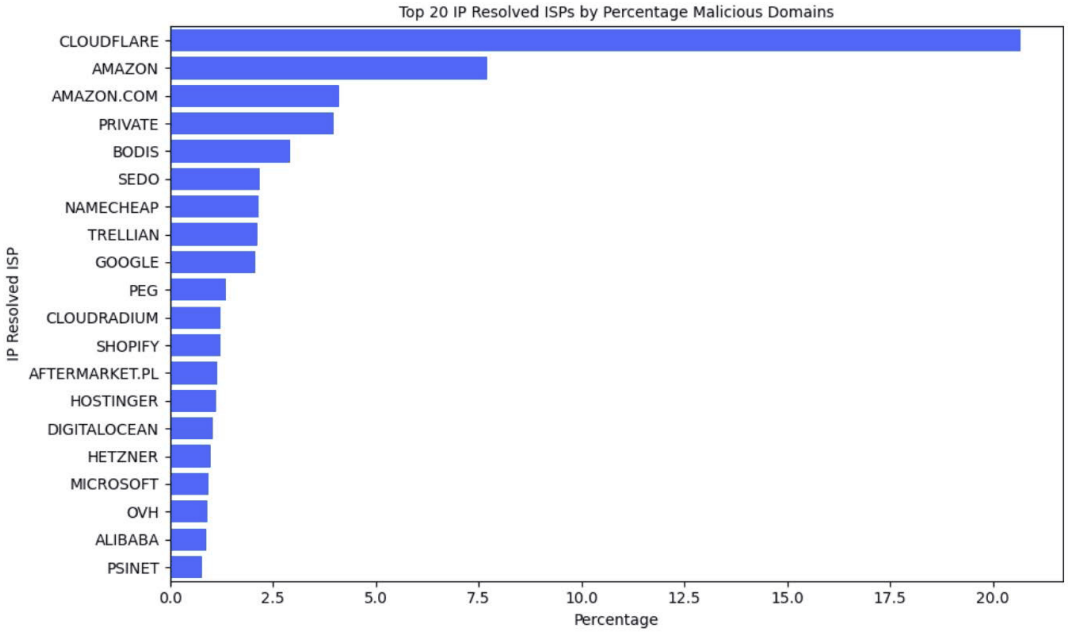
Trends in Registration and Resolution Selections

The disproportionate use of certain providers may indicate preferred platforms for malicious actors or those offering easier account setup. This observation highlights the need for increased scrutiny of domains associated with these providers in security assessments. It could also reflect user preferences, ease of configuration, or even ineffective or easily undermined fraudulent account and abuse mitigations within those platforms enabling malicious actors to continue operating with impunity.

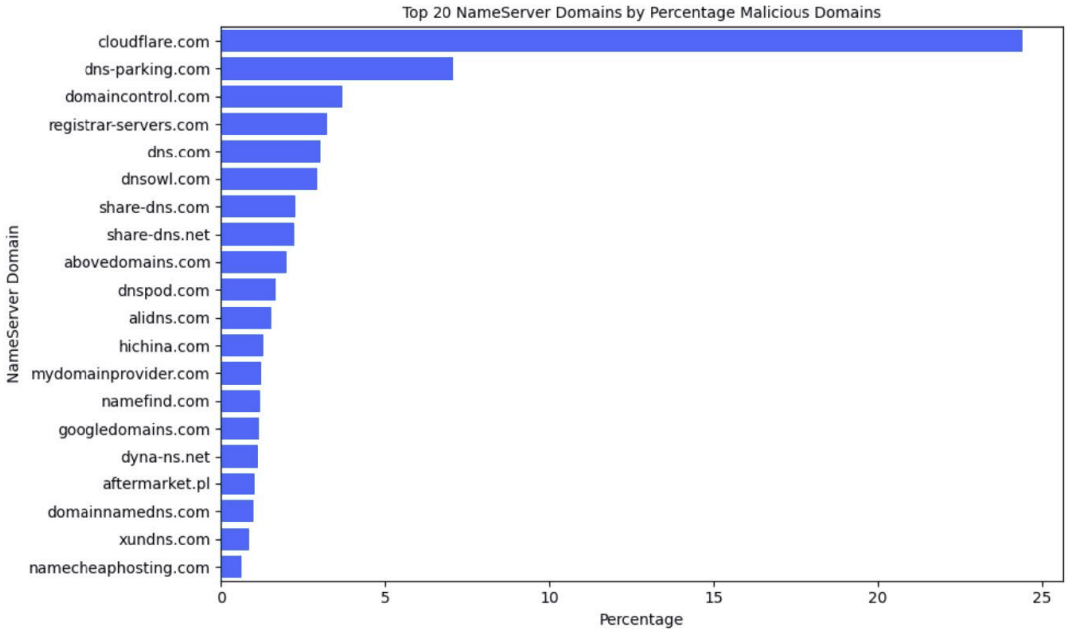
Top 20 Registrars



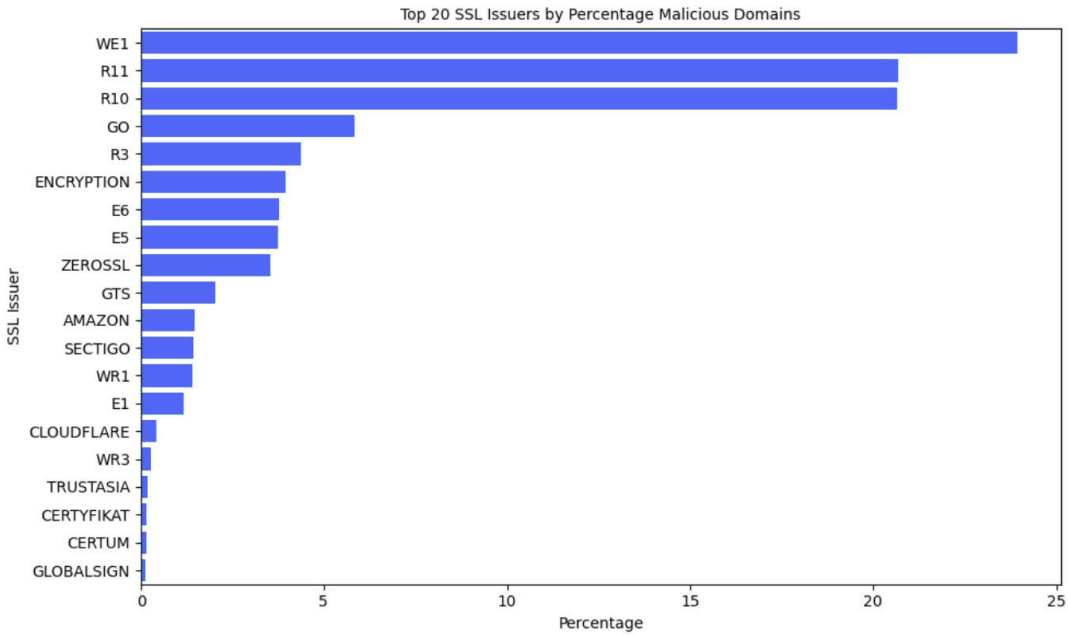
Top 20 IP Resolved Internet Service Providers



Top 20 NameServer Domains



Top 20 SSL Issuers



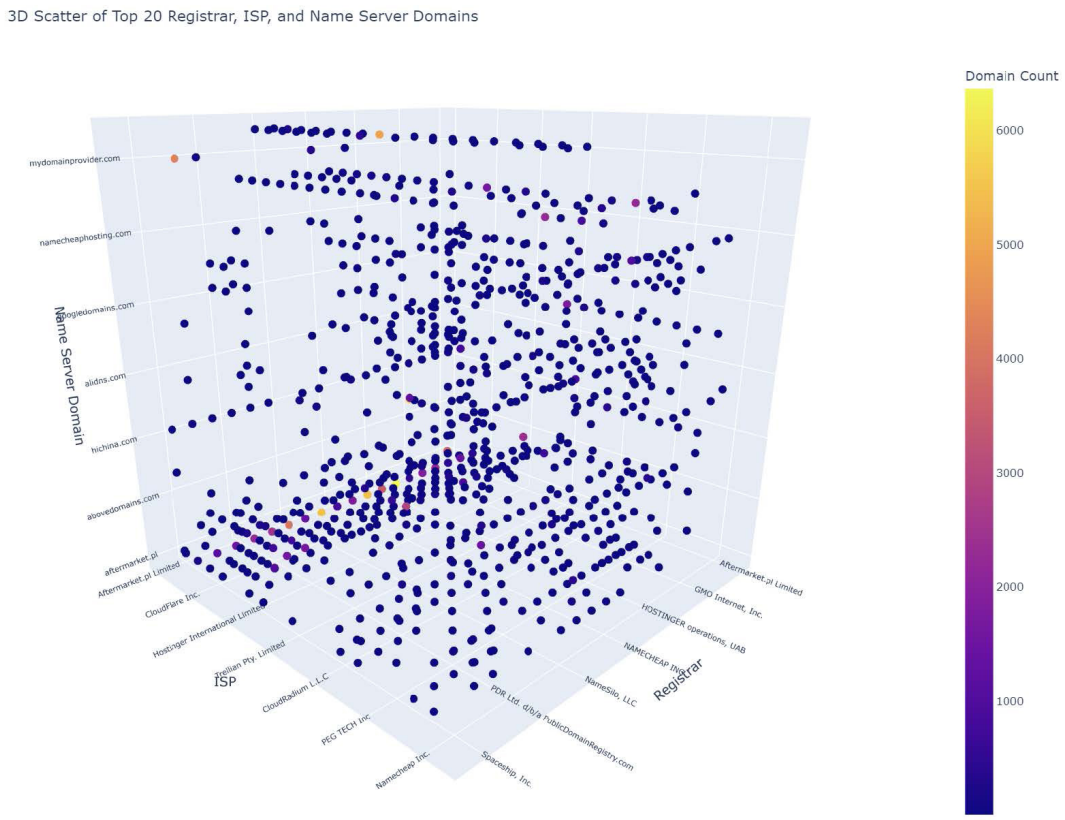
Domain registrars face ongoing challenges in mitigating malicious domain registrations, despite their pivotal role in the Internet ecosystem. For example, specific registrars have been repeatedly implicated in high-profile malicious activities. The New York Times reported on an online disinformation campaign originating in Iceland, and The Record detailed Russian disinformation campaigns surrounding U.S. elections, both of which involved domains registered through these services. Although registrars are subject to ICANN policies that require them to address fraudulent and illegal domain use, the complex nature of Internet infrastructure can sometimes obscure clear lines of responsibility.

However, registrars remain legally responsible for the domain registration itself. Inline with the shared responsibility model prevalent in online services, registrars provide the registration service, while customers are responsible for the content associated with their domains. Acceptable Use Policies (AUPs) outline permissible activities, but the primary obstacle is the timely and effective identification and mitigation of malicious outliers. The sheer volume of domain registrations makes proactive enforcement an extremely difficult task, rather than a matter of policy deficiency.

Assessing for Convergence

The convergence of high-risk attributes elevates threat levels, especially when linked to service providers with high concentrations of malicious activity. This creates a dynamic where malicious actors leverage prominent hosting providers to blend in with legitimate traffic, exploiting the sheer volume to obscure their activities. Providers, aware of this, must continually refine their mitigation strategies to maintain trust and avoid reputational damage. Annual retrospectives play a vital role in ensuring this ongoing adaptation.

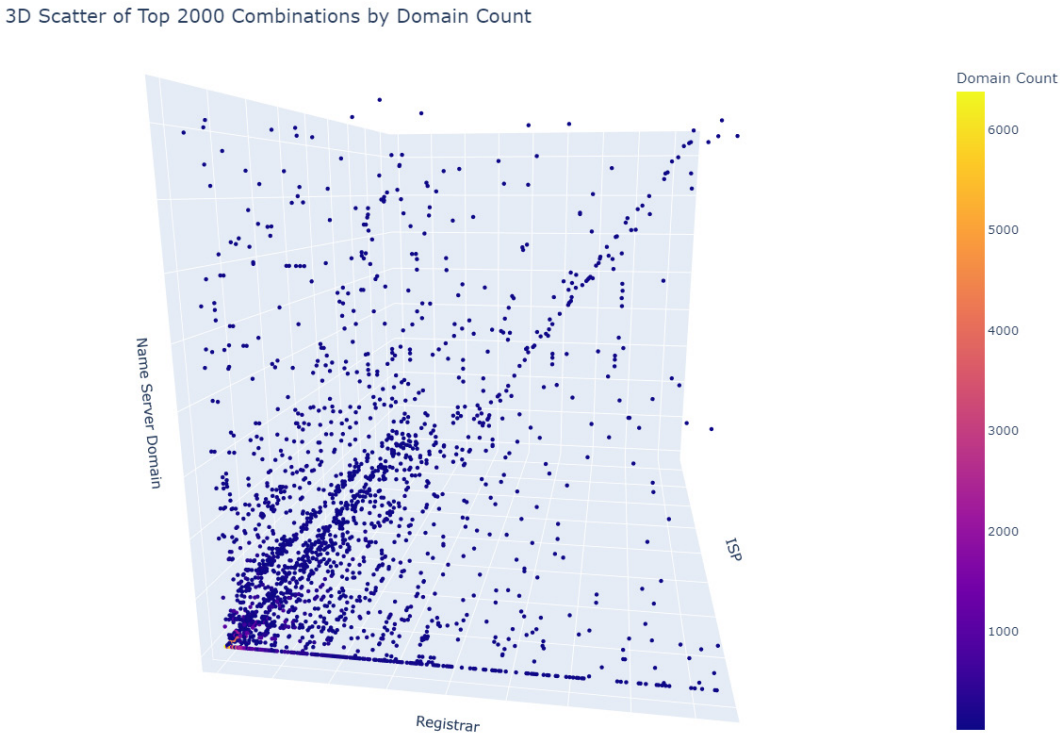
To assess for convergence in service providers within the large sampling of malicious domains, one approach taken was to see if the service providers that are individually the most popular in their own category (the Top 20 for registrars, the Top 20 for ISPs, and the Top 20 for name server domains) also tend to be the ones that are frequently used together. The 3D scatterplot (below) visualizes the three categories each in its own axis. Each point within the 3D space represents a specific combination of one registrar, one ISP, and one name server domain. The point's color tells us how often this specific combination appears in the data. Bright yellow means the combination is frequent (high volume), and dark blue means it appears rarely (low volume). If there is convergence, we would expect to see a lot of yellow points clustered in a certain area of the graph where these "top" providers intersect.



This 3D scatter plot (above) shows that despite observing a tendency for malicious domains to favor specific Name Server Domains, ISPs, and Registrars, these top providers do not consistently appear together in combined analyses, suggesting a lack of strong inter-provider clustering. In other words, the "top choices" in each individual category don't seem to be consistently chosen together when it comes to the higher volumes observed.

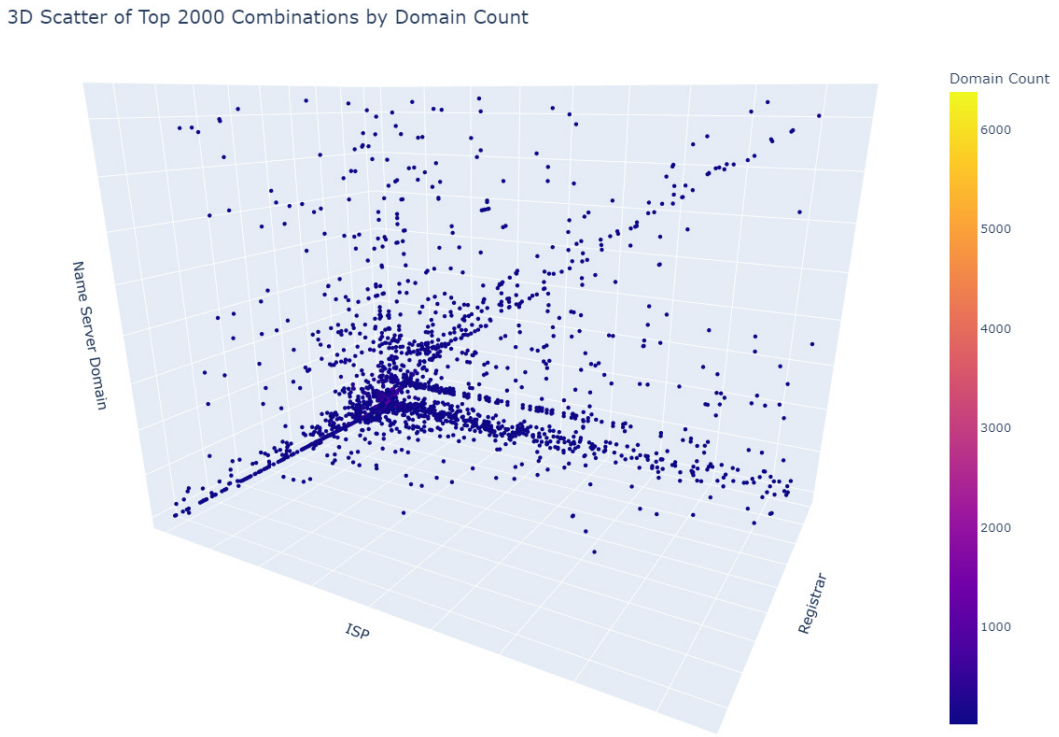
Another assessment approach investigated the most frequent combinations of Registrars, Internet Service Providers (ISPs), and Name Server Domains observed in the large sampling of malicious domain data. This was achieved by identifying the Top 1000 unique combinations of these three fields based on their co-occurrence in our dataset.

Perspective 1:



The results of this analysis are depicted in the 3D scatter plot (above) and a different perspective of the same data (below). In both diagrams, each point represents one of the Top 1000 most frequent combinations. The position of the point is determined by the specific Registrar, ISP, and Name Server Domain in the combination. The color of the point, as in the previous plot, indicates the volume of occurrences for that specific combination within the Top 1000.

Perspective 2:



The 3D scatterplots show two key observations:

1. A concentration of domains on a few service providers
2. A significant spread indicating diverse provider combinations.

The concentration suggests dominant provider groupings, while the spread highlights the need to address domains with unique configurations. Both patterns inform risk assessment and proactive threat detection. This also suggests that while there is a tendency for some providers to be favored (as seen by the convergence), the landscape is not entirely dominated by a small number of rigid combinations. The variety of other frequently occurring combinations indicates a need for flexible detection strategies that can identify threats across a diverse infrastructure.

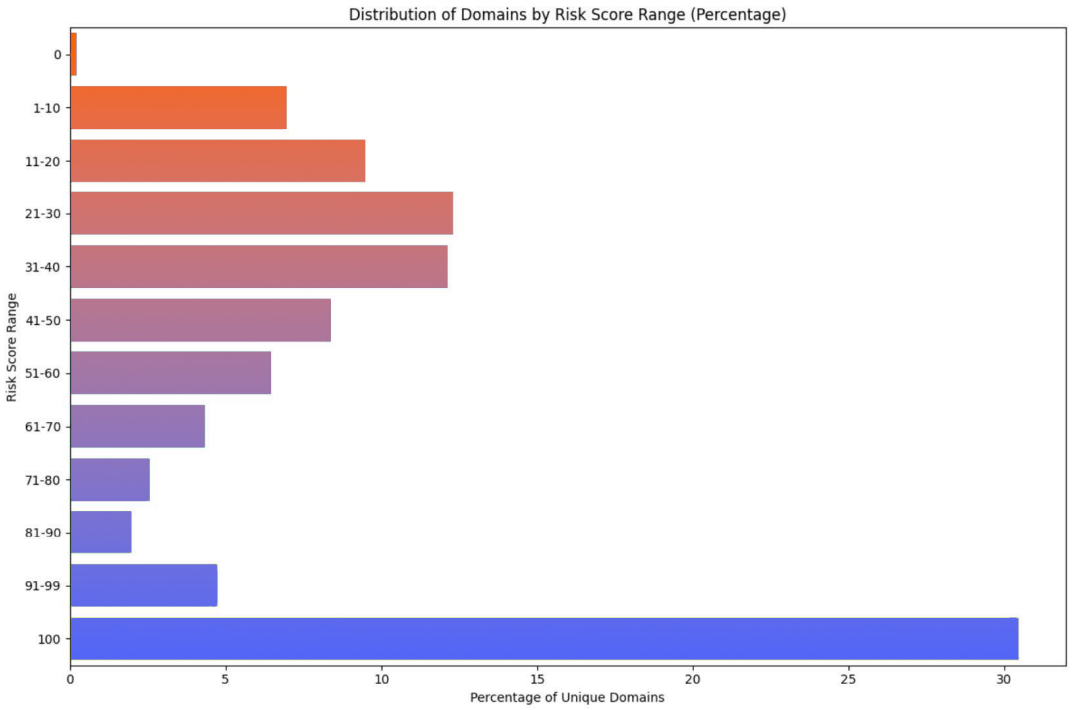
The significant spread observed in the 3D scatter plot, beyond the concentration of dominant provider groupings, also suggests another key inference. It may be that malicious domains frequently utilize relatively less common combinations of ISPs, Registrars, and Name Server Domains.

If true, these less frequent, yet still observable combinations within our malicious domain sample set could serve as valuable “pivot points” for further analysis. Identifying these unique configurations could allow security analysts to proactively search for other potentially malicious domains exhibiting the same unusual combination of providers. This approach could be particularly impactful in uncovering coordinated campaigns or identifying emerging threat patterns. Furthermore, expanding on the analysis by including other relevant data points such as SSL Issuer, server type and email domains could yield even more valuable insights.

Ranking Risk Scores Across Registrars and ISPs

Domain risk scores offer a numerical estimate of a domain's potential threat. These scores are calculated by examining factors like spam keywords, domain length, DGA detection, brand likeness, and proximity to known malicious domains. Risk scores aid security teams in sifting through the noise to identify potential threats and inform security mitigations.

Distribution of domains by risk score range



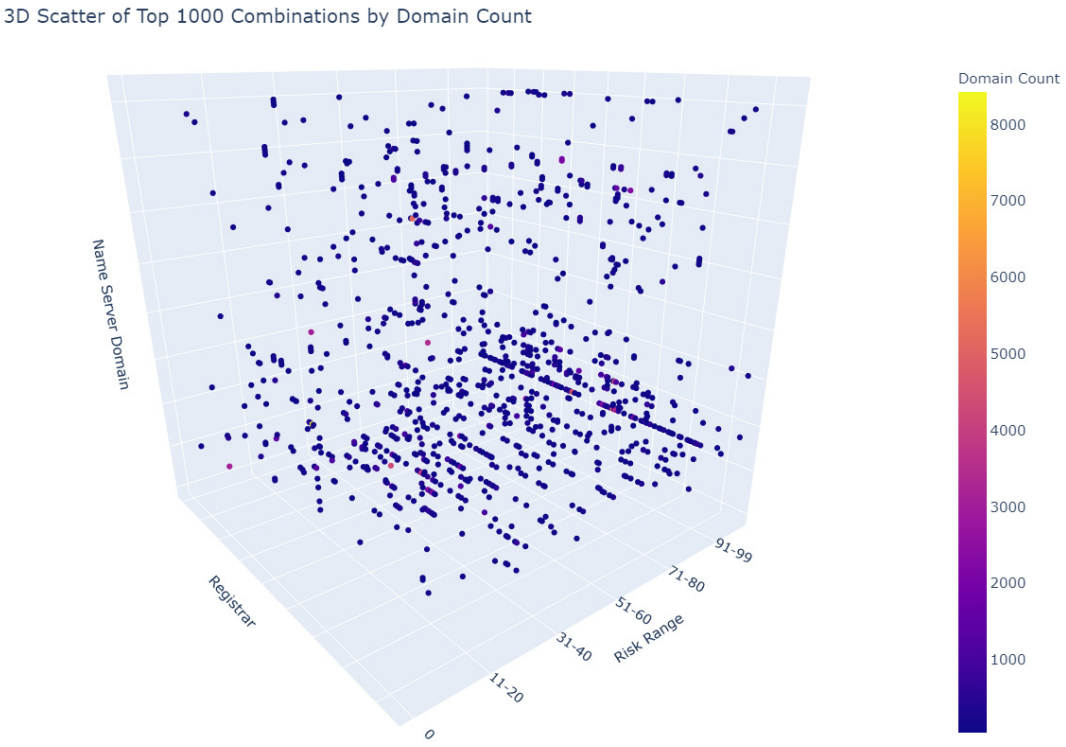
over

30%

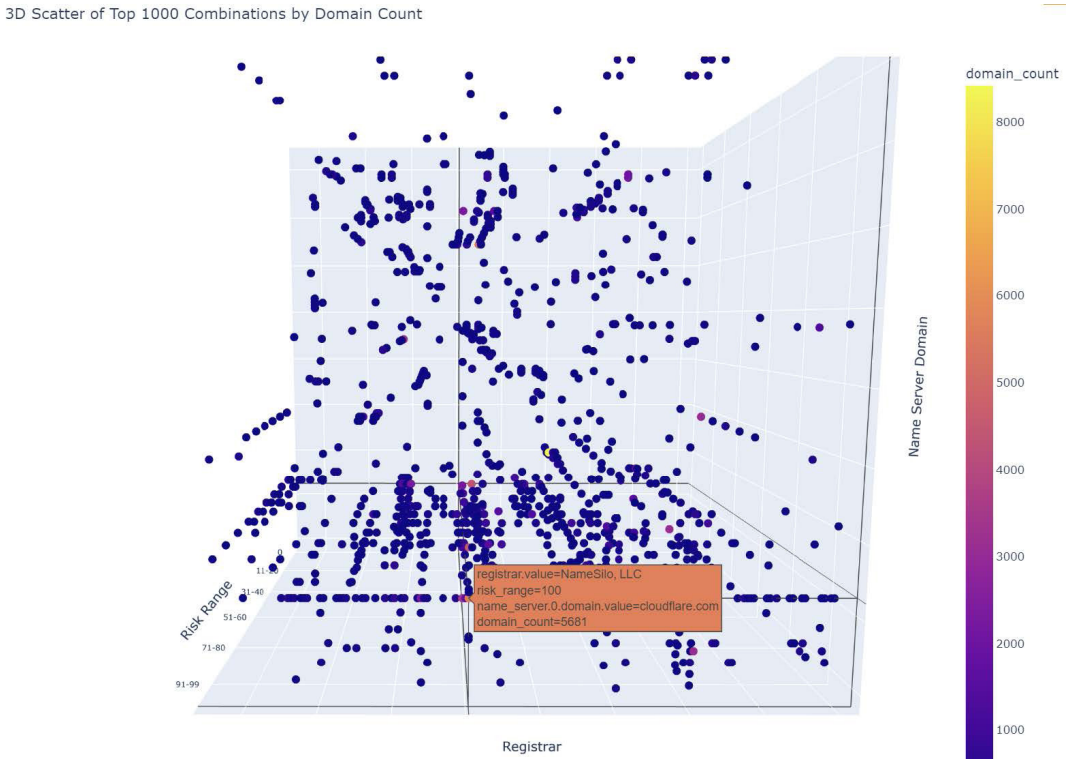
of malicious domains had a risk score of **100**

The chart above sets risk score in ranges and shows the percentage of malicious domains in each range. For instance, over 30% of malicious domains had a risk score of 100, the max score, followed by a broad distribution of malicious domains across all ranges except 0.

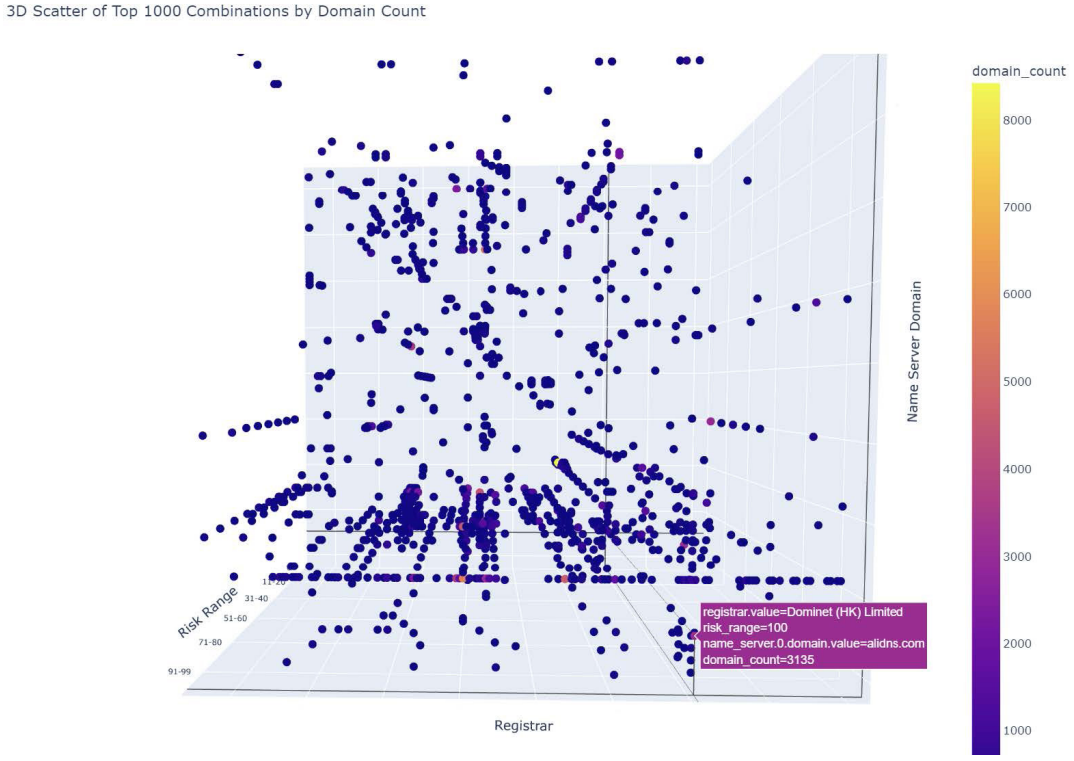
To further understand the infrastructure associated with higher risk levels, we analyzed the relationship between domain risk scores and the registrars and name server domains utilized. In the following 3D scatterplot graph, each point represents a specific combination of a risk score range, a registrar, and a name server domain. The aim is to pinpoint if certain combinations of registrars and name server domains tend to be associated with a higher volume of high-risk domains.



The aggregated data presented in the 3D scatter plot (above) and the two plots (below) visually pinpoints concentrations of high-risk domains within specific registrar-nameserver combinations, particularly within the 91-100 risk score range. While malicious domains are observed across a broad distribution of providers, distinct 'hot spots' emerge where particular pairings of registrars and name server domains exhibit a higher volume of these high-risk domains. This suggests that certain combinations of these infrastructure providers may be more frequently associated with malicious activity.



The graph (above) highlights one of the points in which the registrar “NameSilo LLC” and Name Server “cloudflare[.]com” have malicious domains appearing in relatively high frequency together with a risk score of 100.



The graph (above) highlights one of the points in which the registrar “Dominet (HK) Limited” and Name Server “alidns[.]com” have malicious domains appear in relatively high frequency together with a risk score of 100.

Using Shannon Entropy to Identify DGA-Created Domains

Shannon entropy is a valuable method for identifying DGA-created domains. By computing a value score for each letter in the second level domain name and dividing it by its total length, it establishes an entropy score that when compared with normal ranges can assist in the identification of abnormal naming patterns such as unreadable domain names - domains that do not contain words.

Domains with Entropy Significantly Below the Standard Deviation (Low Entropy Outliers):

- Likely Simpler and More Predictable: These domains tend to have less character variation and potentially more repetition.
- In some security contexts, extremely low entropy domains could be associated with automatically generated domains that are designed to be simple and disposable, but this is not always the case.

Domains with Entropy Significantly Above the Standard Deviation (High Entropy Outliers)

Likely More Complex and Random-Looking

These domains have a wider variety of characters and less predictable patterns.

Algorithmically Generated Domains (DGAs)

Domains generated by DGA's, often used in malware or botnet communication to evade blocking. These domains intentionally have high entropy to make them harder to predict and block.

Suspicious by Design

High entropy domains are often flagged as more suspicious, especially in security contexts, as they can be indicative of DGAs or other automated domain generation techniques.

Context is Key: Entropy alone is not a definitive indicator of whether a domain is malicious or benign; It's just one feature. Other factors should be considered such as domain registration details, domain age, website content, and DNS records.

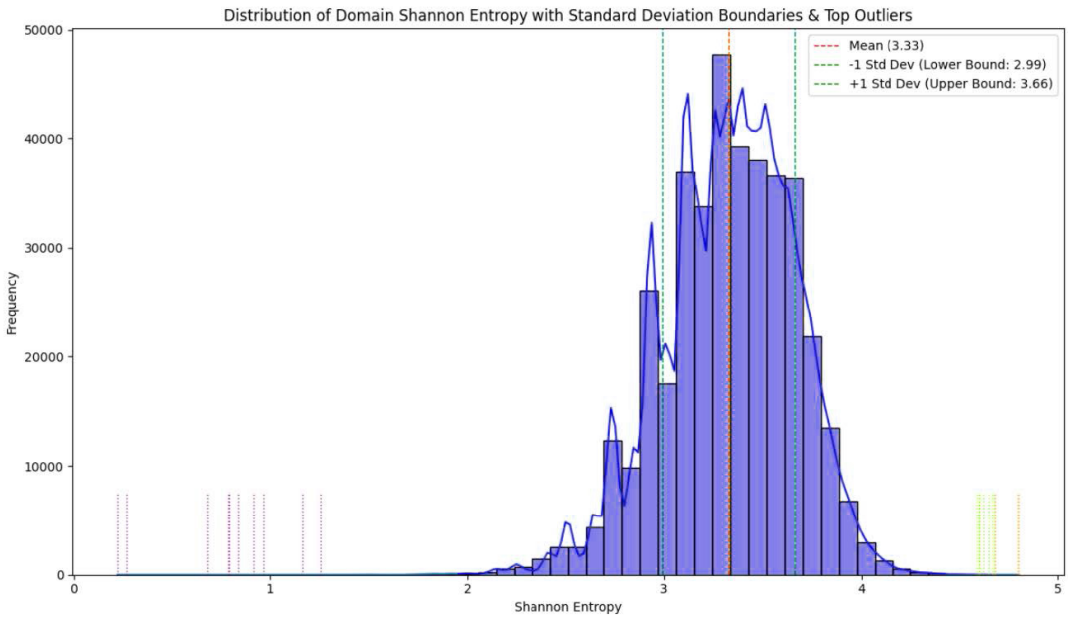
Threat Indicator Domain Entropy Scores For 2024:

The height of each bar chart below shows how common domain names with a particular entropy score are distributed in our data. The peak in the middle represents the most common level of randomness in domain names. The mean line shows the average entropy score across all the domain names in our dataset. It gives us a central point of reference for the typical level of randomness. Two green vertical lines show one standard deviation from the average in both directions. This tells us the range within which most domain names' entropy scores fall. Domains within these lines are considered to have a fairly typical level of randomness. On each side, 10 small lines in purple (low entropy side) and yellow (high entropy side) show a sampling of the lowest and highest entropy values within the set. On the low entropy side, these are domain names much more predictable or structured than the average. On the high entropy side, these are domain names much more random-looking than the average.

Average Shannon Entropy: 3.3274

Standard Deviation of Shannon Entropy: 0.3337

Number of domains outside 1 standard deviation: 125,051



Example outliers:

Sample 10 Low Entropy Outlier Domains (Shown as purple vertical lines in graph)
Entropy: 0.2223, Length: 28, Domain: ooooooooooooooooooooooo[.]ooo
Entropy: 0.2668, Length: 22, Domain: ooooooooooooooooo[.]ooo
Entropy: 0.6822, Length: 39, Domain: aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa[.]com
Entropy: 0.7875, Length: 51, Domain: lllllllllllllllllllllllllllllllllllllllllll[.]online
Entropy: 0.7936, Length: 32, Domain: aaaaaaaaaaaaaaaaaaaaaaaaaaaa[.]com
Entropy: 0.8384, Length: 47, Domain: lllllllllllllllllllllllllllllllllllll[.]online
Entropy: 0.9142, Length: 18, Domain: xxxxxxxxxxxxxx[.]ru
Entropy: 0.9654, Length: 33, Domain: lllllllllllllllllllllllll[.]site
Entropy: 1.1635, Length: 19, Domain: 000000000000000[.]com
Entropy: 1.2577, Length: 17, Domain: 6666666666666[.]com

Sample 10 High Entropy Outlier Domains (Shown as orange vertical lines in graph)
Entropy: 4.8020, Length: 33, Domain: urytwegjsb0953kflqwdn1249aiai[.]com
Entropy: 4.7951, Length: 69, Domain: 5rtr5iwxukegeojg20vh4gohf6k6bd03nl0liegdh0flfhqpyrxefkuac2mjwyy[.]store
Entropy: 4.6808, Length: 33, Domain: buyabcdefghijklmnopqrstuvwxyz[.]com
Entropy: 4.6751, Length: 45, Domain: islamic-87olvp0i0x2cjvfnbjhr4eslgcqaqkb[.]cloud
Entropy: 4.6696, Length: 69, Domain: ztdiyhmsc82kgtrvjmleycib2ppe27wxsgvhejnnncqi1qp9t6vjrxchvphedm8[.]click
Entropy: 4.6482, Length: 46, Domain: bc1qx0anrq4v2aftl3eg22rfnyump7wxln2e7ld60a[.]com
Entropy: 4.6232, Length: 68, Domain: kraken19at-shop-catalog-64cdd504-a32b-5e2c-9a8f-c9a6730fc119vhr[.]pics
Entropy: 4.6048, Length: 46, Domain: bc1qp2we64k79237y0npqehprfgynlz02fwpktlwte[.]com
Entropy: 4.5984, Length: 63, Domain: com-lnsubdpkclwkdwgdsqs8k6knzkdabqh9lyo8ogbatgov8xtzayindex[.]com
Entropy: 4.5884, Length: 46, Domain: hlkrhl56u7erjy468algbdx0346tsjkbdgkjasaiai[.]com

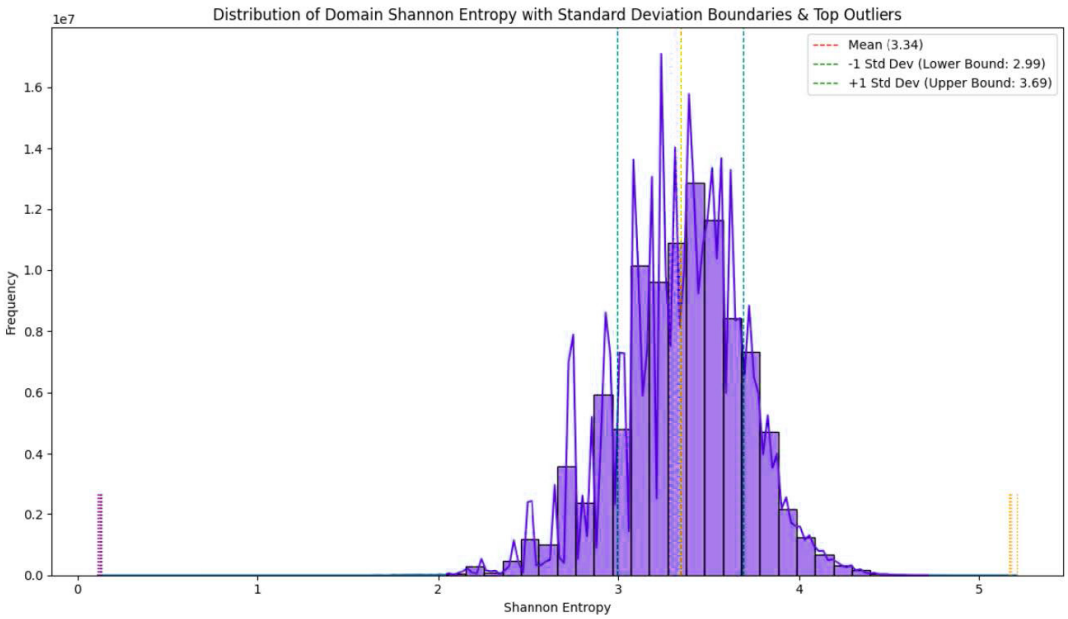
All 106 Million Newly Observed Domain Entropy Scores For 2024:

Average Shannon Entropy: 3.3436

Standard Deviation of Shannon Entropy: 0.3498

Number of domains below 1 standard deviation (low entropy outliers): 16,342,140

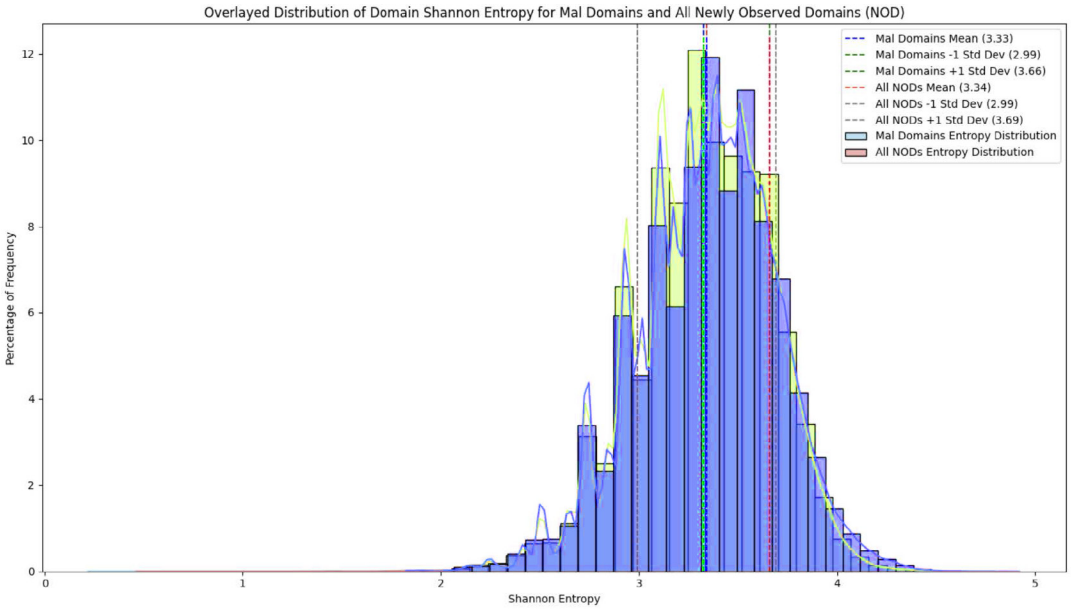
Number of domains above 1 standard deviation (high entropy outliers): 15,498,273



Sample 10 Low Entropy Outlier Domains (Shown as purple vertical lines in graph)
Entropy: 0.1133, Length: 66, Domain: ggggggggggggggggggggggggggggggggggggggggggggggggggggggggggg[.]gg
Entropy: 0.1133, Length: 66, Domain: sssssssssssssssssssssssssssssssssssssssssssssssssssssssssss[.]ss
Entropy: 0.1161, Length: 64, Domain: ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc[.]cc
Entropy: 0.1239, Length: 59, Domain: ccccccccccccccccccccccccccccccccccccccccccccccccccccc[.]cc
Entropy: 0.1257, Length: 58, Domain: cccccccccccccccccccccccccccccccccccccccccccccccccccc[.]cc
Entropy: 0.1274, Length: 57, Domain: ccccccccccccccccccccccccccccccccccccccccccccccccccc[.]cc
Entropy: 0.1292, Length: 56, Domain: cccccccccccccccccccccccccccccccccccccccccccccccccc[.]cc
Entropy: 0.1292, Length: 56, Domain: cccccccccccccccccccccccccccccccccccccccccccccccccc[.]cc
Entropy: 0.1311, Length: 55, Domain: ooooooooooooooooooooooooooooooooooooooooooooooooo[.]ooo
Entropy: 0.1330, Length: 54, Domain: ccccccccccccccccccccccccccccccccccccccccccccccccc[.]cc

Sample 10 High Entropy Outlier Domains (Shown as orange vertical lines in graph)
Entropy: 5.2112, Length: 41, Domain: abcdefghijklmnopqrstuvwxyz-1234567890[.]com
Entropy: 5.2112, Length: 41, Domain: 0123456789-abcdefghijklmnopqrstuvwxyz[.]top
Entropy: 5.1828, Length: 39, Domain: qwertyuiopasdfghjklzxcvbnm1234567890[.]cn
Entropy: 5.1719, Length: 40, Domain: 1234567890poiuytrewqasdfghjklmnbvcxz[.]com
Entropy: 5.1719, Length: 40, Domain: pjbxh4yt2mn5crkfvs81gzow9u0liad7q36e[.]art
Entropy: 5.1719, Length: 40, Domain: svf8rgcw2tjk17obx6niam0uhz543qpye9dl[.]art
Entropy: 5.1719, Length: 40, Domain: nev9ut7hw6iydalfxkz2pq1m0bcjg8ors453[.]art
Entropy: 5.1719, Length: 40, Domain: pncibavuswr8md90zjxtql746gof1ye5h3k2[.]art
Entropy: 5.1719, Length: 40, Domain: btpomvxfu7ry6qhk2leag89jd3z1isc540wn[.]art
Entropy: 5.1719, Length: 40, Domain: oix0wf3csbkudqna745yrmzthp81j6elvg29[.]art

Showing the side-by-side of the Shannon entropy scores for all newly observed domains from 2024 as distribution 1, and the malicious domain as distribution 2. By converting the Y axis to a ratio of their distribution size we can view them more relatively. Notably, while the mean entropy scores are close in proximity to each other at 3.34 and 3.33, the malicious domains in distribution 2 show a higher grouping of domains on the high outlier ranges suggesting a concentrated ratio of similar DGA domains in its distribution.



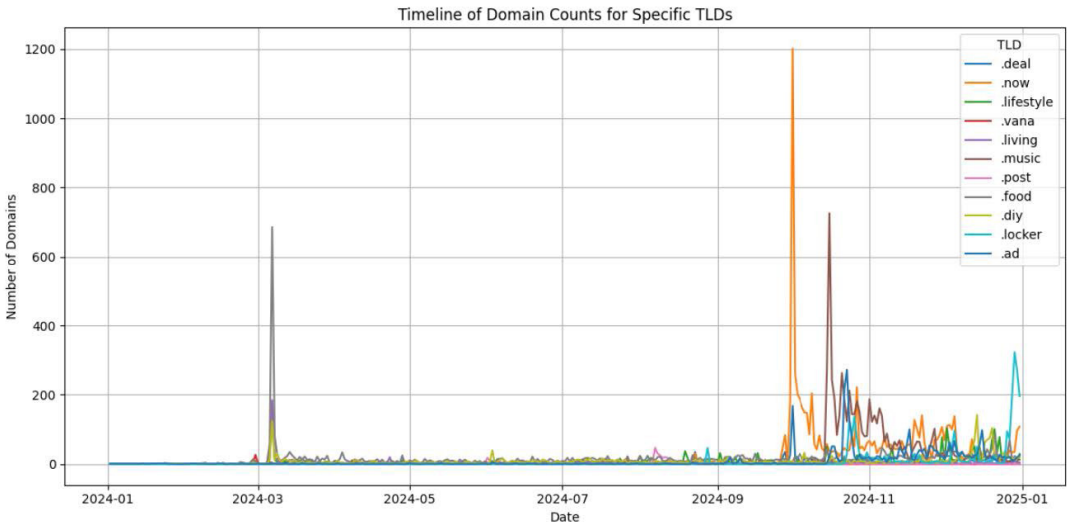
New Top Level Domains in 2024

New top level domains (TLDs) drive surges in domain registrations, as resellers, businesses, and individuals capitalize on the opportunity to acquire relevant domain names. Unfortunately, many security systems utilize static, hard-coded TLD lists, leaving them susceptible to missing potentially malicious domains registered under these new TLDs.

TLD General Availability Date and Total Domain Registrations with Resolving IP Addresses in 2024 After General Availability:

New TLD	General Availability Date	Observed Domains with TLD 2024
.lifestyle	March 6, 2024	2474
.vana	March 6, 2024	4
.living	March 6, 2024	1625
.music	March 6, 2024	6124
.post	March 6, 2024	129
.food	March 6, 2024	4961
.diy	March 6, 2024	2644
.locker	September 25, 2024	2355
.deal	September 30, 2024	460
.now	September 30, 2024	7035
.ad	October 16, 2024	2694
.tr	June 14, 2024	67,556

Newly Observed Domains Count with Newly Available TLDs Over 2024



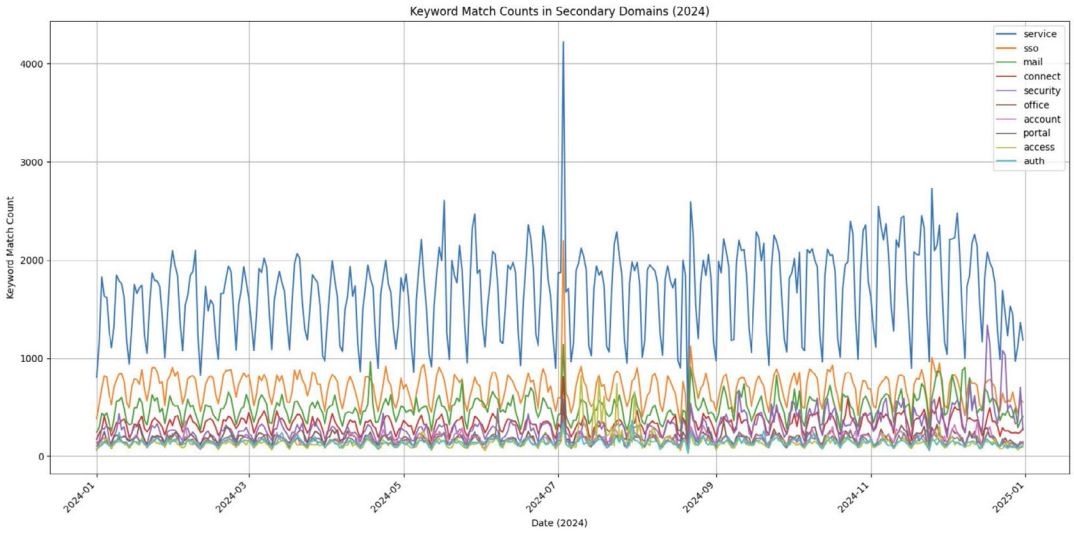
Common Credential Harvesting Domain Name Keywords

Domain names used for credential harvesting often contain specific keywords related to authentication processes. To facilitate the monitoring of these potentially malicious domains, we analyzed keyword frequency throughout 2024. The diagram below shows the top credential-related keywords observed, ranked by their occurrence in domain names.

Domain name contains

‘login’	‘account’	‘security’	‘confirm’
‘signin’	‘access’	‘update’	‘validate’
‘sso’	‘portal’	‘reset’	‘service’
‘mfa’	‘webmail’	‘password’	‘moncompte’
‘2fa’	‘mail’	‘auth’	‘facturacion’
‘verify’	‘secure’	‘authentication’	

Credential Harvesting Related Keywords – All Newly Observed Domains 2024 (100 million)



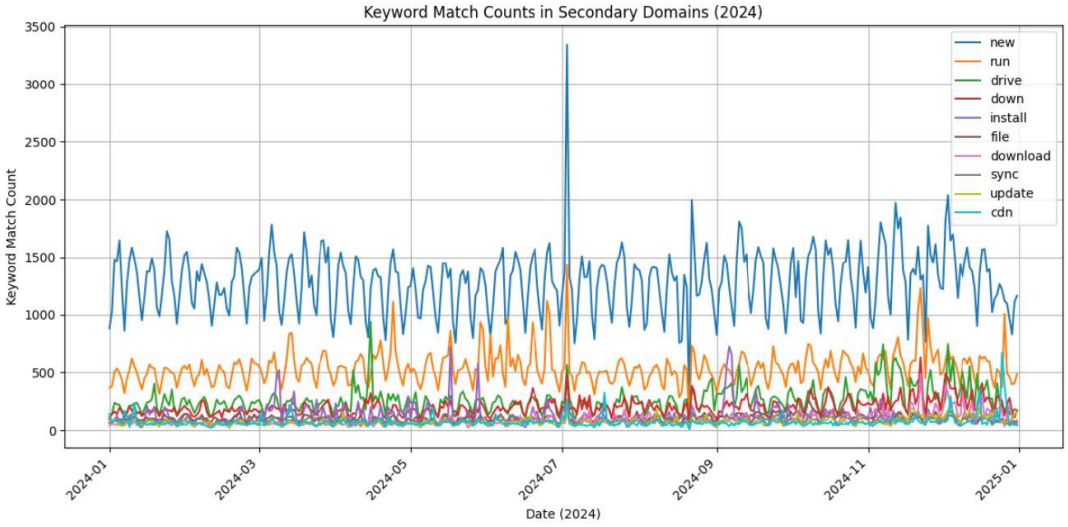
Common Malware Delivery Domain Name Keywords

Domain names used for malware delivery often contain specific keywords related to malicious files and downloads. To facilitate the monitoring of these potentially dangerous domains, we analyzed keyword frequency throughout 2024. The diagram below shows the top malware-related keywords observed, ranked by their occurrence in domain names.

Domain name contains

‘update’	‘latest’	‘run’	‘version’
‘verify’	‘down’	‘patch’	‘browser’
‘download’	‘cdn’	‘new’	‘java’
‘install’	‘sync’	‘critical’	‘protector’
‘file’	‘vpn’	‘urgent’	‘antivirus’
‘document’	‘flash’	‘alert’	‘drive’

Common Download Lure Related Keywords – All Newly Observed Domains 2024 (100 million)

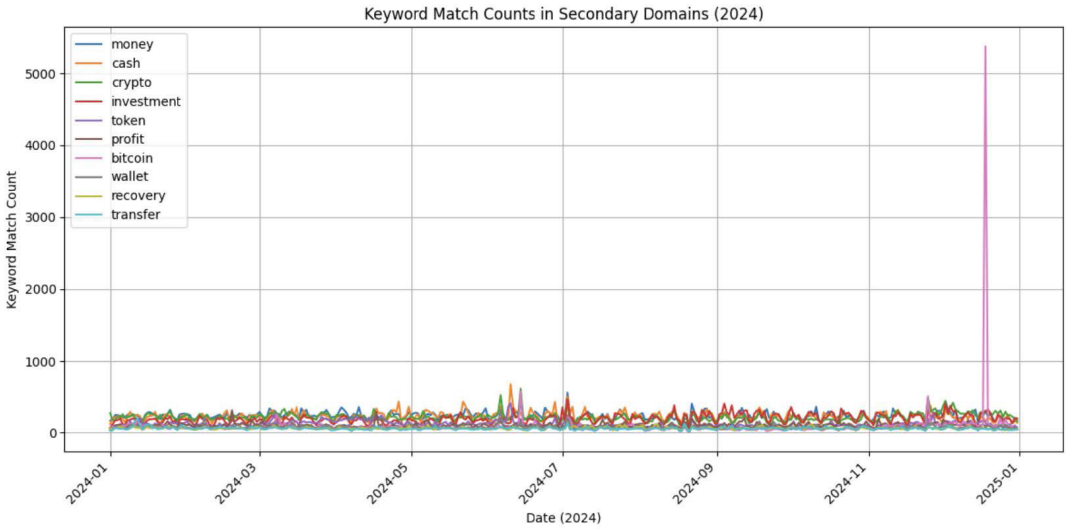


Scam, Fraud and Financial Theft Related Domain Name Keywords

Domain names used for scams, fraud, and financial theft often contain specific keywords related to deceptive practices. To facilitate the monitoring of these potentially harmful domains, we analyzed keyword frequency throughout 2024. The diagram below shows the top 10 scam-related keywords observed, ranked by their occurrence in domain names.

Domain name contains

'phishing'	'duplicate'	'sweepstakes'	'guaranteed'	'recovery'
'fraud'	'airdrop'	'crypto'	'cash'	'unlock'
'scam'	'pre-sale'	'bitcoin'	'money'	'bypass'
'fake'	'virus'	'ethereum'	'funds'	'unblock'
'spoof'	'malware'	'investment'	'transfer'	'token'
'clone'	'lottery'	'profit'	'wallet'	



The growing threat of online scams and fraud, particularly within the cryptocurrency and online gambling/betting sectors, is often observed through significant domain registration spikes. Keyword analysis of these spikes reveals patterns associated with malicious activity. In June 2024, DomainTools Investigations reported on a notable surge of domains containing the keyword 'AirDrop,' demonstrating the direct link between domain registration patterns and potential fraudulent activities.

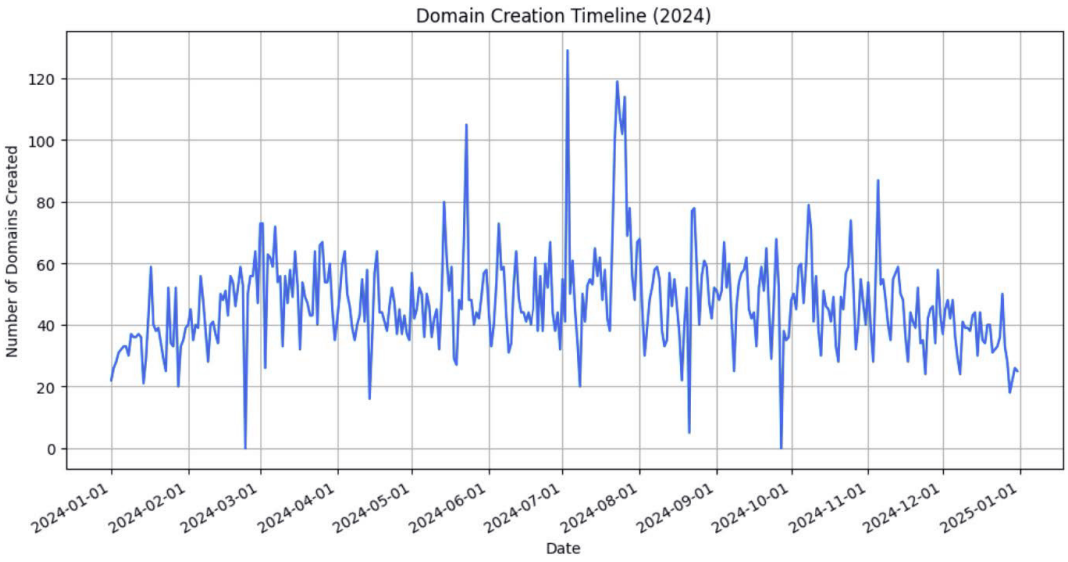
High Publicity Events

Exploiting high-profile events, opportunistic threat actors often create deceptive look-alike domains and websites to target unsuspecting users. To identify these threats, we analyzed 180 major 2024 events across various categories: Political and Elections, Technological Advancements, Natural Disasters, Social Movements, Popular Culture, Global Conflicts, and other significant international events. Keyword lists were generated for each category, and Levenshtein distance methods were used to detect domain names with close likenesses. While not exhaustive, this analysis provides an overview of newly observed domains potentially related to these highly publicized media events.

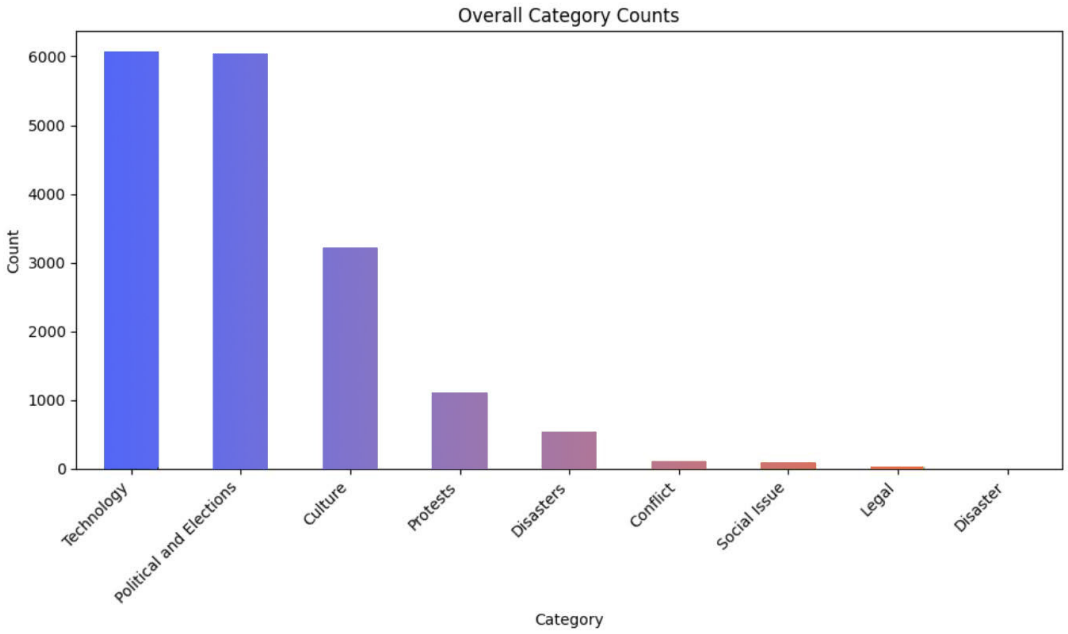
High Publicity Event Categories

- Political and Elections
- Technological Advancements and Disruptions
- Environmental and Natural Disasters Crises
- Popular Culture and Entertainment
- Significant Global Conflicts and International Relations
- Social Movements and Protests
- Other Significant International Events

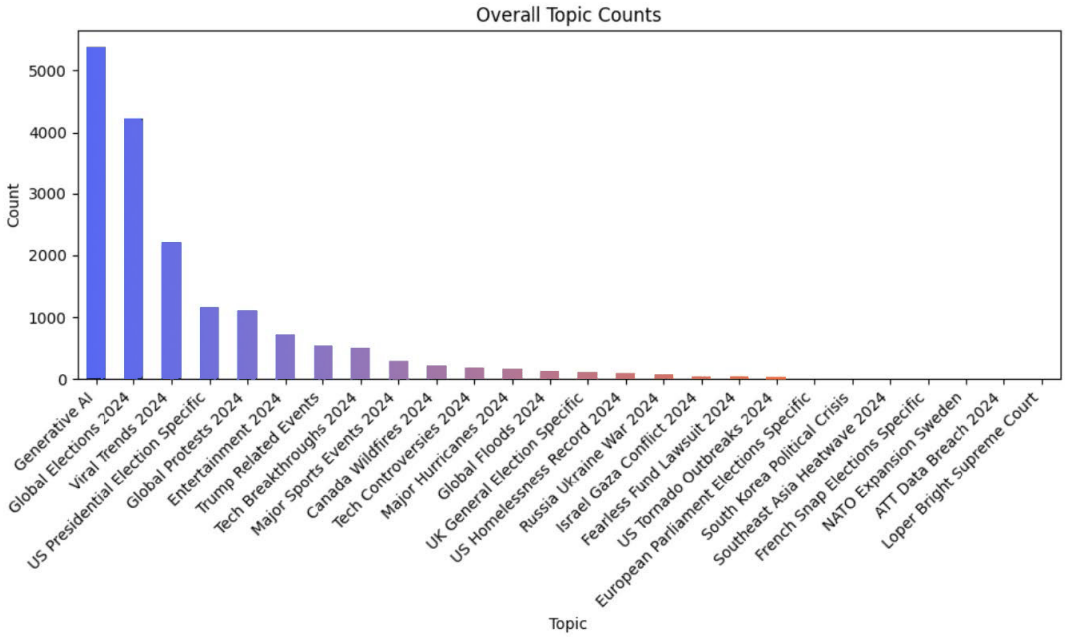
Overall counts of newly observed domain names containing keywords that are potentially related to highly publicized media events.



Overall category counts of newly observed domain names containing keywords that are potentially related to highly publicized media events.

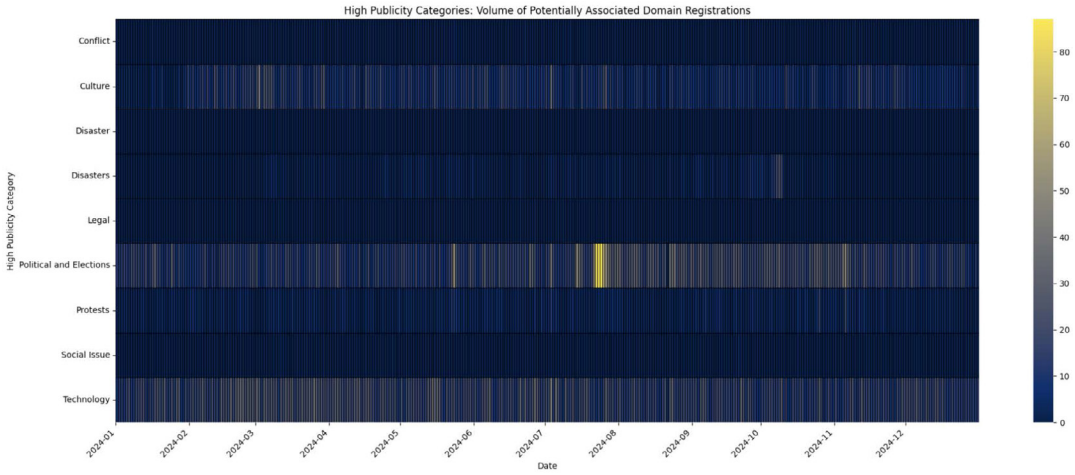


Overall topic counts of newly observed domain names containing keywords that are potentially related to highly publicized media events.



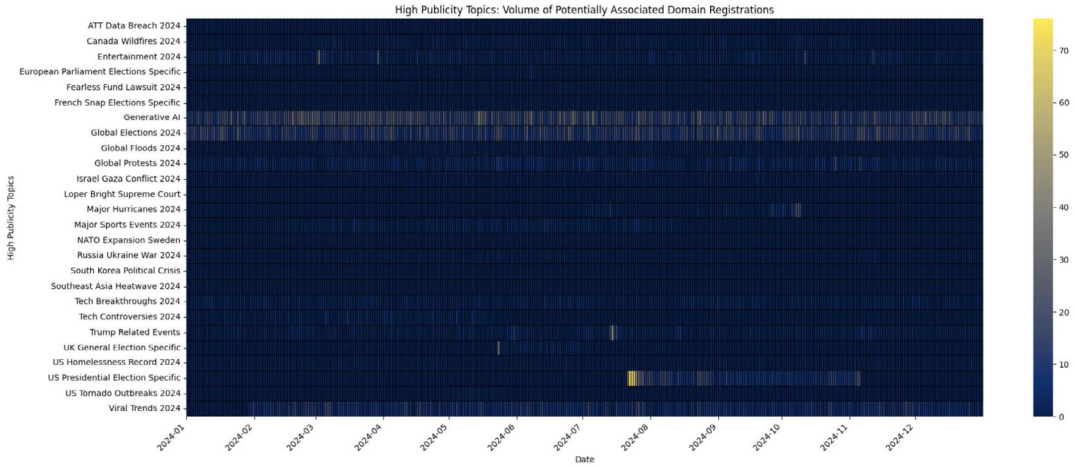
The following heatmaps show the higher concentrations of domains being registered that may appear similarly to keywords associated with high profile media events. Dark blue indicates a low concentration of domains in that given time period and category, yellow indicates a high concentration of domains that may appear related to that category. Using heatmaps in this way can help visually identify the concentrations or patterns in domain associations to certain media events. For instance, technology and specifically AI-related domain names were pervasive throughout 2024.

Heatmap of the Volume of Domains Associated with High Publicity Categories Over 2024



Likewise, politics and election topics were also prevalent, but there was likely a high concentration of newly observed domains in August 2024 associated with US Presidential Elections.

Heatmap of Domains Associated with Topics Over 2024

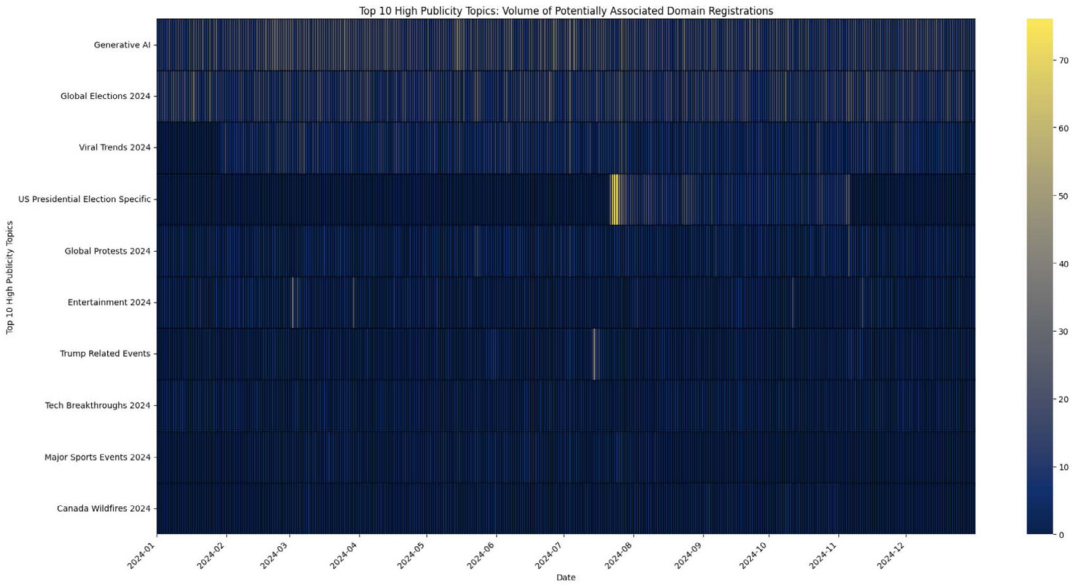


The heatmap (above) visually represents individual topics within the high-publicity categories, with high-frequency events highlighted in yellow, revealing several notable trends. Throughout 2024, a consistently high volume of newly observed domains related to both Generative AI and Global Elections was evident.

Furthermore, the data shows a significant concentration of new domains specifically concerning the US Elections, accompanied by a trailing pattern of such domains that persisted until shortly after the US Election dates in November. This temporal pattern strongly suggests a coordinated and sustained effort to activate a large number of domains daily over several months.

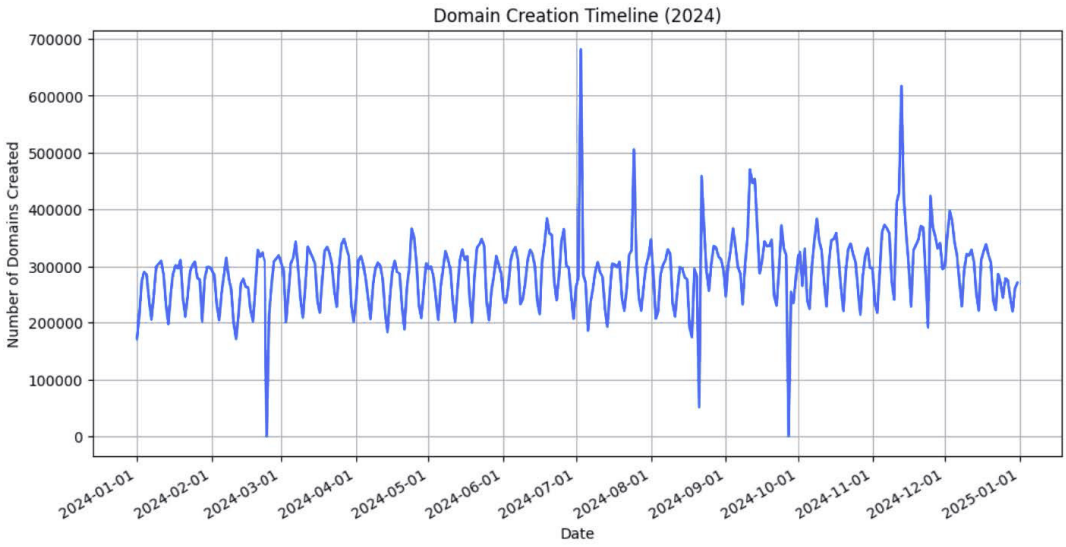
Overall, the prevalence of newly observed domains throughout the year indicates a primary focus on topics related to artificial intelligence and elections.

Heatmap of the Top 10 Topic by Overall Volume of Associated Domains



Anomalous Events Impacting Newly Observed Domains

There are notably two large spikes in new domains in all newly observed domains. One in July 2024 and one in November 2024, just after the US elections.



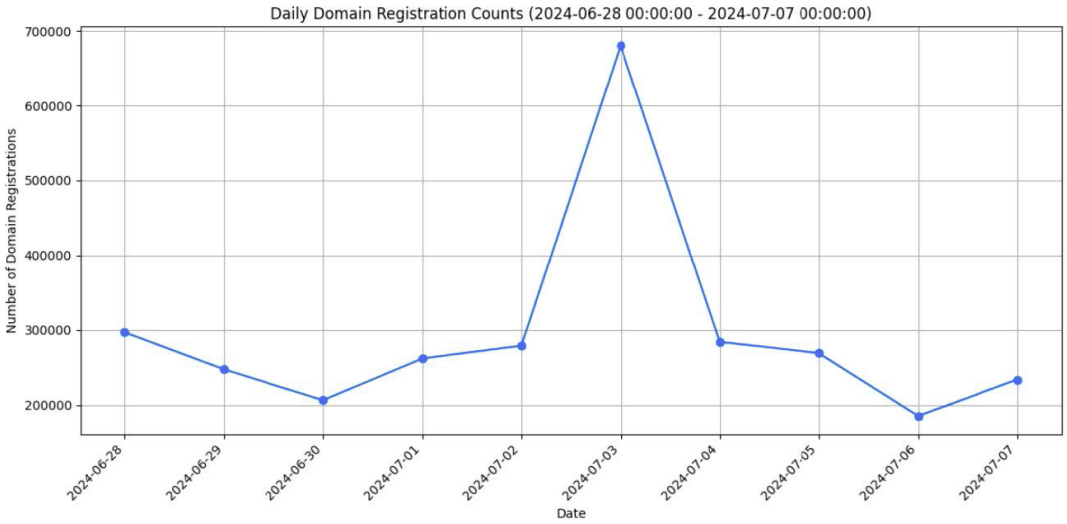
By monitoring for domain registration spikes and applying various analytical techniques, security researchers can investigate and pinpoint anomalous domains. We demonstrate some example methodologies below.

Newly observed domain spikes were defined as days with new registrations 1.5 times higher than the rolling daily average and identifies the following dates:

Investigating the Largest Spike on 2024-07-03:

Identified a spike in domain registrations on 2024-07-03

Total Number of newly observed domains on the day of the spike: **681,099**

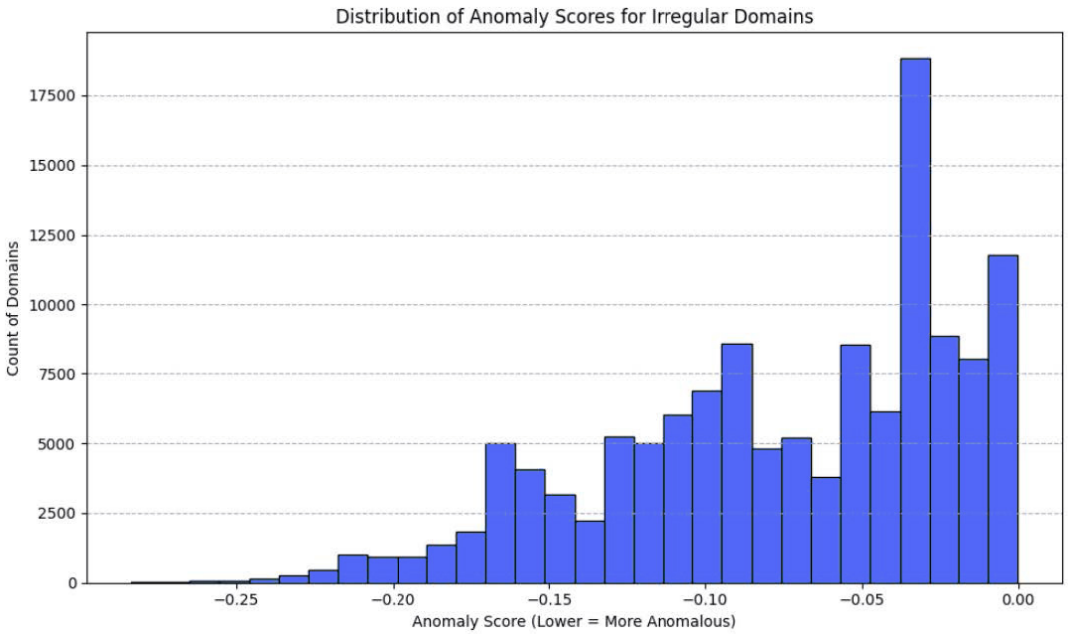


We applied a variety of methods to understand the domain names registered in the spike to attempt to isolate on the domains that are likely contributing to the spike versus likely normal domain registrations.

We used anomaly detection techniques with Isolation Forest to detect irregular domain registrations during spike periods, analyzing domain features like length, structure, and character composition. It analyzes non-spike days to set a baseline then compares domains in spikes to look for irregularities using anomaly scores. The resulting anomaly scores, displayed in the histogram below, revealed the extent to which these domains differed from typical registrations.

This process helps researchers understand if spikes are driven by anomalous domain types, which can indicate potential threats, enabling security researchers to effectively target their investigations and mitigate risks.

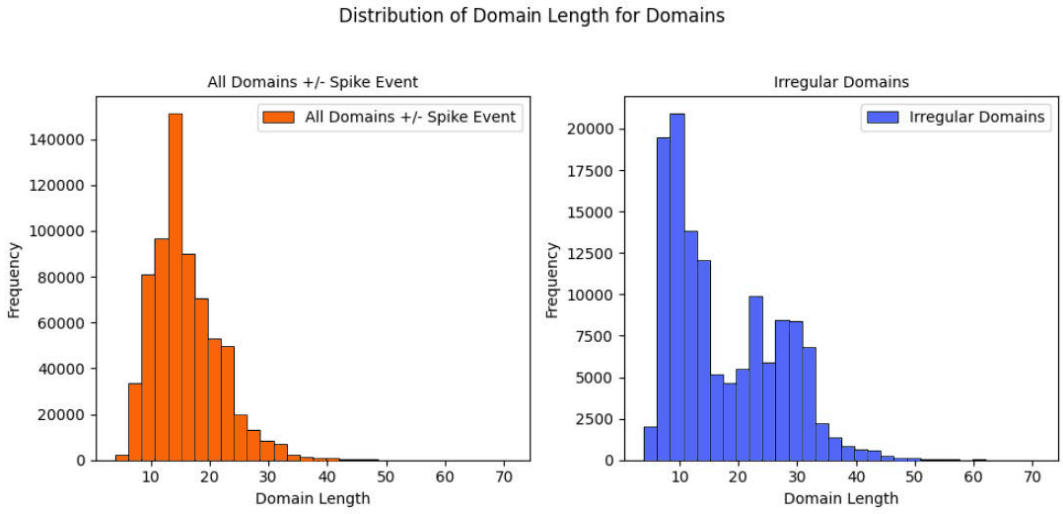
In total, the number of irregular domains identified on the July 3rd spike was 129,154 or approximately 19% of the total newly observed domains for that day.

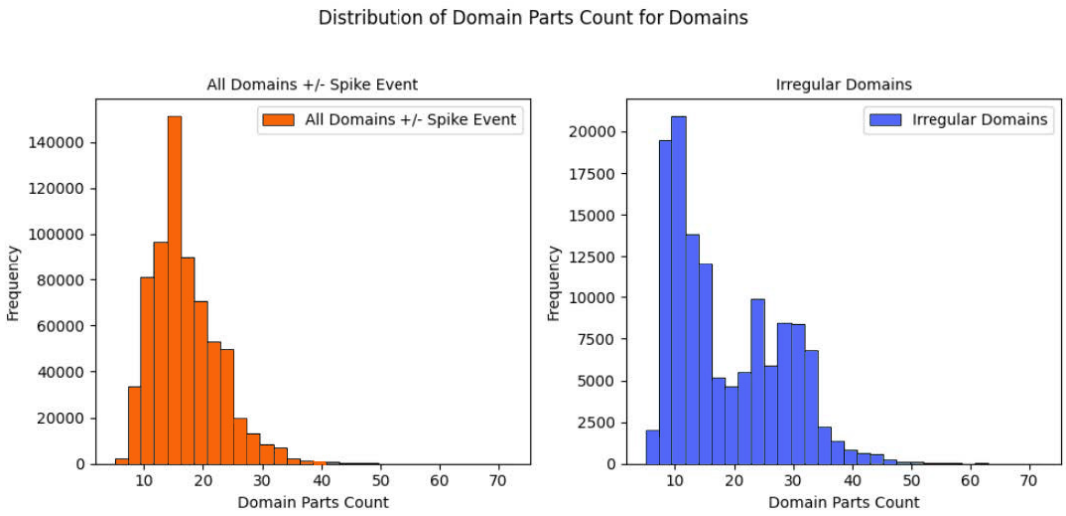


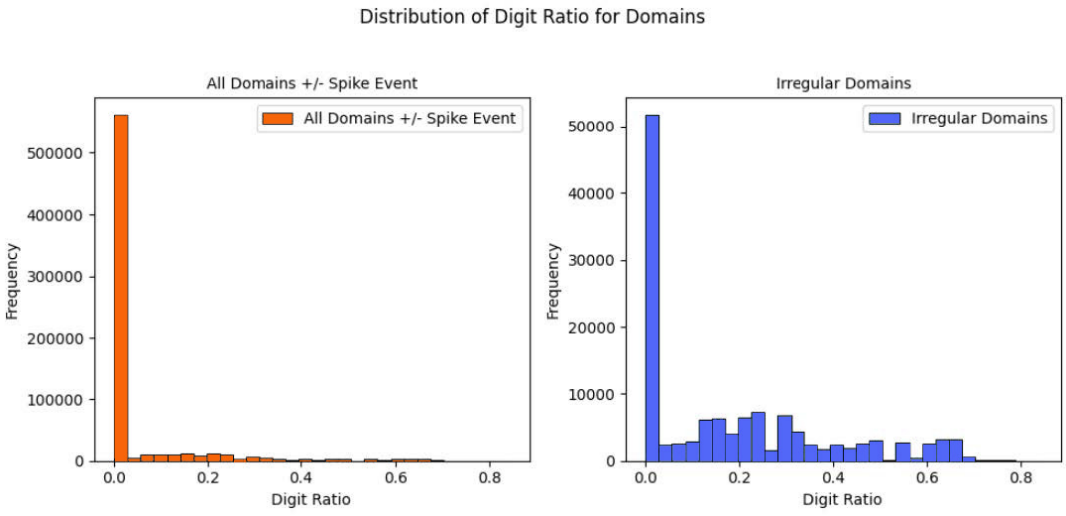
The diagram illustrates the spread and concentration of anomaly scores for the irregular domains. Domains further to the left (lower anomaly scores) indicate that they are quite significantly different from normal registrations and are therefore unusual. Domains further to the right (higher anomaly scores) indicate they are less different to normal domains.

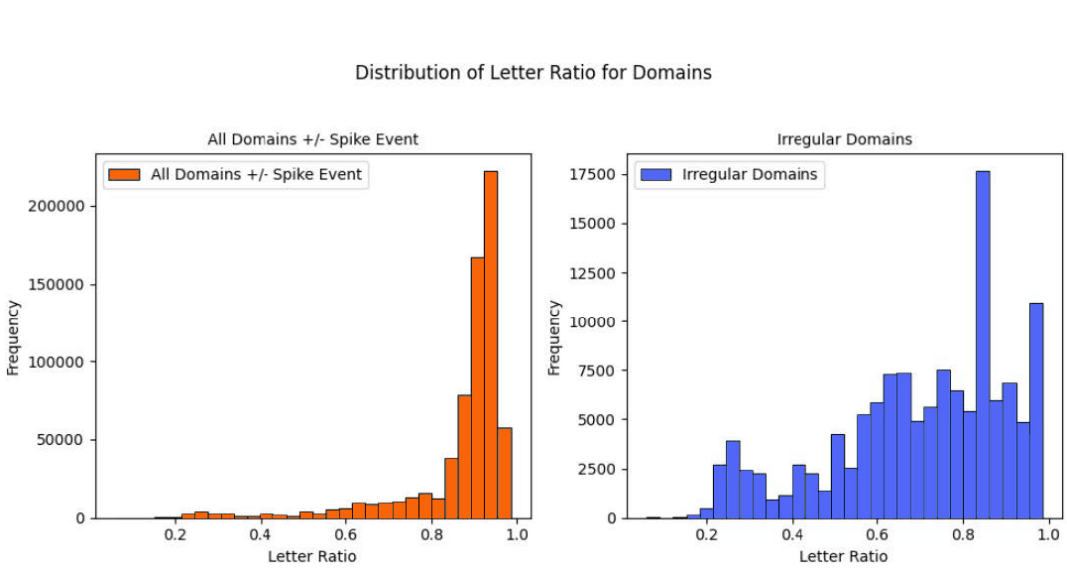
In this case, we see for the July 3rd volume spike, the majority of the unusual domains are a higher volume of medium to slightly off-pattern from normal domain registrations.

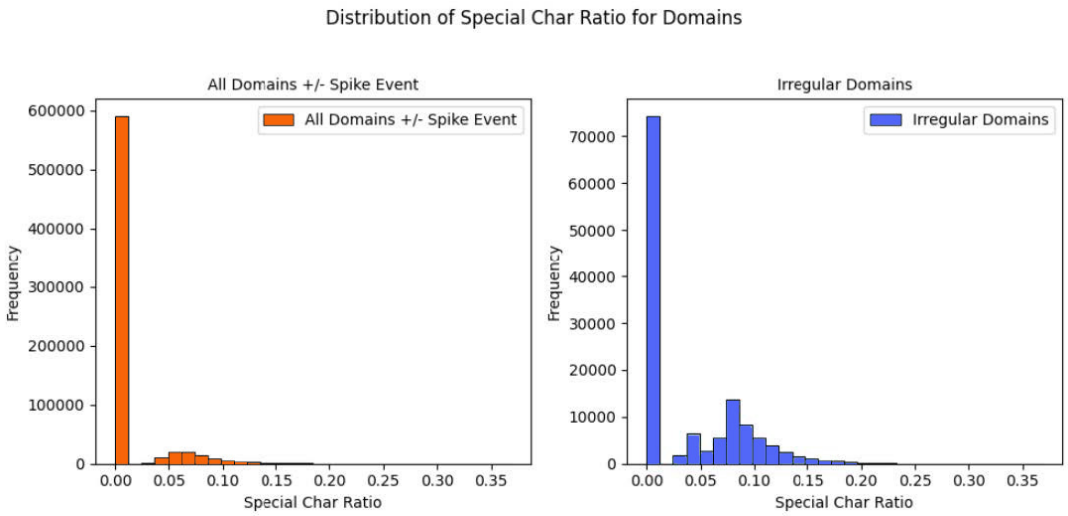
For comparison, the following diagrams show elements of the normal domain patterns a few days before and after the spike on the left and those of the irregular domains during the spike event on the right.




Distribution of Domain Parts Count for Domains
All Domains +/- Spike Event
All Domains +/- Spike Event
Frequency
Domain Parts Count
Irregular Domains
Irregular Domains
Frequency
Domain Parts Count


Distribution of Digit Ratio for Domains
All Domains +/- Spike Event
All Domains +/- Spike Event
Frequency
Digit Ratio
Irregular Domains
Irregular Domains
Frequency
Digit Ratio


Distribution of Letter Ratio for Domains
All Domains +/- Spike Event
All Domains +/- Spike Event
Frequency
Letter Ratio
Irregular Domains
Irregular Domains
Frequency
Letter Ratio

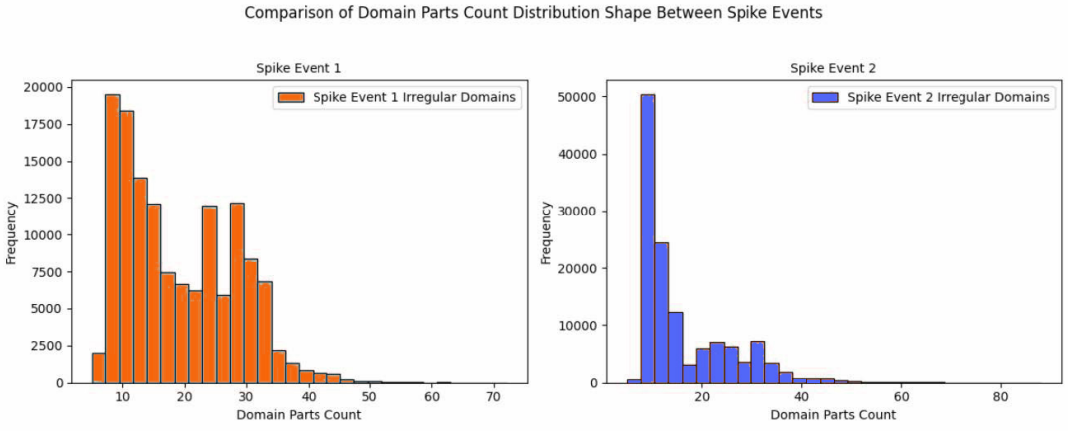


Searching For Patterns Across Multiple Spikes in Newly Observed Domains

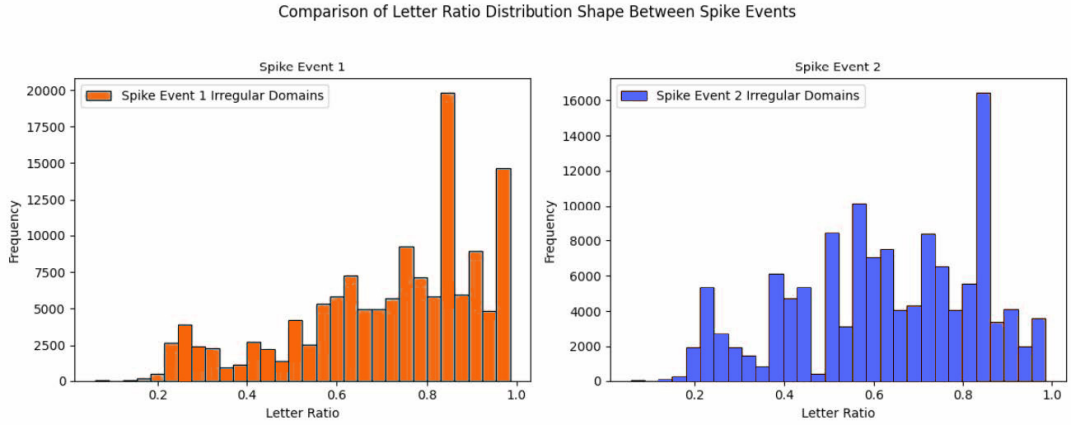
Tuning for the combination of irregularities may enable further efforts to identify similar patterns prior to and proceeding the spike events. For instance, the following diagram compares elements of irregular domains isolated from two different spike events to assess for statistical similarities.

In this diagram, the Kolmogorov-Smirnov (KS) method is used to produce a statistical measure of the distance between the two distributions for each element of comparison. The resulting KS statistic value essentially quantifies the difference. A smaller KS statistic value suggests they are more similar, a larger KS statistic value suggests they are more different. Notably, this approach can be highly sensitive to differences in volume between the two distributions so additional measures were taken to de-emphasize the weight of volume differences in order to focus on the similarity of the underlying patterns or shapes of the distributions.

Comparing Spike Event 1: July 3, 2024 Compared With Spike Event 2: November 13, 2024



Feature: Domain Parts Count KS Statistic (0.1968) suggests distributions may have practical differences (KS >= 0.1)

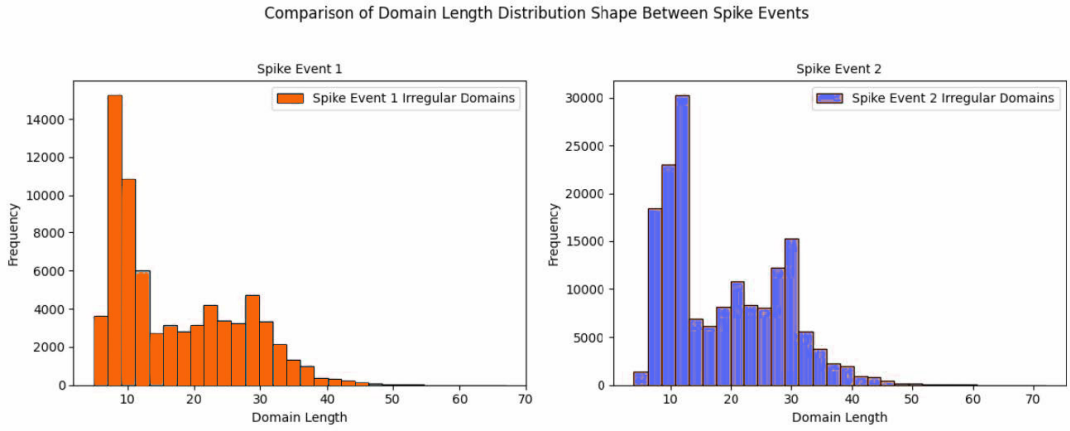


Feature: Letter Ratio KS Statistic (0.2058) suggests distributions may have practical differences (KS >= 0.1)

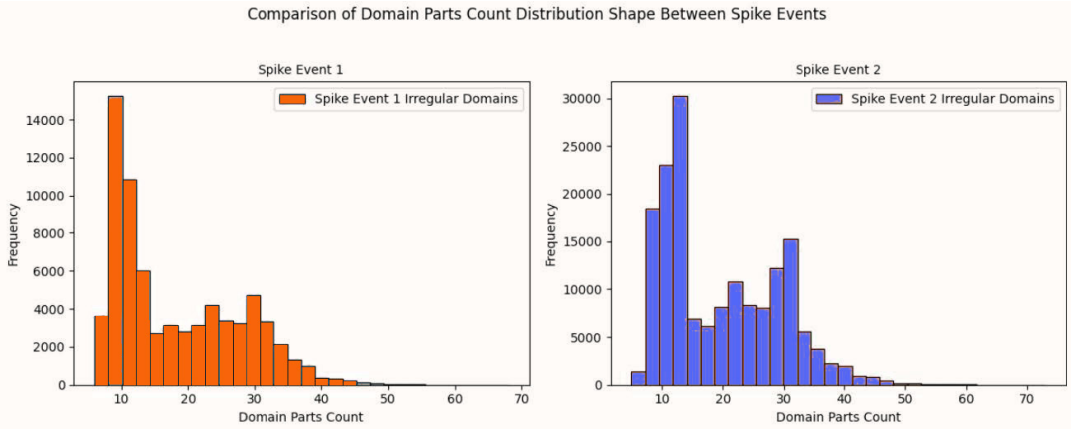
A summary of the statistical similarities between the spike on July 3, 2024 and November 13, 2024 suggests that they were practically different.

Although we present a comparison sampling of some of the elements above, it's crucial to understand that increasing the number of comparative elements improves the reliability and depth of distribution analysis.

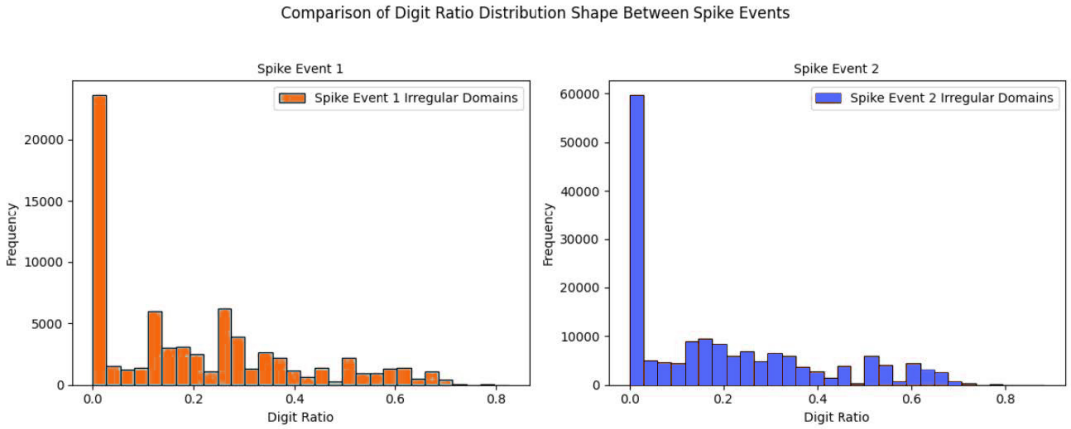
Comparing Spike Event 1: July 25, 2024 Compared With Spike Event 2: August 22, 2024



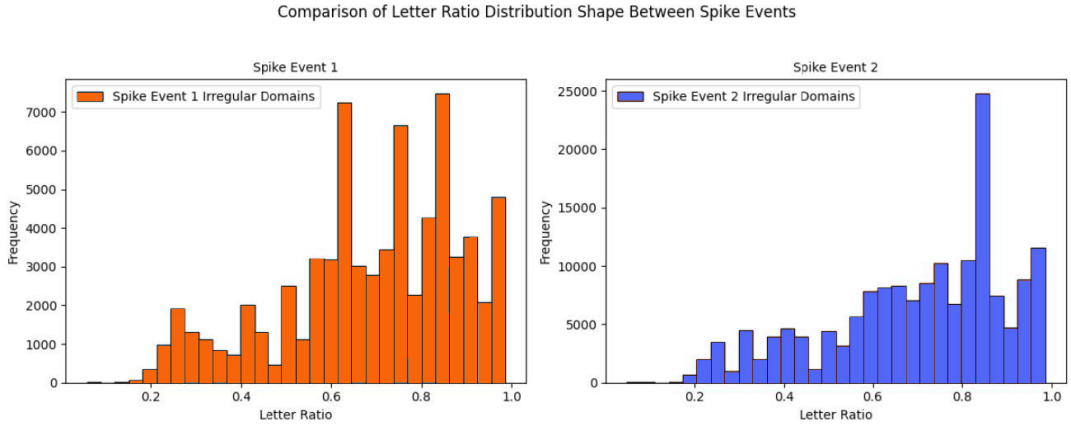
Feature: Domain Length KS Statistic (0.0950) suggests distributions are practically similar (KS < 0.1).



Feature: Domain Parts Count KS Statistic (0.0950) suggests distributions are practically similar (KS < 0.1)



Feature: Digit Ratio KS Statistic (0.0585) suggests distributions are practically similar (KS < 0.1).



Feature: Letter Ratio KS Statistic (0.0707) suggests distributions are practically similar (KS < 0.1).

Overall, the observed statistical similarities between the irregular domains in spike events on July 25, 2024 and August 22, 2024 could suggest for example that they were created from the same scripts, operated by the same group or part of a recurring campaign of activity deploying large volumes of inactive domains such as commonly exhibited by spam campaigns.

Domain Intelligence and the Cyber Landscape

This report offers a glimpse into the potential of domain intelligence for understanding and navigating the ever-evolving cyber landscape. We explored a variety of analytical methods—from examining domain attributes and risk scores to detecting DGAs and assessing keyword likeness—to illustrate how these techniques can reveal patterns and insights within domain registrations. Think of this as a survey of tools and approaches that security researchers, threat intelligence analysts, brand protection teams, and incident responders can employ.

By comparing a large dataset of newly observed domains with known malicious examples, we highlighted how concentrations of activity, keyword trends, and event-driven domain registrations can inform our understanding of threats. This data isn't just about identifying bad actors; it's about building a shared knowledge base that helps us collectively improve our defenses.

A strong security posture is built on community collaboration. Sharing insights, observed techniques, and lessons learned is essential. This report is intended to spark conversations and inspire further exploration of these methods. By working together, we can enhance our ability to identify risky domains and proactively mitigate threats.

Ultimately, the goal is to make the Internet safer for everyone. We encourage you to view this report as a starting point; a collection of ideas and methods that can be adapted and refined. By leveraging domain intelligence and fostering a spirit of collaboration, we can empower each other to better understand and address the challenges of the digital age.

Future Domain Intelligence Reporting

Domain intelligence is foundational to our mission and offerings. As the threat landscape evolves, domains and their associated infrastructure remain fundamental resources for threat actors as well as persistent attack vectors. Malicious actors consistently leverage domains to lure victims, control malicious operations, distribute harmful files, and facilitate email-based attacks. We are committed to empowering security researchers, brand protection teams, incident responders, and threat intelligence analysts with robust tools for proactive threat hunting and investigations. Recognizing the collaborative nature of security, we strive to expand our support for the community, collectively strengthening our defenses against external threats.