Investigating Phishing Attacks using the Registration Data Access Protocol (RDAP)

Hauke Jan Lübbers[1]

[1]CSIS Security Group A/S, Vestergade 2B, 1456 Copenhagen, Denmark

Abstract

The Registration Data Access Protocol (RDAP) is a successor to the WHOIS protocol and enables programmatic access to the registration data of internet resources. Using RDAP, we investigated phishing attacks observed over four days, focusing on DNS domain names. In this paper, we present the opportunities and problems identified in the process. We find that low RDAP adoption among ccTLD registry operators, strict rate limiting, differences in data representation, and the (un-)availability of data due to privacy regulations continue to be hindrances to the widespread use of RDAP in cybercrime investigations. While these issues are currently preventing security researchers from solely relying on RDAP for accessing domain name registration data, we recognize its potential as a valuable data enrichment source for investigating phishing attacks at scale.

Keywords

Registration Data Access Protocol (RDAP), WHOIS, domain names, phishing, cybercrime investigation

1. Introduction

Phishing is a common attack vector employed by threat actors. While the attackers' motives and sophistication may vary, phishing attacks continue to be a relatively simple method for threat actors to gather credentials for later exploitation [1].

When investigating phishing attacks, a common first step is to look up the domain registration data, or "WHOIS information", of the involved domain names suspected of hosting a phishing website or of sending phishing emails [2]. This information can include details of the registrant, the time of registration and renewals, the registrar, and nameservers used. With it, phishing attacks might be correctly identified as such, classified, attributed to previously observed threat actors based on similar modi operandi, and actively defended against, by sending takedown requests to the given abuse contacts.

The commonly used method of looking up domain registration data is the established WHOIS protocol, first standardized in 1982 [3]. Being a relatively old protocol, it comes with several shortcomings. These were addressed in the standardization of a new protocol: The Registration Data Access Protocol (RDAP) [4] [5].

However, RDAP is a comparatively new standard and presents some challenges: It is not available for all Top Level Domains (TLDs) and, if available, is often provided with restrictive terms of service. Registration data is often redacted due to privacy regulations. In this paper, we investigate how these issues affect the use of RDAP in the context of phishing attack investigations.

APWG.EU Technical Summit and Researchers Sync-Up 2023, Dublin, Ireland, June 21 & 22, 2023

hjl@csis.com (H. J. Lübbers)

0009-0007-3481-7616 (H. J. Lübbers)

2. On the Registration Data Access Protocol

RDAP is a Registration Data Directory Service (RDDS) standard that enables programmatic access to information about different internet resources, such as DNS domain names, DNS name servers, IP addresses, and Autonomous System Numbers (ASNs). The protocol was first standardized by the Internet Engineering Task Force in March 2015 and has since been extended [6].

RDAP defines a RESTful API to be provided by TLD registry operators and domain registrars. Accessible via HTTP over TLS, these APIs can be queried for registration details of, for example, DNS domain names, as shown in 1 [7] [5].

Listing 1: Example RDAP API request to query information on example.com

```
GET https://rdap.registry.example/v1/domain/example.com
```

The expected response, assuming that the RDAP service has information about the queried object, is given in JSON following a schema defined by RFC 9083 [8]. Thus, the response should be given in a machine-readable format under the mime type `application/rdap+json`.

With RFC 9224, the IETF standardized a way to identify the authoritative RDAP service for the internet resource that should hold valid registration information on a domain name, IP address, or ASN. The so-called "bootstrap services" associate TLDs, IP ranges, and ASN ranges with their respective authoritative RDAP services [9]. Bootstrap services are provided by the Internet Assigned Numbers Authority (IANA) [10].

Some registry operators do not hold all of the required domain name registration data for all their registered domain names ("thin" registries). Instead, they direct RDDS queries to the sponsoring registrar through which the domain name was registered [11]. In the case of RDAP, this is done in the link section of the RDAP response, in which a link to the RDAP API endpoint of the sponsoring registrar is given with the relation-type `rel` "related" and the (MIME-)`type` "application/rdap+json".

With its properties of machine-readability, transport-encryption, and internationalization, RDAP is a significant improvement over its RDDS predecessor, the WHOIS protocol. WHOIS does not define structured responses and instead returns unstandardized plain text, is not transport-encrypted, and does not standardize handling of encodings other than ASCII, which is a problem for languages with non-ASCII character sets [4].

3. Domain name registration data in the fight against phishing

Domain name registration data, accessed either via WHOIS or RDAP, can be applied in the fight against phishing for multiple purposes: The age of a domain name, as given by its registration date and subsequent renewal dates, is often used as an indicator of its trustworthiness; with older domains being considered more trustworthy than more recently registered ones. This practice has led to some threat actors waiting for a while between registering a domain name and starting to use it for malicious purposes, to "age" their domain names and thereby giving them a higher chance of evading detection [12] [2].

If a domain name is confirmed to be involved in a phishing attack, its age can be used to differentiate between domain names that were registered solely for this malicious purpose and benign domain names merely pointing to a host that was compromised by the threat actors [13]. A third option to be considered is the misuse of legitimate file- or web-hosting services for phishing purposes. This classification is important, as the counter-actions taken by security researchers differ based on the type of phishing at hand.

While the registrant data provided by threat actors is often falsified, it can be used to cluster domains that were registered in bulk, and thereby help to detect new phishing domains [14] [15]. This assumes that the registrant data is publicly available, which often is not the case, as discussed in section 5.4.2 "Data Redaction". In the absence of registrant data, other information on a domain name's registration process, like the sponsoring registrar or reseller used, can help to cluster attacks through similar modi operandi, and potentially attribute it to the same threat actor, often connected with other data.

Finally, the domain name registration data should contain the abuse contacts of the sponsoring registrar. If a domain name is deemed to be registered with malicious intent, security researchers can report it to the registrar, which should have a procedure in place to react to these reports and suspend the domain name [15].

4. Methodology

To understand whether RDAP can be used to gather domain name registration data to be subsequently employed in the fight against phishing as described in 3, we built a software system that analyzes the availability and data quality of domain name registration data and tested it against a realistic workload of domain names suspected to be involved in phishing.

4.1. The "rdapper" software system

The purpose of the "rdapper" software system was to correctly process requests for registration data of domain names following the RDAP standard, temporarily store results in a database for caching purposes, and schedule the requests to RDAP services in order to comply with the providers' terms of service.

Key operations of the system were configured to send telemetry data to a monitoring and visualization system. This allowed us to oversee the operation of the system and extract metrics after the experiment had concluded.

The RDAP standard, which is heavily based on well established web-standards and technologies, makes it straightforward to implement an RDAP client that identifies the authoritative RDAP service for a given internet resource [9], sends an HTTP request to the RESTful API [7] [5], and parses the JSON response [8]. But in this scenario, we would be ignoring the rate limits that are defined in the "Terms of Service" or "Acceptable Use" policies of the many RDAP services the client might connect to. To adhere to these rate limits, we built a scheduling system for our RDAP queries. It took into account the time of the last lookup to this RDAP host and possible back-off requests in the form of HTTP responses with status code 429 ("Too many requests") and their `Retry-After` response header values if they were defined by the server. This delay was individually configurable for each RDAP host. When building such a scheduling system, it is necessary to track queries per RDAP host, not per RDAP service, as some RDAP services for different TLDs are hosted on the same server and track RDAP queries across all services.

We chose a default delay of five minutes between queries to the same host based on the longest observed required delay when we began this project. After each run of the experiment, we identified the RDAP services

that had accumulated the longest delay between the request of domain registration data and the execution of the RDAP query. We then tried to find stated rate limits in the Terms of Service documents of these registry operators or registrars. If we could not find any public rate limit information, we would reach out to the RDAP service provider and ask them for safe rate limits for their RDAP API. Then, we would adjust the delays for those RDAP hosts accordingly.

The system identified itself to the RDAP service providers by sending request headers with a unique `User-Agent` and an email address in the `From` header, in order to give RDAP service providers the ability to contact us should our queries cause issues or breach any terms of service. [1] We did not rotate our clients' IP addresses or implemented any other attempts at evading any potential terms of service enforcement by the RDAP service providers.

4.2. Analyzed domain names

As part of this experiment, we analyzed 53,270 unique domain names. All domain names were labeled by CSIS Security Group's Cyber Intelligence Platform to be potentially involved in phishing attacks. This does not mean that all of the domains were confirmed to host phishing pages: they include (1) false positive reports, (2) domain names from URLs referred to in phishing emails that are not malicious, (3) domain names solely registered for phishing purposes, (4) domain names of compromised benign websites, and (5) domain names of popular file- or web-hosting services misused for phishing attacks.

Data from four individual days [2] was used to aggregate all domain names suspected to be involved in phishing over the course of the past 24 hours (relative to the individual day). The TLD distribution in the test dataset is shown in Table 3. They were analyzed by the "rdapper" system on the following days, again over the course of 24 hours. This was done to simulate a realistic workload of a system that processes RDAP requests for an anti-phishing system.

5. Four days of phishing: The results

We analyzed the RDAP availability and, if available, the results of 53,270 unique domains that were suspected to be involved in phishing attacks over the four selected days. In the following, we highlight and comment on the challenges that we observed related to the use of RDAP to fight phishing.

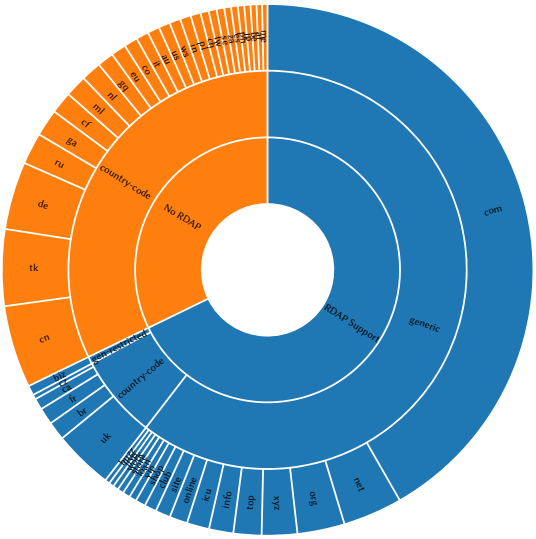

Figure 1: Visualization of the 50 most popular TLDs based on approximate numbers of registered domain names [16]

, grouped by TLD type and official RDAP support as of May 12th 2023.

5.1. RDAP coverage in the domain name ecosystem

gTLD registry operators and registrars are required by ICANN to provide an RDAP service as of the 26th of August 2019 [17]. ccTLD registry operators are not required to implement RDAP, but a number of them do support it, as shown in Figure 1.

At the time of writing, 1192 of all 1479 Top Level Domains (80.59%) have an authoritative RDAP service assigned in IANAs RDAP bootstrap file for Domain Name System registrations [10] [18]. In addition to these officially announced RDAP services, we identified seven RDAP services for ccTLDs which were not published in the IANA bootstrap file. Some of these services are in a testing phase and state this as the reason for their absence from the bootstrap file. These unofficial RDAP services bring the total of RDAP-supporting ccTLDs up to 34, or 11% of all ccTLDs.

These percentages show how far registry operators have come in adopting RDAP. But for practitioners, these numbers need to be adjusted by the distribution of domain names registered across the TLDs – or even better, by the distribution of domains of interest across TLDs; in our case these are domain names suspected of being connected to phishing attacks.

Using approximate numbers of registered domains for each TLD based on domainnamestats.com data accessed

[1] We did not receive any emails.

[2] The data collection days were May 10th 2023, May 22rd 2023, May 31st 2023, and June 5th 2023, with the processing days being the following day respectively.

Table 1
RDAP coverage for active (assigned) TLDs according to the IANAs RDAP bootstrap file for Domain Name System registrations by type as of May 11, 2023.

Type of TLD	Active TLDs	Official RDAP support	RDAP Support Percentage
Generic & generic-restricted TLDs	1155	1155	100%
Sponsored & infrastructure TLDs	15	10	66.67%
Country code TLDs	309	27	8.73%
Total	1479	1192	80.59%

on May 12th 2023 resulted in a calculated coverage of 67.45% of registered domain name having an official authoritative RDAP service assigned to them [16]. This coverage can be expected when the distribution of TLDs in a workload is similar to the distribution of all registered domain names across TLDs.

For our workload of domain names suspected to be involved in phishing, the distribution across TLDs differs slightly from the general population. In practice, and including the seven unofficial RDAP services, we reached a coverage of 78.86% of our domain name test dataset, the TLD distribution of which can be seen in Table 3.

5.2. RDAP service availability

Depending on the setup as a thin or thick registry, the registration data lookup for one domain name might involve RDAP queries to one or two RDAP services run by a registry operator or a registrar. Generic and generic-restricted TLD registry operators and registrars are required by ICANN to operate RDAP services [17]. But these services are no profit centers and are not "mission-critical" for most paying customers of these organizations. During our experiment, we saw 399 authoritative RDAP services run by registry operators and 229 RDAP services run by registrars, which we were pointed to by RDAP responses of registry operators.

We observed two RDAP services run by registry operators that continuously returned HTTP 500 ("Internal server error") responses. One of those services also started to serve an expired TLS certificate for around five days, and returned to serving HTTP 500 responses afterwards. Both RDAP services have since been fixed.

Of the 229 RDAP services run by registrars, we observed eight that served an invalid or expired TLS certificate or that were only available via HTTP without TLS on port 80.

25 registrar-run RDAP services were consistently unavailable, meaning that they never served valid RDAP responses. If we only contacted these services on two distinct days or less, we manually confirmed the unavailability of the services after the experiment phase had concluded, before categorizing them as consistently unavailable. Through this method, we excluded six registrar-run services from this list.

One registrar-run RDAP service was served behind a bot detection service of a content delivery network provider, making programmatic access impossible.

One registrar had restricted access to the individual domain name lookup endpoint, an example of which is shown in 1. When notified about this, they pointed us to their RDAP endpoint to search for domains instead. This is not a viable solution, as the thin registry RDAP service continues to re-redirect queries to the RDAP query endpoint for individual domain lookups.

5.3. RDAP service rate limiting

As described in Section 4.1, scheduling RDAP queries to comply with the terms of service of the RDAP services and, in particular, the rate limits is a crucial aspect when developing an RDAP client to operate at a certain scale. We found that our default delay of 300 seconds between requests generally seemed to work, and for the most part, did not result in back-off responses or even permanent IP blocks by RDAP service providers.

We identified rate limits either in public terms of service documents or in other parts of the the web presence of eight RDAP service providers. One of those has since removed the document, but the rate limits of the remaining seven RDAP service providers are listed in Table 4.

We contacted eleven RDAP service providers, prioritizing those who had accumulated the longest processing delays during our experiment. Two of those, both gTLD registry operators, refused to share the rate limits they enforce. Four shared their rate limits after being contacted, with the highest rate limit being one query per second and the lowest five queries per minute. In four cases, we never got an answer to our contact attempts via contact forms on the organization's websites or support email addresses. In one case, we did not manage to reach people with knowledge about the RDAP service of that organization.

The lowest rate limit we observed was determined through experimentation, as the registrar did not react to our contact attempts: Their RDAP service was configured to allow just one query per hour per IP, before responding with an HTTP 429 "Too many requests" response.

Table 2
RDAP service availability. One host may host multiple RDAP APIs for different TLDs.

Type of RDAP service	Observed hosts	Invalid TLS certificates	Unavailable or invalid API	Total	RDAP queries
Registry operator	399	0	2	2 (0.5%)	46 (0.01% of 42,014)
Registrar	229	8	25	33 (11%)	1101 (4.3% of 25,593)
Total	628	8	27	35 (5.6%)	1147 (1.7% of 67,607)

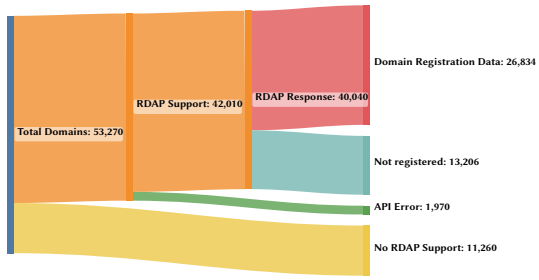

Figure 2: Visualization of RDAP availability and results for domains in the test set.

5.4. RDAP responses

In total, we sent 67,607 queries to RDAP services. This is because 25,873 (48.57%) of all domain names in the test dataset caused us to make two or more queries, the first to the TLD registry operator and the following ones to the sponsoring registrars that we were pointed to in the RDAP response of the registry operator. We had wrongly assumed that only thin WHOIS registries would include links to the RDAP services of sponsoring registrars, but this was incorrect: Some thick WHOIS registries, like CentralNic for the .xyz TLD, include links to the RDAP service of the sponsoring registrars, too.

In 13 cases (six .org and seven .info domain names) we found "related" links to a third RDAP service in the response of the second RDAP service. None of the RDAP services behind these links responded to our queries.

Of the 42,010 domain names with RDAP support, 13,206 (31.44%) were not registered, according to the received RDAP responses. This might be an artifact of the test dataset, which includes domain names that were parsed from spam emails and SMS. As the goal of this process is to extract any potential link from the messages, some of the extracted domain names, while being technically valid domain names, might not actually be registered. It might also be caused by the fact that we were querying the registration data of these domains 24 hours after they were reported, and by that time, they might have already been suspended by the registry operator or sponsoring registrar. However, in these cases, the status of the domain name is usually changed to "server hold" if the action was taken by the registry operator, and "client hold" if the action was taken by the registrar [19]. We observed 1,364 domain names with a "server hold" status, 2,447 domain names with a "client hold" status, and 437 domain names with both statuses, in total 7.97%. [3] Another explanation could be that some RDAP services might have a significant delay between the registration of a domain name and the update of the registration data database, on which the Registration Data Directory Services rely. We investigate this option in the following section.

5.4.1. Data Freshness

46,034 of the recorded RDAP responses include an event called "last update of RDAP database", which, according to the ICANN RDAP Response Profile specification, must contain "a value equal to the timestamp when the RDAP database was last updated" [20].

We compared this self-reported timestamp to the time of our system storing the RDAP result. This timestamp is recorded just after the HTTP response of the RDAP service has been received. Because of that, and to account for minor time-keeping offsets between our database server and the RDAP servers, we allowed for a 10-second buffer time.

866 RDAP database update timestamps were excluded because they lay significantly in the future without specifying a time zone. 1,220 timestamps were excluded because the RDAP service was likely implemented incorrectly, as the supposed timestamp of the RDAP database update always matched the "last changed" event, stating when the information about the object was last changed, or the "registration" event of the domain name [8].

36,260 RDAP database update timestamps (82.5% of the available and correct timestamps) were within 30 seconds of our request, which indicates that they configured their RDAP service with a "live" registration data database setup. In this case, the time of the request is taken as the time of the last RDAP database update.

[3] This includes variations like "client_hold" and "ServerHold"

5.4.2. Data Redaction

The General Data Protection Regulation of the European Union (GDPR), which took effect in 2018, required ICANN to update its policies for gTLD registry operators, in order to enable them to stay compliant both with the law in these jurisdictions and their responsibilities as registry operators [21]. As the GDPR also applies to data controllers not based in the EU who are storing data of EU citizens, this also affected non-EU registry operators and registrars [22]. ICANN agreed on a "Temporary Specification for gTLD Registration Data" specifying which registration data must be redacted by gTLD registry operators and registrars [21].

As it is not always possible to programmatically distinguish between redacted, removed, and un-available registration data, we check for the existence of markers required by ICANNs RDAP Response Profile specification [20] to indicate truncation or redaction of entity objects.[4]

17,716 RDAP responses for 16,546 distinct domain names contained at least one entity that was truncated or redacted. This constitutes 61.66% of the domain names for which we got successful RDAP responses. Not all remaining RDAP responses necessarily contain unobfuscated registrant information, as some RDAP service providers do not declare the information as redacted, even if it is obfuscated.

Apart from attempts to cluster domain names registered with malicious intent via the registrant information, another use case for RDAP in the fight against phishing is the identification of abuse contacts of the sponsoring registrar. For 19,250 unique domains, we got a thick RDAP response containing a contact entity with an "abuse" role. This constitutes 71.72% of all domains for which we got successful thick RDAP responses.

5.4.3. Schema adherence

We focused on parsing three sections of the RDAP response: the events, to gain information on the age of the domain name, the related entities, to identify the abuse contact, and links to follow the potential "related" links to the RDAP service of the sponsoring registrar.

We covered the correctness and plausibility of the "Last update of RDAP database" event in section 5.4.1.

1,503 RDAP responses did not use jCard, a JSON representation [23] for the vCard standard [6], which the RDAP response standard prescribes [8]. These responses were sent by two registrar-run RDAP services. Instead, entity objects were represented using a similar schema, which might more closely represent the underlying data structures of their RDDS backend. An anonymized example of such an entity object representation is shown in Appendix Listing 2.

Another two registrar-run RDAP services returned HTTP responses in which JSON arrays, as used in the RDAP standard to list the events, were instead represented by JSON objects with the array indexes as names and the array elements as the respective values [8]. These two registrars were only responsible for eight domain names in our dataset.

6. Conclusion

Based on the collected data, we conclude that RDAP can be a useful source of registration data of domain names for the fight against phishing, but it cannot currently be the only source for this type of information. This is mainly due to the slow adoption among ccTLD registry operators.

The RDAP standard itself is well-suited to cover the registration data needs of security researchers, and we observed acceptable adherence to the standard among RDAP service providers, also taking into account data freshness.

Restrictive terms of service or acceptable use policies of RDAP service providers present a challenge to the adoption of RDAP by security researchers. These policies were, in many cases, seemingly copied from the WHOIS terms of service and do not reflect the nature of RDAP as a machine-readable API that allows for more fine-grained access control compared to WHOIS. This often results in very aggressive, IP-based rate limits for RDAP queries, which hinders the fight against phishing, as threat actors are often registering domain names in bulk.

Because of privacy regulations, data on the actual registrant of the domain name is commonly not available to un-authenticated RDAP clients. This is not an RDAP-specific issue and also affects other RDDS standards. But in contrast to WHOIS, the RDAP standard enables RDAP service operators to implement delegated authorization mechanisms, like OAuth 2.0, using a number of centralized identity providers.

These providers could ensure that interested parties have a legitimate use case for accessing domain name registration data. Instead of having to prove the legitimacy of their registration data access request to each individual RDAP service provider, security researchers would just have to prove this to a smaller number of identity providers.

Because of RDAP's extensive use of established web standards and its resulting extendability, we think that it has the potential to become an important tool in the fight against phishing.

[4] Required in the specification is a remark of type "object truncated due to authorization", but we also included variations like "object redacted due to authorization" and "object redacted due to privacy laws".

References

[1] Anti-Phishing Working Group, Phishing activity trends report, 4th quarter 2022, 2023. URL: https://docs.apwg.org/reports/apwg_trends_report_q4_2022.pdf.

[2] A. Oest, Y. Safei, A. Doupé, G.-J. Ahn, B. Wardman, G. Warner, Inside a phisher's mind: Understanding the anti-phishing ecosystem through phishing kit analysis, in: 2018 APWG Symposium on Electronic Crime Research (eCrime), 2018, pp. 1–12. doi:10.1109/ECRIME.2018.8376206.

[3] K. Harrenstien, V. White, NICNAME/WHOIS, RFC 812, 1982. URL: https://www.rfc-editor.org/info/rfc812. doi:10.17487/RFC0812.

[4] L. Daigle, WHOIS Protocol Specification, RFC 3912, 2004. URL: https://www.rfc-editor.org/info/rfc3912. doi:10.17487/RFC3912.

[5] S. Hollenbeck, A. Newton, Registration Data Access Protocol (RDAP) Query Format, RFC 9082, 2021. URL: https://www.rfc-editor.org/info/rfc9082. doi:10.17487/RFC9082.

[6] A. Newton, S. Hollenbeck, JSON Responses for the Registration Data Access Protocol (RDAP), RFC 7483, 2015. URL: https://www.rfc-editor.org/info/rfc7483. doi:10.17487/RFC7483.

[7] A. Newton, B. Ellacott, N. Kong, HTTP Usage in the Registration Data Access Protocol (RDAP), RFC 7480, 2015. URL: https://www.rfc-editor.org/info/rfc7480. doi:10.17487/RFC7480.

[8] S. Hollenbeck, A. Newton, JSON Responses for the Registration Data Access Protocol (RDAP), RFC 9083, 2021. URL: https://www.rfc-editor.org/info/rfc9083. doi:10.17487/RFC9083.

[9] M. Blanchet, Finding the Authoritative Registration Data Access Protocol (RDAP) Service, RFC 9224, 2022. URL: https://www.rfc-editor.org/info/rfc9224. doi:10.17487/RFC9224.

[10] IANA, Bootstrap service registry for domain name space, 2023. URL: https://www.iana.org/assignments/rdap-dns/rdap-dns.xhtml, accessed May 11, 2023.

[11] S. Liu, I. Foster, S. Savage, G. M. Voelker, L. K. Saul, Who is. com? learning to parse whois records, in: Proceedings of the 2015 Internet Measurement Conference, 2015, pp. 369–380.

[12] G. Aaron, R. Rasmussen, Global phishing survey: Trends and domain name use in 2016, 2016. URL: https://docs.apwg.org/reports/APWG_Global_Phishing_Report_2015-2016.pdf.

[13] S. Maroofi, M. Korczyński, C. Hesselman, B. Ampeau, A. Duda, Comar: Classification of compromised versus maliciously registered domains, in: 2020 IEEE European Symposium on Security and Privacy (EuroS&P), 2020, pp. 607–623. doi:10.1109/EuroSP48549.2020.00045.

[14] T. Vissers, J. Spooren, P. Agten, D. Jumpertz, P. Janssen, M. Van Wesemael, F. Piessens, W. Joosen, L. Desmet, Exploring the ecosystem of malicious domain registrations in the .eu tld, in: M. Dacier, M. Bailey, M. Polychronakis, M. Antonakakis (Eds.), Research in Attacks, Intrusions, and Defenses, Springer International Publishing, Cham, 2017, pp. 472–493.

[15] G. Aaron, L. Chapin, D. Piscitello, C. Strutt, Phishing landscape 2021, 2021. URL: https://interisle.net/PhishingLandscape2021.pdf.

[16] Domain Name Stat, LLC, Domain name registrations, by tld, 2023. URL: https://domainnamestat.com/statistics/tld/others, accessed May 12, 2023.

[17] ICANN, Registration data access protocol timeline, 2018. URL: https://www.icann.org/resources/pages/rdap-background-2018-08-31-en, accessed May 11, 2023.

[18] IANA, Root zone database, 2023. URL: https://www.iana.org/domains/root/db, accessed May 11, 2023.

[19] S. Hollenbeck, Extensible Provisioning Protocol (EPP) Domain Name Mapping, RFC 5731, 2009. URL: https://www.rfc-editor.org/info/rfc5731. doi:10.17487/RFC5731.

[20] ICANN, RDAP Response Profile v2.1, 2019. URL: https://www.icann.org/en/system/files/files/rdap-response-profile-15feb19-en.pdf.

[21] ICANN, Temporary Specification for gTLD Registration Data, 2018. URL: https://www.icann.org/en/system/files/files/gtld-registration-data-temp-spec-17may18-en.pdf.

[22] European Commission, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016.

[23] P. Kewisch, jCard: The JSON Format for vCard, RFC 7095, 2014. URL: https://www.rfc-editor.org/info/rfc7095. doi:10.17487/RFC7095.

A. Appendix

Listing 2: Shortened and anonymized example a non-standard RDAP entity object representation

```
"entities": [
  {
    "objectClassName": "entity",
    "vcardArray": {
      "properties": [
        {
          "name": "FN",
          "value": {
            "stringValue": "Domain Administrator",
            "typeName": "text"
          }
        },
        {
          "name": "ADR",
          "value": {
            "components": [
              {
                "name": "street",
                "value": {
                  "values": [
                    {
                      "stringValue": "1337 Lowland Ave.",
                      "typeName": "text"
                    },
                    {
                      "stringValue": "PMB# 333",
                      "typeName": "text"
                    }
                  ],
                  "typeName": "text"
                }
              },
              {
                "name": "locality",
                "value": {
                  "values": [
                    {
                      "stringValue": "Example City",
                      "typeName": "text"
                    }
                  ],
                  "typeName": "text"
                }
              },
            ],
            "typeName": "text"
          }
        },
        {
          "name": "TEL",
          "parameters": {},
          "value": {
            "stringValue": "tel:+0.13371337",
            "typeName": "uri"
          }
        },
        {
          "name": "EMAIL",
          "value": {
            "stringValue": "example@example.com",
            "typeName": "text"
          }
        }
      ]
    },
    "roles": [
      "REGISTRANT"
    ]
  }
```

Table 3
Top 50 TLDs in the test data set and their (in some cases unofficial) RDAP support.

TLD	Number of domain names	Percentage of dataset	RDAP Support
.com	19943	37.44 %	RDAP Support
.top	3322	6.24 %	RDAP Support
.xyz	2348	4.41 %	RDAP Support
.net	1643	3.08 %	RDAP Support
.cn	1431	2.69 %	No RDAP
.org	1049	1.97 %	RDAP Support
.info	1019	1.91 %	RDAP Support
.ru	908	1.70 %	No RDAP
.tk	844	1.58 %	No RDAP
.online	748	1.40 %	RDAP Support
.ml	730	1.37 %	No RDAP
.br	724	1.36 %	RDAP Support
.site	677	1.27 %	RDAP Support
.de	603	1.13 %	Unofficial RDAP Support
.ga	601	1.13 %	No RDAP
.shop	567	1.06 %	RDAP Support
.cf	547	1.03 %	No RDAP
.pl	536	1.01 %	No RDAP
.uk	512	0.96 %	RDAP Support
.in	509	0.96 %	No RDAP
.live	482	0.90 %	RDAP Support
.gq	432	0.81 %	No RDAP
.co	421	0.79 %	No RDAP
.au	382	0.72 %	No RDAP
.stream	363	0.68 %	RDAP Support
.cc	360	0.68 %	RDAP Support
.us	348	0.65 %	Unofficial RDAP Support
.fr	347	0.65 %	RDAP Support
.icu	338	0.63 %	RDAP Support
.club	325	0.61 %	RDAP Support
.cyou	303	0.57 %	RDAP Support
.it	268	0.50 %	No RDAP
.eu	267	0.50 %	No RDAP
.id	246	0.46 %	RDAP Support
.bid	239	0.45 %	RDAP Support
.nl	233	0.44 %	No RDAP
.buzz	221	0.41 %	RDAP Support
.me	218	0.41 %	No RDAP
.click	212	0.40 %	RDAP Support
.space	206	0.39 %	RDAP Support
.za	184	0.35 %	No RDAP
.win	178	0.33 %	RDAP Support
.ca	177	0.33 %	RDAP Support
.io	173	0.32 %	No RDAP
.store	171	0.32 %	RDAP Support
.asia	164	0.31 %	RDAP Support
.cloud	163	0.31 %	RDAP Support
.pw	163	0.31 %	RDAP Support
.biz	159	0.30 %	RDAP Support
.cl	157	0.29 %	No RDAP
277 other TLDs	3690	6.93 %	RDAP Support
142 other TLDs	2419	4.54 %	No RDAP
Total RDAP Support	42010	78.86 %	RDAP Support
Total No RDAP Support	11260	31.14 %	No RDAP

Table 4
Incomplete list of public rate limits of RDAP service providers for un-authenticated RDAP clients

RDAP host	Type	RDAP query rate limits	Normalized query delay	Source
centralnic.com	Registry operator	7,200/hour	0.5s	https://registrar-console.centralnic.com/pub/whois_guidance
godaddy.com	Registrar	100/hour	36s	https://img1.wsimg.com//Sitecore/3/B/GDR-RDAP-Access-Policy-0.2.pdf
isnic.is	Registry operator	50/30min	36s	https://www.isnic.is/en/rdap
nic.tatar	Registry operator	30/min	2s	https://domain.tatar/users/docs/WhoisTermsOfUse_en.php
nominet.uk	Registry operator	1000/day and 5/s	86.4s	https://media.nominet.uk/wp-content/uploads/2019/06/gTLD-Acceptable-Use-Policies-version-1.pdf
norid.no	Registry operator	300/day and 10/min	288s	https://teknisk.norid.no/en/integrere-mot-norid/rdap-tjenesten/
tucows.com	Registrar	1/min	60s	https://tucowsdomains.com/rdap/help/