Detecting Malicious Domain Registration Batches: Patterns, Prevalence, and Security Implications

Samuel Cheadle, Carlos H. Gañán, Siôn Lloyd, Samaneh Tajalizadehkhoob
Security, Stability, and Resiliency Research, Office of the CTO
ICANN
Email: {sam.cheadle, carlos.ganan, sion.lloyd, samaneh.tajali}@icann.org

***Abstract*—The registration of domains in large, time-bound batches is a well-known tactic among cybercriminals seeking to enable DNS abuse at scale. This paper presents a comprehensive study of batch domain registrations, focusing on their detection, prevalence, and correlation with malicious activity. We introduce a clustering-based methodology leveraging domain creation time, registrar and authoritative nameserver data; analyze millions of recent gTLD registrations, and cross-reference these with security feeds to assess abuse rates. Our results indicate that batch registrations are prevalent, significantly predict overall abuse rates, and are useful for pivoting and expanding from known malicious "seed" domain sets, particularly in certain TLDs and registrar environments. We discuss the implications for defenders and propose directions for further research, including the challenges posed by privacy regulations and evolving attacker tactics.**

***Index Terms*—domain registration, abuse detection, clustering, security feeds, batch analysis, gTLD, registrar, DNS, cybercrime**

I. Introduction

The Domain Name System (DNS) is a foundational component of Internet infrastructure, providing the essential mapping between human-readable domain names and machine-routable IP addresses. However, the openness and scalability of the DNS has also made it a persistent target for abuse. Malicious actors routinely exploit the DNS by registering large numbers of domains in rapid succession—a practice referred to as *bulk registration*—to facilitate phishing campaigns, malware distribution, and other forms of cybercrime. The ability to easily register domains in bulk, often through automated interfaces or Application Programming Interfaces (APIs), enables attackers to quickly establish disposable infrastructure, evade detection, and scale their operations with minimal overhead.

From a security perspective, the identification and characterization of bulk domain registrations are of critical importance for abuse mitigation. Detecting such patterns can provide early warning of malicious campaigns and, in some cases, enable intervention before domains are weaponized. The operational reality is that attackers, seeking efficiency, overwhelmingly prefer automated, high-volume registration methods [1]. This behavior creates observable patterns—notably, bursty, time-bound spikes in registration activity—that can be leveraged by defenders to identify coordinated malicious activity.

979-8-3315-8969-1/25/$31.00 ©2025 IEEE

Despite the clear relevance to abuse mitigation efforts, the systematic detection and analysis of batch registrations has become increasingly challenging. Prior to the introduction of comprehensive privacy regulations, such as the General Data Protection Regulation (GDPR), defenders could more easily utilize registrant level data to group domains by owner, enabling identification of coordinated activity. In the current landscape, registrant information is frequently redacted or obscured by privacy and proxy services, significantly limiting the scope of registration features available for analysis. As a result, defenders must now rely on alternative signals, such as registrar, authoritative nameservers and creation timestamps, to infer relationships among domains.

Many past studies have investigated the potential predictive value of registration patterns for malicious domain detection [2]–[7]. However, these approaches often depend on data sources or features that are no longer widely accessible in the current regulatory environment. Moreover, there is a lack of large-scale, empirical studies that quantify the prevalence, characteristics, and security implications of bulk registrations across the global domain registration landscape, using only post-GDPR accessible features.

Bulk domain registrations typically consist of multiple registrations, occurring over a short period of time, linked to the same individual registrant. Bulk registrations may also include asynchronously registered domains; domains registered in series, as part of a campaign, with fixed or jittered intervals between subsequent registrations. While this latter class is important to consider for addressing the broader issue of bulk registrations, here we constrain the focus to *time-bound bulk registrations* only - referred to as *batch* registrations throughout the report.

This research addresses the following question: *How can batch domain registrations be reliably detected using publicly available thin*[1] *registration data, and what quantitative relationships exist between batch registrations and DNS abuse?* By focusing on publicly available registration features, this study aims to fill the gap in understanding the operational scale and security relevance of batch registrations.

The main contributions of this paper are as follows:

[1]Including only technical registration data such as nameservers, creation date and domain status, rather than detailed registrant contact information (registrant, admin and billing details)

- We propose a robust clustering methodology for the detection of batch domain registrations, utilizing only registrar, authoritative nameservers, creation timestamp and lexical domain string features.
- We conduct a large-scale empirical analysis of 10 million gTLD registrations, quantifying the prevalence of batch activity and its correlation with various forms of DNS abuse.
- We identify TLD- and registrar-specific risk profiles, demonstrating that certain TLDs and registrars are disproportionately associated with malicious batch registrations.
- We show that expanding detection to include all domains within batches containing known malicious domains can substantially increase the identification of potentially malicious domains.
- We provide an open, actionable framework for real-time batch detection, equipping defenders with GDPR-compliant tools for pre-emptive domain abuse mitigation.

II. Background and Motivation

The domain registration ecosystem is a complex, multi-stakeholder environment that underpins the operation of the DNS. At its core are registries, which manage top-level domains (TLDs), and registrars, which act as intermediaries between registrants (individuals or organizations) and registries. In addition, a significant portion of domain registrations are facilitated by resellers—entities that operate under the umbrella of registrars, often providing domain registration as part of a broader suite of web services.

To efficiently manage high volumes of registrations, both registrars and resellers commonly employ automated systems and APIs. These APIs enable bulk operations, such as registering, renewing, or configuring domains at scale, which is essential for legitimate businesses managing portfolios of domains, as well as for hosting providers serving large customer bases.

However, the same automation capabilities can be exploited by malicious actors. Attackers seeking to launch large-scale abuse campaigns—such as phishing, spam, or malware distribution—will often register domains in bulk [8]. Manual registration of hundreds or thousands of domains is impractical and inefficient; instead, attackers utilize registrar or reseller APIs to programmatically acquire large batches of domains over short time frames. This operational efficiency is a common feature of modern abuse campaigns, often reflected in the bursty, time-bound patterns observed in registration data.

From a defender's perspective, these batch registrations generate distinctive patterns, particularly in the timing of registrations, the choice of registrar or reseller, and the DNS infrastructure (such as authoritative nameservers) employed. Early identification of suspicious batches can enable defenders to flag, monitor, or even preemptively block malicious domains before they are weaponized, thereby reducing the window of opportunity for attackers.

However, the detection of batch registrations is complicated by several factors:

- **Privacy Regulations:** The current regulatory landscape, leading to the increased use of privacy / proxy services has significantly limited the availability of registrant data, making it harder to group domains by owner.
- **Shared Infrastructure:** Large registrars and DNS providers often serve many unrelated customers, complicating efforts to cluster domains based on shared infrastructure.
- **Evasion Tactics:** Sophisticated attackers may intentionally jitter registration times or diversify infrastructure to evade detection.

Despite these challenges, certain features—such as registrar, authoritative nameservers, and creation timestamp—remain available and can be leveraged for effective clustering.

III. Related Work

A. Ecosystem: Registrars and Registration

The practices of registrars, registries, and resellers play a critical role in the domain abuse landscape. Coull et al. [9] provided an early examination of registration abuse phenomena such as domain tasting, speculation and front-running, drawing attention to how market incentives and registrar practices can facilitate abuse. Liu et al. [10] analyzed the impact of registrar-level interventions and registry policy changes on the takedown of illicit online pharmacy domains, finding that while such interventions can temporarily disrupt abuse, adversaries often adapt rapidly. Alrwais et al. [11] provided a comprehensive study of the abuse ecosystem, revealing the interplay between registrars, resellers, and abuse mitigation efforts. Noroozian et al. [12] hypothesized the potential effectiveness of registrar-based interventions, showing that proactive registrar policies can significantly reduce abuse, but also highlighting the challenges posed by reseller channels and the use of bulk registration APIs. The role of APIs is particularly salient, as they enable both legitimate bulk operations and facilitate automated, large-scale acquisitions by attackers. Tajalizadehkhoob et al. [13] and Korczyński et al. [14] further highlighted the heterogeneity of the hosting and registrar ecosystem and its impact on the prevalence of malicious domains, with certain providers and TLDs disproportionately associated with high rates of abuse [1].

B. Predicting malicious registrations

A wide body of work has focused on the potential value of post-registration features, such as DNS traffic analysis for abuse detection [15]–[19]. Here, we focus only on past studies using features available at, or close to, the time of domain registration. These features typically include domain registration data, such as registrar and creation date, in addition to basic DNS infrastructure features.

Existing literature has established the predictive value of registration patterns for malicious domain detection. In 2010 Felegyhazi et al. [2] explored the potential value of clustering newly registered domains, leveraging information contained in DNS zone files and WHOIS registration data, highlighting

the effectiveness of the approach for the identification of associated domains.

Past work has also highlighted the importance of detailed registration records, including personal *registrant* level information such as name, address, email and phone number, for the detection of large-scale ongoing abuse campaigns in the .eu TLD [5]. A number of promising approaches have been developed by European ccTLD registries in this area [4], [5], [20]–[22]. However, visibility of such data is only granted to registrars and (some) registries due to the widespread use of privacy protection services. Therefore, although the theoretical value of clustering domains based on registrant information has been established, the required data typically cannot be accessed by external security researchers, hindering broad cross-sector analysis.

Other approaches to proactive abuse detection have focused on the use of public domain registration and infrastructure features. PREDATOR [3] was designed to detect potentially malicious domains shortly after registration, utilizing thin registration features such as registrar and authoritative nameservers. Relevant to the current study are the "batch correlation" features included within the model input, based on analysis of neighboring domains within a <registrar, five-minute epoch> tuple. This input serves to highlight 1) large volumes of registrations or 2) high name cohesiveness (lexical domain name similarity) registrations, occurring over a short time period. Similarly, recent investigations into bulk domain registration patterns have focused on the use of a relatively coarse temporal window for capturing batch registrations [8], [23]. The existing approaches described above face limitations due to the following reasons - 1) Batch registrations are defined only per *registrar*, rather than per *registrar-nameserver* combination, 2) The course temporal granularity of batch groupings is likely to fail to capture smaller (lower volume) batch registrations, and 3) Domains are not explicitly clustered into registration batches – False positives are likely to be generated if independently registered domains (sharing a common registrar) co-occur with real batch registrations, across a relatively wide temporal window (e.g. 5 – 10 minutes).

In this study, we use a principled approach to feature selection and grouping, based on the assumption that bulk registrations are typically associated with identical *registrar–nameserver* groupings. This a priori assumption enables more efficient clustering analysis by decreasing the search space, as well as improving the explainability and transparency of the algorithm.

IV. Methodology

A. Data Sources

Our analysis is based on a comprehensive dataset of generic Top-Level Domain (gTLD) registrations. All registration data is sourced from the ICANN compiled 'BRDA' (Bulk Registration Data Access) dataset, comprising thin registration features supplied directly by gTLD registry operators [24], covering approximately 17 million domains registered over a three-month period (01/01/25 to 31/03/25). For each domain, we compiled the following features:

- Registrar name/IANA ID
- Authoritative name server(s)
- Creation timestamp

Due to the focus of the present study on thin registration features (excluding registrant information) we refer to all listed features as publicly available (i.e. accessible via standard RDAP/WHOIS queries). However, it should be noted that the compiled BRDA dataset is currently available for internal ICANN analysis only.

To assess reported domain abuse, we cross-referenced this dataset with multiple threat intelligence and reputation blocklists (RBLs), including commercial and open-source feeds. The full set included Spamhaus [25], SURBL [26], WMC-Global [27], PhishTank [28], Urlscan [29], the Anti-Phishing Working Group (APWG) [30], and Abuse.ch's Urlhaus [31]. These sources provide labels for domains associated with spam, phishing, malware distribution, and botnet command-and-control (C&C) infrastructure.

While spam domains do not fall under the technical definition of DNS abuse as used in ICANN policy [32], which includes only spam domains that act as a delivery mechanism for other threats, we include spam in the exploratory analysis below to support comparative insights. However, all subsequent abuse analyses excludes spam-related classifications, unless stated otherwise.

B. Batch Detection Algorithm

We define a *batch* as a group of domains registered close in time, through the same registrar and served by the same authoritative nameservers. The detection process involves:

1) Grouping domains by unique combinations of *registrar + authoritative nameservers*
2) Applying density-based clustering (DBSCAN) to the creation timestamps within each group to identify time-bound clusters
3) Applying a set of filtering criteria to exclude low reliability clusters

Density-based spatial clustering of applications with noise (DBSCAN [33]) is a popular clustering algorithm suitable for segregating data into regions of high and low density, well-suited to the task of identifying dense clusters of registrations against a sparse background of unrelated activity. Unlike centroid based or hierarchical clustering methods, which typically do not consider the density of datapoint distribution, DBSCAN segregates regions of high density (clusters) from lower density regions (noise). This quality makes the algorithm suitable for batch registration detection – groups of domains sharing identical infrastructure, created at, or close to, the same time will be clustered together within regions of high-density.

DBSCAN relies on two primary hyperparameters: *epsilon* and *min-samples*. *Epsilon* defines the minimum distance between datapoints required for them to be considered neighbours - initial testing explored a range of epsilon values, from

0.1 to 10 seconds. Although maximum precision is achieved for low *epsilon* values (for which DBSCAN captures only batches of domains registered near simultaneously), we were interested in capturing expanded batch registrations taking place over the course of multiple seconds or minutes. Therefore, an *epsilon* value of 5 seconds was selected. A *min-samples* value of 2 was selected, enabling analysis of all batch sizes.

C. Filtering criteria

Filtering is applied to generate a high confidence subset of batch clusters for detailed analysis. We use a set of filters based on batch size and lexical properties (domain string consistency), with the aim of ensuring a transparent and justifiable approach. The method represents a conservative filtering approach, likely to also exclude many real batch registrations. The current analysis aims to investigate batch registrations linked to strong evidence of association only, justifying this stringent level of filtering. Specific filtering criteria are listed and justified below:

- Exclusion of small clusters ($n < 10$). Validation tests on the raw batch results highlighted the existence of false positives (clustered but non-related domains) for small batch sizes (see Validation IV-E section). Therefore, we chose the conservative approach of including only larger batch sizes ($n \geq 10$) in our analysis, for which we observed a zero false positive rate in validation testing.
- Exclusion of large clusters ($n > 1000$). Although rare, large clusters do occur. These may be linked to drop catch registrations or other batch update operations performed by the registrar. Manual checking of a sample of these larger batches revealed evidence of heterogeneous sets of domains, lacking additional evidence of association. Although further work is required to understand the causes of large ($n > 1000$) batch registrations, we chose to exclude large sizes in order to maintain the focus on high confidence batches only.
- Exclusion of clusters containing high-variance domain strings. Homogeneity of string properties provides additional evidence of association. We employ two separate string similarity metrics described below.

D. Domain string similarity

Although lexical similarity between batch members is not a prerequisite, it is a common feature. For instance, many batches contain domains sharing common keywords (e.g. login123[.]com, 456login[.]com) or higher order lexical properties linked to algorithmic domain generation. In order to select batches with strong evidence of association, we used two forms of string similarity validation, similar to the techniques employed in the PREDATOR study [34] :

First order similarity - N-grams of domain strings (excluding TLD) were calculated (3-grams), followed by jaccard similarity scoring. Scores were calculated by comparing the initial domain within each cluster, to all other domains within the cluster. Only clusters with an average (mean) $jaccard\ score \geq 0.3$ were selected for inclusion based on first order string similarity.

Second order similarity - Some clusters lack first order feature similarity, such as identical sub-strings, but nevertheless contain domains that share common second order features, commonly linked to domain generating algorithms (DGAs). To capture similarity of this type, the following set of second order metrics were generated for all domains within the sample: *string length, % high frequency characters, % distinct characters, % digits, % hyphens* (see Appendix A for definitions). Following the calculation of these scores at the domain level, intra-batch statistics were calculated, capturing the mean and variance for each of the five separate metrics. Anomalous batches were identified via comparison to a baseline population of domains (randomly sampled Tranco top 1M domains, n=500). Mann-Whitney U tests (non-parametric test of variance) were performed on all metrics, on a per cluster basis, in order to highlight statistically significant second order string differences. Clusters were selected for inclusion if two or more metrics exhibited a statistically significant difference (p <0.05) from the baseline population.

E. Validation

Collaborative work was carried out with a single registrar to validate a sample of batch results, via cross-checking account holder records. Account holder information was considered ground truth, and used for assessing the homogeneity of cluster output. Collaborative work was carried out in a privacy preserving manner, ensuring that only aggregated and annonymised data was transferred between parties. A true positive (TP) was defined as a correctly identified batch (all domains belong to the same account holder). A false positive (FP) was defined as an incorrectly identified batch (domains belong to multiple different account holders). Appendix A, Fig. 7 plots precision [$TP/(TP+FP)$] as a function of batch size. Validation results demonstrate high precision for larger batches ($n \geq$8).

F. Abuse Labeling

Domains were labeled as *malicious* if they appeared in any reputation block list (RBL) during the main observation period (January - March 2025), plus an additional two month period (April-May 2025), allowing for a minimum 60 day interval between domain registration and RBL report. We further categorize abuse by threat type (spam, phishing, malware and C&C) where possible.

Batches categorized as "unknown" exhibit no overlap with RBLs (including spam, phishing, malware and C&C).

V. Results

A. Prevalence of Batch Registrations

Applying our clustering algorithm to the dataset of 16.6M newly registered gTLD domains, we found that just over 42% of all registrations during the study period could be grouped into batches, from the initial clustering step. After filtering

TABLE I
BATCH REGISTRATION COVERAGE. COMPARISON OF RAW AND FILTERED RESULTS

	Raw batches	Filtered batches
All	42.0% (7.0M)	16.0% (2.7M)
Spam	65.1% (907K)	45.7% (637K)
Phishing	46.7% (112K)	24.2% (58K)
Malware	40.8% (938)	22.2% (511)
C&C	31.3% (754)	14.3% (344)

based on batch size and string similarity, this figure dropped to 16% of newly registered gTLD domains (Table I).

Based on domains reported as abused and newly registered within the analysis period (approximately 1.5M), the batch registered percentage varies by threat type. Focusing on filtered batches, spam-related domains exhibit the highest batch registration rate (45.7%) and C&C related domains the lowest (14.3%), revealing large differences in batch registration prevalence between threat types. However, the relatively small sample size for some threat types (Malware and C&C) should be noted.

All subsequent analyses focus on technical DNS abuse only (phishing, malware and C&C), excluding spam unless otherwise stated.

B. Batch Size Distribution By Reported Abuse Type

Raw batches exhibit a wide range of sizes, from small batches, containing just a few domains, up to large batches containing over 3500 domains (Fig. 1, top panel, grey), with some batch sizes linked to disproportionately high counts (e.g. batch size 1000, raw data). We choose a conservative approach, excluding both small ($n < 10$) and large batches ($n > 1000$) batches from subsequent detailed analysis (See Method - Filtering Criteria for details).

Fig. 1 (bottom panel) focuses on batch sizes of 100 or less, highlighting clear preferences for regular batch sizes, spiking at 10, 15, 20, 30, 40, etc, particularly for abused (RBL) batch domains.

Fig. 2 displays cumulative frequency distributions (CFDs) for filtered batches, split by threat type. Smaller batches are more common for abused groups (phishing, C&C), relative to batches with no evidence of abuse (Fig. 2). The results also highlight the prevalence of smaller batches in general; 26% of domains were contained in batches of size 20 or less.

C. Registrar Level Overview

Fig. 3 illustrates batch registration time-course trends over a 1-month period (February, 2025), revealing substantial variation between the 25 registrars with the highest batch registration rates (R1 - R25). In Fig. 3, results for individual registrars are represented by a series of circles along the horizontal axis, with the area of each circle reflecting the volume of batch registrations per hour. Higher reported abuse rates are shown in red.

Fig. 3 highlights the following points:

- Variability in batch sizes and frequency. Some registrars exhibit infrequent high volume batch registrations (over 3.5K per hour), while others display a more consistent profile of regular small volume batch registrations.
- The proportion of batch registered domains contained in RBL feeds is noticeably high for a small number of registrars (e.g. registrar 7 has a consistently high proportion RBL flagged domains), while the majority of registrars exhibit a more mixed abuse profile, with high abuse concentrations occurring infrequently.
- Some registrars display evidence of batch registrations over a short time period only. For instance, registrar 24 has no batch registrations until February 17th, after which there is a regular stream. This may indicate the presence of campaigns, initiated by a limited numbers of registrants.

See Appendix A, Fig. 8, for an equivalent plot including the spam category.

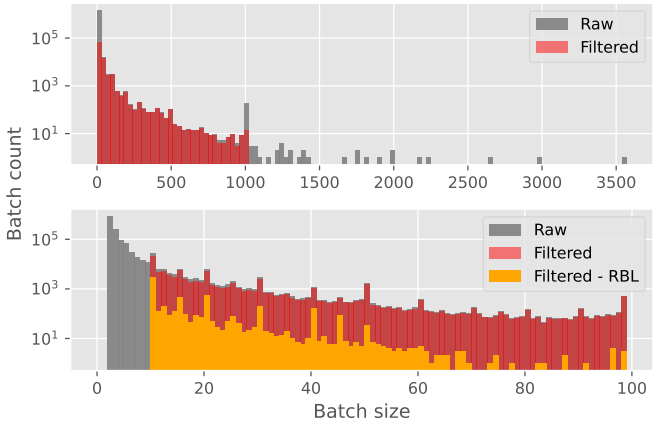

Fig. 1. Batch size histograms split by type (Raw, Filtered and Filtered-RBL). Top panel: Full distributions split by filter type (raw vs filtered). Bottom panel: Distributions focused on 0 - 100 size range, with the addition of filtered batch domains reported as abused (Filtered-RBL). Both distributions are plotted on a logarithmic scale.

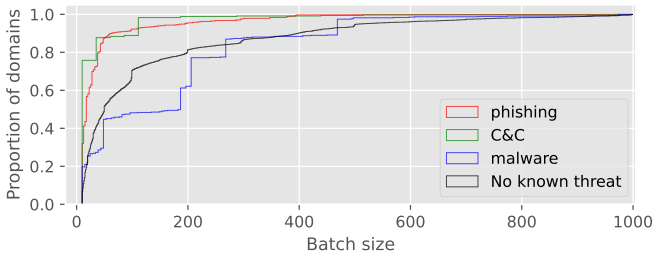

Fig. 2. Cumulative frequency distribution of domain counts as a function of batch size, split by threat type (phishing, C&C,malware, no known threat).

D. TLD Composition

Focusing on gTLDs only (due to the limitations of our source data) reveals that certain TLDs are disproportionately associated with malicious batch registrations. For example, TLDs with a history of abuse are over-represented among both malicious and 'unknown' (unlabeled) batches, but this

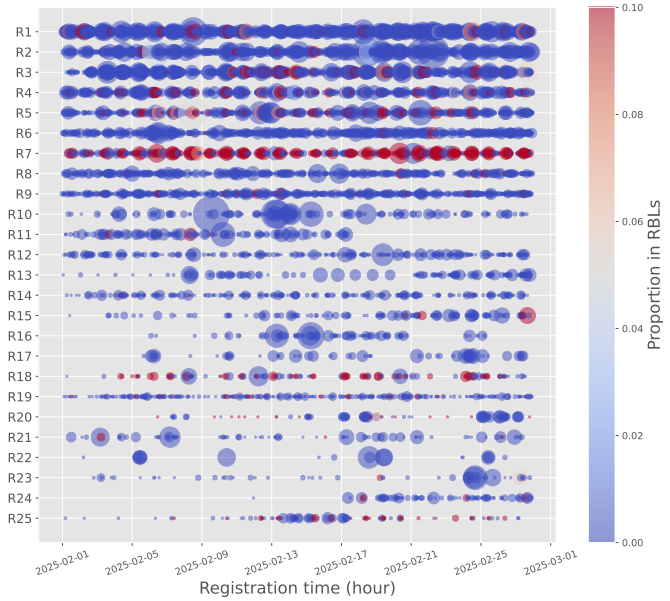

Fig. 3. Registrar time-course plot of batch registrations over a 1 month period (February, 2025). The top 25 registrars with the highest overall volumes of batch registered domains are listed along the vertical axis, with the area of individual circles representing batch registration volume per hour (min: 10, max: 3623 registrations per hour). Colour indicates abuse rates, ranging from blue (0% of domains in RBLs) to red (10% or more of domains in RBLs).

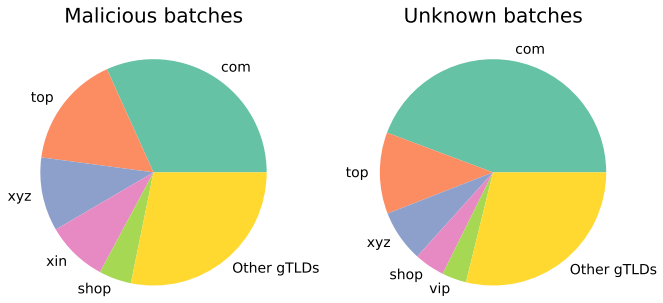

Fig. 4. gTLD composition (top 5) for batches containing at least one RBL domain (malicious batches, left panel) and batches containing no RBL domains (unknown batches, right panel)

over-representation is most pronounced for malicious batches where four gTLDs (.top, .bond, .xyz and .xin) cover 45% of the sample. See Fig. 4, left panel. In contrast, the same four TLDs comprise less than 25% of unknown category batches in this sample (Fig. 4, right panel), with .vip replacing .xin in the top 5 TLD list.

E. Relationship Between Abuse and Batch Registration Rates

Some registrars exhibit high abuse and batch registration rates (Fig. 3). Recent analysis [1] demonstrates that registrar level practices help to explain abuse rate variance across registrars - for instance, the degree to which API restrictions are implemented to control automated bulk domain registrations by unverified account holders. Rather than examining specific registrar level policies or practices (see [1] for a detailed examination of this topic), we tested the relationship between abuse levels and batch registration rates[2] per registrar.

Based on technical DNS abuse (phishing, malware, C&C) only, OLS linear regression demonstrated a weak positive relationship between abuse levels and batch registration rates ($beta = 0.12$, $p < 0.01$). The strength of relationship between batch and reported abuse rates increases after the inclusion of spam as an abuse type (OLS linear regression, $beta = 0.63$, $p < 0.001$). See Appendix A, Fig. 9 for a scatter plot. This result indicates that, after the inclusion of spam, for each 1% increase in batch registration rates, corresponding abuse rates increase by approximately 0.63%.

Preliminary work indicates that batch registration rates are a stronger predictor of abuse rates than many other registrar level characteristics. This finding is based on replication of the random effects regression analysis conducted for the INFERMAL study [1] (Model 2), including a selection of the main registrar characteristics as predictors from the original analysis (API restrictions, registration restrictions and discount options), together with the addition of the newly defined "batch registration rate" metric. This preliminary analysis indicates that, after accounting for other registrar level characteristics, batch registration rates remain a strong predictor of overall abuse rates.

Despite the significant positive relationship between batch and reported abuse rates, our analysis highlights a lack of uniformity; some registrars exhibit high batch registration rates (>50%), while maintaining low reported abuse rates (<1%). This may be due to a number of factors, such as the bulk creation of domains for legitimate purposes, for spam, or for abuse types/campaigns that are not captured within our RBLs. For example, manual checks on a subset of domains linked to a single registrar with high batch (>50%) but low reported abuse rates (<1%) revealed evidence of a large number of Chinese language e-commerce stores, serving content for only a short time period. This example is consistent with the existence of fake web-shops in some batches, a form of abuse that is not covered in our current RBL data.

F. Expansion Analysis

By examining batches containing at least one malicious domain, we infer that many unlabeled domains within the same batch are likely to also be malicious (See [2] for a similar expansion method). Expanding the set of flagged domains to include all members of these batches almost doubles the number of reported abused domains in our sample (80% expansion rate). Fig. 5 illustrates the known threat type composition of a randomly selected set of 300 batches from the main dataset (max size, $n < 100$), containing at least one reported RBL domain. The plot displays the number of reported abused (black) and suspicious (grey) batch members. Expansion analysis labels all suspicious domains (Fig. 5, blue) as inferred malicious. The figure highlights that many of the

[2]The proportion of all domains under management, registered within the sample time period, that were batch registered.

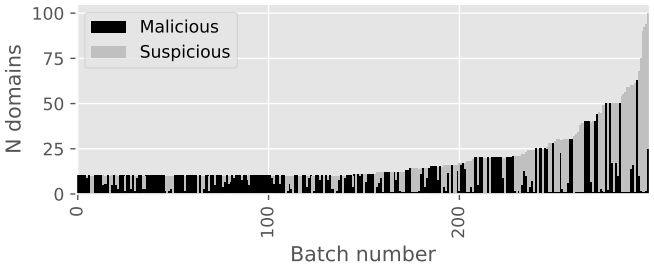

Fig. 5. Composition of a sample of 300 malicious batches, illustrating the proportion of domains per batch contained in RBL feeds (black), together with all remaining unlabeled suspicious domains (grey). Each vertical bar represents a single batch, sorted by size.

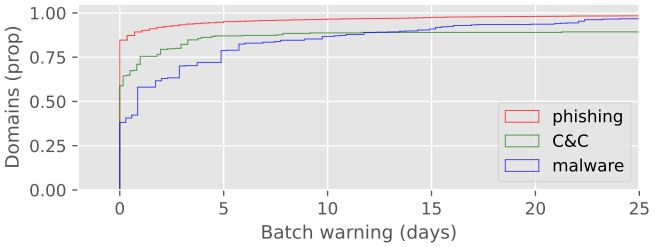

Fig. 6. RBL flagging time-course for domains tagged as malicious via batch inference, and subsequently included in RBLs (alerted domains). Zero mark on x-axis indicates the time at which the initial evidence of batch abuse is received.

larger batches ($n > 50$) contain a high proportion of suspicious domains, responsible for the majority of the expanded set of results.

Based on a total of 102K batch registered malicious domains (phishing, malware or C&C) in our sample, 23K could be detected earlier based on batch inference. This calculation assumes immediate access to registration data and does not account for real world delays associated with the compilation of bulk registration data (such as BRDA). Table II illustrates the concept by listing RBL detection dates for a single batch of malware domains all registered on 2nd January 2025. After the initial RBL report (for 24-gg123[.]sbs) on the 2nd January, the remaining seven RBL domains are flagged after a delay of between 2 and 5 days. These inferred malicious domains, included in RBLs at a later date, are referred to as the "alerted" set. Based on the full dataset, Fig. 6 displays the cumulative frequency plot for this set, split by threat type.

Additionally, 266K domains not listed in RBLs (within the sample time-span) were tagged as malicious via batch inference, referred to as the "expanded" set. For example, the domains listed in Table II with *Unknown* threat type (1-gg123[.]sbs, 20-gg123[.]sbs, etc).

Based on the calculation of these two sets, an initial seed domain volume of 79K (RBL batch registered - alerted) yields an additional 289K domains (alerted + expanded), reflecting a batch registered RBL expansion rate of 365%, and a total RBL expansion rate of 80%. The latter result indicates that for every three newly registered malicious domains reported through

TABLE II
SAMPLE OF DOMAINS FROM A MALICIOUS BATCH, SHOWING THE FIRST 10 DOMAINS, TOGETHER WITH THREAT TYPE AND RBL REPORT DATE

DOMAIN	Threat label	Report date
26-gg123[.]sbs	Malware	2025-01-02
28-gg123[.]sbs	Malware	2025-01-07
27-gg123[.]sbs	Malware	2025-01-08
24-gg123[.]sbs	Malware	2025-01-08
9-gg123[.]sbs	Malware	2025-01-05
7-gg123[.]sbs	Malware	2025-01-03
1-gg123[.]sbs	Unknown	
25-gg123[.]sbs	Malware	2025-01-03
20-gg123.sbs	Unknown	
21-gg123.sbs	Malware	2025-01-03

RBL feeds, batch expansion (using conservative filtering) identifies an additional two neighboring domains.

Manual validation of a sample of the expanded dataset revealed no false positive batches, but we note that it is possible to implement a percentage threshold for inclusion (e.g. 10% minimum batch-RBL overlap) in order to generate higher confidence results.

VI. DISCUSSION

Our findings highlight the prevalence of batch registrations within the gTLD landscape. At least 16% of newly registered gTLD domains, in the first quarter of 2025, exhibit batch registration patterns. The true rate is likely to be higher, due to the conservative filtering employed in the current study, which excludes many valid batches.

For the majority of reported abuse types we observe batch registration rates above 20%, up to a maximum of 45% (spam). At the level of individual registrars, we further demonstrate a significant positive relationship between batch registration rates and overall abuse rates. However, this relationship is weak when limited to technical DNS abuse (phishing, malware and C&C), strengthening only after the inclusion of spam. Results are consistent with the frequent use of batch registrations for low-value, disposable domain abuse (spam), for which the consequences of detection are limited. For higher severity abuse types, such as C&C, batch registrations are less frequent. This may reflect deliberate attacker strategies, aiming to evade detection and suspension of high-value domains.

Results are consistent with the importance of policies and practices put in place by individual registrars (e.g. regarding API availability/restrictions) for limiting high-volume registrations. While highlighting the link between batch registrations and abuse rates, our findings are also consistent with the legitimate use of many batch registered domains. There are many possible benign use cases for batch registered domains, including defensive registrations, advertising networks or traffic distribution systems (TDS). The current analysis has not explored the nature of the many, presumed legitimate, batch domains which are not reported as abused in our RBL data, but this represents a valuable line of future investigation.

Batch expansion analysis demonstrates that the process of batch inference leads to a substantial increase in RBL reported

abuse volumes; an 80% increase in the volume of newly registered RBL domains. The expansion methodology is likely to require refinement before adoption in an operational setting (for instance, by imposing a minimum threshold of evidence prior to expansion), but represents a promising approach for proactive detection and suspension of to-be-abused domains.

A. Limitations

The batch clustering technique used for this analysis relies only on publicly available domain registration and infrastructure data, beneficial for wide-scale centralized analysis of registration patterns across multiple gTLDs. However, the limitations of this approach include:

- **Attribution Challenges:** Shared infrastructure and privacy protections limit the ability to attribute batches to specific actors. The proposed analysis uncovers *associated* domains only, and does not directly provide new data for enhanced attribution. However, the technique can expand attribution from known seed domains, to previously unknown associated domains.
- **Evasion:** Attackers may adapt by distributing registrations over longer time periods or across multiple registrars. The proposed clustering technique relies on the assumption that bulk registrations are associated with identical registrar–nameserver groupings. Although commonly true, this is not necessarily always the case. While attackers may attempt to evade detection by batch registering across multiple registrars and nameservers, this process is more effortful, time consuming and costly.
- **Validation:** Validation analysis highlighted two issues that require further consideration. Firstly, based on ground truth data held by a single registrar, examining inter-batch account holder consistency, we observe evidence of low reliability for small clusters. In these cases, co-occurrence of unrelated domains is most likely, particularity for common registrar-nameserver combinations. This issue has been addressed in the current work via exclusion of small batches ($n<10$). Secondly, we observed evidence of large clusters ($n>1000$) containing heterogeneous sets of domains, which do not appear to share other common characteristics, other than occurring in a batch registration. One possibility, under investigation, is that resellers may introduce *artificial* batch signals, through delays in the processing pipeline, leading to skewed registration timestamps and incorrect batch classification.

B. Future Work

Key directions for future research include:

- **Broader bulk registration coverage**: The current analysis focuses on detection of time-bound batches of domain registrations. Future work will aim to capture ongoing campaign registrations, occurring over a longer time frame. In addition to time-series analysis of creation dates, follow-up analysis will leverage lexical analysis of domain string properties, to cluster domains sharing similar lexical properties, facilitating campaign detection.
- **Public sharing of batch registration data:** Ongoing publishing of batch registration data may be beneficial to the community in a number of ways. Firstly, to increase visibility, ensuring the community has access to reliable registrar and TLD level batch registration statistics. The current data source (centralised, ICANN compiled, gTLD registration data) should be sufficient for this type of historical trend reporting[3]. A number of options are being explored for the publication of batch registration data, including trend reporting, associated domain search (looking up related domains based on a seed domain) and metadata linked to RBL domains provided to registrars and registries through Domain Metrica [35].
- **Faster Batch detection:** For proactive abuse mitigation work, involving pivoting on a known malicious seed domain, it is important to minimse update delays to the underlying registration data. This may be achieved through the use of external 'newly observed domain' (NOD) feeds, in combination with bulk RDAP queries to rapidly compile domain registration data. It should be noted, however, that large scale collection of registration data through RDAP is often constrained by rate limiting imposed by the registry/registrar operators. Future work should explore options for rapid publication of registration data, enabling efficient analysis of registration data for batch tagging, as well as other security related use cases. Additionally, the use of alternative NOD source data, such as certificate transparency logs, may enable coverage to be extended to include ccTLDs [36].
- **Collaborative work with Registrars:** The ground truth data necessary for comprehensive evaluation of the batch clustering technique is held by registrars or registries (i.e. detailed registrant / account holder details). Although we have performed validation analysis via collaboration with a single registrar, to increase reliability, ongoing collaboration between ICANN and multiple registrars is required - an area of active discussion within the ICANN community.

VII. Conclusion

Batch domain registrations are a prevalent and significant feature of the modern domain landscape, closely linked to various forms of abuse. By leveraging simple, robust features and clustering techniques, batch registrations can be detected at scale across all gTLDs. This wide-scale analysis enables batch registration rates to be computed for individual registrars or TLDs, important for improved visibility and trend reporting, helping to enable proactive mitigation work by individual registrars and TLD operators. Batch clustering results also have the potential to assist defenders with efficient detection and blocking of malicious campaigns. Continued research and

[3]The update delays inherent to the BRDA dataset (between 1 and 7 days) may be deemed acceptable

collaboration are needed to refine these methods and adapt to evolving attacker strategies.

Acknowledgments

We would like to thank Realtime Register for collaborative work, assisting with the validation of a sample of batch data.

References

[1] Y. Nosyk, M. Korczynski, S. Maroofi, J. Bayer, Z. Odgerel, A. Duda, S. Tajalizadehkhoob, and C. Gañán, "INFERMAL: Inferential analysis of maliciously registered domains," Dec. 2024. [Online]. Available: https://www.icann.org/resources/pages/inferential-analysis-maliciously-registered-domains-infermal-2024-12-03-en

[2] M. Felegyhazi, C. Kreibich, and V. Paxson, "On the potential of proactive domain blacklisting," in *Proceedings of the 3rd USENIX Conference on Large-scale Exploits and Emergent Threats (LEET)*, 2010.

[3] S. Hao, A. Kantchelian, B. Miller, V. Paxson, and N. Feamster, "PREDATOR: proactive recognition and elimination of domain abuse at time-of-registration," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 1568–1579.

[4] J. Prins, "Proactive recognition of domain abuse," Thesis, University of Twente, Enschede, October 2020. [Online]. Available: https://purl.utwente.nl/essays/84073

[5] J. Spooren, T. Vissers, P. Janssen, W. Joosen, and L. Desmet, "Premadoma: an operational solution for DNS registries to prevent malicious domain registrations," in *Proceedings of the 35th Annual Computer Security Applications Conference*. San Juan Puerto Rico USA: ACM, Dec. 2019, pp. 557–567. [Online]. Available: https://dl.acm.org/doi/10.1145/3359789.3359836

[6] L. Berenschot, "Early warning system for newly registered malicious domains : A machine learning and certificate transparency approach," Thesis, University of Twente, Enschede, August 2024. [Online]. Available: https://purl.utwente.nl/essays/102379

[7] M. Weber, J. Wang, and Y. Zhou, "Unsupervised Clustering for Identification of Malicious Domain Campaigns," in *Proceedings of the First Workshop on Radical and Experiential Security*. Incheon Republic of Korea: ACM, May 2018, pp. 33–39. [Online]. Available: https://dl.acm.org/doi/10.1145/3203422.3203423

[8] Interisle Consulting Group, "Phishing Landscape 2025: An Annual Study of the Scope and Distribution of Phishing," Sep. 2025. [Online]. Available: https://interisle.net/insights/phishing-landscape-2025-an-annual-study-of-the-scope-and-distribution-of-phishing

[9] S. E. Coull, A. M. White, T.-F. Yen, F. Monrose, and M. K. Reiter, "Understanding domain registration abuses," in *Computers & security*, vol. 31, no. 7. Elsevier, 2012, pp. 806–815.

[10] H. L. Liu, K. Levchenko, M. Félegyházi, C. Kreibich, G. Maier, and G. M. Voelker, "On the effects of registrar-level intervention," in *4th USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET 11)*, 2011.

[11] S. Alrwais, K. Yuan, E. Alowaisheq, Z. Li, and X. Wang, "Understanding the dark side of domain parking," in *23rd USENIX Security Symposium (USENIX Security 14)*, 2014, pp. 207–222.

[12] A. Noroozian, M. Korczyński, C. H. Gañan, D. Makita, K. Yoshioka, and M. Van Eeten, "Who gets the boot? analyzing victimization by ddos-as-a-service," in *Research in Attacks, Intrusions, and Defenses: 19th International Symposium, RAID 2016, Paris, France, September 19-21, 2016, Proceedings 19*. Springer, 2016, pp. 368–389.

[13] S. Tajalizadehkhoob, M. Korczyński, C. H. Gañán, G. C. M. Moura, C. Hesselman, and M. van Eeten, "Apples, oranges and hosting providers: Heterogeneity and security in the hosting market," *Journal of Cybersecurity*, vol. 3, no. 3, pp. 265–279, 2017.

[14] M. Korczynski, M. Wullink, S. Tajalizadehkhoob, G. C. Moura, A. Noroozian, D. Bagley, and C. Hesselman, "Cybercrime after the sunrise: A statistical analysis of DNS abuse in new gTLDs," in *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, 2018, pp. 609–623.

[15] P. B. Danzig, K. Kumar, and K. Obraczka, "An Analysis of Wide-Area Name Server Traffic: A Study of the Internet Domain Name System," in *Proceedings of the ACM SIGCOMM*, 1992, pp. 281–292.

[16] J. Jung, E. Sit, H. Balakrishnan, and R. Morris, "DNS Performance and the Effectiveness of Caching," in *Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement (IMW)*, 2002, pp. 153–167.

[17] M. Antonakakis, R. Perdisci, D. Dagon, W. Lee, and N. Feamster, "Building a Dynamic Reputation System for DNS," in *Proceedings of the 19th USENIX Security Symposium*, 2010, pp. 273–290.

[18] L. Bilge, S. Sen, D. Balzarotti, E. Kirda, and C. Kruegel, "EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis," in *Proceedings of the 18th Annual Network and Distributed System Security Symposium (NDSS)*, 2011.

[19] M. Antonakakis, R. Perdisci, Y. Nadji, N. Vasiloglou, S. Abu-Nimeh, W. Lee, and D. Dagon, "Kopis: detecting Malware Domains at the Upper DNS Hierarchy," in *Proceedings of the 20th USENIX Security Symposium*, 2011, pp. 513–528.

[20] A. Del Soldato, D. Sartiano, and M. Martinelli, "Reads: Enhancing .it domains integrity with registrant's anomalies detection system," in *2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)*, 2024, pp. 1–7.

[21] T. Wabeke and T. van den Hout, "Assessing the risk of new .nl registrations using RegCheck," Jan. 2023. [Online]. Available: https://www.sidnlabs.nl/en/news-and-blogs/assessing-the-risk-of-new-nl-registrations-using-regcheck

[22] Nominet, "Domain watch," Jul. 2025. [Online]. Available: https://registrars.nominet.uk/uk-namespace/security-tools-and-protection/domain-watch/

[23] DNS Research Federation, "Bulk Registrations Uncovered: Legitimate Uses vs. Cybercriminal Exploits," Feb. 2025. [Online]. Available: https://dnsrf.org/blog/bulk-registrations-uncovered--legitimate-uses-vs--cybercriminal-exploits

[24] ICANN, "FAQ for Implementing the Temporary Specification for gTLD Registration Data," Jul. 2018. [Online]. Available: https://www.icann.org/en/contracted-parties/registry-operators/faq-for-implementing-the-temporary-specification-for-gtld-registration-data-22-06-2018-en

[25] The Spamhaus Project, "Spamhaus Blocklists," https://www.spamhaus.org/, real-time threat intelligence on spam and botnet activity.

[26] SURBL, "SURBL: URI Reputation Data," https://www.surbl.org/, reputation blocklist for detecting malicious URLs in unsolicited messages.

[27] WMC Global, "Threat Intelligence Services," https://wmcglobal.com/, feeds on phishing, SMS-based scams, and threat actor infrastructure.

[28] PhishTank, "PhishTank: Join the Fight Against Phishing," https://www.phishtank.com/, community-driven phishing verification and blocklist.

[29] Urlscan.io, "urlscan.io - Website Scanning and Threat Intelligence," https://urlscan.io/, public URL scanner used to detect phishing and malicious behavior.

[30] Anti-Phishing Working Group, "APWG eCrime Exchange (eCX)," https://apwg.org/ecx/, global industry, law enforcement, and NGO cooperative sharing phishing data.

[31] Abuse.ch, "URLhaus: Malware URL Database," https://urlhaus.abuse.ch/, community project to collect and share malware distribution URLs.

[32] ICANN, "What is DNS Abuse?" https://www.icann.org/resources/pages/dns-abuse-2020-03-31-en, 2020, iCANN's definition of DNS abuse, including malware, phishing, botnets, and pharming.

[33] M. Ester, H.-P. Kriegel, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, p. 226–231, 1996.

[34] S. Hao, A. Kantchelian, B. Miller, V. Paxson, and N. Feamster, "PREDATOR: Proactive Recognition and Elimination of Domain Abuse at Time-Of-Registration," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. Vienna Austria: ACM, Oct. 2016, pp. 1568–1579. [Online]. Available: https://dl.acm.org/doi/10.1145/2976749.2978317

[35] ICANN, "ICANN Domain Metrica: A Measurement Platform - ICANN," 2024. [Online]. Available: https://www.icann.org/octo-ssr/metrica-en

[36] R. Sommese, R. v. Rijswijk-Deij, and M. Jonker, "This Is a Local Domain: On Amassing Country-Code Top-Level Domains from Public Data," Sep. 2023, arXiv:2309.01441 [cs]. [Online]. Available: http://arxiv.org/abs/2309.01441

APPENDIX A

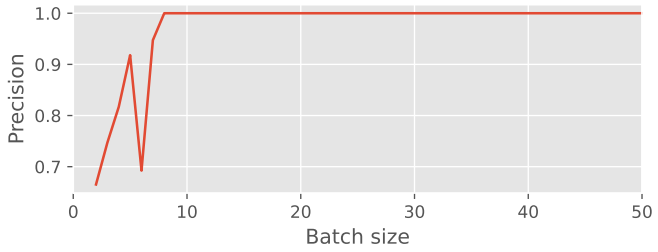

Fig. 7. Results of validation of the batch results, based on data from a single medium sized registrar. Precision [TP / (TP+FP)] is plotted as a function of batch size.

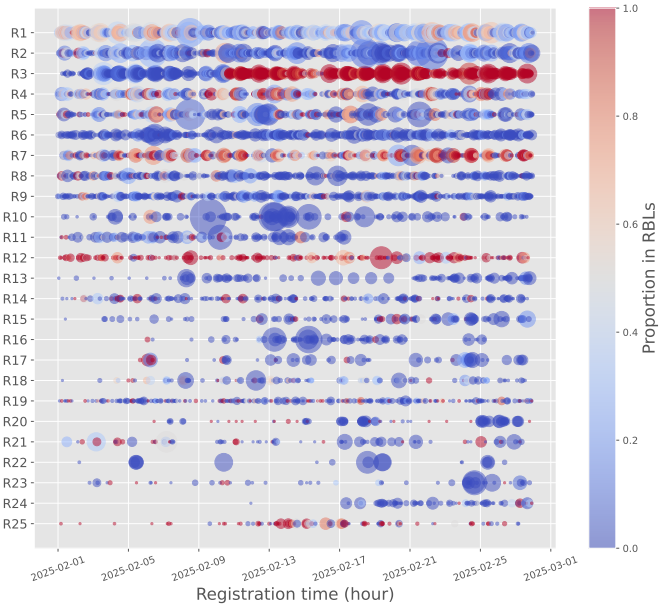

Fig. 8. Registrar time-course plot of batch registrations over a 1 month period (February, 2025), **including spam**. The top 25 registrars with the highest overall batch registration rates are listed along the vertical axis, with the size of individual dots representing batch registration rates per hour (min: 10, max: 3623 registrations per hour). Colour indicates abuse rates, ranging from blue (0% of domains in RBLs) to red (100% of domains in RBLs).

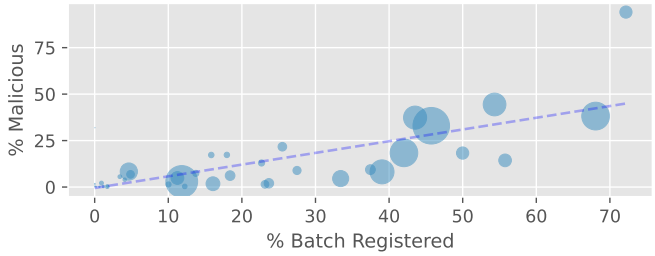

Fig. 9. Relationship between batch registration and abuse rates, **including spam**, at the registrar level. Each circle represents a single registrar, with the area of each circle reflecting the total number of batch registered domains per registrar. The dashed blue line shows the OLS best fit.

Second order lexical features - The following five features were included in the second order lexical domain string analysis. Where n specifies the character count, and n_total the total character count within the second level domain string (excluding the TLD):

$$n\_total = count\ of\ characters\ (exc\ TLD)$$

In addition to n_total, the following second order lexical features were included in the model:

The percentage of the domain string composed of the most common (highest frequency) character $(\%\ max\ frequency\ character)$:

$$count\ of\ highest\ frequency\ character/n\_total$$

The percentage of the domain string composed of distinct characters, occurring only once in the string $(\%\ distinct\ characters)$:

$$count\ of\ distinct\ characters/n\_total$$

The percentage of the domain string composed of digits $(\%\ digits)$:

$$count\ of\ digits/n\_total$$

The percentage of the domain string composed of hyphens $(\%\ hyphens)$:

$$count\ of\ hyphens/n\_total$$