October 18, 2018

To: Goran Marby, CEO, ICANN Cherine Chalaby, COB, ICANN Rod Rasmussen, Chairman, ICANN SSAC

From: Dave Jevans, on behalf of the Anti-Phishing Working Group Members and Board of Directors

Dear Sirs,

The AntiPhishing Working Group (APWG) is an international coalition of private industry, government and law-enforcement actors and NGO communities who focus on financial fraud and related cybercrime identification and mitigation.

The APWG membership continues to follow ICANN community's efforts to define an interim plan to ensure that the existing public display (disclosure) of domain name registration data complies with the impending European Union General Data Protection Regulation (GDPR). To assist ICANN community in this effort, the APWG has collaborated with the Messaging, Malware, and Mobile Anti Abuse Working Group (M3AAWG) to conduct a survey of cyber investigators to understand how ICANN's *Temporary Specification for gTLD Registration Data* has affected their access and usage of domain name registration information and their ability to mitigate abuse.

From our analysis of over 300 survey responses, we find that the changes to WHOIS access following ICANN's implementation of the EU GDPR, the Temporary Specification for gTLD Registration Data[1] ("Temp Spec", adopted in May 2018), is significantly impeding cyber applications and forensic investigations and allowing more harm to victims.

The policy has introduced delays to investigations and the reduced utility of public WHOIS data is a dire problem. Delays favor the attacker and criminal, who can claim victims or profit over longer windows of opportunity. The loss of timely and repeatable access to complete WHOIS data is impeding investigations of all kinds, from fraud activities such as phishing and ransomware, to the distribution of fake news and subversive political influence campaigns.

From the responses of cybercrime investigators and anti-abuse service providers, we find that implementation of ICANN's Temp Spec impedes cyber security investigations: specifically,

Cyber-investigations and mitigations are impeded because investigators are unable to access complete domain name registration data.

The mitigation or triage of cyber incidents cannot be accomplished in a timely manner.

WHOIS has become an unreliable or less meaningful source of threat intelligence.

Requests to access non-public WHOIS by legitimate investigators for legitimate purposes are routinely refused

[1] Temporary Specification for gTLD Registration Data, https://www.icann.org/resources/pages/gtld-registration-data-specs-en

Those who protect Internet resources are also making more coarse blocking or mitigation decisions in the absence of what was formerly reliable data.

The utility of WHOIS has been severely damaged.

The redaction of WHOIS data is excessive

The body of our report offers an analysis of the 327 survey responses from the combined APWG and M^{3}AAWG surveys.

Based on these findings, we encourage the ICANN organization and community to consider these recommendations during their ongoing deliberation of WHOIS policy:

Recommendation 1: There must be an accredited access mechanism, providing tiered or gated access to qualified security actors.

Recommendation 2: ICANN should *not* allow redaction of the contact data of legal entities.

Recommendation 3: ICANN should adopt a contact data access request specification that will ensure consistency across all accredited registrars and gTLD registries.

Recommendation 4: ICANN should ensure that the accredited access to redacted WHOIS data does not introduce delays in collecting or processing WHOIS data, and further, that the access not be encumbered by per request authorizations.

Recommendation 5: ICANN should reconsider the current redaction policy.

Recommendation 6: We ask that ICANN publish point of contact email addresses to provide investigators with an effective means of identifying domains associated with a victim or person of interest in an investigation.

We respectfully request that ICANN organization, community and board consider the attached survey report. The report is also published at the Anti-Phishing Working Group web site, http://www.apwg.org.

We recognize that ICANN is likely aware of several of these issues. We also realize that ICANN organization and Board of Directors are awaiting the Expedited Policy Development Process for answers to many issues; however, we believe that the ICANN Board of Directors and ICANN organization have the ability to update the Temp Spec to fix the problems that this survey and others have identified as most pressing or egregious while the EPDP work continues.

APWG or M3AAWG members welcome the opportunity to brief the EPDP panel, any ICANN community members, or ICANN organization on the survey or its results. We would welcome this opportunity to share, anecdotally, additional field experiences.

Thank you in advance for your consideration,

David Jevans

Dave Jevans,

Chairman, Anti-Phishing Working Group (apwg.org)

ICANN GDPR and WHOIS Users Survey

A Joint Survey by the

Anti-Phishing Working Group (APWG)

and the

Messaging, Malware and Mobile

Anti-Abuse Working Group (M^{3}AAWG)

Principal Investigator

David Piscitello, InterIsle Consulting Group / APWG

www.apwg.org

www.m3aawg.org

Table of Contents

Executive Summary

The Anti-Phishing Working Group (APWG) and the Messaging, Malware and Mobile Anti-Abuse Working Group (M^{3}AAWG) have collaborated to conduct a survey of cyber investigators and anti-abuse service providers to understand how ICANN's application of the European Union's General Data Protection Regulation (GDPR) to the distributed WHOIS service has affected their access and usage of domain name registration information and their ability to mitigate abuse.

From our analysis of over 300 survey responses, we find that the changes to WHOIS access following ICANN's implementation of the EU GDPR, the Temporary Specification for gTLD Registration Data[1] ("Temp Spec", adopted in May 2018), is significantly impeding cyber applications and forensic investigations and allowing more harm to victims.

The policy has introduced delays to investigations and the reduced utility of public WHOIS data is a dire problem. Delays favor the attacker and criminal, who can claim victims or profit over longer windows of opportunity while investigators struggle to identify perpetrators or strip them of their assets (i.e., domain names) with limited or no access to the data that had previously been obtained or derived from WHOIS data. The loss of timely and repeatable access to complete WHOIS data is impeding investigations of all kinds, from cybercrime activities such as phishing and ransomware, to the distribution of fake news and subversive political influence campaigns.

In their responses, cybercrime investigators and anti-abuse service providers overwhelming report that the implementation of ICANN's Temp Spec introduces the following impediments to cyber security investigations:

Cyber-investigations and mitigations are impeded because investigators are unable to access complete domain name registration data through public WHOIS services in (near) real-time, as they had before the implementation of the Temp Spec. The partial data that are available through the public WHOIS services after redaction are insufficient to investigate or to respond to incidents.

The mitigation or triage of cyber incidents cannot be accomplished in a timely manner. Specifically, the need to request access to the non-public data elements introduces, at a minimum, delays of days in circumstances where mitigation prior to the adoption of the Temp Spec was often accomplished in hours or one day. Such delays allow attacks to remain active longer. Extended windows of operation put more Internet users in harm's way.

WHOIS has become an unreliable or less meaningful source of threat intelligence. The WHOIS contact data that is most relevant to investigators and has evidentiary value to law enforcement and prosecutors is generally not available through public WHOIS services since the implementation of the Temp Spec. (Note: even fraudulently composed, pseudonymous, incomplete, or inaccurate data is

[1] Temporary Specification for gTLD Registration Data, https://www.icann.org/resources/pages/gtld-registration-data-specs-en

useful for assigning reputations or creating correlations; for instance, in tracing known perpetrators' latest criminal excursions in establishing spoof domain names.) Investigators have been using alternative data sources with mixed success.

Requests to access non-public WHOIS by legitimate investigators for legitimate purposes are routinely refused. Investigators indicate that the implementation of "WHOIS reveal" is largely not working. The Temp Spec is unspecific, implementation is not uniform, and the processes are poorly understood by investigators, domain name registrars, and domain name registries. The majority of survey responders report that investigators do not know how to request access to non-public WHOIS data. Registrars and registries disclose redacted WHOIS data at their individual discretion, often without reasonable justification.

Those who protect Internet resources are also making more coarse blocking or mitigation decisions in the absence of what was formerly reliable data. Network operators and protective service providers, in fact, are blocking entire Top-level Domains. Investigators report that in circumstances where they are unable to use non-public WHOIS data to make blocking decisions about individual domains, and where they cannot establish associations across (very large) sets of suspicious domain names, they are exercising an abundance of caution and blocking more aggressively.

The utility of WHOIS has been severely damaged. Four months after the Temp Spec implementation, an alarming 17% of responders claim that public WHOIS service is no longer useful or reliable, and 13% have ceased using WHOIS entirely.

The redaction of WHOIS data is excessive. Investigators do not believe that it is necessary to redact legal entity point of contact data or point of contact data for data subjects outside the EU to comply with the EU GDPR.

The body of this report offers an analysis of the 327 survey responses from the combined APWG and M^{3}AAWG surveys. Note that the surveys asked identical questions; however, the APWG survey included additional questions.

Based on these findings, we encourage the ICANN organization and community to consider these recommendations during their ongoing deliberation of WHOIS policy:

Recommendation 1: There must be an accredited access mechanism, providing tiered or gated access to qualified security actors. A unified access program is necessary to restore predictable, automatable, swift access that balances privacy with legitimate use under GDPR. The technical mechanism would be RDAP (Registration Data Access Protocol).

Recommendation 2: ICANN should *not* allow redaction of the contact data of legal entities. Other WHOIS operators, such as the Regional Internet Registries and some European ccTLDs (Country Code Top Level Domains), do not redact data as aggressively as the Temp Spec allows.

Recommendation 3: ICANN should adopt a contact data access request specification that will ensure consistency across all accredited registrars and gTLD registries. The policy should be specific regarding the legitimate uses for which a timely completion of response is appropriate. These should align with the legitimate uses as described in the GDPR. A clear definition of "timely" should be included in the policy. Approved access requests should accommodate repeated access. Further, the specification should be specific with respect to:

A) Format of request (for both forms of WHOIS services);
B) Identification of the information required to be set forth in the request;
C) Email addresses where requests can be sent;
D) Specification of the documentation required for authenticating request; and
E) Time limitation for a response to requests. We recommend that responses be processed in 24 hours or less.

Recommendation 4: ICANN should ensure that the accredited access to redacted WHOIS data does not introduce delays in collecting or processing WHOIS data, and further, that the access not be encumbered by per request authorizations: these simply do not scale for the volumes and purposes that investigators identified in their responses. We further ask that ICANN consider a framework wherein an accredited party be granted timely access and persistent access to complete WHOIS data.

Recommendation 5: ICANN should reconsider the current redaction policy. In the interest of serving data protection *and* legitimate needs to access complete WHOIS data, we urge ICANN to consider using secure hashes rather than redacting data. We call your attention to the proposal, *Public WHOIS Attributes, Securely Hashed (WhASH): Hashing Point of Contact Details in Public Domain Name WHOIS*, submitted 4 June 2018, by APWG's Board of Directors to ICANN CEO and Chairman of the Board[2].

Recommendation 6: We ask that ICANN publish point of contact email addresses to provide investigators with an effective means of identifying domains associated with a victim or person of interest in an investigation. Some European ccTLDs and the RIRs (Regional Internet Registries) are publishing email addresses in their public WHOIS. Consistency across WHOIS services will facilitate investigations across identifier systems.

[2] https://www.icann.org/en/system/files/correspondence/jevans-to-marby-et-al-04jun18-en.pdf

Analysis of Responses to Survey Questions

The questions that comprised this survey were prepared by M^{3}AAWG and APWG members and their Boards of Directors members. Our survey concentrated on cybersecurity practitioners, which include personnel responsible for maintaining protective services and products, personnel responsible for directly defending the network of their employers, and academic researchers. Several questions we selected for our survey are similar to questions asked in a Law Enforcement Survey published in the Registration Directory Service (RDS)-WHOIS2 Review[3].

Some questions are demographic. No personal data were collected. The remaining questions allow responders to characterize their WHOIS usage and purpose, and to describe whether and how the implementation of ICANN's Temporary Specification for gTLD Registration Data (Temp Spec) has affected their usage.

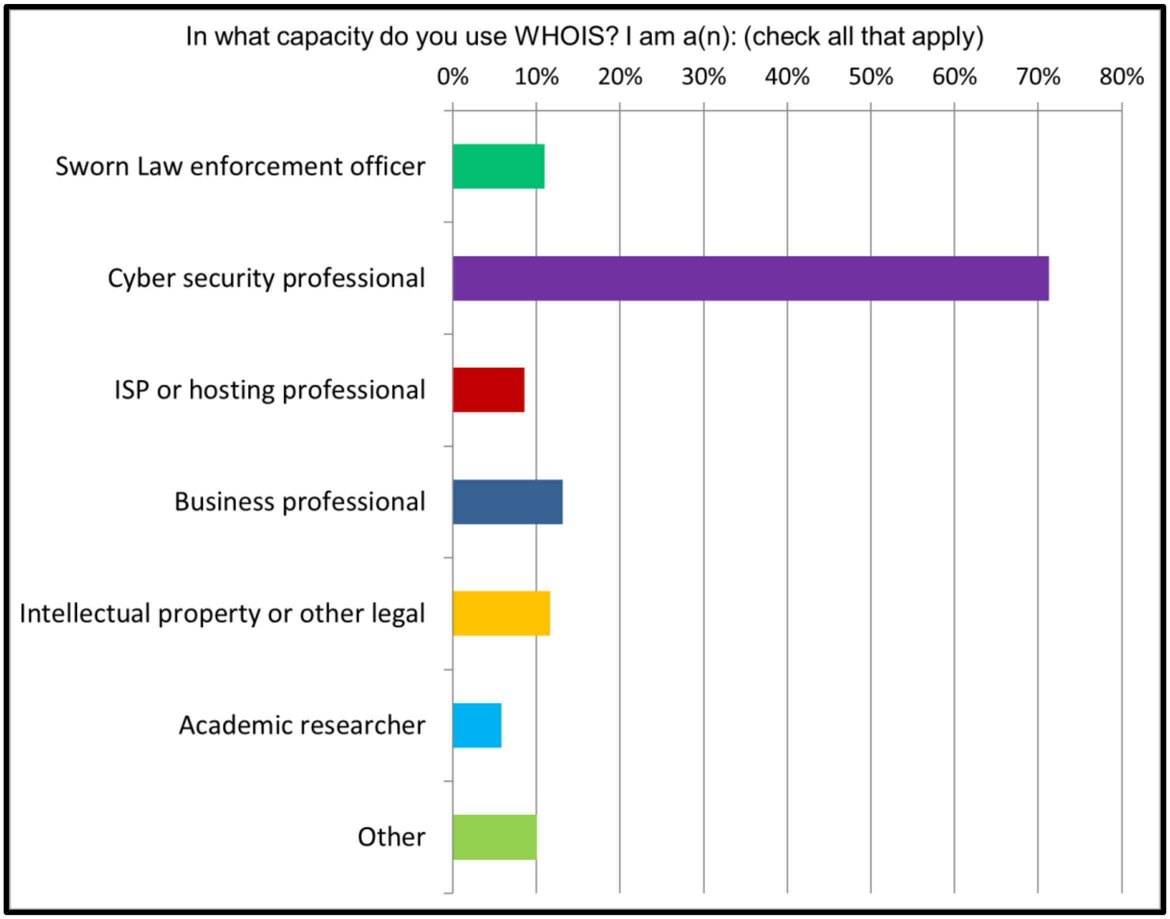
Our sample size is 327 responses. We estimate that the mailing lists used to announce the survey were delivered to a population of between 2500 and 5000 cybersecurity investigators. Assuming a Confidence Level of 95%, we calculate the Confidence interval to be 5.25%. We believe the results of this survey are statistically significant.

It is important to note that the survey attempts to understand how investigations have been affected since ICANN's application of the European Union's General Data Protection Regulation to the distributed WHOIS service. It was not intended to critique the Regulation itself.

[3] See Section 5.3.1, Law Enforcement Survey, pp 89-100, https://www.icann.org/public-comments/rds-whois2-review-2018-09-04-en

Question 1: In what capacity do you use WHOIS?

This survey was intended for security professionals who have roles or participate in cyber investigations, cyber incident response, or who provide threat intelligence data that are relevant to cybercrime or cyber investigations. The survey was distributed to APWG and M^{3}AAWG members through mailing list notifications. The survey was also posted to trust collaboration mailing lists used by security practitioners and law enforcement. APWG posted a public notice announcing the survey on the APWG web site. Responses indicate that the survey was completed by the intended audiences.

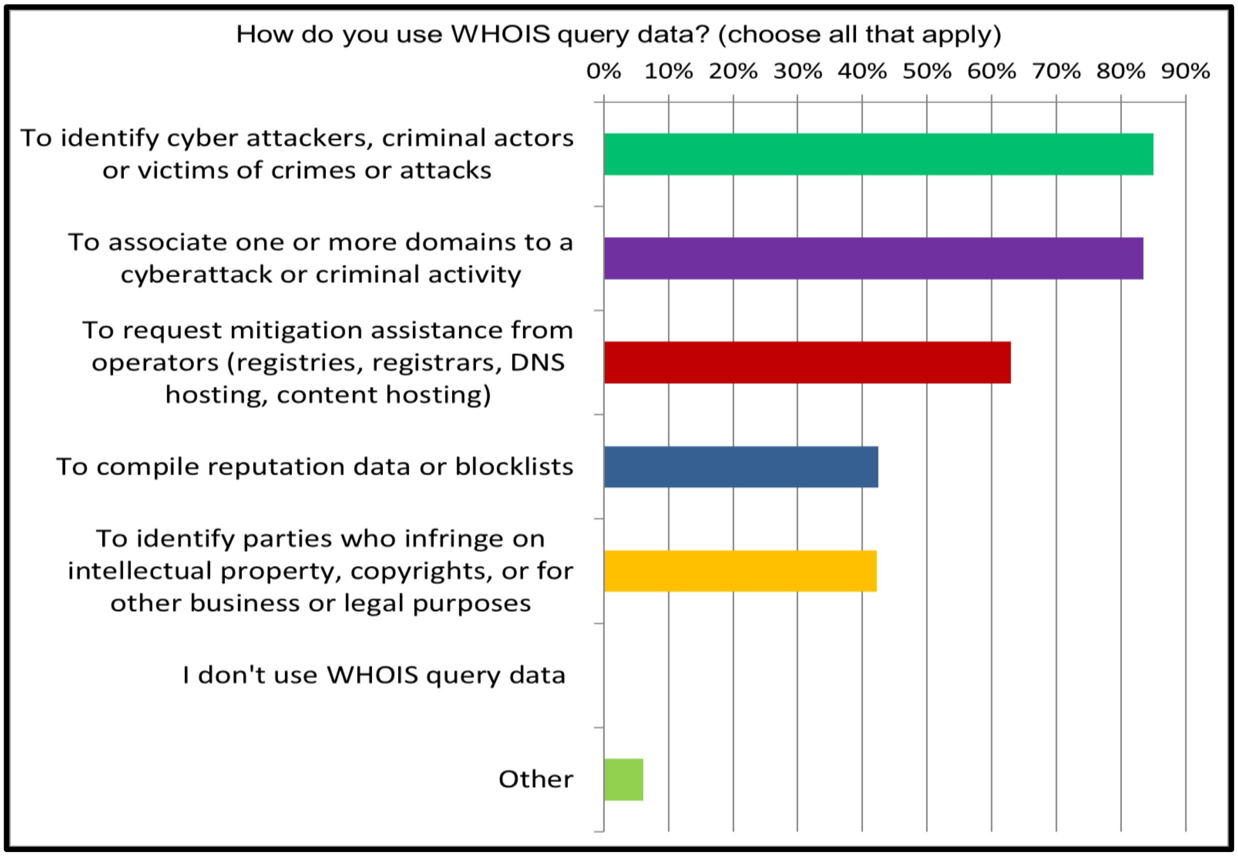


Note that the responses to this question were not mutually exclusive; for example, certain sworn law enforcement officers actively engage in cybersecurity investigations.

Question 2: How do you use WHOIS query data?

The responses to this question help us characterize the responder's (legitimate) purposes for querying WHOIS services. The answer choices are again not mutually exclusive.

The parties who responded to the survey are involved in all aspects of cyber investigations. They participate in initial detection and mitigation, victim identification and notification. They then continue to participate or collaborate in remediation, criminal attribution and prosecution. Eighty-five percent of investigators use WHOIS query data for attribution purposes; i.e., to identify attackers, criminal actors, or victims of crimes or attacks. Eighty-four percent of investigators use WHOIS query data to associate one or more domains to a cyberattack or criminal activity. In these cases, investigators search across multiple WHOIS query responses or large sets of collected responses using some identity data; e.g., an email address, as a search argument.

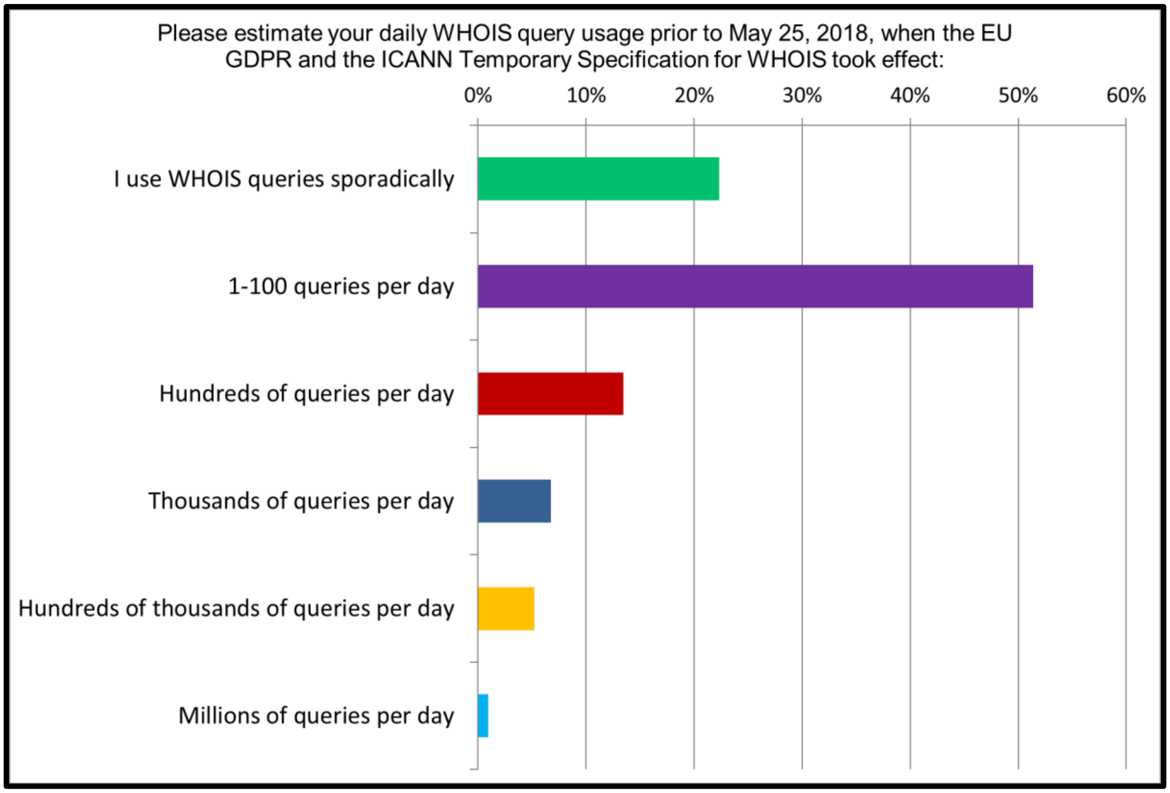


Sixty-three percent of investigators use WHOIS data to request mitigation assistance from operators. In such cases, operators may require identity data to take an action or comply to a court order; e.g., to associate the identified actor with content they host. Another 42% of investigators use WHOIS data to identify actors who infringe on intellectual property, and a similar percentage use WHOIS as part of the process of compiling domain name block lists.

Question 3: Please estimate your daily WHOIS query usage prior to May 25, 2018, when the EU GDPR and the ICANN Temporary Specification for WHOIS took effect.

The responses to this question show that WHOIS query usage prior to May 25, 2018 was diverse. The largest percent of responses (51%), 1-100 queries per day, merits some explanation. Cyber attackers or cyber criminals often register large numbers (hundreds, even thousands) of domain names over a period and then use them in a single "campaign." When investigators determine that a domain name is associated with an attack, abuse or criminal activity, they query WHOIS (often using programmatic methods and equally often, querying multiple domain names) to obtain the point of contact data, such as the email address submitted by the registrant. This, or generally any point of contact data, is then used as an *attribution* search argument to find other domains registered by the same party.

To effectively mitigate an attack, it is essential that investigators identify all or as many domains used by attackers, and this identification can be efficiently accomplished using point of contact data. Other WHOIS data such as name server, is useful, but not nearly as useful as point of contact data. It is important to recognize that the point of contact data that are redacted are also data that are commonly obtained from email, websites or social media; again, as an example, an email that is found on a public social media page can be used in a search of domain names. Conversely, the registrant email address of a domain that is used for inauthentic news can be used to search through public social media posts or group pages to identify persons of interest in a global criminal conspiracy.



The responses that chose the answer, *I use WHOIS queries sporadically,* (22%), is also interesting. Some investigations start with a single WHOIS query that points investigators toward a direction to look next. For other investigations, for example investigations of certain phishing attacks, investigators may concentrate on a single domain, e.g., a compromised website that hosts a phishing page: here, the investigator is interested in obtaining the contact data for the victim. For investigations into such activities as inauthentic news or terrorist radicalization or recruitment sites, investigators may begin with a single domain and track a small number of domain names over time that share one or more point of contact datum in common.

Twelve percent of respondents also show that certain investigators are high volume users. They, and the automated scripts under their control, may make hundreds, thousands or millions of WHOIS queries daily. Such volume is necessary to identify all the domain names that are used for spam or botnet criminal infrastructures. Daily queries at this volume are essential, since these infrastructures often use large numbers of domain names or domain names generated daily using a domain generation algorithm (DGA). Researchers or abuse monitoring programs also query WHOIS at very high volumes daily so that they can conduct longitudinal studies of abuse. Lastly, certain high volume users are investigators who compile domain block lists that protect Internet users everywhere.

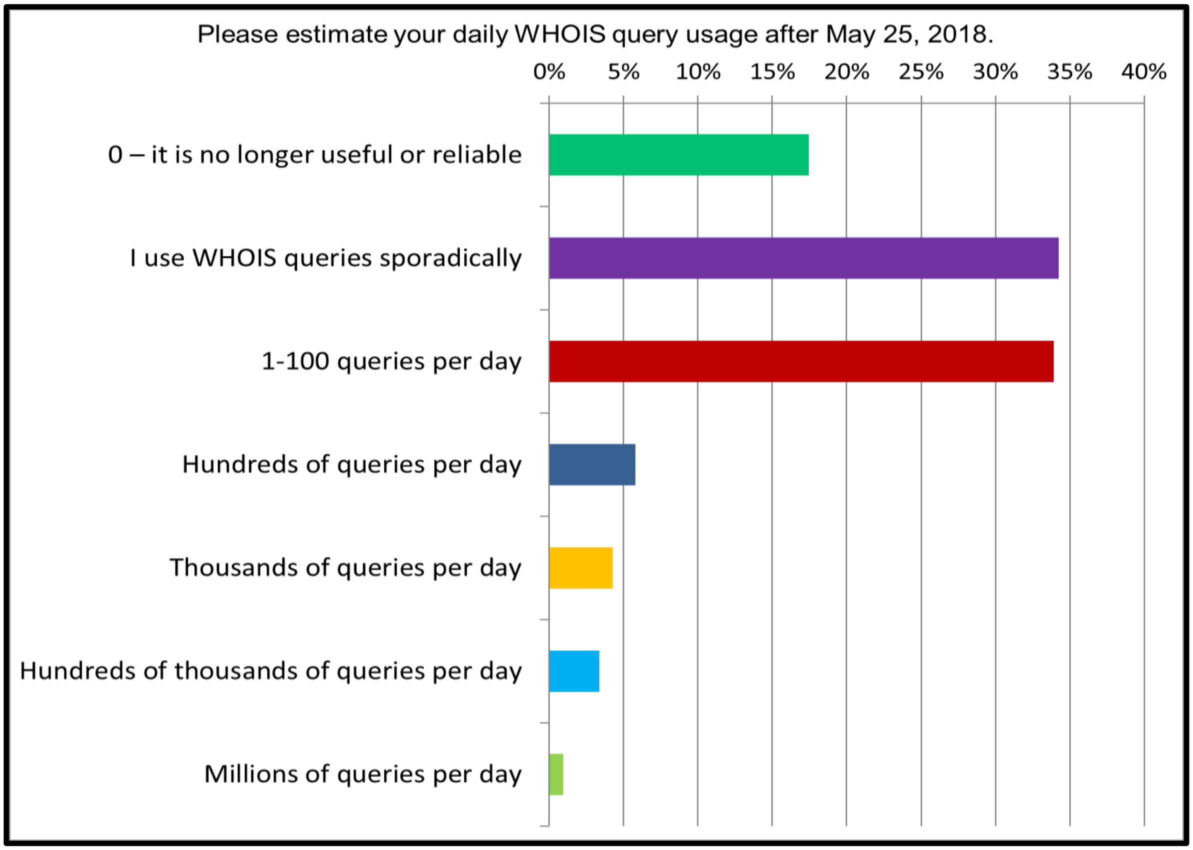
Question 4: Please estimate your daily WHOIS query usage after May 25, 2018.

We use questions 3 and 4 to observe whether WHOIS query behavior has changed since the adoption of the EU GDPR and implementation of the ICANN Temporary Specification for gTLD Registration Data on May 25.

Comparing responses to questions 3 and 4, we observe that their WHOIS query volume has decreased.

Four months after the Temp Spec implementation, an alarming 17% of responders claim that WHOIS services are no longer useful or reliable.

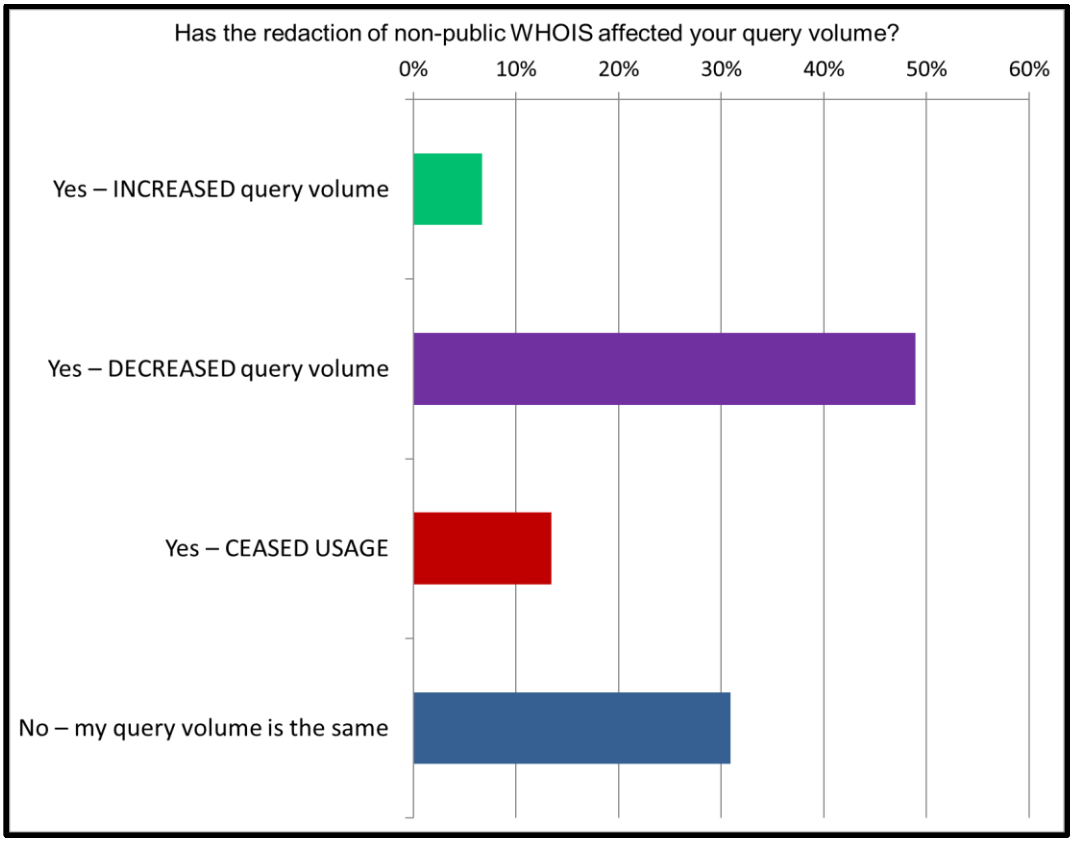
Use has fallen across the board. The percentage of investigators who use WHOIS just sporadically has increased from 22% to 34%. The percentage who made 1-100 queries per day has decreased *dramatically* from 51% to 34%. Those that made thousands of queries per day has nearly halved. High volume users have been similarly affected. Prior to May 25, 2018, responders who made hundreds, thousands, hundreds of thousands or millions of WHOIS queries daily cumulatively represented 26% of the responses. With the Temp Spec implemented, this cumulative value dropped to 15%.



Subsequent questions help us better understand the reasons behind the decreases.

Question 5: Has the redaction of non-public WHOIS affected your query volume?

Redaction of non-public WHOIS data suppresses the display of data that can be used for *attribution* and for assigning trustworthiness or reputation. Here, we try to understand whether the redaction of non-public WHOIS has influenced WHOIS query volume change.



Forty-nine percent of respondents indicated that they have decreased WHOIS usage and an additional 13% have *ceased usage altogether* because of redaction. Investigators can no longer search for common attributes across large numbers of WHOIS records using point of contact data as search arguments. One submitted comment explains a typical use case:

"Some of the registrants of BEC [Business Email Compromise] domains which look visually similar to the genuine email fill in identity details in a consistent manner. This allows us to find one domain and then find others from the same miscreant — which gives the possibility of issuing warnings. Redaction of details means we have abandoned, for the time being, our academic research efforts in this area."

Other submitted comments corroborate our finding that redaction impedes or impairs cyber investigations:

"The ability to identify malicious domains registered by a known actor is no longer possible. Identifying one domain in the past allowed us to search for other domains registered by the same 'person'. This is no longer possible and greatly impairs our ability to find additional attack infrastructure."

"historically we could link information related to fake registrants, and/or organized malicious activity, based on the results of WHOIS. Some examples would show thousands of related domains registered at the same time/date and/or organization, and/or other relevant markers which could be used to detect criminal and malicious activity."

"Due to the redaction of registrant data, it has become massively more difficult, and in many cases impossible, to proactively identify phishing domains purchased in bulk by large-scale and well-organized phishing groups."

"The biggest impact has been to determine who has registered a criminal/fraudulent domain, and the ability to use that information to find other domains registered by the same actor. That devastates our ability to find all of the fraudulent domains registered by the same entity."

The limited data now available through public WHOIS has limited attribution value; investigators can use name server and techniques such as passive DNS to identify servers that host zone data associated with abuse domains, but investigators have less intelligence about the registrants of the domains used for abuse purposes than they had prior to May 25, 2018.

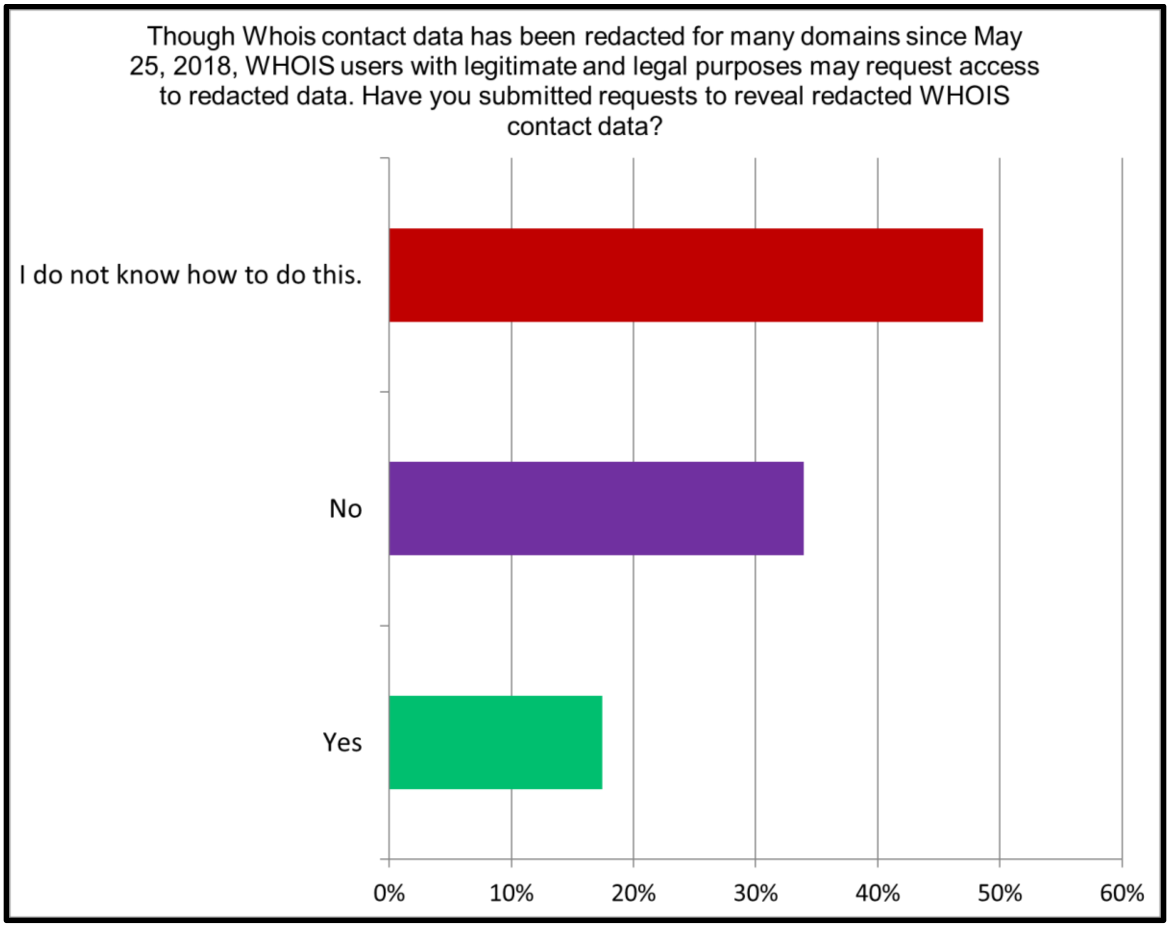
Redaction also has a profound effect on academic researchers and threat intelligence platforms. Research and threat intelligence require timely and repeated WHOIS access to provide data for longitudinal studies or to attain the necessary confidence levels to perform a preemptive or mitigation action: any study that attempts to monitor any attribution data such as an email address, patently obvious faked personal data, or point of contact data generally, can no longer be performed without requesting access to volumes of data from hundreds of registries or registrars. Several comments corroborate this observation:

"There are many threat actor groups who can be tracked using infrastructure…With information redacted, I can no longer find the domains that threat actors will use in the future. Historically, I've been able to find future infrastructure and block it 6-8 months prior to it being operational, which prevents compromises to networks from occurring. Now I can only block things when I see it already doing malicious things, a very reactive posture which ultimately means data loss from some victim…"

Even if access requests are satisfied, the studies are affected by the time lapse. Longitudinal studies often rely on daily or more frequent query patterns: if the data cannot be provided in near real time, then the study cannot be conducted in the manner intended or manner most likely to produce insights or critical findings.

Question 6: Though WHOIS contact data has been redacted for many domains since May 25, 2018, WHOIS users with legitimate and legal purposes may request access to redacted data. Have you submitted requests to reveal redacted WHOIS contact data?

The responses to question 6 corroborate other surveys or anecdotal data: few investigators are requesting access to redacted WHOIS data. Thirty-four percent indicate that they have not requested access. Forty-nine percent of the responders indicated that they do not know how to do this. We can only speculate the reasons based on our own experience with access requests: The Temporary Specification for gTLD Registration Data does not specify the mechanisms, which are left to the parties who collect and publish WHOIS data. Implementation across registrars and registries lacks uniformity. They define their own request mechanisms and formats and document or publish information about these mechanisms as they see fit.

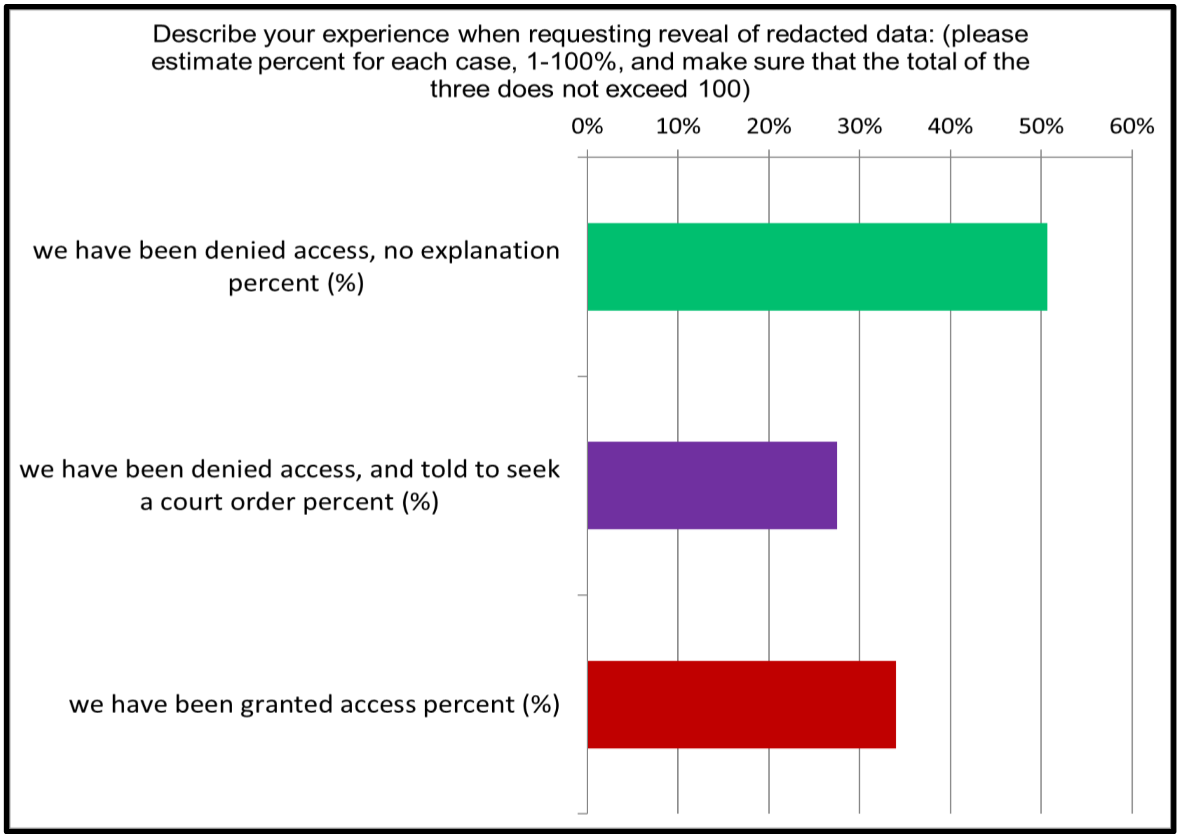


One responding investigator observed:

"It's hard to find out how to obtain non-public WHOIS from most registrars and registries and when we do find it, it seems to be different for each one and the huge majority of them don't provide WHOIS info."

Question 7: Describe your experience when requesting reveal of redacted data.

Seventy-nine percent of respondents indicate that they have either been denied requests for access out-of-hand (51%), or they have been asked to obtain a court order (28%). These outcomes call attention to the urgent need for an access program that accredits or credentials legitimate investigators and respects the needs for access as legitimate.



One comment demonstrates the challenges investigators face:

"Since June of 2018 we have submitted redacted WHOIS reveal requests for 20 malicious domains to registrars. 18 of the 20 requests have no response whatsoever from the registrar despite multiple requests. 1 sent the WHOIS data after six weeks and multiple asks. 1 other registrar sent this response after weeks and multiple requests:

> *'Dear Sir or Madam,*
>
> *We are struggling with spam and fake identities every day and therefore, due to data protection, have to identify the company / person, to which we send details of our customers. In this case we can not identify the sender and therefore ask you to send us an official writing, which you just can add as attachment in your next e-mail.'"*

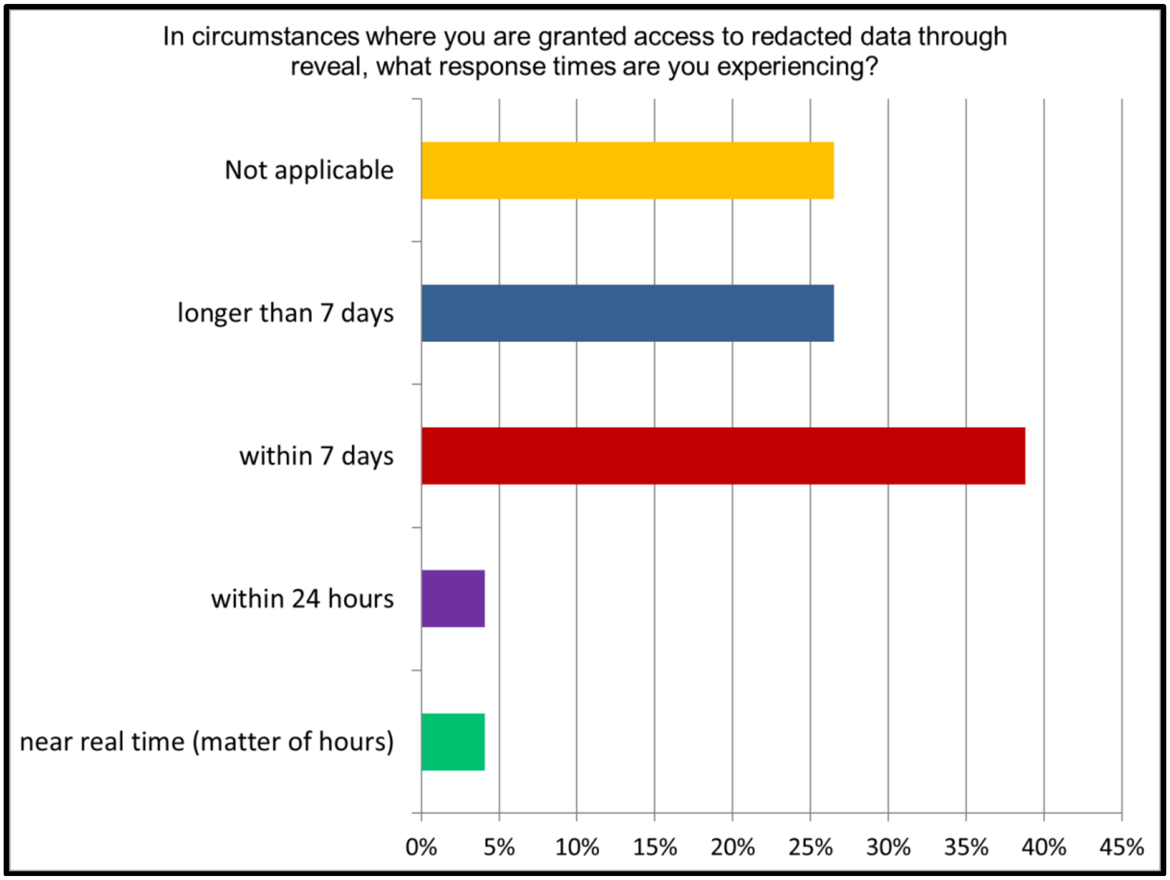
"Responses to redacted WHOIS data requests for phishing sites are not being honored."

Insistence on court orders underscores the extraordinary change away from timely access to access that renders WHOIS nearly useless. Obtaining a court order from a local jurisdiction is challenging in and of itself. Out of jurisdiction investigators must obtain court orders through protracted Mutual Legal Assistance Treaty (MLAT) processes that take months, at great expense, if they are able to obtain them at all. This is true even for law enforcement, when they can avail themselves of MLAT assistance procedures.

Question 8: In circumstances where you are granted access to redacted data through reveal, what response times are you experiencing?

Only 4% of respondents reported that they were granted access to redacted data in a matter of hours, and 4% received access within 24 hours. Only 39% were granted access within 7 days, and 27% waited more than 7 days. These percentages are deceivingly *low* since 27% responded *Not applicable*.

For context, it is imperative to compare the response times of days against an incident response objective of 2-4 hours, versus WHOIS query/response times which only take a matter of seconds. **Delayed criminal or victim identification prolongs attack windows, delays victim notification, and jeopardizes research that relies on timely access for longitudinal studies of threat data.**



For example, consider a scenario where an investigator identifies a phishing page at the URL http://www.example.com/bank/login.html. Prior to the Temp Spec implementation, the investigator would query WHOIS, identify example.com's registrant or the Web administrator, notify the victim of the phishing page, and request that the page be removed, in a matter of 1-4 hours. This same activity now requires a reveal request. While the request is processed, the phishing page remains active, exposing those Internet users who bank at the targeted financial institution to fraud and financial loss. Further, the domain name example.com may be added to blocklists, effectively victimizing the registrant twice. Lastly, the reputation of the Web hosting provider may be damaged.

Several comments explain how the inability to acquire data in near real time impedes attack mitigation and exposes Internet users to more harm or loss today than before May 25, 2018:

"In some cases we are investigating a malicious-looking domain and we don't know if it was owned by a criminal or by a compromised victim. WHOIS information in the past helped us decide if we were dealing with innocent people or criminals. This is putting innocent people at risk by making investigators assume they are criminals, when they are simply victims of compromises."

"A decade ago, phishing attacks happened over 72 hours. Today, they happen over, at most, an hour. Having WHOIS to identify additional and future attacks is mission critical to protecting hundreds of millions of people from identity theft and massive loss of personal information."

"We cannot quickly identify useful links in bad infrastructures or get an alert if a bad actor creates new domains for criminal activities."

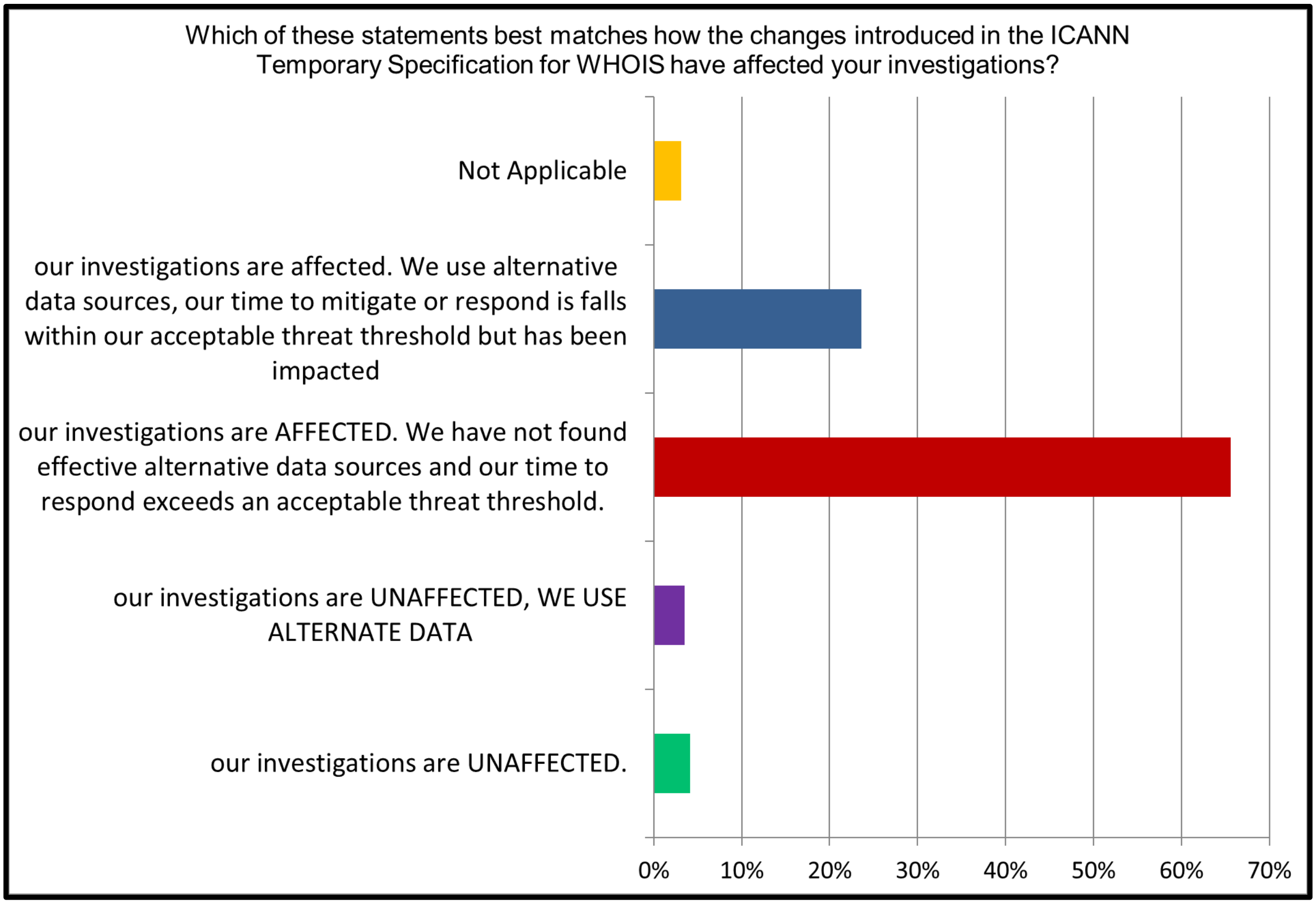
"Malicious actors create thousands of domains daily. Not being able to access registrant information severely limits our ability to identify bad actors before they harm significant number of users."

"A timely takedown of malware and phishing on a compromised system can even cancel out a cybercrime campaign. Killing notification ensures cybercrime continues unabated."

"Daily investigations for phishing takedowns are slowed because of no useful WHOIS data."

Question 9: Which of these statements best matches how the changes introduced in the ICANN Temporary Specification for WHOIS have affected your investigations?

Only 4% of responses indicate that investigations have been unaffected since the implementation of the Temp Spec. An alarming 66% of responders report that their investigations have been affected since May 25, 2018: they have not found effective alternative data sources and their time to respond exceeds the acceptable threat threshold they, their organizations or local regulations prescribe. Only 27% report that they have found alternative data, with varying effects to their acceptable threat thresholds: 3% have found alternative data and are able to maintain acceptable threat thresholds but 24% claim that their investigations are affected despite having found alternative data, perhaps because alternative data may not be as quickly collected as WHOIS prior to May 25, 2018, or that extracting or deriving data as useful as non-public WHOIS is time consuming.



Investigator responded that:

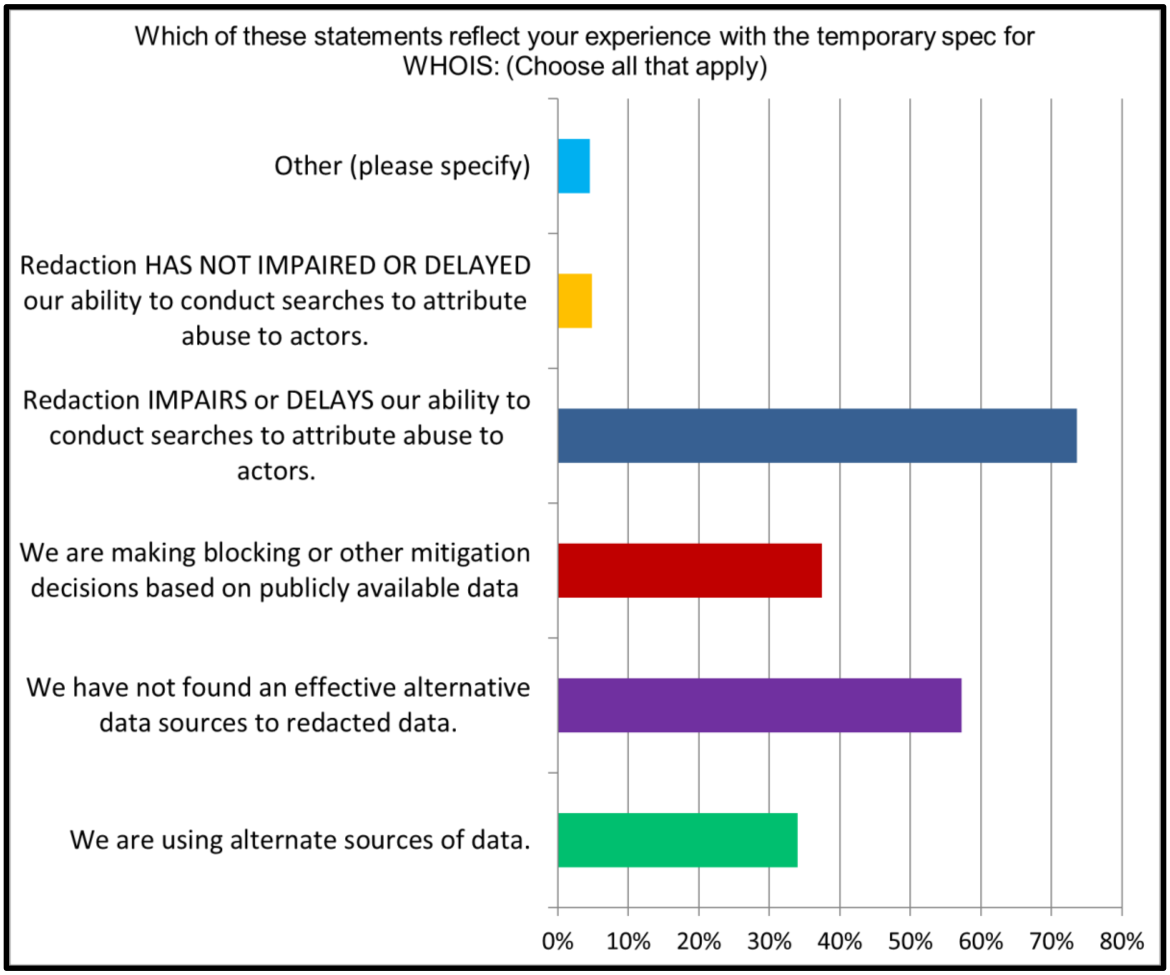
"In general having the information redacted has been [made] more difficult to connect entities together that may be involved in fraud. While we can still make connections with other data, it makes it much more difficult to establish who owns a domain at a particular point in time, especially after the redactions."

Question 10: Which of these statements reflect your experience with the temporary spec for WHOIS

Question 10 allows responders to share experiences while conducting investigations since the Temporary Specification for gTLD Registration Data has been implemented. Responders were asked to select all answers that apply.

Seventy-four percent of respondents indicated that redaction impairs or delays an investigator's ability to conduct searches to attribute abuse to actors. Only 5% indicate that redaction has not impaired or delayed their ability to conduct searches to attribute abuse to actors.

What measures have investigators taken to compensate for the loss of access to data that is redacted? Some have no answers (yet): 58% of responders indicate that they have been unable to find effective alternatives to the redacted WHOIS data. 34% have workarounds.



Thirty-eight percent of the respondents are making coarser blocking decisions; for example, blocking entire TLDs.

Several comments corroborate this more aggressive policy:

“Denied domain name WHOIS registrant info of bad actors so can’t use as search attributes in other data sets that point to criminal activity; bad actors win, consumers lose. Thanks ICANN. Without fulsome WHOIS info, we're protecting consumers by blocking domains if available info is in anyway suspicious = blocking a lot more domains.”

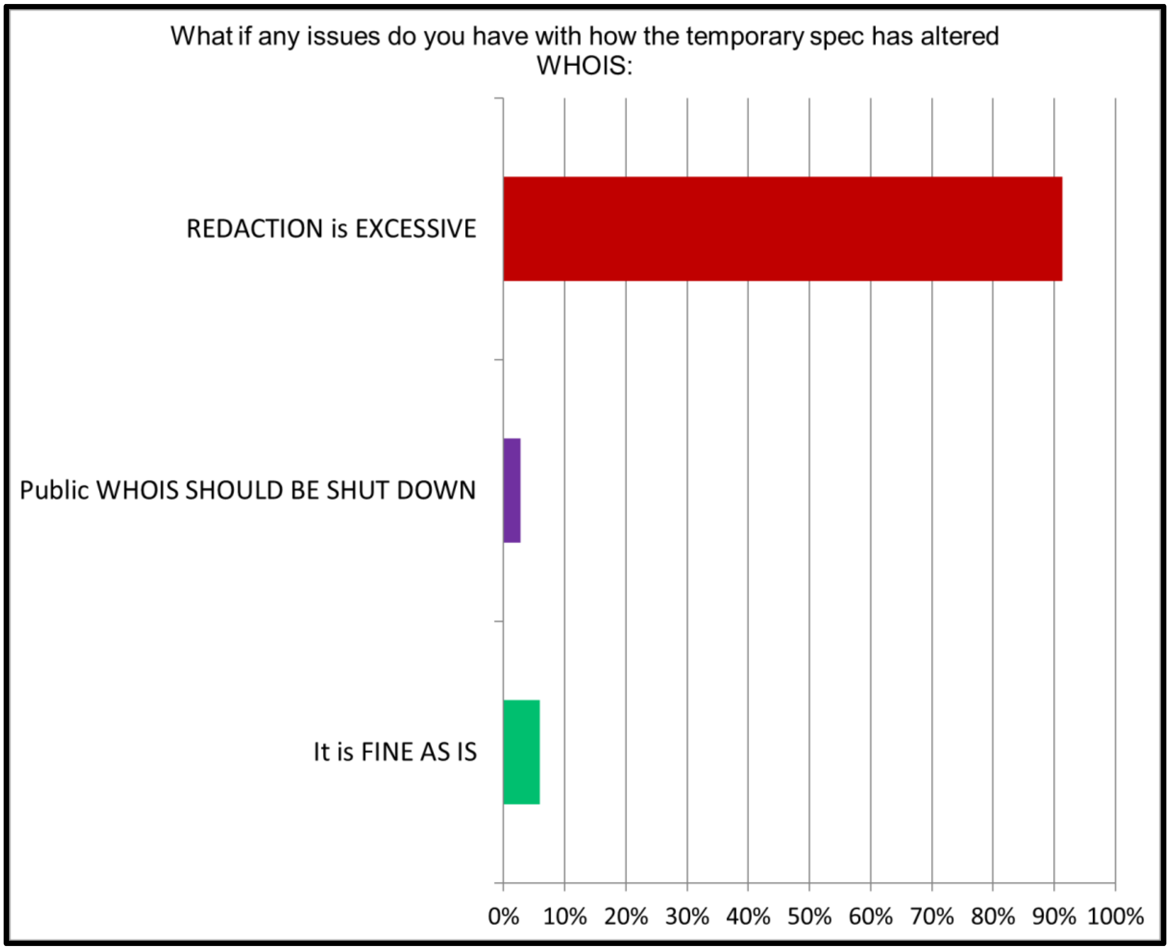
“In addition to more difficulty identifying malicious actors, redaction has increased the likelihood that innocent domains are impacted by overly aggressive enforcement (because relation to an innocent party cannot be verified).”

“We have blocklisted a domain and subsequently their hosting is canceled. They write us to revoke the complaint of spam against them. If we can't identify the owner of the domain and verify they are an innocent party, the complaint stands and they may not get their hosting back.”

“We calculate reputation over millions of domains. Anything that is not an automated process does not scale.”

Question 11: What if any issues do you have with how the Temporary Spec has altered WHOIS?

The overwhelming response to question 11 (91%) is that the Temp Spec redaction of data is excessive. Some 3% suggested that public WHOIS be shut down and 6% think that WHOIS is fine as is.

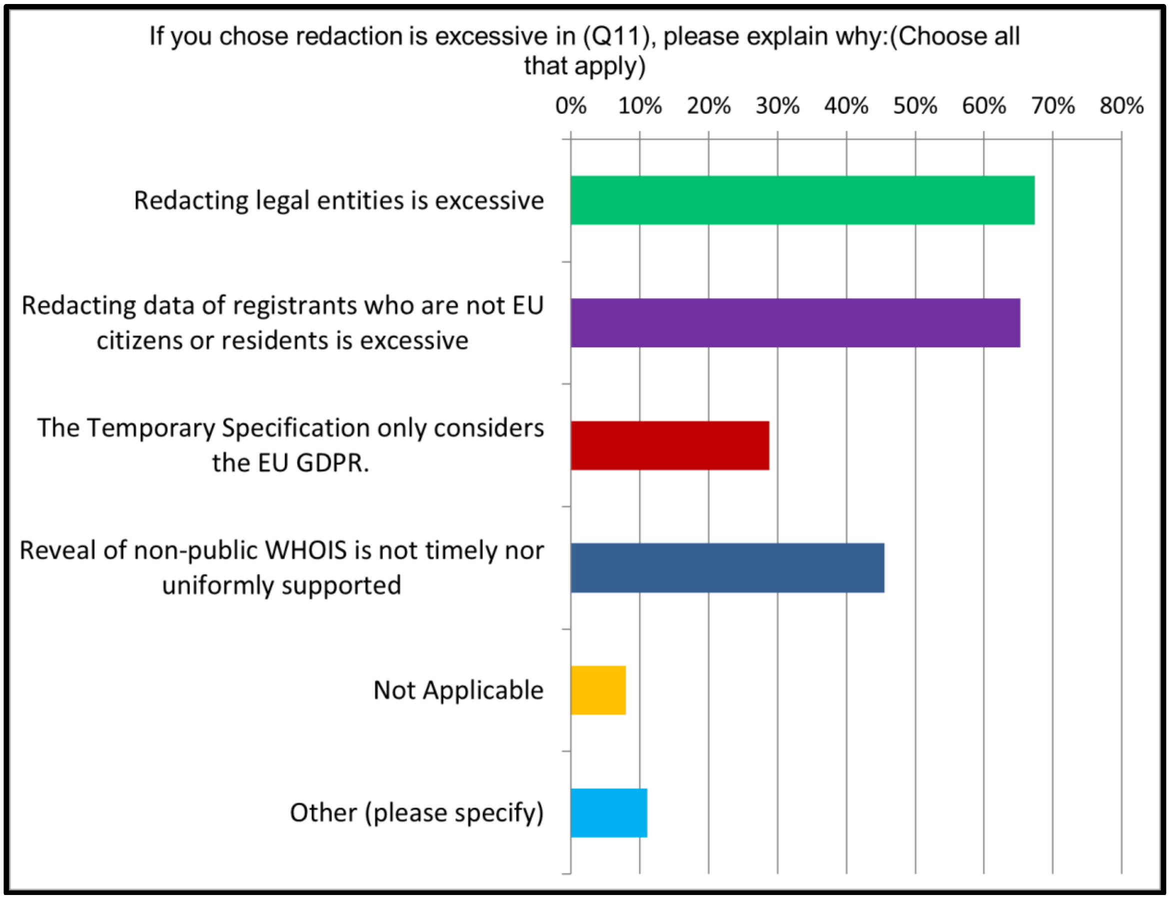


To better understand why investigators respond so uniformly against redaction, we asked investigators a follow-on question (12).

Question 12: If you chose redaction is excessive in (11), please explain why.

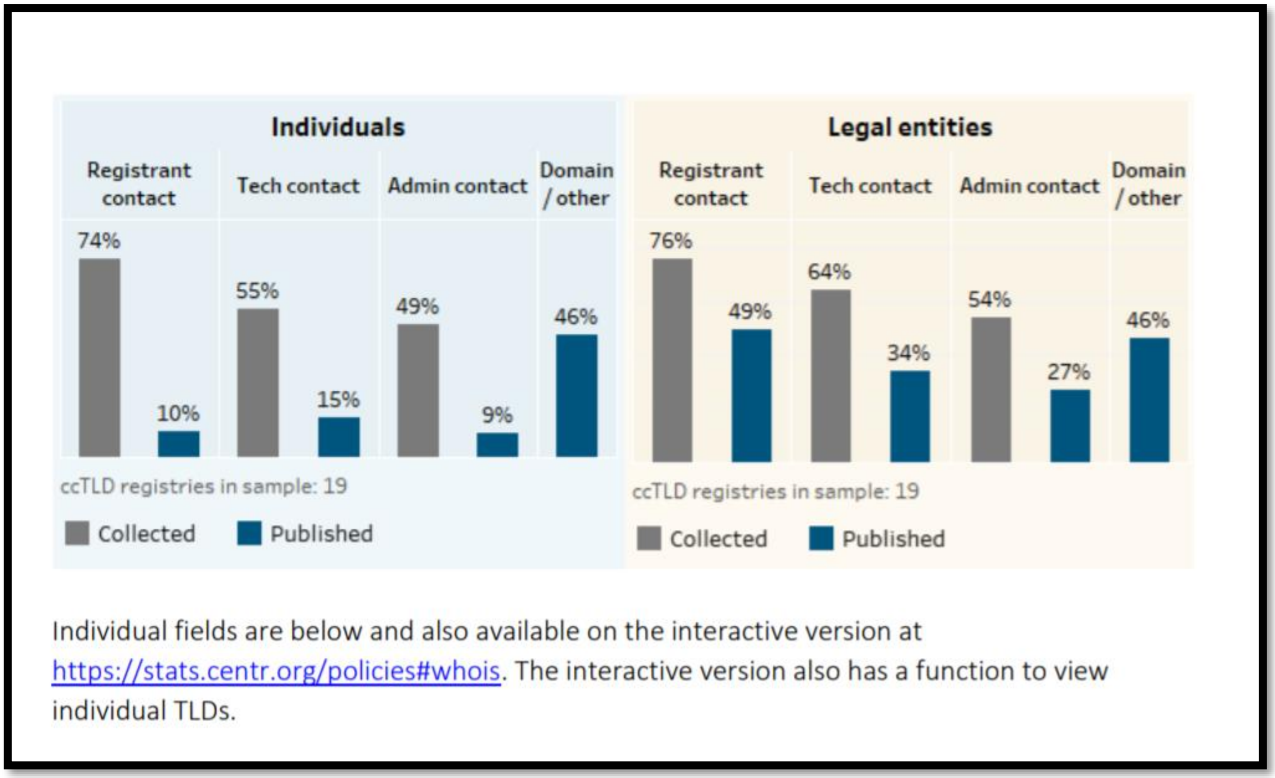
For this question, investigators were asked to choose all that apply.

When asked to explain what makes redaction excessive, 67% of responders feel that redacting legal entities is excessive. The EU GDPR is intended to protect EU data subjects; unsurprisingly, 65% of investigators believe that redacting data of registrants who are not EU data subjects is excessive.



Redacting legal entity data is ostensibly due to the reluctance of registrars and registries to trust that registrants have provided complete and accurate registration data. It is worth noting that a number of EU ccTLDs collect and publish legal entity contact data. A recent Centr Survey, *WHOIS Status and Impacts of GDPR*[4], reports on registrant data that is collected and published in WHOIS by ccTLDs based in the European Union. A table from the Centr Survey illustrates the collection and publication practices of registrants as private individuals and legal entities as reported by 19 ccTLDs.

[4] CENTR report - Whois status and impacts from GDPR, https://centr.org/library/library/survey-report/centr-report-whois-status-and-impacts-from-gdpr.html



Individual fields are below and also available on the interactive version at https://stats.centr.org/policies#whois. The interactive version also has a function to view individual TLDs.

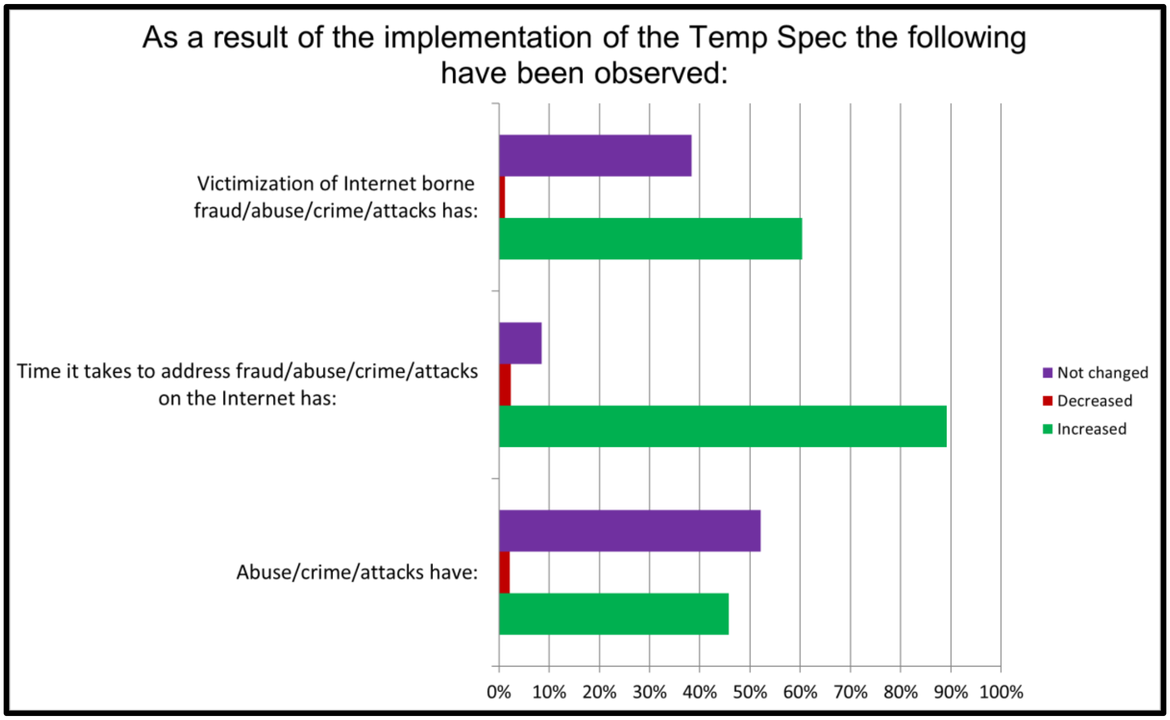
Responses to earlier survey questions document the lack of uniformity across registries and registrars with respect to timely and uniform access to non-public WHOIS data. Forty-five percent of investigators reaffirm the frustration with the current state of reveal expressed in earlier questions.

Twenty-nine percent of investigators express concern that the Temp Spec implementation cannot accommodate the adoption by another nation of a regulation that conflicts with the EU GDPR; for example, if a country were to insist that data currently considered non-public be published.

Question 13: As a result of the implementation of the Temp Spec the following have been observed.

The responses to question 13 offered investigators an opportunity to share whether they have observed increases or decreases in attacks, crime or abuse, times to address fraud, abuse, crime or attacks, and victim volume.

The percent of investigators that report increases in abuse, crime or attacks (47%) is nearly the same as the percent of investigators that report no change (combined, 52%). We have only four months of post-GDPR attack data. There are too many variables to confidently attribute increases or decreases that are measured in this manner. Still, nearly one-half of investigators who responded are reporting increases.



The responses are clearer when investigators are asked to describe the effect of the Temp Spec on the time it takes to address fraud, abuse, crime, or attacks. **Investigators overwhelmingly respond that the time it takes to address fraud, abuse, crime, or attacks on the Internet has increased (89%).** In addition, 60% of investigators reported that victimization has increased.

Comments from investigators relating to victimization include:

"In addition to more difficulty identifying malicious actors, redaction has increased the likelihood that innocent domains are impacted by overly aggressive enforcement (because relation to an innocent party cannot be verified)."

"Investigating fake bank and courier websites (used in 419 scams) usually reveals multiple websites by the same fraudster. This is now impeded and puts more victims of crime at risk ."

"In some cases we are investigating a malicious-looking domain and we don't know if it was owned by a criminal or by a compromised victim. WHOIS information in the past helped us decide if we were dealing with innocent people or criminals. This is putting innocent people at risk by making investigators assume they are criminals, when they are simply victims of compromise."

"Massive uptick in phishing campaigns with inability to proactively research domains linked to bad actors or their infrastructure. Huge negative impact to users. Criminals are abusing the heck out of a mandate that was supposed to improve privacy. While optional domain privacy services have hindered investigations for years, this level of new obfuscation has brought anti-phishing efforts to a serious crossroads."

"There are a variety of data sources that are synthesized together to assist in domain anti-abuse efforts. However, complete WHOIS information is among the quickest and most useful. Predictable WHOIS by criminals enables flagging of malicious domains via WHOIS in some cases. It also enables analysts to attempt to identify whether a domain is registered to researchers, a legitimate organization, or some other non-threat actor. This reduces the chances of false positives that lead to inappropriate and unfortunate suspensions that then need to be reversed. As a registry, we do have access to WHOIS data… presuming registrars are not obfuscating the data or not transmitting all of it. There are cases where registrars are labeling data as GDPR Redacted and thus essentially providing privacy shields to all WHOIS. This means a reduced ability to make connections between related domains in a campaign. Instead of neutralizing all the related abuse domains, we can only address a subset."

"I have seen a huge rise in phishing campaigns, whilst my ability to respond/trace and abate them has been removed. This has exposed Government Agencies to more risk and may have facilitated Nation State Actor intrusion and possible Data Loss."

Findings

From the survey responses, we find that:

Cyber-investigations and mitigations are impeded because investigators are unable to access complete domain name registration data through public WHOIS services in (near) real-time, as they had before the implementation of the Temp Spec. The partial data that are available through the public WHOIS services after redaction are insufficient to investigate or to respond to an incident.

The mitigation or triage of cyber incidents cannot be accomplished in a timely manner. Specifically, the need to request access to the non-public data elements introduces, at a minimum, delays of days in circumstances where mitigation prior to the adoption of the Temp Spec were often accomplished in hours or one day. Such delays allow attacks to remain active longer. Extended windows of operation put more internet users in harm's way.

WHOIS has become an unreliable and less meaningful source of threat intelligence. The WHOIS contact data that is most relevant to investigators and has evidentiary value to law enforcement and prosecutors is generally not available through public WHOIS services since the implementation of the Temp Spec. (Note: even fraudulently composed, pseudonymous, incomplete, or inaccurate data is useful for assigning reputations or creating correlations, for instance, in tracing known perpetrators' latest criminal excursions in establishing spoof domain names.) Investigators have been using alternative data sources with mixed success.

Requests to access non-public WHOIS by legitimate investigators for legitimate purposes are routinely refused. Investigators indicate that the implementation of "WHOIS reveal" is largely not working. The Temp Spec is unspecific, implementation is not uniform, and the processes are poorly understood by investigators, domain name registrars, and domain name registries. The majority of survey responders report that investigators do not know how to request access to non-public WHOIS data. Registrars and registries disclose redacted WHOIS data at their individual discretion, often without reasonable justification.

Those who protect internet resources are also making more coarse blocking or mitigation decisions in the absence of what was formerly reliable data. Network operators and protective service providers, in fact, are blocking entire Top-level Domains. Investigators report that in circumstances where they are unable to use non-public WHOIS data to make blocking decisions about individual domains, and where they cannot establish associations across (very large) sets of suspicious domain names, they are exercising an abundance of caution and blocking more aggressively.

The utility of WHOIS has been severely damaged. Four months after the Temp Spec implementation, an alarming 17% of responders claim that the public WHOIS service is no longer useful or reliable, and 13% have ceased using WHOIS entirely.

The redaction of WHOIS data is excessive. Investigators do not believe that it is necessary to redact legal entity point of contact data or point of contact data for data subjects outside the EU to comply with the EU GDPR.

Recommendations

Based on these findings, we encourage the ICANN organization and community to consider these recommendations during their ongoing deliberation of WHOIS policy:

Recommendation 1: There must be an accredited access mechanism, providing tiered or gated access to qualified security actors. A unified access program is necessary to restore predictable, automatable, swift access that balances privacy with legitimate use under GDPR. The technical mechanism would be RDAP (Registration Data Access Protocol).

Recommendation 2: ICANN should not allow redaction of the contact data of legal entities. Other WHOIS operators, such as the Regional Internet Registries and some European ccTLDs, do not redact data as aggressively as the Temp Spec allows.

Recommendation 3: ICANN should adopt a contact data access request specification that will ensure consistency across all accredited registrars and gTLD registries. The policy should be specific regarding the legitimate uses for which a timely completion of response is appropriate. These should align with the legitimate uses as described in the GDPR. A clear definition of "timely" should be included in the policy. Approved access requests should accommodate repeated access. Further, the specification should be specific with respect to:

A) Format of request (for both forms of WHOIS services);
B) Identification of information required to be set forth in the request;
C) Email addresses where requests can be sent;
D) Specification of documentation required for authenticating request; and
E) Time limitation for response to requests. We recommend that responses be processed in 24 hours or less.

The APWG or M^{3}AAWG members welcome the opportunity to work with ICANN and contracted parties to draft this specification in very short order. We would further welcome the opportunity to work with ICANN compliance and contracted parties to report aggregate statistics on access successes and challenges to help drive towards a better operational success rate.

Recommendation 4: ICANN should ensure that the accredited access to redacted WHOIS data does not introduce delays in collecting or processing WHOIS data, and further, that the access not be encumbered by per request authorizations: these simply do not scale for the volumes and purposes that investigators identify in their responses. We further ask that ICANN consider a framework wherein an accredited party be granted timely access and persistent access to complete WHOIS data.

Recommendation 5: ICANN should reconsider the current redaction of public WHOIS data policy. In the interest of serving data protection *and* legitimate needs to access complete WHOIS data, we urge ICANN to consider secure hashes rather than redacting WHOIS data. We call your attention to the proposal, *Public WHOIS Attributes, Securely Hashed (WhASH): Hashing Point of Contact Details in*

Public Domain Name WHOIS, submitted 4 June 2018, by APWG's Board of Directors to ICANN CEO and Chairman of the Board[5].

Recommendation 6: We ask that ICANN publish point of contact email addresses to provide investigators with an effective means of identifying domains associated with a victim or person of interest in an investigation. Some European ccTLDs and the RIRs are publishing email addresses in in their public WHOIS. Consistency across WHOIS services will facilitate investigations across identifier systems.

[5] Letter from David Jevans, APWG, to ICANN CEO and COB conveying a proposal, *Public WHOIS Attributes, Securely Hashed (WhASH): Hashing Point of Contact Details in Public Domain Name WHOIS*
https://www.icann.org/en/system/files/correspondence/jevans-to-marby-et-al-04jun18-en.pdf

Final Comments

We recognize that ICANN is likely aware of several of these issues. We intend that the survey results will help all parties to recognize and focus attention on the aspects of the Temp Spec that have the most serious impacts on cyber applications and forensic investigations; particularly, the issues relating to the implementation of reveal processes by contracted parties and the critical need for timely access.

We also realize that the ICANN organization and Board of Directors are awaiting the Expedited Policy Development Process for answers to many issues. We acknowledge that the EPDP will address many of these issues during consensus deliberation. Both APWG and M^{3}AAWG have member organizations represented on the EPDP panel. They are committed to lending expertise and to sharing practical field experience during these deliberations. APWG or M^{3}AAWG members are also willing to brief the EPDP panel, any ICANN community members, or ICANN organization on the survey and its results. We would welcome this opportunity to share, anecdotally, additional field experiences.

We believe that the ICANN Board of Directors and ICANN organization have the ability to update the Temp Spec to fix the problems that this survey and others have identified as most pressing or egregious while the EPDP work continues. Ignoring these problems serves to further erode the trust and respect for ICANN processes that is already damaged in the broader security community. This is likely to drive members of the security community to look outside of ICANN for solutions to their issues, from petitioning governments, to implement regulations that result in conflicts with the EU GDPR (and Temp Spec assumptions), to obtaining redacted data via non-transparent methods; e.g., sealed court orders, specific arrangements with some contracted parties, or alternative data collection methods.