Study on
Domain Name System (DNS)
Abuse
Final Report

Appendix 1 – Technical Report

Jan Bayer[1], Yevheniya Nosyk[1], Olivier Hureau[1],
Simon Fernandez[1], Ivett Paulovics[2], Andrzej Duda[1],
Maciej Korczyński[1]

October 30, 2021

[1]Grenoble INP-UGA Institute of Engineering
[2]Fasano Paulovics Società tra Avvocati

Contents

List of Figures

List of Tables

List of Acronyms

APWG Anti-Phishing Working Group.

AS Autonomous System.

ASN Autonomous System Number.

BEC Business Email Compromise.

BGP Border Gateway Protocol.

C&C Command and Control.

ccTLD Country-Code Top-Level Domain.

CSAM Child Sexual Abuse Material.

CZDS Centralized Zone Data Service.

DBL Domain Block List.

DDoS Distributed Denial-of-Service.

DKIM DomainKeys Identified Mail.

DMARC Domain-based Message Authentication, Reporting and Conformance.

DNS Domain Name System.

DNSSEC Domain Name System Security Extension.

DRDoS Distributed Reflective Denial-of-Service.

ESP Email Service Provider.

FQDN Fully Qualified Domain Name.

gTLD Generic Top-Level Domain.

IANA Internet Assigned Numbers Authority.

ICANN Internet Corporation for Assigned Names and Numbers.

IDN Internationalized Domain Name.

IOC Indicators of Compromise.

IP Internet Protocol.

IPv4 Internet Protocol Version 4.

IPv6 Internet Protocol Version 6.

ISP Internet Service Provider.

ISSP Information Society Service Providers.

KYBC Know Your Business Customer.

NTP Network Time Protocol.

RDAP Remote Data Access and Protocol.

SAV Source Address Validation.

SMTP Simple Mail Transfer Protocol.

SNMP Simple Network Management Protocol.

SPF Sender Policy Framework.

SSDP Simple Service Discovery Protocol.

TLD Top-Level Domain.

URL Uniform Resource Locator.

WHOIS protocol for querying databases that store the registered users or assignees of an Internet resource.

Executive Summary

The European Commission contracted Fasano Paulovics Società tra Avvocati and Institut Polytechnique de Grenoble for carrying out a Study on Domain Name System (DNS) abuse, EC reference VIGIE 2020/0653. This document is Appendix 1 - Technical Report and provides the analysis of the scope and magnitude of DNS abuse.

Definition of DNS Abuse

A safe and secure Domain Name System (DNS) is of paramount importance for the digital economy and society. Malicious activities on the DNS, generally referred to as "DNS abuse" are frequent and severe problems affecting online security and undermining users' trust in the Internet. The proposed definition of DNS abuse is as follows:

> *Domain Name System (DNS) abuse is any activity that makes use of domain names or the DNS protocol to carry out harmful or illegal activity.*

DNS abuse exploits the domain name registration process, the domain name resolution process, or other services associated with the domain name (e.g., shared web hosting service). Notably, we distinguish between:

- **maliciously registered domain names**: domain name registered with the malicious intent to carry out harmful or illegal activity

- **compromised domain names**: domain name registered by bona fide third-party for legitimate purposes, compromised by malicious actors to carry out harmful and illegal activity.

DNS abuse disrupts, damages, or otherwise adversely impacts the DNS and the Internet infrastructure, their users or other persons.

DNS abuse can be categorized into three main types that can also appear combined:

Type 1: Abuse related to **maliciously registered domain names**.

Type 2: Abuse related to the **operation of the DNS and other infrastructures**.

Type 3: Abuse related to **domain names distributing malicious content**.[1]

---

[1]This type of abuse may take advantage of maliciously registered or compromised domain names.

The three types also differ in terms of which relevant entities are responsible and/or best positioned to put in place mitigation measures:

1. **Abuse related to maliciously registered names (Type 1)** is usually best addressed by resellers (if any), registrars, and registries with the following proper remediation path:

   `Domain Reseller (if any) → Registrar → TLD Registry (at the DNS level)`

2. **Malicious content** can be distributed using a **maliciously registered domain name** (Type 1 and 3) or it can be distributed using a compromised website (Type 3), where the domain under which the malicious content is made available is registered by an unaware third-party, which uses it legitimately.

   (a) In case of **illegal/harmful content distributed using a maliciously registered domain name (Type 1 and 3)** (e.g., typosquatting domain name serving phishing content), the following remediation path is to be followed to effectively mitigate this abuse:

   `Hosting Reseller → Hosting Provider (at the hosting level)`

   and

   `Domain Reseller (if any) → Registrar → TLD Registry (at the DNS level)`

   Mitigating the abuse only at the hosting or DNS level will prevent access to malicious content but will not block all elements of the malicious infrastructure. Therefore, both levels have to be involved in the mitigation of this kind of abuse.

   (b) While it is also possible for the reseller (if any) / registrar / TLD registry to take action in case of **malicious content hosted on compromised websites** (Type 3), addressing abuse at the DNS level can be counterproductive, as it can cause collateral damage to legitimate registrants and website visitors. In this case, the site operator, hosting provider (and where it exists, its reseller) are well positioned to take action to curb the abuse. Remediation path is as follows:

   `Site operator → Registrant (if different from site operator) → Hosting reseller → Hosting provider (at the hosting level)`

Mitigating abuse at the hosting level includes removing malicious content from the hacked website and patching the vulnerability. Site operators are best positioned to mitigate abuse in case of so-called unmanaged dedicated servers that they are in complete control and are responsible for their hosting servers and software. Hosting companies are best positioned to mitigate abuse in case of so-called managed shared hosting as they maintain the operating system and application infrastructure.

3. All entities related to the DNS infrastructure (registrars, registries, resellers, operators of authoritative name servers, and DNS resolvers) are concerned with **abuse related to DNS operations** (Type 2).

To illustrate the types of DNS abuse and responsible entities, we give below the examples of DNS abuse cases:

- maliciously registered domain names serving phishing content:

    - Type 1 and 3 → mitigation action at the DNS level and the hosting level

- compromised websites serving phishing content:

    - Type 3 → mitigation action at the hosting level

- compromised websites used to distribute (deliver) malware:

    - Type 3 → mitigation action at the hosting level

- maliciously registered domain name used to distribute (i.e., deliver) spam—emails containing malicious content:

    - Type 1 → mitigation action at the DNS level

- maliciously registered domain name (e.g., algorithmically generated domain name - AGD) used for the malicious command-and-control (C&C) communication (between compromised hosts and a malicious actor):

    - Type 1 → mitigation action at the DNS level

- file sharing system abused to distribute child sexual abuse material (CSAM):

    - Type 3 → mitigation action at the hosting level

3

- maliciously registered domain name used to distribute CSAM:

  - Type 1 and 3 → mitigation action at the DNS level and the hosting level

- DDoS attack against a DNS server (e.g., NXNSAttack):

  - Type 2 → mitigation action at the DNS level

- DDoS attack against a web server using DNS open resolvers as amplifiers/reflectors:

  - Type 2 → mitigation action at the DNS level

- hijacked domain name (e.g., using zone or cache poisoning)

  - Type 2 → mitigation action at the DNS level

Report Structure

In this Technical Report, we first present the datasets as well as passive and active measurement methods used in the analysis of the scope and magnitude of DNS abuse. The analysis involves the following blacklist feeds: Spamhaus Domain Block Lists, the lists of SURBL, APWG, PhishTank, OpenPhish, URLHaus, and ThreatFox. We have also gathered the list of domain names for each TLD from zone files whenever available and with active web content crawling. We have identified over 251 million active domain names for generic TLDs, new gTLDs, European Union country-code TLDs, and non-EU ccTLDs, and collected the following additional data:

- 'A' resource records to calculate the reputation metrics for hosting providers,

- registration information using RDAP/WHOIS protocols to calculate the reputation metrics for domain registrars,

- TLD sizes to express the "overall health" of TLD ecosystems,

- information about the deployment of DNSSEC ('DS', 'DNSKEY', and 'RRSIG' resource records),

- open DNS resolvers,

4

- SPF and DMARC entries in 'TXT', records to measure the deployment of email security extensions.

Finally, we measured uptimes, i.e., how long a malicious URL (or a domain name) has been active since it appeared on one of the blacklists.

In the second part, we show the results of measurements and data analysis related to DNS abuse. We start with high-level statistics describing the distributions of the malicious resources and abuse rates per TLDs, hosting providers, and countries for different abuse types. We distinguish between compromised and maliciously registered domain names. While some domains are registered exclusively for malicious purposes, others are benign but get compromised and abused to, for example, serve malicious content. This part also provides key statistics on maliciously registered and compromised sites concerning TLDs and types of abuse. Since registrars cannot prevent abuse of vulnerable hosting software (unless they provide registration and hosting services), we calculate registrar reputation metrics based on domain names categorized as maliciously registered. Finally, we discuss targeted brands and special domain abuse.

The last part concerns the abuse of DNS infrastructure—how technical DNS infrastructure can be abused to perform different types of illegal activities. We first discuss the deployment of DNSSEC, followed by DNS resolver vulnerabilities, the study on SPF and DMARC adoption, the analysis of RFC-compliant email aliases, and inbound Source Address Validation that enables a great variety of attacks on DNS infrastructure. The study covers the second quarter of 2021.

Key findings of the measurements

The key findings of the measurements are as follows:

1. **"Overall health" of TLDs**

    (a) In relative terms, **new generic TLDs (gTLDs), with an estimated market share of 6.6%, are the most abused group of TLDs**. In the second quarter of 2021, 20.5% of all abused domain names representing phishing, spam, botnet C&C (Command and Control), and malware distribution combined were registered in new gTLDs (Section 4, pp. 27-28).

(b) However, not all new gTLDs suffer from DNS abuse to the same extent. The **two most abused new gTLDs combined account for 41% of all abused new gTLD names** (Section 8.2 p. 39).

(c) **European Union country-code TLDs (EU ccTLDs) are by far the least abused** in absolute terms, relative to their overall market share. Only 0.8 percent of all abused (compromised and maliciously registered) domain names were registered under EU ccTLDs (Section 4, pp. 27-28).

2. **Malicious vs. compromised domains: where does the abuse occur?**

(a) The vast **majority of spam and botnet C&C domain names are maliciously registered**, which is expected given the nature of the abuse (Section 9.3 p. 48).

(b) In the analysed data, **about 25% of phishing domain names and 41% of malware distribution domain names are presumably registered by legitimate users, but compromised at the hosting level**. In these cases, trying to address abuse at the DNS level can be counterproductive, as it can cause collateral damage to their legitimate users (Section 9.3 p. 48).

(c) When looking at compromised domain names, it emerged that **for highly used TLDs such as European ccTLDs, there is a higher incidence (42%) of hacked websites**. In TLDs with lower usage rates such as new gTLDs, attackers have a much stronger tendency to register directly the domains they intend to use for their malicious activities (Section 9.3 p. 49.

(d) TLD registries and registrars can **prevent malicious registrations (proactive measures) and mitigate maliciously registered domains (reactive measures) at the DNS level**. However, they have no control over the hosting infrastructure (unless they also provide a hosting service). Therefore, the authors have computed reputation metrics for domain names that are found to be maliciously registered and exclude domains that are likely compromised at the hosting level (Section 10 p. 50).

(e) The **top five most abused registrars account for 48% of all maliciously registered domain names** (Section 10.2 pp. 54-55).

(f) The study reveals **hosting providers with a disproportionate concentrations of spam domains** reaching 3,000 abused domains per 10,000

registered domain names (Section 11.3 pp. 57-61).

(g) **Phishers make heavy use of free subdomain and hosting providers** because they incur no cost, which makes them practical for serving malicious content. These services are less suitable for distributing spam and botnet C&C communication (Section 12 pp. 63-64).

3. **Adoption of DNS security extensions and mail protection protocols**

(a) The overall level of **DNSSEC adoption remains low**. In a large sample of 227 million domain names, only 9.4 million domains have all the required DNSSEC resource records (DNSSKEY, RRSIG and DS). 98.1% of these are correctly signed and have been correctly validated (Section 14.3 pp. 73-74).

(b) As for EU ccTLDs, **.cz (59%), .se (55%), .nl (51%), and .sk (48%) have the highest percentage of domain names signed with DNSSEC**. The registry operators of these domains provide price incentives and technical support for DNSSEC adoption (Section 14.3 p. 74-75).

(c) The measurements revealed **2.5 million open DNS resolvers** worldwide that can be effectively used as amplifiers in Distributed Denial-of-Service attacks (Section 15.4 p. 84).

(d) In large sample of 247 million domain names, more than **60% of domain names are without SPF and 97% of domains are without DMARC records** that prevent email spoofing, one of the techniques used in Business Email Compromise scams (Section 16.3 pp. 87-88).

Recommendations

Based on the analysis and measurements, the authors propose the following set of recommendations.

1. **Better DNS Metadata (for identifying resources and their attribution to intermediaries)**

(a) Likewise gTLDs, ccTLDs registries should provide a **scalable and unified way of accessing complete registration (WHOIS) information**

(in compliance with data protection laws), using the Registration Data Access Protocol (RDAP), necessary to attribute abused and vulnerable domain names to their respective registrars and obtain their contact information (Section 5 p. 30).

(b) In the same manner as gTLDs, ccTLDs registries should consider **publishing DNS zone file data** through DNS zone transfer or a system similar to the Centralized Zone Data Service maintained by ICANN (Section 4 p. 28).

2. **Contact Information and Abuse Reporting**

(a) The **email addresses of registrants and domain name administrators** that are not visible in the public WHOIS could be displayed as **anonymized email addresses** to ensure privacy and the ability to contact domain owners and administrators directly to notify security vulnerabilities and abuses (Section 17.4 p. 96).

(b) With no direct contact with domain name registrants and administrators via the public WHOIS database, domain name administrators should **maintain RFC 2142 specific email aliases** for given domain names (e.g., abuse, hostmaster, webmaster) and an email in the DNS SOA record so that they can be contacted directly in the event of vulnerabilities and domain name abuse (Section 17.4 p. 95).

(c) A **standardized (and potentially centralized) system for access to registration data (WHOIS data)** should be set up, identifying the minimum information necessary to process disclosure requests. The reaction time to such requests shall be clearly defined.

(d) The study also recommends to set up a **standardized (and potentially centralized) system for abuse reporting**, identifying the minimum information necessary to process such report. The receipt of abuse reports is to be acknowledged. The reaction time to such reports shall be clearly defined and the abuse reporter should be provided with information on the actions taken. The DNS service providers shall provide for an appeal proceeding against their decisions to a third neutral party.

(e) We encourage the exchange of information on threats between parties involved (e.g., CERTs, security organizations) using **collaborative platforms such as Malware Information Sharing Platform (MISP)** to report and mitigate abuse in a more effective and timely way.

3. **Improved Prevention, Detection and Mitigation of DNS Abuse Type 1**

   (a) TLD registries, registrars, and resellers should **verify the accuracy of the domain registration (WHOIS) data**. The identification of the registrants could be implemented through possibly harmonised Know Your Business Customer (KYBC) procedures. In case of registrants from the EU, KYBC could be carried out through eID authentication in accordance with the eIDAS Regulation, as amended by the forthcoming Regulation on the European Digital Identity. KYBC procedure shall use cross-checks in other publicly available and reputed databases (Section 8.2 p. 41).

   (b) TLD registries are encouraged to **develop or improve existing similarity search tools or surveillance services** to enable third-parties to identify names that could potentially infringe their rights (Section 10.2 p. 53).

   (c) TLD registries are encouraged to offer, directly or through the registrars/resellers, **services allowing intellectual property rights (IPR) holders to preventively block infringing domain name registrations** (similar to services already existing on the gTLD market).

   (d) The use of **predictive algorithms to prevent abusive registrations** by TLD registries and registrars is also encouraged.

   (e) The study recommends that the **abuse rates of TLD registries or registrars be monitored** on an ongoing basis by independent researchers in cooperation with institutions and regulatory bodies (e.g., ICANN, European Commission, European Union Agency for Cybersecurity – ENISA or national authorities). Abuse rates should not exceed predetermined thresholds. If thresholds are exceeded and the abuse rates do not improve within a given time period, accreditation may be revoked (Section 8.2 p. 43).

   (f) **TLD registries and registrars with lower abuse rates may be financially rewarded**, e.g., through a reduction in domain registration fees, to

align economic incentives and raise barriers to abuse (Section 8.2 p. 43).

(g) TLD registry operators are encouraged to:

- **maintain access to existing domain/URL blacklists**,

- identify the **registrars with the highest and lowest concentrations and rates of DNS abuse** in their ecosystems,

- propose incentive structures to **encourage their registrars to develop methods to prevent and mitigate malicious registrations** effectively (Section 10.2 p. 54).

4. **Improved Detection and Mitigation of DNS Abuse Type 3**

(a) In a similar way with respect to the TLD registries and the registrars, the **abuse rates of hosting providers should be monitored** on an ongoing basis by independent researchers in cooperation with institutions and regulatory bodies (e.g., European Commission, European Union Agency for Cybersecurity – ENISA or national authorities). Abuse rates should not exceed predetermined thresholds. Incentive structures should be studied to induce hosting providers to develop technical solutions that effectively curb hosting and content abuse (Section 11.3 p. 61).

(b) Since free services (e.g., free hosting and subdomains) are commonly exploited in phishing attacks, their operators should **employ advanced prevention and remediation solutions to quickly curb abuses of subdomain names and hosting infrastructure**. They should proactively detect suspicious domain names containing keywords of the most frequently targeted brands and names and work closely with the most heavily attacked companies and develop trusted notifier programs (Section 12 p. 66).

5. **Better Protection of the DNS Operation and Preventing DNS Abuse Type 2**

(a) Similarly to gTLD registries, registry operators of ccTLDs should be required to **sign TLD zone files with DNSSEC** and facilitate its deployment according to good practices (Section 14.3 p. 71).

(b) To facilitate the implementation of DNSSEC, domain administrators (registrants) should have **easy access to DNSSEC signing of domain names within the TLD**. TLD registries should require all registrars that offer domain names in the TLD to support DNSSEC signing for registrants (Section 14.3 p. 73).

(c) As an incentive to the deployment of DNSSEC, TLD registries might **offer discounts for DNSSEC-signed domain names** (Section 14.3 p. 75).

(d) Internet Service Providers (ISP) that operate DNS resolvers should **configure DNSSEC validation** to protect end users from cache poisoning attacks and ensure the integrity and authenticity of domain name resolutions (Section 15 p. 80).

(e) National CERT teams should subscribe to data sources that identify open DNS resolvers. National governments and Computer Emergency Response Team (CERT) teams should intensify **notification efforts to reduce the number of open DNS resolvers** (and other open services), which are among the root causes of distributed reflective denial-of-service (DRDoS) attacks (Section 15.4 p. 86).

(f) Security community should intensify efforts to continuously **measure the adoption of SPF and DMARC** protocols, especially for high risk domain names and raise awareness of the domain spoofing problem among domain owners and email service providers. Correct and strict SPF and DMARC rules can mitigate email spoofing and provide the first line of defence against Business Email Compromise scams (Section 16.4 p. 90).

(g) Network operators should deploy **IP Source Address Validation** not only for outgoing but also for incoming traffic at the edge of a network. It provides an effective way of protecting closed DNS resolvers from different external attacks against DNS infrastructure, including possible zero-day vulnerabilities within the DNS server software (Section 18 p. 97).

6. **DNS Abuse Awareness, Knowledge Building, and Mitigation Collaboration**

(a) At the EU level, the study recommends the harmonisation/approximation

of the practices of EU ccTLDs by the **adoption of the good practices** available at the European and international levels.

(b) The study recommends to require the **DNS service providers to collaborate with EU and Member States' institutions, law enforcement authorities (LEA) and so-called trusted notifiers or trusted flaggers**. Where informal collaborations exist, they are to be further strengthened and formal processes are to be set up for the parties to interact.

(c) The study encourages **awareness-raising and knowledge-building activities** to make the consumers, IPR holders, or other affected parties aware of existing measures tackling DNS abuse.

(d) The study encourages **knowledge-sharing and capacity-building** activities between all intermediaries and stakeholders involved in the fight against DNS abuse.

Part I

Background

The Domain Name System (DNS), along with the IP protocol, is a vital service of the Internet, mapping applications, hosts, and services from names to IP addresses.

ICANN delegates responsibility to registry operators to maintain an authoritative source for registered domains within the TLD (e.g., Donuts is the registry for the .lawyer TLD). TLDs are divided in gTLDs and ccTLDs, the first group is generic, the second group is reserved to countries, territories and geographical locations identified in the ISO 3166-1 country codes list. The governance of the gTLD namespace by ICANN is contractual. The management of the ccTLD namespace varies from informal to formal contracts between some countries or territories and ICANN.

Domain registrars manage the registration of domain names. They are generally accredited by TLD registries and can be accredited by ICANN. Domain resellers are third-party organizations that offer domain name registration services through a registrar but may not be accredited by a TLD registry or ICANN. Registrants are individuals or organizations that register domain names through domain registrars or resellers for a specific time period.

DNS providers operate zone files and authoritative DNS servers that map domain and hostnames to the corresponding IP addresses. While registrars usually maintain zone files, registrants may choose to delegate the responsibility to third-party authoritative DNS services such as Cloudflare or their own servers.

Web hosting providers maintain the server infrastructure used to host content for a given domain. Hosting providers may sell their services to individuals or other Web hosting providers—hosting resellers. Web hosting providers may offer so-called managed hosting. A hosting company handles the setup, management, and support of a server and/or applications (such as content management systems). Often, multiple (sometimes thousands) domain names might be hosted on the same physical server sharing the same IP address. Hosting providers may also offer unmanaged hosting, for example, a dedicated server with, for example, only an operating system installed. A user (webmaster) needs to install all the necessary software and keep the software up-to-date.

Note that the registrant/webmaster may choose to buy a reverse proxy service that

can hide the characteristics of an origin (i.e., backend) server such as its IP address. Finally, ISPs (access providers) typically maintain recursive DNS resolvers that resolve domain names on behalf of the end user's computer willing to access an application, a host, or a service associated with the domain name. Note that the same entity can provide different services (e.g., it is common that registrars offer authoritative DNS services and Web hosting plans).

Because the DNS encompasses a large ecosystem of different types of intermediaries that maintain the technical DNS infrastructure and hosting, the role of intermediaries in addressing abuse depends on both the type of abuse and the services they provide. In this sense, we should first formulate a comprehensive definition of DNS abuse and then detail the types of abuse to understand better the role of each intermediary involved in the DNS abuse mitigation process.

1 Definition of DNS Abuse

To start with, we propose the following definition of DNS Abuse:

> *Domain Name System (DNS) abuse is any activity that makes use of domain names or the DNS protocol to carry out harmful or illegal activity.*

DNS abuse consists of three possibly overlapping types:

Type 1: Abuse related to **maliciously registered domain names**.

Type 2: Abuse related to the **operation of the DNS and other infrastructures**.

Type 3: Abuse related to **domain names distributing malicious content**.[2]

The three types also differ in terms of which relevant entities are responsible and/or best positioned to put in place mitigation measures:

1. **Abuse related to maliciously registered names (Type 1)** is usually best addressed by domain resellers, registrars, and TLD registries with the following proper remediation path:

   Domain Reseller → Registrar → TLD Registry (at the DNS level)

---

[2]This type of abuse may take advantage of maliciously registered or compromised domain names.

2. **Malicious content** can be distributed using a **maliciously registered domain name** (Type 1 and 3) or it can be distributed using a compromised website (Type 3) where the domain under which the malicious content is made available is registered by an unaware third-party, which uses it legitimately.

   (a) In case of **illegal/harmful content distributed using a maliciously registered domain name (Type 1 and 3)** (e.g., typosquatting domain name serving phishing content), the following remediation path is to be followed to effectively mitigate this abuse:

   `Hosting Reseller → Hosting Provider (at the hosting level)`

   and

   `Domain Reseller → Registrar → TLD Registry (at the DNS level)`

   Mitigating the abuse only at the hosting or DNS level will prevent access to malicious content but will not block all elements of the malicious infrastructure. Therefore, actors at both levels have to be involved in the mitigation of this kind of abuse.

   (b) While it is also possible for the reseller (if any) / registrar / TLD registry to take action in case of **malicious content hosted on compromised websites** (Type 3), addressing abuse at the DNS level can be counterproductive, as it can cause collateral damage to legitimate registrants and website visitors. In this case, the site operator, hosting provider (and where it exists, its reseller) are well positioned to take action to curb the abuse. Remediation path is as follows:

   `Site operator → Registrant (if different from site operator) → Hosting reseller (if any) → Hosting provider (at the hosting level)`

   Mitigating abuse at the hosting level includes removing malicious content from the hacked website and patching the vulnerability. Site operators are best positioned to mitigate abuse in case of so-called unmanaged dedicated servers that they are in complete control and are responsible for their hosting servers and software. Hosting companies are best positioned to mitigate abuse in case of so-called managed shared hosting as they maintain the operating system and application infrastructure.

3. All entities related to the DNS infrastructure (registrars, registries, resellers, oper-

ators of authoritative name servers, and DNS resolvers) are concerned with **abuse related to DNS operations** (Type 2). This type of abuse is to be addressed at the DNS level.

To illustrate the types of DNS abuse and responsible entities, we give below the examples of DNS abuse cases:

- maliciously registered domain names serving phishing content:

  - Type 1 and 3 → mitigation action at the DNS level and hosting level

- compromised websites serving phishing content:

  - Type 3 → mitigation action at the hosting level

- compromised websites used to distribute (deliver) malware:

  - Type 3 → mitigation action at the hosting level

- maliciously registered domain name used to distribute (i.e., deliver) spam—emails containing malicious content:

  - Type 1 → mitigation action at the DNS level

- maliciously registered domain name (e.g., algorithmically generated domain name - AGD) used for the malicious command-and-control (C&C) communication (between compromised hosts and a malicious actor):

  - Type 1 → mitigation action at the DNS level

- file sharing system abused to distribute child sexual abuse material (CSAM):

  - Type 3 → mitigation action at the hosting level

- maliciously registered domain name used to distribute CSAM:

  - Type 1 and 3 → mitigation action at the DNS level and hosting level

- DDoS attack against a DNS server (e.g., NXNSAttack):

  - Type 2 → mitigation action at the DNS level

- DDoS attack against a web server using DNS open resolvers as amplifiers/reflectors:

– Type 2 → mitigation action at the DNS level

- hijacked domain name (e.g., using zone or cache poisoning)

  – Type 2 → mitigation action at the DNS level

The rationale of this approach resides in taking into account the perspectives of different actors along the entire abuse chain. Indeed, the following actors should be taken into consideration. Concerning abuse involving domain names, the following actors should be considered:

- the **abuser/attacker** – the registrant of the maliciously registered domain name or the actor compromising legitimately registered domain names (for example, by exploiting vulnerable websites),

- the **abused party** – the Internet users and/or third parties affected by the abuse causing physical, psychological, or economic harms, such as minors in case of child sexual abuse material (CSAM), consumers that are the victims of online frauds, the owners of intellectual property rights, etc.

- the **intermediaries** – the DNS operators (notably registries and registrars) and information society service providers (ISSPs)[3], including providers of hosting, access, and online platforms operators, as well as regular Internet users of the misused infrastructures that facilitate the distribution of illegal content. They should also be considered as victims (unless they are willingly facilitating cybercrime) because DNS infrastructure and content abuse affect their reputation and impose economic costs related to abuse handling. At the same time, this third group of actors plays a key role in effective abuse prevention and mitigation.

Recent definitions of DNS abuse[4,5] generally categorize phishing, botnet C&C, malware, pharming, and spam (when used as a delivery mechanism to other forms of abuse) as DNS "technical" abuse and therefore the registries and registrars "must" act upon this category. On the other hand, child abuse material, controlled substances, regulated goods, or intellectual property are considered website content abuses.

---

[3]https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016DC0872&from=DE
[4]https://dnsabuseframework.org/media/files/2020-05-29_DNSAbuseFramework.pdf
[5]https://www.internetjurisdiction.net/uploads/pdfs/Papers/Domains-Jurisdiction-Program-Operational-Approaches.pdf

In practice (and depending on the type of abuse), there is a great deal of overlap between DNS technical and content-related abuse. For example, an attacker may register a domain name to launch a phishing campaign to deceive potential victims into disclosing passwords to their bank accounts. The fake website may use the official logo of the bank to look more trustworthy. Therefore, we deal with content abuse (phishing of credentials, trademark, and copyright infringement of the bank) and DNS technical abuse (domain name is maliciously registered). Both a hosting provider and a DNS service operator (i.e., registrar) have to react in such a case. As long as the hosting provider removes the malicious content, but the malicious domain is not suspended, the attacker can purchase another hosting service from another provider and reuse the maliciously registered domain name. Let us assume that only the registrar (or a TLD registry operator) removes the domain name from the zone file. In that case, the attacker may reuse the hosting and register a new domain name with another operator to continue malicious operations.

In the example given, mitigating the abuse at the hosting or domain name level interrupts the malicious actions of the attacker. However, mitigating the problem at both hosting and DNS levels is required because the attacker abused hosting and DNS technical infrastructures. Such an approach also leads to an increased cost for the attacker and thus creates higher barriers to DNS abuse.

Consider another example of an attacker that compromises a legitimate website using a vulnerable content management system and uploads malware to distribute it and infect end users. This example represents content abuse (distributing malicious software), but it is not the abuse of the DNS infrastructure. The domain name is registered by a benign user and abused (the site is hacked). In this case, the malicious content should be removed and the hosting infrastructure (vulnerability) patched by the webmaster or hosting provider (depending on whether the hosting is managed or unmanaged).

Generally, in such cases, DNS service providers should not intervene at the DNS level, since suspending a benign domain name can cause collateral damage and disrupt legitimate activities of the domain owner and its users. However, certain types of abuse, such as online distribution of CSAM materials or human trafficking, can cause serious psychological and physical threats. Even if the DNS operator determines that the malicious user is not abusing the DNS infrastructure, the DNS operator must in-

form about the abuse the intermediaries involved in hosting (website operator, domain owner, and hosting provider). The DNS operator should also have the legal authority to temporarily suspend the domain if the content is not removed promptly. If, on the other hand, the DNS service operator concludes that, for example, CSAM material is being distributed using a maliciously registered domain name (DNS infrastructure abuse), the domain must be suspended as well as the content removed by the hosting provider operating the infrastructure.

From the perspective of the domain name holder who registered a domain with malicious intent, all kind of abuses are prohibited and no distinction is made between "technical" and "content-related" abuses (see ICANN New gTLD Registry Agreement Specification 11 3a "*prohibiting Registered Name Holders from distributing malware, abusively operating botnets, phishing, piracy, trademark or copyright infringement, fraudulent or deceptive practices, counterfeiting or otherwise engaging in activity contrary to applicable law*"; ICANN Registrar Accreditation Agreement 1.13 *"Illegal Activity" means a conduct involving use of a Registered Name sponsored by Registrar that is prohibited by applicable law and/or exploitation of Registrar's domain name resolution or registration services in furtherance of conduct involving the use of a Registered Name sponsored by Registrar that is prohibited by applicable law"*).

From the perspective of the abused or affected party, the focus is on the harm suffered and not on the categorisation whether the origin of the harm is of the "technical" and/or "content-related" nature.

Moreover, in many cases such a clear-cut distinction (technical vs. content-related) is not possible at high level and the borderline is blurred due to the great deal of overlap between different types of abuse. Recent definitions of DNS abuse generally categorise, for example, phishing as "technical" abuse and therefore the registries and registrars "must" act upon this category.[6] However, previous research [1] showed that in the sample of manually labelled phishing domains (gathered from blacklisted URLs), 58% of domaines were registered by seemingly malicious actors indicating DNS technical abuse, whereas all of the URLs served abusive content affecting, in the first place, Internet users (by tricking them into revealing sensitive personal or financial information), but also third parties (by incorporating well-known trademarks in the phishing websites), and finally, ISSPs whose infrastructure was misused to host malicious content.

---

[6]https://dnsabuseframework.org/media/files/2020-05-29_DNSAbuseFramework.pdf

The remaining 42% of domains were seemingly compromised, meaning that benign registrants registered the underlying domains, but an abuser most probably exploited vulnerable web hosting. Therefore, the registries and registrars should respond (by forwarding the abuse complaint to another intermediary such as the hosting provider) rather than being obliged to react at the DNS level because these domains do not directly abuse DNS infrastructure (unless the benign domain name was compromised at the DNS level, e.g., it was hijacked using zone poisoning [2]).

URLs used to distribute malware is another example indicating that the clear-cut distinction between "technical" and "content-related" is not appropriate. Malware distribution URLs serve harmful content (malicious software) to infect end users. In the sample of manually labeled malware URLs [1], as many as 57% of domains were compromised [1] by exploiting, for example, web vulnerabilities, and therefore again, in most cases they do not directly abuse the DNS technical infrastructure. Therefore, registries and registrars should respond (following the principle of victims first) by forwarding the abuse complaint to the hosting provider rather than must react at the DNS level. Moreover, the recent definitions of DNS abuse generally categorise domains used to host websites offering, for example, counterfeit goods, pirate content, or CSAM material as "content-related" abuse, and therefore, fall outside the recently proposed definitions of DNS abuse. However, similarly to phishing or malware, the abusers may use the DNS infrastructure and in particular, maliciously registered domain names to distribute such content. In such cases, the registries and registrars as well as content providers must react to these types of abuse.

2 Role of Intermediaries in Abuse Handling

This complex ecosystem requires a **bottom-up** approach in handling DNS abuse. In respect to DNS abuse involving domain names, such as phishing, malware, IPR infringement, CSAM, etc., the DNS intermediary that detects or is notified about abuse must first assess if a given incident is related to DNS infrastructure and/or content abuse, identify and inform an appropriate party that might be in a better position to make such assessment, and address abuse.

Let us assume that a DNS operator (e.g., registry or registrar) receives an abuse notification and concludes that the domain name is registered for malicious purposes

based on the collected evidence or the evidence provided by the (trusted) notifier. In this case, the domain name must be blocked by the DNS service operator, i.e., registry, registrar, domain reseller, or authoritative DNS service provider (if different from the registry or registrar) according to applicable policies. Assume that, in addition, the domain name reveals illegal/harmful content, i.e., is used to distribute malware, hosts CSAM material, or a phishing website. In this case, the DNS service operator should, in addition to the takedown at the DNS level, identify and contact a hosting provider using, for example, WHOIS information. The domain name must be blocked within the period specified by applicable laws, but also, the hosting is suspended. Otherwise, as mentioned earlier, after the domain name suspension, the attacker may register another domain name and map the newly registered domain name to the operational hosting service. The hosting operator must suspend the hosting or (if not possible) contact the responsible hosting reseller that must suspend the service. Note that if the domain name uses a reverse proxy service (e.g., Cloudflare), the proxy provider must suspend the service and contact the hosting provider of the back-end infrastructure, which must suspend the hosting service.

On the other hand, if a DNS operator receives an abuse notification and concludes, based on the collected evidence, or evidence provided by the notifier, that the domain name is legitimate but compromised (hacked), the DNS operator generally should not suspend the domain name. The notified DNS service operator, i.e., registry, registrar, domain reseller, or authoritative DNS service provider (if different from the registry or registrar), should contact the hosting provider. A provider should not suspend the hosting server (especially a shared hosting server) but it should block access to a malicious website used to host or facilitate the distribution of illegal content. The hosting provider must contact the webmaster (possibly the domain owner) and inform about the incident. For unmanaged services, the hosting provider must contact the webmaster directly to mitigate abuse. If the hosting provider is not in a direct relationship with the end user, it must identify the hosting reseller, who must take appropriate steps.

Part II
Datasets

In this part, we present the datasets as well as passive and active measurement methods used in the analysis of the scope and magnitude of DNS Abuse.

We first leverage the list of abused domain names and URLs (Section 3) and the complete list of all domain names for certain TLDs, or the large sample of domains for the TLDs that do not make their zone files available (Section 4). We collect the registration information of the active domain names (Section 5), collect DNS records to evaluate the deployment of security technologies (Section 6) and finally regularly collect Web content, DNS and WHOIS information for blacklisted domain names/URLs to assess the time required to mitigate abuse (Section 7).

3   Blacklists

To estimate the prevalence and persistence of DNS infrastructure and content abuse, we use **sixteen** distinct blacklists generously provided to us by six blacklist providers: Spamhaus [3], SURBL [4], ABUSE.ch [5], the Anti-Phishing Working Group (APWG) [6], PhishTank [7], and OpenPhish [8]. They represent URL, fully-qualified domain name (FQDN)/IP address or domain name blacklists of spam, malware distribution, command-and-control (C&C), phishing, and IPR infringement. The blacklists are typically used by different intermediaries such as Internet Service Providers (ISPs) in operational environments to block malicious communications.

- **Spamhaus** Domain Block List (DBL) provides malicious domains obtained from URLs enumerated in spam email payloads, senders of spam emails, known spammers, phishing, malware-related websites, or suspicious domain names sharing common patterns with domains involved in DNS infrastructure and content abuse [9]. We use eight data feeds provided by Spamhaus grouped based on the type of malicious activity the domain is associated with[7]:

  - **Spam (SP)**: is a group of seemingly maliciously registered domains used for spam distribution (i.e., for sending unsolicited emails).

---

[7]https://www.spamhaus.org/faq/section/Spamhaus%20DBL#291

- **Phishing (PH)**: is a group of domains seemingly registered by malicious actors and used for phishing attacks.

- **Malware (MW)**: is a group of domains seemingly registered by malicious actors and used for distributing malicious software and infecting end users to gain access to private computer systems and sensitive information.

- **Botnet command-and-control (C&C)**: is a group of seemingly maliciously registered domains used for botnet C&C communication between C&C servers and compromised machines (bots).

- **Abused-legit spam (AL-SP)**: is a group of domains that are mostly registered by benign users (businesses and individuals) but compromised and abused by malicious actors to distribute spam.

- **Abused-legit phish (AL-PH)**: is a group of domains that, similarly to abused-legit spam domains, are labeled as benign but compromised and abused by malicious actors in phishing attacks.

- **Abused-legit malware (AL-MW)**: is a group of domains registered by legitimate users but compromised and abused to distribute malware.

- **Abused-legit botnet C&C (AL-C&C)**: is a group of benign domains abused by malicious actors to perform botnet C&C communication.

- **SURBL** list is composed of domain names observed in unsolicited email messages and external blacklists [10]. The combined list provides domain names categorized as phishing, malware, or spam.

  - **SURBL phishing (PH)** contains phishing domains supplied by several organizations including PhishTank, OITC phishing, PhishLabs, RiskAnalytics, and internal feeds maintained by SURBL. These blacklists collide maliciously registered domain names with domains of compromised websites abused to launch phishing attacks.

  - **SURBL malware (MW)** contains data from multiple sources that cover domains used to host malware websites, payloads, or associated redirectors. Note that compromised domains may also be included in the list since there is no distinction between compromised websites and maliciously registered domain names.

- **SURBL abuse (SP)** comprises domain names involved in spam activity. The list is mostly based on SURBL internal research, however, ISPs, Email Service Providers (ESPs), and other entities also contribute to this list. It may contain both maliciously registered domain names and domains of compromised websites.

- **APGW** contains phishing URLs submitted by accredited users through the eCrime Exchange (eCX) platform.[8] Blacklisted URLs are accompanied by metadata, including the confidence level and the **target brand name**. There is no distinction between maliciously registered domains and compromised websites.

- **PhishTank** provides a community-based phishing verification system, which contains phishing URLs submitted and verified manually by its contributors as malicious. Blacklisted ULRs contain additional metadata such as the **target brand name** but no indication if the registered domain name is malicious or the website is compromised.

- **OpenPhish** feed contains phishing URLs (and **targeted brands**) identified by OpenPhish or reported to OpenPhish and then verified as phishing. The feed collides malicious registrations and compromised websites.

- **Abuse.ch** is a non-profit organization combating malware and helping ISPs and network operators protecting their infrastructure from malware activities. We use two following sources:

  - **URLHaus**[9] is a service provided by `abuse.ch` that publishes URLs (containing either domains or IP addresses) used for malware distribution.
  - **ThreatFox**[10] is a platform from `abuse.ch` with the goal of sharing indicators of compromise (IOCs) associated with malware. The database contains URLs and domains used to communicate with command-and-control servers, the malware family and confidence level.

For this study, we collect newly blacklisted resources from March/April to June 2021 (second quarter of 2021) and gather relevant metadata related to registration or hosting infrastructure using active and passive methods.

---

[8]https://apwg.org/ecx/
[9]https://urlhaus.abuse.ch
[10]https://threatfox.abuse.ch

Dataset	Type	# Entries	# Entries (with domains)	% Entries (with domains)	Unique domains	Period
SURBL	FQDN/IP	797,162	595,383	74.69%	475,925	Mar-June 2021
SP		340,611	340,513	99.97%	338,363	
MW		205,839	8,933	4.34%	7,276	
PH		253,169	248,384	98.11%	132,838	
Spamhaus	Domain	1,088,863	1,088,863	100.0%	1,087,770	Apr-June 2021
SP		946,068	946,067	100.0%	945,397	
PH		105,254	105,254	100.0%	105,025	
MW		4,143	4,143	100.0%	4,143	
C&C		6,798	6,798	100.0%	6,798	
AL-SP		25,166	25,165	100.0%	25,065	
AL-PH		6,147	6,147	100.0%	6,070	
AL-MW		4,258	4,258	100.0%	4,257	
AL-C&C		139	139	100.0%	139	
APGW	URL	248,457	248,432	99.99%	42,748	Mar-June 2021
OpenPhish	URL	177,330	170,055	95.9%	49,418	Mar-June 2021
PhishTank	URL	54,975	54,156	98.51%	22,717	Mar-June 2021
URLHaus	URL	349,834	14,085	4.03%	6,401	Mar-June 2021
ThreatFox	URL	3,724	2,930	78.68%	1,290	Apr-June 2021

Table 1: Summary of blacklists including the total number of entries (FQDNs/IP, Domains, URLs), the number of entries including a domain name and its ratio in regard to the total number of entries, the number of unique domain names, and the collection period per feed.

Table 1 shows the summary of sixteen datasets. The **Type** refers to one of three types of entries that the blacklists provide, i.e. FQDN/IP, Domain, or URL. The FQDN/IP type refers to fully-qualified domain names (e.g., `malicious.example.com`) or IP addresses used in attacks, whereas the Domain type refers to 2nd–level domain names (e.g., `malicious.pl`) or sometimes 3rd–level domains if a given TLD registry provides such registrations (e.g., `malicious.com.pl`). Finally, the URL type provides URLs that serves malicious content (e.g., `http://example.com.pl/paypal-account.php`). The number of **Entries** corresponds to the identified abuse incidents (i.e., the number of unique blacklisted FQDNs/IPs, domains, or URLs). Note that the total number of entries for a given abuse feed (e.g., SURBL) might be different from the total number of entries from individual datasets combined (e.g., SURBL SP, SURBL MW, and SURBL PH) because some FQDNs may appear in two or more blacklists (e.g., SURBL PH and SURBL MW) at the same time.

In this study, we analyzed over **2.7 M** incidents (almost **2.17 M involving domain names**) and **1.68 M abused domain names**. While for some types of attacks, such

as phishing, malicious actors typically use domain names to induce the victim to reveal sensitive personal or financial information, other types of attacks, such as malware distribution, generally do not need to involve domain names to be successful. Therefore, Table 1 presents the number and percentage of entries that involve (abuse) domain names (**Entries with domains**). Note that SURBL MW and URLHaus datasets, representing malware distribution incidents, contain **only 5% and 7% of entries with domain names**. Moreover, not all of those attacks (involving domain names) abuse DNS infrastructure as both feeds collide maliciously registered and compromised domains (i.e., benign domain names whose websites have been compromised) or, in the case of URLHaus, free services such as online file storage services like Google drive.

4   Domain Names

To calculate the abuse rates per different actors involved in the domain name registration and hosting, and the deployment of DNS security technologies, we first need the list of domain names for each TLD. We use two sources of data: i) zone files whenever available, and ii) active web content crawling.

- **DNS Zone Files**. For the purposes of the study, we collect .com, .net, .org, .biz, .tel, .info legacy gTLDs and new gTLDs made available to us by the ICANN Centralized Zone Data Service (CZDS),[11] as well as .se and .nu ccTLD zone files.[12] Zone files provide the most complete list of registered domain names in each TLD and are therefore, the most accurate source of TLD sizes needed to calculate abuse rates for individual TLDs reliably.

- **Active Web Scans**. Apart from TLDs, for which we have access to their zone files through CZDS, to have a more exhaustive list of domains, especially for ccTLDs without having access to zone files, we develop the scanning platform to crawl all the websites of the domains in our database actively (only scanning homepages looking for hyperlinks) to retrieve newly observed domain names.

Using the approaches, we can collect a total of **251,585,395** active domain names related to 1,376 distinct TLDs including new gTLDs (16,081,977 domains related to

---

[11] https://czds.icann.org/home
[12] https://internetstiftelsen.se/en/domains/tech-tools/access-to-zonefiles-for-se-and-nu/

1,057 distinct TLDs), legacy gTLDs (172,796,509 domains related to 22 distinct TLDs), European Union ccTLDs (26,280,324 domains related to 32 distinct TLDs), other European ccTLDs (10,516,931 domains related to 28 distinct TLDs), and non-European ccTLDs (25,909,654 domains related to 237 distinct TLDs).

Although our collection of domain names is not complete, it provides a representative sample of registered domains per TLD registry to evaluate the deployment of security technologies (DNSSEC, SPF, DMARC) that prevent DNS infrastructure abuse such as DNS cache poisoning attacks (DNSSEC) or domain/email spoofing (SPF, DMARC).

We also use the enumerated domain names to calculate sizes of domain registrars and hosting providers. We further collect additional information about hosting infrastructure (by resolving the registered domains to their IP addresses and respective autonomous systems) and domain registrars (by collecting the registrar information using the WHOIS or RDAP protocols). However, not having access to the full list of domain names is a limitation of this work: we do not have a complete picture of the market share of registrars and hosting providers, so we cannot precisely calculate their abuse rates. More generally, such data allows interested parties to conduct research and develop new insights into the security practices of hosting providers or domain registrars, verify their policies or create reliable reputation metrics.

On the other hand, the TLD size information, relevant to the calculation of abuse rates, does not necessarily involve the requirement of having access to the full list of domain names. We use available zone files as the most accurate source of sizes. If these are not available, we use the sizes of ccTLDs affiliated with the Council of European National Top-Level Domain Registries (CENTR),[13] whose members have explicitly agreed to make this information available for the purpose of this study. For all other TLDs, we used the approximate sizes provided by DomainTools.[14]

Figure 1a, shows the market share of five groups of TLDs: legacy gTLDs (53%), new gTLDs (6%), European Union ccTLDs (14%), other European ccTLDs (7%), and other non-European ccTLDs (20%). By comparing the market share with the distribution of blacklisted domains (Figure 1b), we conclude that EU ccTLDs are the least abused in both absolute and relative terms to market share. In relative terms, new gTLDs (with the estimated market share of 6.6%) are the most abused group of TLDs (20.5% of abused domain names).

---

[13]https://www.centr.org
[14]https://research.domaintools.com/statistics/tld-counts

(a) All domains   (b) Abused domains
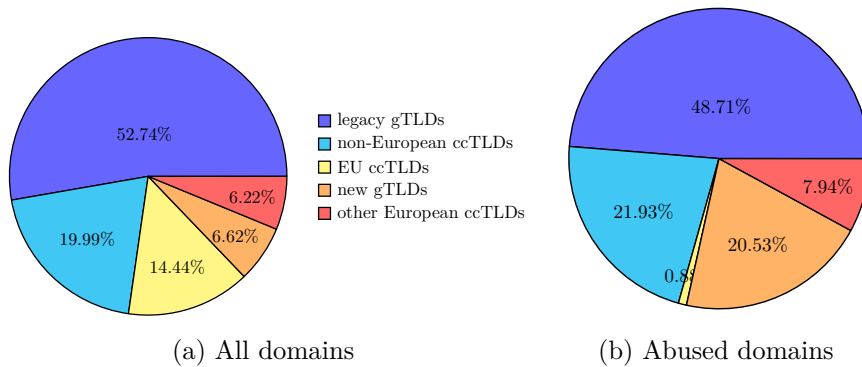
Figure 1: Divison of the domain namespace per TLD type

As one of the safeguards introduced by ICANN intended to mitigate the abusive and criminal activity, all new gTLDs and some legacy gTLDs make their zone files available to third parties using the Centralized Zone Data Service (CZDS). Some ccTLD registry operators make their zone files available and easily accessible to the public either voluntarily (e.g., .se),[15] or as a result of local regulations (e.g., .ch).[16] Both zones are available through DNS zone transfer (AXFR).[17,18] Some others make zone files available to the vetted parties (e.g., .nl), for example, for the research purposes[19] and after signing a data sharing agreement. However, the majority of ccTLD registry operators are against making their zone files available to third parties. One argument is that open access to a list of domain names may have unforeseen negative consequences for their security and stability.

> **Recommendation**: The registry operators of ccTLDs should consider publishing DNS zone file data through DNS zone transfer or a system similar to the Centralized Zone Data Service (CZDS) maintained by ICANN.

5 Registration Information

To calculate the reputation of each registrar, first, we have to calculate their sizes, i.e., the number of domain names registered with each registrar. By the term 'registrar',

---

[15]https://internetstiftelsen.se/en/domains/tech-tools/access-to-zonefiles-for-se-and-nu/

[16]https://www.switch.ch/de/open-data/#tab-c5442a19-67cf-11e8-9cf6-5254009dc73c-3

[17]https://zonedata.iis.se/

[18]https://www.switch.ch/de/open-data/

[19]https://openintel.nl/coverage

we refer to the ICANN-accredited list of registrars.[20] It is a challenging task since the only publicly available data source that can be queried and obtained at scale is the WHOIS information of the registered domain names. Using our data collection method described in Section 4, we enumerated 251,585,395 domain names for different TLDs.

Two general approaches are available for collecting WHOIS information at scale: i) the Registration Data Access Protocol (RDAP),[21] and ii) the WHOIS protocol. Whenever the registry operator supports the RDAP protocol, we use it to collect registration information since the result of RDAP is in the JSON format (easy for machines to parse) and generally does not need an extra parsing step to extract the information. Occasionally, we observe cases in which RDAP implementations are not fully compliant with the RFC specification and recommended format. We informed one of the European ccTLD registries (also managing a new gTLD) about such a problem. We received a reply confirming the existence of the issue and assuring that the operator will fix it. At the time of writing, only 17 ccTLDs (such as .vg, .no, .cz) have support for RDAP.[22] Therefore, for most of the ccTLDs, we had to retrieve the data from a TLD WHOIS server where there is no specific and consistent format for serving data. Therefore, we parsed the collected data and used regular expressions to extract relevant fields, including the registrar name, IANA ID (for ICANN-accredited registrars only), and the registration date. The approach required extensive data processing as the formats differ among different entities.

In total, we collected WHOIS data for 241,618,279 (approximately 96%) domain names in our list. Note that for some TLDs, it is not feasible to collect WHOIS information since either there is no TLD WHOIS server (e.g., in case of .gr) or the access is restricted to authorized IP addresses (e.g., in case of .es TLD). Moreover, for some TLDs (e.g., .de), the information displayed in the public WHOIS is very limited. It does not contain the registrar name, IANA ID, administrative or abuse email address or the domain name creation date. Instead, the registry operator provides the web-based WHOIS service that cannot be queried at scale. Access to the registration data *at scale* via WHOIS or (preferably via) RDAP is essential for different security tasks. For example, large-scale vulnerability or domain name abuse notifications require mapping abused or vulnerable resources to their respective intermediaries, such as reg-

---

[20]https://www.icann.org/en/accredited-registrars
[21]https://datatracker.ietf.org/doc/html/rfc3912
[22]https://data.iana.org/rdap/dns.json

istrars, that are in the best position to mitigate specific vulnerabilities or can takedown maliciously registered domain names. It also requires collecting (at scale) e-mail information to contact registrars directly. One limitation of this work is that we have no information about specific TLD ecosystems, such as for .es, .gr, or .de domain namespaces, for which the WHOIS information is not available, restricted, or provided only via the web-based service. In these cases, we cannot map domain names (particularly maliciously registered) to relevant registrars in specific TLD ecosystems and calculate abuse rates.

> **Recommendation**: Likewise gTLDs, ccTLDs registries should provide a scalable and unified way of accessing complete registration (WHOIS) information (in compliance with data protection laws), using the RDAP protocol, necessary to attribute abused and vulnerable domain names to their respective registrars and obtain their contact information.

After collecting and parsing WHOIS information, for each record, we use the following algorithm to identify the registrar name:

1. if the IANA ID field exists and is valid, automatically extract the corresponding registrar name from the ICANN-accredited registrars list.

2. in the absence of the IANA ID or the presence of an invalid value (not present in the ICANN list), extract the raw registrar name from the WHOIS record and match with the registrar names in the ICANN list to get the correct IANA ID. The second step required manual labeling.

While the approach mentioned above provides us with the common registrar names, some cases still exist that cannot be handled in this way. For example, looking at Table 2 for the Godaddy[23] registrar, there are three ICANN-accredited registrars with IANA ID 146, 3786, and 1659[24] all with different registrar names in WHOIS data (the first column of the table). This case is the result of selling domain names by re-sellers or merging registrar companies to expand their businesses [11]. For such cases, we manually labeled registrar names based on the ICANN list and checked the registrar websites.

---

[23]https://godaddy.com/
[24]Note that we do not merge entities whenever the IANA ID is different, even though they may belong to one parent company.

Our final list consists of 3,733 registrar names, which have 1,222 variants of registrar names for those in the ICANN-accredited list.

#	Registrar name	ICANN Registrar name	IANA ID
1	Go Daddy, LLC	GoDaddy.com, LLC	146
2	GoDaddy.com, LLC	GoDaddy.com, LLC	146
3	GoDaddy	GoDaddy.com, LLC	146
4	GoDaddy.com	GoDaddy.com, LLC	146
5	Godaddy.com, LLC	GoDaddy.com, LLC	146
6	GoDaddy.com, LLC.	GoDaddy.com, LLC	146
7	GoDaddy.com LLC	GoDaddy.com, LLC	146
8	GoDaddy.com, Inc.	GoDaddy.com, LLC	146
9	GODADDY	GoDaddy.com, LLC	146
10	GoDaddy.com,LLC	GoDaddy.com, LLC	146
11	GoDaddy.com LLC trading as GoDaddy.com	GoDaddy.com, LLC	146
12	GoDaddy Corporate Domains LLC	GoDaddy Corporate Domains, LLC	3786
13	GoDaddy Corporate Domains, LLC	GoDaddy Corporate Domains, LLC	3786
14	GoDaddy Online Services Cayman Islands	GoDaddy Online Services Cayman Islands Ltd.	1659

Table 2: Fourteen different registrar names in the WHOIS data for three ICANN-accredited registrars all belong to godaddy.com.

While there are still many non-standard registrar names that we cannot reliably label (such as personal names or companies accredited by ccTLD registries locally), we can identify registrars for about 85% of WHOIS records.

We also map each domain name found in an abuse feed to its corresponding registrar name using WHOIS records collected as soon as the domain name gets blacklisted. Having the size of each registrar and the number of the maliciously registered domain for that registrar, we can calculate the abuse rate for each of them.

6 Active DNS Scans

This section explains the method for collecting information regarding different types of DNS abuse, vulnerabilities, and security technologies. We actively collect the following data: i) 'A' resource records to calculate the reputation metrics for hosting providers, ii) information about the deployment of DNSSEC ('DS', 'DNSKEY', and 'RRSIG' resource records), iii) open DNS resolvers, and, iv) SPF and DMARC entires in 'TXT' records to measure the deployment of email security extensions.

- **DNS Resource Records: 'A' Records and Geolocation**

  To calculate the reputation of hosting providers, we need to collect all the 'A' records of all the enumerated benign and malicious domain names using the open-

source ZDNS[25] tool—a fast DNS resource record scanner. Using the MaxMind[26] database, we convert each 'A' record to its corresponding AS number and a country code.

- **DNSSEC-related Records**

  To measure the deployment of DNSSEC, we actively collect three DNS resource records for each domain in our database: 'DNSKEY', 'DS', and 'RRSIG'. The presence of such records does not necessarily mean that domains are correctly signed but rather signifies that domain owners attempted to do so. If the 'DNSKEY' and 'DS' are present, we make an attempt to validate DNSSEC chain using our recursive (validating) resolver. If the validation succeeds, the domain name is correctly signed.

- **Email Security Extensions: SPF and DMARC**

  To measure the deployment of the Sender Policy Framework (SPF) and Domain-based Message Authentication, Reporting and Conformance (DMARC), we collect the 'TXT' record of each registered domain (e.g., example.com) for SPF and the 'TXT' record of the '_dmarc' subdomain (e.g., _dmarc.example.com). Note that the presence of the SPF and DMARC does not necessarily mean that the domain name is safe against email spoofing since these records should be configured correctly to protect the domain name. However, the absence of even one of these records makes the domain name vulnerable to email spoofing attacks [12].

- **Identifying Open Resolvers**

  Various open services (such as DNS, NTP, Memcached, etc.) have long been known as efficient DDoS reflectors and powerful amplificators [13,14]. Open DNS resolvers accept DNS requests from any end host (instead of the predefined ranges of IP addresses). They can be used to either target authoritative nameservers by sending the excessive number of incoming requests or, if combined with IP address spoofing, used to redirect responses to victim end hosts (e.g., high-profile websites). Reducing the number of open resolvers will increase the barriers for cybercriminals to launch reflective DDoS attacks (DRDoS).

---

[25] https://github.com/zmap/zdns
[26] https://www.maxmind.com/

We actively scan for open DNS resolvers in IPv4 and IPv6 address spaces and analyze their distribution across organizations and countries.

Scanning for open resolvers requires sending DNS requests to end hosts and inspecting the received responses. The response codes (`RCODE`), defined in RFC 1035 [15], signal whether the DNS server is configured to process the incoming request. If the query resolution is successful, open resolvers send back the responses to end clients along with `NOERROR` status code.

We use the following three datasets to scan for open resolvers: IPv4 BGP prefixes [16], IPv6 Hitlist Service [17], and IPv6 addresses enumerated in our previous study [18]. All three datasets contain globally reachable IP addresses that may be operational recursive resolvers. Each end host from the list receives an 'A' request for the unique domain name maintained by us. We developed software that allows us efficiently sending DNS packets on a large scale [19].

7 Uptime Measurements

The goal of the uptime measurement is to find the amount of time a malicious URL is accessible from the time it appeared in one of the blacklists. The uptimes (or persistence of abuse) will be calculated for different intermediaries and abuse types (phishing, malware and botnet C&C).

First, we collect a variety of information related to each URL when it gets blacklisted. We download the content of the malicious URL, the content of the homepage of the registered domain name, and the WHOIS information of the domain name for each malicious URL we find in the blacklist feeds. The difficulty of the uptime measurements compared with a single-time measurement is that the platform repeats the single-time measurement twelve times for each blacklisted URL.

Figure 2 shows the diagram of the uptime measurement platform designed for the purpose of this study. The "feed server" (step ①) is synchronized with all the blacklist feeds introduced in Section 3. Whenever a newly blacklisted URL appears in one of the feeds, we receive the update on our feed server. We automatically collect all data for each URL (step ②) once it gets blacklisted, 5 minutes after blacklisting, 30 minutes, 1 hour, 6 hours, 12 hours, 24 hours, 48 hours after blacklisting, and then once per week (12 measurements in total).
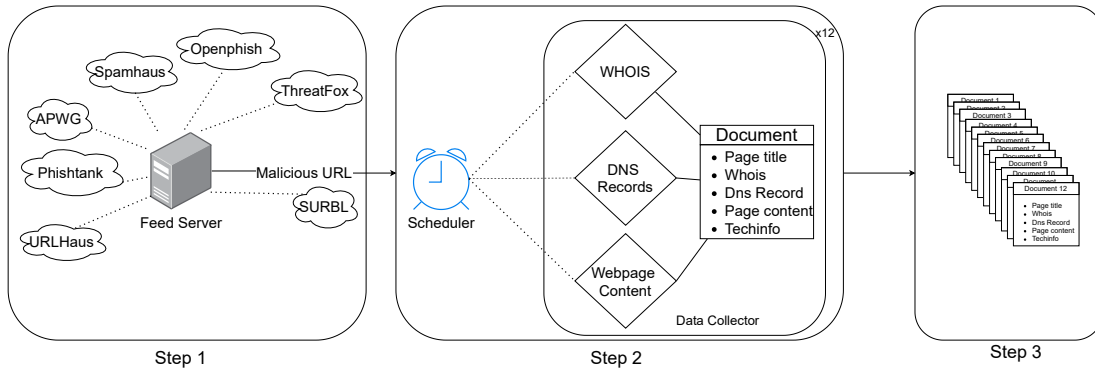
Figure 2: Uptime measurement platform

After having all the 12 measurements for each URL (step ③), we can compare the 12 data collections. The difference in some key features of each data collection (IP address of the registered domain, WHOIS, page content, etc.) can answer when the URL/domain name has been suspended. We show the result of the uptime measurements in Part III.

Part III
Domain Name Abuse Landscape

In this part, we show the results of the domain name abuse measurements and analysis performed in this study. We start with statistics describing the distributions of the malicious resources and abuse rates per TLDs (Section 8). We distinguish between compromised and maliciously registered domain names (Section 9). Since registrars cannot prevent abuse of vulnerable hosting software (unless they provide registration and hosting services), we calculate registrar reputation metrics based on domain names categorized as maliciously registered (Section 10). Then, we present reputation metrics for hosting providers, and countries for different abuse types (Section 11). Finally, we discuss abuse of special domains (Section 12) and targeted brands in phishing attacks (Section 13).

8 TLD Reputation Metrics

8.1 Methodology

To measure the reputation of each TLD [20], we use two security metrics: i) *occurrence* and ii) *ratio*. The *occurrence* metric refers to the number of unique domains extracted from blacklisted URLs. Note that we distinguish the reputation metric from measuring the "security performance" of TLD registries. While some domain names may be maliciously registered (and therefore they abuse DNS and hosting infrastructure), some websites are hacked and used to distribute harmful or illegal content (they abuse hosting but not necessarily DNS infrastructure). In these latter cases, TLD registries generally should not intervene at the DNS level to avoid collateral damage to legitimate owners and visitors of benign but compromised domain names (unless attackers have exploited DNS vulnerabilities). However, note that even hacked websites, distributing malicious content, impair the reputation of legitimate domain names and all intermediaries involved in domain name registration and hosting. Therefore, TLD reputation metrics express the "overall health" of TLD ecosystems consisting of many types of intermediaries such as domain registrars, re-sellers, hosting, content providers, etc.

Previous work [21, 22] proposed the use of two additional complimentary security

metrics to measure TLD health, i.e., the number of blacklisted fully qualified domain names (FQDNs) and unique blacklisted URLs. The reason is that one registered domain name can be used in a single attack while another can be used to distribute different malicious content using subdomains or multiple URLs. Therefore, two complementary metrics provide a more complete picture of the DNS abuse extent. However, they must be interpreted with caution because a single phishing campaign that uses hundreds or even thousands of subdomains can affect the reputation of an entire TLD. More importantly, this research revealed an emerging trend among attackers: one maliciously registered domain name serves the same content but dynamically generates a unique URL each time a potential victim visits it. As such, the complementary reputation metrics may be biased and overestimate the number of security incidents. For example, we observed one malicious domain redirecting users to the same page impersonating the Chase banking website and 5,192 unique URLs blacklisted by APWG, OpenPhish, and Phishtank. Therefore, in this study, we only use the *occurrence* of unique domain names as a TLD reputation metric.

Note that, reliable reputation metrics need to account for sizes, as bigger providers tend to suffer from more security incidents [22,23]. To calculate the abuse rates of each TLDs, we use the following formula:

$$ratio = \frac{occurrence}{size\ of\ a\ TLD} \times 10,000 \tag{1}$$

This rate expresses the number of abused domains per 10,000 registrations.

We extract registered domain names from blacklisted URLs using the modified version of the public suffix list maintained by Mozilla.[27] For the purpose of this study, we only use the public part of the suffix list.

8.2 Results

Table 3 shows the occurrence of the top 10 most abused TLDs in seven blacklist feeds (and the corresponding sixteen blacklists), namely: APWG, OpenPhish, PhishTank, Abuse.ch (URLHaus, ThreatFox), SURBL, and Spamhaus. The table shows all TLDs, including new gTLDs, legacy gTLDs, and ccTLDs. In Table 4, we aggregate different blacklists per abuse type (botnet C&C, malware, phishing, and spam).

---

[27]https://publicsuffix.org

APWG		OpenPhish		PhishTank		URLHaus		ThreatFox	
TLD	# Domains	TLD	# Domains	TLD	# Domains	TLD	# Domains	TLD	# Domains
com	18,662	com	23,253	com	9,161	com	3,265	com	424
ir	3,509	xyz	2,418	xyz	1,324	in	249	top	213
xyz	2,375	tk	2,125	tk	1,139	br	244	xyz	61
top	1,456	ml	1,473	top	793	net	219	ru	56
net	1,123	ga	1,318	cn	662	org	207	in	40
ru	946	net	1,178	net	525	top	113	net	38
org	835	cf	1,093	shop	492	uk	112	br	35
shop	806	org	1,055	org	420	xyz	111	club	29
site	753	gq	1,016	ml	332	za	111	org	26
ml	561	top	897	ga	319	ru	90	space	22

SURBL combined		SURBL mw		SURBL ph		SURBL abuse	
com	146,955	com	3,551	com	52,134	com	91,818
ru	48,202	xyz	729	ml	9,142	ru	46,086
cn	42,738	top	668	tk	7,455	cn	41,105
su	30,713	za	210	xyz	7,425	su	30,691
xyz	25,065	net	200	ga	5,223	xyz	17,007
work	16,736	shop	180	cf	5,102	work	16,130
top	10,807	org	128	gq	4,368	us	9,598
us	10,193	in	111	net	2,784	top	8,258
ml	9,541	ru	106	cn	2,443	me	6,888
tk	8,520	online	94	ru	2,184	info	6,269

Spamhaus combined		Spamhaus SP		Spamhaus PH		Spamhaus MW		Spamhaus C&C	
com	470,877	com	425,632	com	27,547	com	853	com	2,594
cn	144,785	cn	136,284	cn	8,015	buzz	664	info	673
net	66,077	net	63,123	xyz	7,047	xyz	634	net	497
work	58,094	work	57,673	ga	5,511	top	235	top	442
xyz	43,353	xyz	35,390	tk	5,505	vip	172	org	225
ru	23,025	ru	20112	shop	4,959	cloud	136	ru	198
info	16,625	info	13,944	ml	4,029	br	98	me	189
tk	15,658	online	12,280	press	3,971	ru	91	su	142
online	13,666	biz	12,239	cf	3,434	monster	83	eu	142
biz	12,709	tk	10,139	bar	3,008	live	82	tk	137

Spamhaus AL-SP		Spamhaus AL-PH		Spamhaus AL-MW		Spamhaus AL-C&C	
com	11,927	com	3,399	com	2,253	com	54
ru	1,468	org	196	in	207	ru	32
net	884	br	195	br	190	org	7
org	748	net	167	org	156	kr	5
tk	546	in	146	net	146	cz	3
br	452	uk	136	za	72	net	3
cn	433	cn	129	uk	71	pl	3
in	403	au	126	au	54	it	3
de	392	za	76	co	46	ua	3
uk	256	co	69	cn	41	ir	2

Table 3: Top 10 TLDs with the highest occurrence of blacklisted domains for APWG, OpenPhish, PhishTank, URLHaus, SURBL, and Spamhaus datasets.

In different types of abuse, .com and (generally) .net legacy gTLDs are among the most abused TLDs. The distribution of abused legacy gTLDs domains is mainly driven by .com domains, which is not surprising given the TLD market share. For phishing abuse, five out of ten most abused TLDs (.ml, .tk, .ga, .cf, and .gq) are operated by Freenom.[28] Freenom provides domain registrations for free for .ml, .tk, .ga, .cf, and

---

[28]https://www.freenom.com

.gq ccTLDs. Freenom is also an ICANN-accredited registrar and provides registrations for most of the other top-level domains. Registrants of free domain names have no "licensee" status, but they are considered as domain name "users". Therefore, Freenom can take down any domain name without further notice. The lack of registration fees is the most likely reason Freenom's TLDs are widely abused by malicious actors.

Botnet C&C		Malware		Phishing		Spam	
TLD	# Domains	TLD	# Domains	TLD	# Domains	TLD	# Domains
com	2,763	com	6,788	com	91,587	com	519,965
info	676	xyz	1,373	xyz	15,658	cn	177,374
top	567	top	894	ml	12,644	net	66,565
net	523	buzz	668	tk	11,912	ru	64,462
ru	261	net	405	ga	10,459	work	64,235
org	238	br	371	cn	10,259	xyz	48,790
me	189	org	357	cf	8,241	su	32,974
xyz	174	in	298	shop	6,511	info	19,460
su	145	za	270	gq	6,235	top	15,724
eu	142	shop	191	top	5,652	us	15,499

Table 4: Top 10 TLDs with the highest occurrence of blacklisted domains per abuse type.

Botnet C&C		Malware		Phishing		Spam	
TLD	# Domains	TLD	# Domains	TLD	# Domains	TLD	# Domains
ru	261	ru	190	ru	3,787	ru	64,462
me	189	uk	143	me	1,918	su	32,974
su	145	**de**	57	uk	1,826	me	8,900
eu	142	**ro**	48	**de**	857	рф	4,129
at	91	me	39	**fr**	600	**eu**	1,948
ua	18	su	38	**eu**	513	uk	1,518
pl	15	**it**	35	nl	444	**nl**	1,067
nl	15	**fr**	34	it	413	**de**	1,026
de	15	es	32	**pl**	386	**pl**	853
it	14	**nl**	32	es	271	**fr**	778
fr	12	**eu**	29	**ro**	204	**it**	488
hu	10	**pl**	27	**be**	171	ua	380
be	10	**gr**	21	ch	157	**be**	329
uk	4	rs	20	ua	135	**se**	236
cz	3	**hu**	20	**hu**	135	es	218
pt	3	**pt**	20	**gr**	121	ch	193
im	1	ch	18	**pt**	98	**at**	191
bg	1	**hr**	12	**cz**	94	**ro**	147
by	1	**be**	12	**dk**	88	**cz**	119
lt	1	**at**	7	su	85	**dk**	110

Table 5: Top 20 European ccTLDs with the highest occurrence of blacklisted domains per abuse type.

Botnet C&C			Malware			Phishing			Spam		
TLD	# Domains	Rate	TLD	# Domains	Rate	TLD	# Domains	Rate	TLD	# Domains	Rate
su	145	8.08	**ax**	1	2.78	me	1,918	17.50	su	32,974	1,837.10
me	189	1.72	su	38	2.12	al	26	11.63	ru	64,462	131.48
at	91	0.66	rs	20	1.73	im	23	8.06	me	8,900	81.21
ru	261	0.53	**hr**	12	1.14	ru	3,787	7.72	рф	4,129	63.11
eu	142	0.39	бел	1	0.98	md	19	7.33	va	1	24.39
im	1	0.35	**ro**	48	0.90	rs	77	6.67	al	25	11.18
ua	18	0.31	**gr**	21	0.49	mk	16	5.29	im	31	10.87
bg	1	0.15	**ee**	6	0.45	**hr**	50	4.74	by	87	6.73
hu	10	0.12	al	1	0.45	su	85	4.74	ua	380	6.65
by	1	0.08	ru	190	0.39	gg	11	3.84	md	16	6.17
pl	15	0.06	me	39	0.36	**ro**	204	3.84	**eu**	1,948	5.33
be	10	0.06	gg	1	0.35	li	21	3.72	ad	1	4.78
lt	1	0.05	mk	1	0.33	**bg**	22	3.39	sm	1	4.49
it	14	0.05	by	4	0.31	**si**	36	2.84	li	24	4.25
fr	12	0.03	**hu**	20	0.25	**gr**	121	2.82	**fo**	2	3.84
nl	15	0.02	ie	6	0.19	by	35	2.71	**pl**	853	3.44
cz	3	0.02	**lv**	2	0.17	**mt**	4	2.46	**lv**	36	3.11
pt	3	0.02	es	32	0.16	ua	135	2.36	mk	9	2.97
рф	1	0.02	**bg**	1	0.15	is	18	2.36	rs	34	2.95
de	15	0.01	**pt**	20	0.15	бел	2	1.96	**bg**	19	2.93

Table 6: Top 20 European ccTLDs with the highest abuse rate per abuse type.

We also observe that malicious actors extensively misuse .xyz, .work (for spam) and .top new gTLDs. One of the plausible reasons might be their pricing, as suggested already in the previous studies (e.g., [21]). The retail pricing of .xyz and .top domains oscillates around US $1. In 2020, we observed that the price of .xyz domains proposed by Alibaba Cloud was as low as US $0.18. Figure 1 in Section 4 shows that new gTLDs suffer from the highest concentration of abused domains relative to their market share. It does not mean that the entire market of new gTLDs is extensively abused. The two most abused new gTLDs (.work and .xyz) together account for 41% of all abused new gTLD names in the second quarter of 2021, while abused domain names from the top five most abused new gTLDs (.work, .xyz, .top, .online, .site, .icu) account for about 60% of all abused new gTLD domains.

Note that European Union ccTLDs are not among the most abused TLDs (with the exception of .eu domains for the botnet C&C feed). In general, EU ccTLDs are proactive in deploying (voluntarily) preventive and reactive security measures to combat DNS abuse more effectively (e.g., .eu, .nl, .dk, .se), and do not need to aggressively compete in pricing (as opposed to some new gTLDs such as .top, or .xyz).

Table 5 shows the top 20 European ccTLDs (European Union ccTLDs are marked in bold) with the highest occurrence of abused domains for four different abuse types (botnet C&C, malware, phishing, and spam). The Russian Federation top-level domain (.ru) is the most abused European ccTLD of all different abuse types (in terms of

occurrence). As many as 64 K .ru and 33 K .su (Soviet Union) ccTLD domain names were labeled as spam. In comparison, for the fifth most abused .eu ccTLD, significantly less (2 K) domains were blacklisted in the second quarter of 2021.

If we consider only non-European Union ccTLDs and combined abuse counts (C&C, malware, phishing and spam), abused .ru and .su second-level domain names account for 75% (50.4% and 24.8%, respectively) of all abused domains among non-EU ccTLDs. In comparison, abused domain names in EU ccTLDs, apart from significantly lower absolute abuse counts, are more distributed among different TLDs. Seven most abused EU ccTLDs (.eu, .de, .nl, .fr., .pl, .it, .es, .be) account for about 76% of all abused domains among EU ccTLDs.

The .dk TLD is at the other extreme: only 110 .dk domain names were blacklisted and marked as spam, 88 as phishing. In total, our dataset contains 168 unique, abused .dk domain names over a three-month period. Some domains were reused in both phishing and spam attacks. Therefore, the sum of spam and phishing domains is higher than the number of unique domains. Using the method described in Section 9, we automatically labeled .dk domain names as compromised or maliciously registered. One limitation of the classification is that we could not automatically identify patterns of mass registrations indicating malicious registrations because WHOIS of .dk domains does not provide the registrar name or IANA ID. Only 15 domains were found to be registered with malicious intent. The cheapest registration price for .dk domains during the study oscillated between US $9 and US $10. The price is just one of the possible factors that make attackers not to choose to maliciously register domain names with .dk TLD. There may be other reasons as well, such as verifying the accuracy of the registrant information. If the registry operator verifies the registrant identity (as is the case with .dk domain names[29]), a cybercriminal may choose to register a malicious domain name with a different TLD.

---

[29]https://www.dk-hostmaster.dk/en/id-check

> **Recommendation**: TLD registries, registrars, and resellers should verify the accuracy of the domain registration (WHOIS) data. The identification of the registrants could be implemented through possibly harmonised Know Your Business Customer (KYBC) procedures. In case of registrants from the EU, KYBC could be carried out through eID authentication in accordance with the eIDAS Regulation, as amended by the forthcoming Regulation on the European Digital Identity. KYBC procedure shall use cross-checks in other publicly available and reputed databases.

This study provides anecdotal evidence indicating that cybercriminals choose to abuse TLDs that offer low domain name registration prices and avoid TLDs that strictly verify the registrant identity. However, there is a need for a very comprehensive statistical analysis of factors driving DNS abuse. One malicious actor may prefer lower registration prices. The other may choose to abuse a registrar that offers specific payment methods (e.g, cryptocurrencies) or a free API allowing for domain registration in bulk.

While the absolute number of abused domains is essential, yet it does not show the (relative to the size) *overall health* of each TLD. Table 6 shows the abuse rate for the top 20 European ccTLDs per abuse type (European Union ccTLDs are highlighted in bold). While the number of malicious domains with .ru TLD is the highest (see Table 5), .su suffers from the highest abuse rate in terms of spam and botnet C&C because the size of the .ru zone file is approximately 45 times bigger than .su zone file. Security researchers observe that for years, .su domains are abused to control botnets, spread malware, send spam, or infect end-users to steal banking data. While hosting companies and registrars usually suspend such domains instantly, malicious .su domains run for months.[30] The uptime (the time between the blacklisting/notification and the takedown) is another plausible reason why certain TLDs are abused more than others.

Table 7 shows the top 10 TLDs with the highest abuse rates in four abuse categories (botnet C&C, malware, phishing, and spam). We divide all TLDs in five different groups of TLDs, corresponding to 5 different size ranges. The reason for grouping TLDs is to make a more appropriate comparison by considering the estimated size of each TLD. For example, in the spam category, the abuse rate of the .cleaning TLD is 103 while the number of abused domains for this TLD is 32, which is not comparable to the

---

[30]https://www.theguardian.com/technology/2013/may/31/ussr-cybercriminals-su-domain-space

		Botnet C&C		Malware			Phishing			Spam		
	TLD	# Domains	Rate	TLD	# Domains	Rate	TLD	# Domains	Rate	TLD	# Domains	Rate
(0, 10k]	ug	29	37	date	33	33	date	591	596	ryukyu	151	2,654
	bond	2	7	ug	20	25	gle	1	556	okinawa	295	701
	download	5	6	lr	1	24	ci	303	305	date	525	529
	sb	1	6	krd	1	16	ss	1	286	xn–nqv7f	10	316
	sx	1	4	sd	3	8	saxo	1	118	ss	1	286
	army	1	3	bingo	1	7	aq	1	104	sbs	73	253
				tg	1	5	int	2	97	recipes	106	209
				cd	3	5	makeup	6	70	webcam	44	125
				mz	3	4	ki	3	54	cleaning	32	103
				cv	1	4	delivery	52	54	exposed	32	99
		Botnet C&C		Malware			Phishing			Spam		
	TLD	# Domains	Rate	TLD	# Domains	Rate	TLD	# Domains	Rate	TLD	# Domains	Rate
(10k, 100k]	support	28	10	trade	54	28	press	3,984	1,420	surf	1,979	1,529
	uno	22	9	bid	49	27	support	514	180	cam	5,143	989
	bid	12	7	institute	10	7	casa	349	71	casa	3,675	751
	casa	30	6	pk	54	6	help	113	67	rest	2,459	701
	gdn	3	3	casa	26	5	cyou	608	55	fit	4,563	549
	rocks	19	2	best	34	4	review	61	53	cyou	5,195	467
	golf	3	2	py	10	4	surf	60	46	uno	864	335
	city	5	2	win	24	4	finance	118	33	host	1,450	325
	band	2	1	review	4	3	digital	301	33	trade	428	221
	digital	13	1	bo	4	3	cam	168	32	support	526	184
		Botnet C&C		Malware			Phishing			Spam		
	TLD	# Domains	Rate	TLD	# Domains	Rate	TLD	# Domains	Rate	TLD	# Domains	Rate
(100k, 1M]	su	145	8.00	buzz	668	27	buzz	668	27	su	32,974	1,837
	pw	75	2.00	monster	85	7	monster	85	7	work	64,235	1,123
	space	39	1.00	cloud	137	7	cloud	137	7	email	2,878	261
	website	33	1.00	email	43	4	email	43	4	buzz	5,523	219
	fun	21	1.00	ke	31	3	ke	31	3	life	3,387	154
	cc	69	0.91	link	32	3	link	32	3	fun	3,094	154
	link	9	0.77	ng	30	2	ng	30	2	world	1,906	148
	kz	11	0.70	vip	178	2	vip	178	2	asia	2,587	131
	mobi	21	0.65	live	104	2	live	104	2	website	3,791	129
	life	11	0.50	su	38	2	su	38	2	tech	4,182	128
		Botnet C&C		Malware			Phishing			Spam		
	TLD	# Domains	Rate	TLD	# Domains	Rate	TLD	# Domains	Rate	TLD	# Domains	Rate
(1M, ∞)	top	567	5.00	top	567	5.00	xyz	15,658	52	cn	177,374	170
	me	189	2.00	me	189	2.00	top	5,652	48	xyz	48,790	161
	info	676	2.00	info	676	2.00	ir	3,865	34	top	15,724	132
	at	91	0.66	at	91	0.66	ml	12,644	32	ru	64,462	131
	club	69	0.66	club	69	0.66	tk	11,912	22	biz	14,456	105
	xyz	174	0.57	xyz	174	0.57	site	2226	20	us	15,499	92
	ru	261	0.53	ru	261	0.53	ga	10,459	19	me	8,900	81
	online	94	0.53	online	94	0.53	cf	8,241	19	online	13,561	76
	biz	62	0.45	biz	62	0.45	gq	6,235	18	site	7,313	66
	net	523	0.40	net	523	0.40	club	1,898	18	club	5,706	55

Table 7: Top 10 TLDs with the highest relative concentrations of blacklisted domains classified by their corresponding TLD size and abuse type.

.xyz TLD with 48,790 abused domains for spam with abuse rate of 161. Abuse rates, especially for the smallest TLDs, should be considered with caution. Even a small number of abused domains (or sometimes false positives, i.e., domain names blacklisted by security companies by mistake) can strongly affect a TLD reputation.

Abuse rates should be monitored on an ongoing basis by independent researchers on behalf of regulatory bodies and should not exceed predetermined thresholds (e.g., 30%, e.g., 3 K abused per 10 K registered domain names). Such recommendations were already proposed in recent years [24–26]. If a contracted party (TLD registry or

registrar) exceeds the acceptable abuse rates and does not improve within a given time period, accreditation may be revoked [24, 26]. Note that abuse thresholds should be carefully designed, as even a low absolute number of incidents for small TLD registries may result in thresholds being exceeded. On the other hand, large registries (in terms of registered domain names) may not face elevated abuse rates even if the number of abused domains is high (e.g., .com).

> **Recommendation**: The study recommends that the abuse rates of TLD registries or registrars be monitored on an ongoing basis by independent researchers in cooperation with institutions and regulatory bodies (e.g., ICANN, European Commission, European Union Agency for Cybersecurity – ENISA or national authorities). Abuse rates should not exceed predetermined thresholds. If thresholds are exceeded and the abuse rates do not improve within a given time period, accreditation may be revoked.

Previous reports [24–26] indicate that TLD registries (and registrars) facing disproportionately high rates of abused domain names could be charged higher fees for domain registrations. In contrast, intermediaries with lower abuse rates could be financially rewarded, e.g., by reducing domain registration fees to align incentives with raising barriers to abuse.

Such incentive structures must be very carefully designed to encourage the development of security practices by TLD registries with high abuse rates and those that affect the reputation of the overall market. For example, .xyz does not suffer from high abuse rates (spam rate in Q2 2021: 161 per 10,000 registrations; spam domains: 48,790, Table 7). However, new gTLDs as a class of registry operators suffer from disproportionately high abuse rates (Figure 1), with .xyz accounting for about 20% of all abused domains among new gTLDs.

> **Recommendation**: TLD registries and registrars with lower abuse rates may be financially rewarded, e.g., through a reduction in domain registration fees, to align economic incentives and raise barriers to abuse. Price incentives could be implemented by ICANN for TLD registries, or by TLD registries for locally accredited registrars.

We now analyze uptimes (persistence of abuse), i.e., the time between blacklisting a URL and taking mitigation action. Figure 3 shows the so-called survival analysis of the ten most abused TLDs (in terms of occurrence) associated with malware delivery URLs. The data results from an uptime analysis for 13,249 URLs (5,547 unique domains) identified by URLhaus. The survival analysis and the curves shown illustrate the chance of survival (not being taken down by the intermediaries involved) over a certain period of time. For example, in Figure 3, the chance for an abused domain name with the `.net` TLD to stay alive after four days is about 32%, while the probability for a domain with the `.br` TLD to survive after four days is about 47%. Therefore, the intermediaries involved in the `.net` TLD ecosystem mitigate faster abuse compared to the `.br` TLD.

Figure 4 shows the survival analysis for different TLD types. In general, new gTLDs are the most efficient at cleaning up malicious content compared to legacy gTLDs and (European/non-European) ccTLDs. For example, the chance of accessing a malicious payload on a non-European ccTLD ten days after it has been blacklisted is about 15% higher compared to registered domains of new gTLDs.

Figure 5 shows the survival analysis of the 15 most used TLDs in phishing (APWG, OpenPhish, and PhishTank data), while Table 8 shows the mean and median uptimes. In Table 8, all Freenom TLDs (.tk, .ml, .cf, .ga, and .gq) are in the top 15 TLDs most commonly used for phishing. However, the median for cleaning actions is 1 hour for all of them, which is 12 times faster than for the `.net` and `.com` TLDs. It is probably since, as mentioned earlier, registrants of free domain names do not have the status of "licensee" but are treated as "users" of the domain name. Therefore, Freenom can take down any domain name without notice (see Section 8 for more details).

Interestingly, in the case of the `.ir` ccTLD (the second most abused TLD for phishing) the median is 0, while the average uptime is about 4 hours. It is because for at least 50 percent of the phishing URLs in the `.ir` TLD, the takedown action was carried out within the first 5 minutes after the URL was blacklisted, while the median value is skewed by a few samples that took up to 21 days to be removed. Overall, the `.ir` TLD demonstrated the highest efficiency (Figure 5) in cleaning up phishing content compared to the other abused TLDs in the top 15 list (Table 8).
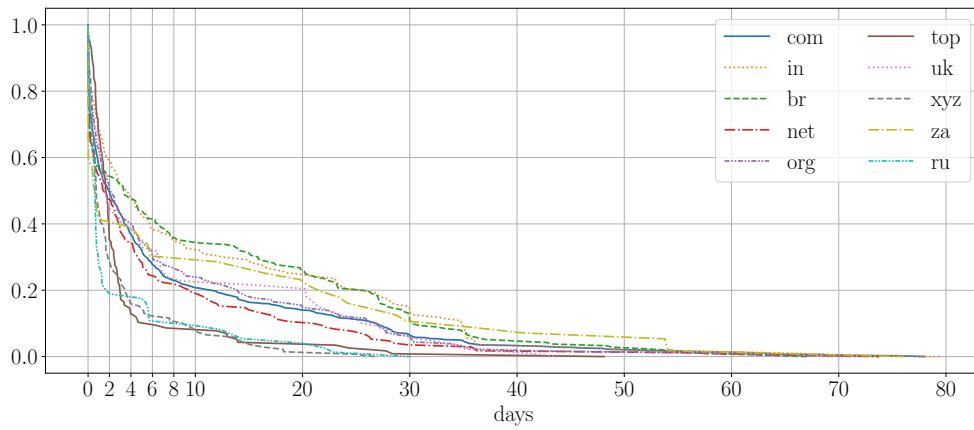
Figure 3: Survival analysis of the top 10 most abused TLDs for malware delivery.
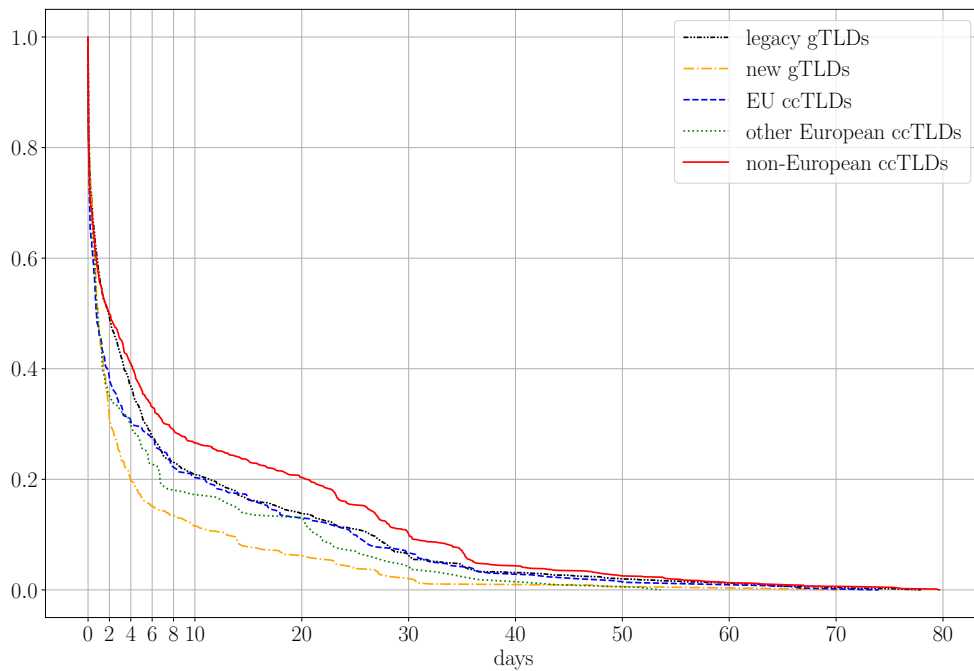
Figure 4: Survival analysis for different types of TLDs for malware delivery.
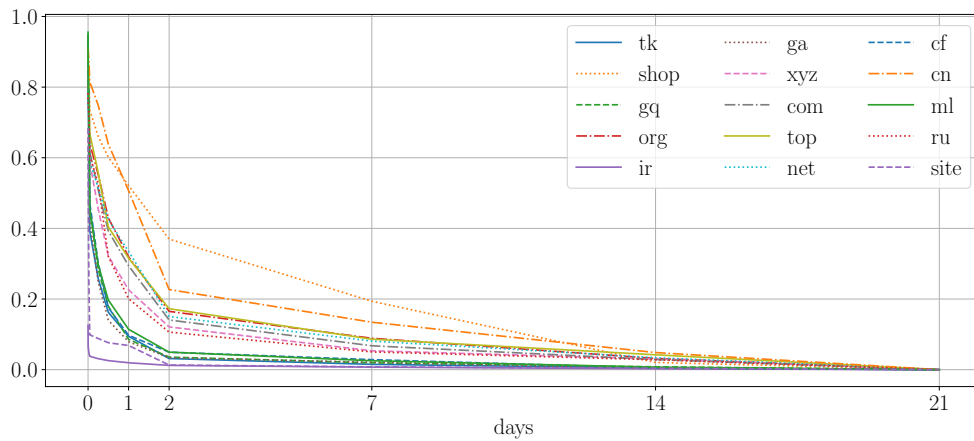
Figure 5: Survival analysis of the 15 most abused TLDs for phishing.

TLD	# of domains	mean	median
com	20,577	2 days 03:44:20	0 days 12:00:00
ir	3,476	0 days 04:18:27	0 days 00:00:00
xyz	2,578	1 days 20:22:06	0 days 06:00:00
tk	1,460	0 days 15:34:26	0 days 01:00:00
net	1,150	2 days 09:14:22	0 days 12:00:00
ml	989	0 days 21:19:05	0 days 01:00:00
org	945	2 days 12:11:45	0 days 12:00:00
ga	859	0 days 16:06:09	0 days 01:00:00
site	760	0 days 07:33:18	0 days 00:30:00
shop	733	4 days 12:34:25	2 days 00:00:00
ru	677	1 days 17:30:50	0 days 12:00:00
cf	667	0 days 20:55:12	0 days 01:00:00
gq	618	0 days 17:56:30	0 days 01:00:00
top	475	2 days 14:16:02	0 days 12:00:00
cn	454	3 days 15:13:22	2 days 00:00:00

Table 8: Uptimes of the phishing URLs for top 15 most abused TLDs.

9 Malicious versus Compromised Domains

9.1 Motivation

The distinction between compromised and malicious domains is essential for intermediaries involved in the registration and hosting. When a malicious URL/domain name is encountered (i.e., a URL/domain serving harmful/illegal content), it is crucial to assess the intentions of the domain registrant, as mitigation actions may vary if the registration is done for malicious purposes or not. While some domains are registered exclusively for malicious purposes, others are benign but get compromised (usually exploiting web vulnerabilities) and misused to serve malicious content. In many cases, there is no clear-cut distinction between DNS technical versus content-related abuse.

In the remainder of this section, we describe the method used to distinguish between compromised and maliciously registered domain names and present key statistics for TLDs and types of abuse. Since registrars cannot prevent the abuse of vulnerable hosting software (unless they provide both services), in the following section, we present registrar reputation metrics based on domain names categorized as maliciously registered.

9.2 Methodology

There exist two main approaches used in distinguishing compromised from maliciously registered domain names. COMAR (Classification of COmpromised versus MAliciously Registered Domains) [1] is a tool based on a machine-learning approach developed at Grenoble INP in close collaboration with and funded by two registry operators SIDN (registry of .nl domains) and AFNIC (registry of .fr domains). COMAR uses a set of 38 features and can achieve high accuracy without using any non-publicly available data, which makes it suitable for use by any organization.

Another alternative approach [27] proposed by Greg Aaron and Rod Rasmussen, used a set of heuristics to distinguish maliciously registered domains from compromised websites. A domain name is considered malicious if it was reported shortly after registration or contained a brand name or was registered in a batch, or there was a pattern indicating common owner or intent. Similar approaches have been used by researchers in various academic papers and industry reports [21, 28, 29].

In this study, we use a method similar to the second approach. We automatically flag a domain as maliciously registered if it was registered in a batch (i.e., among the blacklisted URLs, there are at least two domain names registered with the same registrar and at precisely the same time). We also automatically flag a domain name as maliciously registered if the time between registration and blacklisting does not exceed three months. The period was determined based on a sample of manually labeled spam, phishing, malware, and C&C domains and by tuning the parameter. The parameter giving the highest accuracy is equal to 98 days. We set the threshold to three months because it does not degrade the accuracy on the sample of manually labeled data and is consistent with previous studies [21, 28]. For phishing attacks, we build a list of 230 brand names. From the list of enumerated brand names, we generate a list of misspelled versions of brand names using standard methods such as omission, insertion, character substitution, and homographs. If a given FQDN contains a brand name or its misspelled version, the registered domain is also considered to be maliciously registered. We exclude all free service provider domains from the classification because they are neither compromised nor maliciously registered, and the domains for which we were unable to collect registration information.
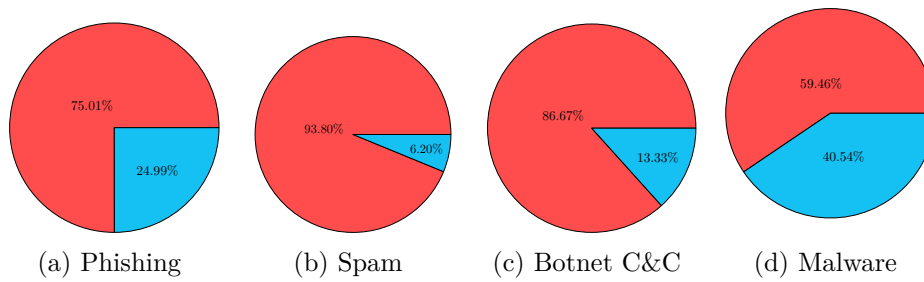
Figure 6: Distribution of compromised (blue) and maliciously registered (red) domain names per abuse type.

9.3   TLD Statistics: Malicious vs. Compromised

Figure 6 shows the distribution of compromised (blue) and maliciously registered (red) domain names per abuse type. The vast majority of spam domains (94%) are presumably maliciously registered, which is expected since they are typically used only to send an email, and attackers need to have complete control over them (maliciously registered domains ensure this). Alternatively, attackers can gain access to a legitimate email server and use it to send emails on behalf of benign (reputed) domain names. Domain name (email) spoofing is another attack vector described in Section 16.

We observe similar results for botnet C&C domain names. Most of them (87%) are maliciously registered. The remaining 13% of alleged compromised domains is more than we expected. That said, there are cases where compromised machines (domain names) are used in C&C communications as intermediate proxy components that act as an interface between the main C&C panel and bots (e.g. LokiBot[31]).

About 25% and 41% of phishing and malware domains are presumably benign but compromised at the hosting level. These proportions may not fully reflect the attacker preferences, as some blacklist providers may develop methods to detect maliciously registered domains rather than compromised websites. Therefore, the second group may be underrepresented.

In the cases of compromised domains, TLD operators generally should not mitigate abuse at the DNS level because it can cause collateral damage to their benign owners and visitors. Cybercriminals, when looking for the suitable means to, for example, distribute malicious content such as phishing or malware, must choose the most convenient strategy depending on their skills and available financial resources. Malicious (and le-

---

[31]https://www.virusbulletin.com/virusbulletin/2020/02/lokibot-dissecting-cc-panel-deployments/

gitimate) users have a vast choice of TLDs whose registrars compete on price, available options such as payment methods using (stolen) PayPal accounts or cryptocurrencies, etc. Therefore, it is not surprising that malicious registrations remain a more attractive option than exploiting website vulnerabilities.

(a) EU ccTLDs   (b) European ccTLDs   (c) other ccTLDs   (d) legacy gTLDs   (e) new gTLDs
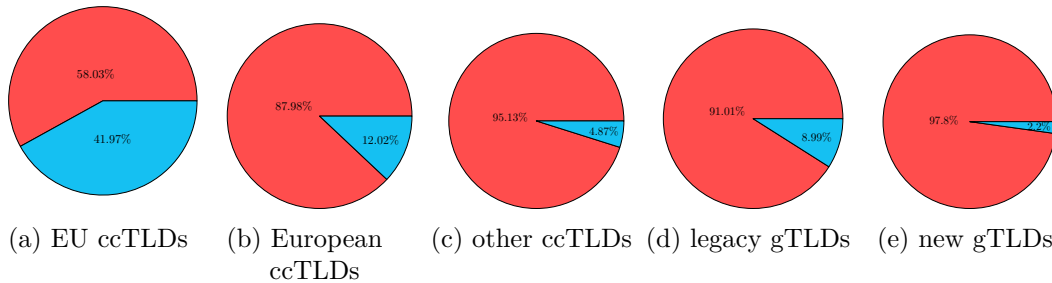
Figure 7: Distribution of compromised (blue) and maliciously registered (red) domain names per TLD type.

Figure 7 shows the distribution of compromised (blue) and maliciously registered (red) domains in different TLD markets. The vast majority of new gTLD domain names (97.8%) were flagged as maliciously registered. One likely reason is the low price of some of the new gTLDs. However, previous research [21, 30] also shows that in general, for new gTLDs, a relatively large proportion of domain names are either parked or have no content (DNS or HTTP errors), compared to legacy gTLDs. Intuitively, only domain names serving content are vulnerable to certain types of exploits and can be compromised.

For ccTLDs in the EU, compromised websites account for 42% of all abused domain names. ccTLDs do not generally aggressively compete on price. Moreover, ccTLDs are proactive in introducing preventive and reactive security measures to combat DNS abuse more effectively (e.g. .eu, .nl, .dk, .se, .fr, .cz, .sk). Consequently, they generally suffer from much less abuse (see Figure 1), in particular malicious registrations. To the best of our knowledge, no research has systematically classified the type of content served by ccTLDs (parked domains, HTTP/DNS errors, etc.). We hypothesize that speculative registrations in EU ccTLDs are less prevalent (fewer parked domains, fewer domains that fail to resolve) and thus more sites serving content and potentially vulnerable to exploits.

10 Registrar Reputation Metrics

10.1 Methodology

To measure the reputation of each registrar, similarly to TLDs, we use the *occurrence* of registered domain names and *rate* as security metrics. We collected and parsed the registration data of domain names of blacklisted URLs (or domains if URLs were not present) as soon as they were blacklisted. As mentioned in the previous section, registrars can only prevent malicious registrations (proactive measures) and remove maliciously registered domains (reactive measures) at the DNS level. They have no control over the hosting infrastructure (unless they also provide a hosting service). Therefore, we compute reputation metrics for domain names that we have determined to be malicious.

To compute abuse rates for each registrar, we had to estimate their sizes. As described in detail in Section 5, using the RDAP and WHOIS protocols, we collected registration information for approximately 241 million domain names (96% of all active domains we enumerated). We were able to parse the registration information and match the domain names of about 85% of the RDAP/WHOIS records to their respective registrars. We calculate rates as the number of maliciously registered domains per 10,000 registrations.

Name	Size	Market share (%)
GoDaddy.com, LLC	63,522,904	30.84
NameCheap, Inc.	10,901,924	5.29
Tucows Domains Inc.	9,492,927	4.61
Network Solutions, LLC	6,393,947	3.10
Alibaba Cloud Computing (Beijing) Co., Ltd.	5,668,641	2.75
Google LLC	5,342,956	2.59
1&1 IONOS SE	4,861,279	2.36
eNom, LLC	4,650,888	2.26
PDR Ltd. d/b/a PublicDomainRegistry.com	4,564,240	2.22
TurnCommerce, Inc. DBA NameBright.com	3,583,210	1.74
GMO Internet, Inc. d/b/a Onamae.com	3,403,676	1.65
OVH sas	3,208,371	1.56
NameSilo, LLC	3,166,460	1.54
Wild West Domains, LLC	2,842,400	1.38
FastDomain Inc.	2,272,984	1.10

Table 9: Top 15 registrars based on the estimated overall domain market share.

10.2 Results

Table 9 shows the estimated market share and the top 15 registrars based on the collected and parsed registration data. Therefore, the reported sizes of each registrar constitute the lower bound of the actual number of registered domain names mainly because of no access to all the domain names globally and not available registration information (WHOIS/RDAP). Godaddy[32] owns 30.84% of the market by having more than 63 M domains registered, followed by NameCheap[33], Tucows[34] and Network Solutions.[35] The presented size estimations are in agreement with the independent market share analysis preformed by `domainstate.com`.[36]

Table 10 shows the top 30 registrars with the highest number of registered (abused) domains and the corresponding abuse rates. NameCheap suffers from the highest amount of abused domains followed by GMO Internet, Inc. d/b/a Onamae.com (with IANA ID 49), and GoDaddy.com, LLC (IANA ID 146). The top five most abused registrars account for 48% of all maliciously registered domain names. However, in terms of abuse rates, Xi'an Qianxi Network Technology Co. Ltd and EIMS (Shenzhen) Culture & Technology Co., Ltd have the highest abuse rates of 6,921 and 2,366 maliciously registered domains per 10,000 registrations, respectively (see Table 11).

The metrics presented here cannot be directly compared to those described in the SADAG study [28] due to differences in data collection and aggregation, and the much shorter time frame of the data analyzed. The SADAG report revealed registrars with disproportionately high concentrations (absolute and relative) of abused domains, such as Nanjing Imperiosus Technology or Alpnames Limited,[37] whose accreditation was revoked by ICANN. Both registrars willingly or unwillingly facilitated cybercrimes. Note that we focused on ICANN-accredited registrars in this study because mapping of domain names to the appropriate registrars is complex and error-prone if they are not ICANN-accredited (there is no IANA ID) or WHOIS data is missing. Consequently, it is much more challenging to identify domain resellers and non-ICANN accredited registrars that potentially facilitate domain name abuse.

We manually analyzed the registration patterns of maliciously registered domain

---

[32]https://godaddy.com
[33]https://www.namecheap.com
[34]https://tucows.com
[35]https://www.networksolutions.com
[36]https://www.domainstate.com/registrar-stats.html
[37]http://domainincite.com/22659-tech-giants-gunning-for-alpnames-over-new-gtld-abuse

Name	IANA ID	# of domains	Rate
NameCheap, Inc.	1068	131,925	121
GMO Internet, Inc. d/b/a Onamae.com	49	93,905	276
GoDaddy.com, LLC	146	53,185	8
NameSilo, LLC	1479	52,188	165
PDR Ltd. d/b/a PublicDomainRegistry.com	303	38,804	85
Alibaba Cloud Computing (Beijing) Co., Ltd.	420	35,242	62
PSI-USA, Inc. dba Domain Robot	151	23,485	181
ALIBABA.COM SINGAPORE E-COMMERCE PRIVATE LIMITED	3775	22,139	321
Xin Net Technology Corporation	120	18,497	110
Hongkong Domain Name Information Management Co....	2251	16,000	800
Key-Systems GmbH	269	15,056	87
Dynadot, LLC	472	14,835	69
Web Commerce Communications Limited dba WebNic.cc	460	11,700	324
Launchpad.com Inc.	955	11,251	154
Eranet International Limited	1868	10,097	623
REGRU-RU	—	9,598	71
Wild West Domains, LLC	440	9,389	33
Hong Kong Juming Network Technology Co., Ltd	3855	8,478	721
Registrar of Domain Names REG.RU LLC	1606	7,396	151
Cloud Yuqu LLC	3824	7,025	298
Hostinger, UAB	1636	7,012	173
Register.com, Inc.	9	6,926	37
eNom, LLC	48	6,534	14
DNSPod, Inc.	1697	6,389	84
Chengdu West Dimension Digital Technology Co., ...	1556	6,385	80
OVH sas	433	6,318	20
Sav.com, LLC	609	6,162	118
Beget LLC	3806	5,283	465
Name.com, Inc.	625	5,040	27
Xiamen 35.Com Technology Co., Ltd.	1316	4,531	247

Table 10: Top 30 registrars with the highest occurrence and rates of abuse.

names for a few the most abused registrars. For NameCheap, Inc., for example, we found no registered domain with the keyword "facebook", which may be a consequence of the disagreement between Facebook and NameCheap last year regarding the sale of suspicious domains similar to that of Facebook.[38] Such proactive anti-cybersquatting measures significantly increase the barriers to abuse, especially for less-skilled cybercriminals. Nevertheless, such measures do not eliminate the problem of deceptive subdomains. An attacker can register a random second-level domain name and then create a subdomain containing deceptive keywords. However, assuming that the registrar controls the zone file, such a subdomain name can be easily detected when added to the zone file. Therefore, to avoid detection, the malicious user must delegate the registered domain name to its authoritative name server and add the deceptive subdomain. In this way, the registrar is unable to detect the malicious subdomain at the DNS level.

Interestingly, for NameCheap, Inc., we identified only three likely maliciously registered domains with subdomains containing the deceptive keyword "facebook" (e.g.,

---

[38]https://www.zdnet.com/article/facebook-sues-namecheap-to-unmask-hackers-who-registered-malicious-domains

Name	IANA ID	# of domains	Rate
Xi'an Qianxi Network Technology Co. Ltd.	3825	454	6,921
EIMS (Shenzhen) Culture & Technology Co., Ltd	2485	2,337	2,366
Tencent Cloud Computing (Beijing) Limited Liabi...	3755	2,315	2,351
Global Domain Name Trading Center Ltd	3792	892	1,231
FLAPPY DOMAIN, INC.	1872	1,538	1,097
DotMedia Limited	1863	925	1,037
DOMAINNAME BLVD, INC.	1870	903	1,001
DOMAIN ORIENTAL LIMITED	3252	428	972
DOMAINNAME FWY, INC.	1871	715	907
MainReg Inc.	1917	182	836
Hefei Juming Network Technology Co., Ltd	3758	3,180	798
Hongkong Domain Name Information Management Co....	2251	16,000	800
NICENIC INTERNATIONAL GROUP CO., LIMITED	3765	987	726
Hong Kong Juming Network Technology Co., Ltd	3855	8,478	721
Shinjiru Technology Sdn Bhd	1741	908	601
Eranet International Limited	1868	10,097	623
BR domain Inc. dba namegear.co	1898	143	586
AppCroNix Infotech Private Limited, d/b/a VEBON...	3844	35	585
Intracom Middle East FZE	1875	1	500
Taka Enterprise Ltd	1726	2	455
Beget LLC	3806	5,283	465
Topnets Group Limited	3805	2	426
Vertex Names.com, LLC	1665	2	408
DomainName Path, Inc.	1907	1,351	401
DOMAIN NAME NETWORK PTY LTD	1527	795	380
Dynadot4 LLC	1652	523	377
Atak Domain Hosting Internet ve Bilgi Teknoloji...	1601	3,710	348
Net-Chinese Co., Ltd.	1336	2,265	345
Dynadot10 LLC	1865	480	333
CyanDomains, Inc.	1899	900	323

Table 11: Top 30 registrars with the highest rate of abuse.

`facebook.com.marketplace-item-id-135921470.com`). A manual analysis of phishing data indicates that NameCheap prevents malicious registration of selected brand names. We found 113 domain names containing the keywords "wells" and "fargo" (e.g., `wells-fargo-page.mobi`, or `wells-fargo-login-security.work`). In comparison, for GoDaddy.com, LLC (IANA ID 146), we found maliciously registered domain names containing "paypal", "facebook" and "wellsfargo" in our dataset.

> **Recommendation**: TLD registries are encouraged to offer, directly or through the registrars or resellers, services allowing intellectual property rights (IPR) holders to preventively block infringing domain name registrations (similar to services already existing on the gTLD market).

To effectively reduce domain name abuse, TLD registry operators should maintain access to existing domain/URL blacklists and partner with trusted notifiers. TLD reg-

istries should identify the registrars with the highest concentration and rates of DNS abuse in their ecosystems.

One limitation of this study is that we focus on ICANN-accredited registrars and have limited access to domain names, especially ccTLDs, and limited access to WHOIS. Operators of ccTLD registries that work with ICANN-accredited registrars, but also accredit registrars locally, have a complete picture of their local market. They should consider developing incentive structures to encourage their registrars to develop methods to prevent malicious registrations effectively, such as reduced domain registration fees for registrars with the least amount of abuse.

> **Recommendation**: TLD registry operators are encouraged to:
> - maintain access to existing domain/URL blacklists,
> - identify the registrars with the highest and lowest concentrations and rates of DNS abuse in their ecosystems,
> - develop incentive structures to encourage their registrars to develop methods to prevent and mitigate malicious registrations effectively.

Table 12, shows uptimes (medians and means) of *maliciously registered* phishing domains for the top 30 most abused registrars (for APWG, OpenPhish, and PhishTank data sources). Figure 8 shows the survival analysis of the top 5 most abused registrars, corresponding to Table 12. NameCheap, Inc. suffers from the highest number of maliciously registered domain names (5,774). However, Figure 8, shows that in the second quarter of 2021, it mitigated abuse faster compared to other registrars. The median uptime for NameCheap is 6 hours, and the average is one day and 6 hours. Although it accounts for the highest number of maliciously registered domain names, the survival analysis shows that it is one of the fastest registrars in terms of remediation.

11 Hosting Providers Reputation

11.1 Motivation

Malware distribution domains, CSAM, hate speech, intellectual property infringement, or phishing attacks spread harmful and illegal content. For example, as mentioned earlier, phishing attacks contain deceptive content asking for sensitive information,

Registrar	count	mean	median
NameCheap, Inc.	5,774	1 days 06:50:06	0 days 06:00:00
NameSilo, LLC	1,928	1 days 12:41:29	0 days 12:00:00
Registrar of Domain Names REG.RU LLC	1,025	2 days 07:57:14	0 days 01:00:00
GoDaddy.com, LLC	705	3 days 16:22:11	1 days 00:00:00
PDR Ltd. d/b/a PublicDomainRegistry.com	587	1 days 08:29:26	0 days 12:00:00
GMO Internet, Inc. d/b/a Onamae.com	475	2 days 00:55:39	1 days 00:00:00
Tucows Domains Inc.	409	1 days 07:43:38	0 days 12:00:00
Wild West Domains, LLC	392	1 days 22:08:03	1 days 00:00:00
REGRU-RU	186	1 days 07:44:13	0 days 12:00:00
Alibaba Cloud Computing (Beijing) Co., Ltd.	169	4 days 16:44:01	2 days 00:00:00
Hostinger, UAB	162	0 days 06:43:49	0 days 01:00:00
Squarespace Domains LLC	151	0 days 15:58:04	0 days 12:00:00
Name.com, Inc.	146	1 days 05:15:45	1 days 00:00:00
Google LLC	129	2 days 14:35:48	2 days 00:00:00
Web Commerce Communications Limited dba WebNic.cc	122	1 days 00:03:31	0 days 06:00:00
Alibaba Cloud Computing Ltd. d/b/a HiChina (www...	110	7 days 03:42:40	2 days 00:00:00
Key-Systems, LLC	109	6 days 20:44:35	2 days 00:00:00
Hosting Concepts B.V. d/b/a Registrar.eu	101	0 days 21:17:55	0 days 06:00:00
West263 International Limited	95	10 days 19:38:31	14 days 00:00:00
Porkbun LLC	93	2 days 13:49:01	0 days 12:00:00

Table 12: Uptimes of maliciously registered domain names used in phishing for the top 20 most abused registrars (in terms of abuse counts).
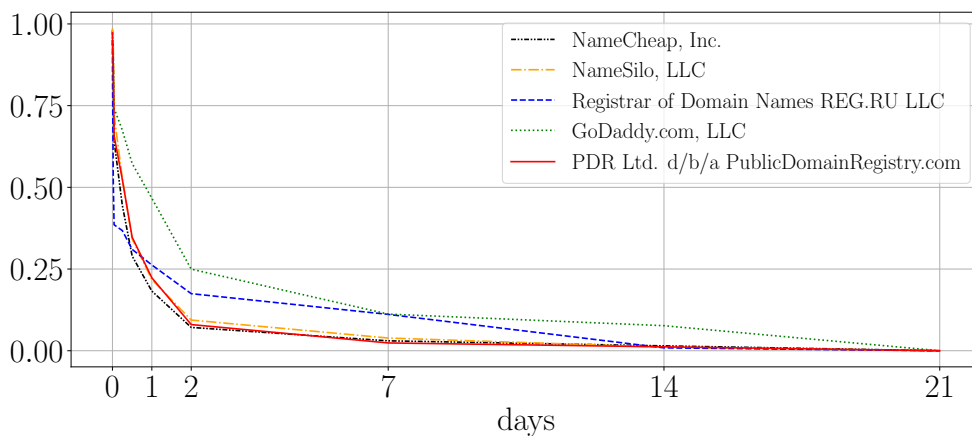
Figure 8: Survival analysis of the top five most abused registrars for phishing.

usually infringing on the trademarks of companies such as banks, payment services, and social networking sites. Malware attacks are used to distribute and infect end users with malicious software. CSAM attacks facilitate access to files that harm minors. Therefore, in addition to potential DNS infrastructure abuse, they often represent content abuse.

Some other domains, such as spam (used to send emails containing, for example, phishing content), botnet C&C, IP fast-flux infrastructures, or domain names redirecting users to other malicious websites (for example, phishing or malware distribution), may not host any malicious or illegal content. Nevertheless, they are involved in attacks and generally resolve to IP addresses and thus, abuse the hosting infrastructure.

Therefore, we resolve each domain name to its IP address and build reputation metrics for hosting providers, more specifically, information society service providers (ISSPs), including access, hosting, and online platform providers.[39] We consider both maliciously registered and compromised domains and exclude legitimate services such as free subdomain providers, which we discuss later.

11.2 Methodology

Reliable reputation metrics must account for the commonly observed trend that larger hosting providers also experience more abuse [31]. A common way to measure the "size" of hosting providers is the number of IP addresses routed through the corresponding AS (Autonomous System). However, not all IP addresses routed through an AS are used to host content. IP addresses may be leased and used for other purposes. Inaccuracies in the size estimation can adversely affect the validity of the metric and lead to erroneous results [31]. Due to the simplicity of computation, the advertised IP space remains a popular choice for size estimation. Another approach is to calculate the portion of the routed IP address space used for hosting as the size estimate. However, this approach is also not free from bias because it favors hosting providers with a disproportionately large volume of shared hosting (e.g., 10,000 domains hosted at a single IP address). Therefore, we use the number of hosted second-level domains as an estimator, which treats shared hosting fairly.

Our approach is also not without bias. For example, infrastructure providers may lease their IP space to other, smaller providers such as hosting services. There may be a chain of resellers difficult to identify even for AS operators. For example, Leaseweb

---

[39]https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52016DC0872

provides services to businesses and generally cannot directly control the end users who may host malicious content. However, victims typically contact them in case of abuse, and AS operators should directly contact the reseller, which should mitigate hosting abuse.

To estimate the reputation of hosting providers, we first collect the 'A' records of each domain. To get the location information of providers we use the MaxMind dataset.[40] It provides us with both the country code and the AS number associated with the 'A' record of a domain name. If the domain name has more than one 'A' record then we collect the set of ASes from which the domain name is hosted.

We also explore the difference between the set of ASes associated with the FQDN and the registered domains. We found that 97% of abused domains have the same set of ASes for FQDN and registered domain names, which means that if an attacker registers a domain name maliciously, the chance of managing different subdomains in different ASes is as low as 3%, as this may increase the cost of, for example, renting servers in different locations.

11.3   Results

Table 13 shows the top 10 ASes grouped into 5 different sizes i.e., ASes with less than 10 K domains, between 10 K and 100 K domains, between 100 K and 1 M domains, and finally greater than 1 M domains. The intuition behind grouping ASes is to compare relatively equal size ASes (in terms of hosted domains) with each other.

Abuse rates for the smallest providers should be considered with caution. Similarly to TLDs or registrars, even a small number of abused domains can significantly affect their reputation. Among the smallest providers, we may notice universities and educational networks. Security researchers may deploy honeypots or, for example, harmless sites simulating phishing that can get blacklisted. For example, SURF B.V.[41] is an ICT collaborative organization in Dutch education and research. Researchers often use their infrastructure in cybersecurity research (e.g., to study phishing [32]).

The second group of providers, in terms of the size, is widely used by spammers. In the top ten most abused providers, the percentage of spam domains ranges from 1,275 to 3,430 per 10,000 registrations (12,75% to 34,3% of all domains). Interestingly, some

---

[40]https://www.maxmind.com
[41]https://www.surf.nl

Botnet C&C			Malware		
AS	# Domains	Rate	AS	# Domains	Rate
ACCESS INTERNET	1	10,000	Dinas Komunikasi dan Informatika Kota Medan	1	10,000
LEGON TELECOMUNICACIONES SAS	1	10,000	VietServer Services technology company limited	1	10,000
GLOBALWEB S.R.L.	1	10,000	Rajabhat Mahasarakham University	1	5,000
SURF B.V.	352	6,132	Dm Lot Infotech Solutions Private Limited	1	5,000
UNICOM JiangSu WuXi IDC network	3	4,286	PT Widara Media Informasi	1	3333
Trans Tel Services SRL	1	3,333	WideBand Communications(Pvt)Ltd	1	2,500
TAQUARANET SERVICOS DE PROVEDOR DE INTERNET LTDA M	1	3,333	university of dhaka	1	2,500
NETCOM Ltd.	3	3,000	Universitas Brawijaya	1	2,500
ERITEL-AS	5	2,500	SAYDA LLC	23	1,667
Bangladesh University of Engineering and Techno...	1	2,000	Electrosim Srl	3	1,667

(0, 10k)

Botnet C&C			Malware		
AS	# Domains	Rate	AS	# Domains	Rate
MICROSOFT-CORP-AS	90	69	DataWeb Global Group B.V.	213	116
CHINATELECOM JiangSu YangZhou IDC networkdescr:...	11	9	CDSI	58	54
HURRICANE	82	9	WII	119	25
Leaseweb Deutschland GmbH	27	7	HOSTROCKET	30	18
Fuzhou	7	6	AZDIGI Corporation	27	15
FRONTIER-FRTR	7	5	Host-Africa-AS	18	14
Claranet Limited	17	4	ZA-1-Grid	39	13
FEDERAL-ONLINE-GROUP-LLC	9	4	IOFLOOD	82	12
AS Number for CHINANET jiangsu province backbone	7	4	Telepoint Ltd	40	12
G-Core Labs S.A.	4	4	Shinjiru Technology Sdn Bhd	13	9

Phishing			Spam		
AS	# Domains	Rate	AS	# Domains	Rate
PT. Jupiter Jala Arta	1,573	1,115	GROUP-IID-01	12,282	3,430
Shinjiru Technology Sdn Bhd	971	703	Equinix Japan Enterprise K.K.	8,205	3,305
G-Core Labs S.A.	658	603	FEDERAL-ONLINE-GROUP-LLC	7,139	3,292
Asiatech Data Transmission company	1,973	571	EONIX-COMMUNICATIONS-ASBLOCK-62904	9,165	3,009
LLC Smart Ape	605	518	Network-Transit	5,592	1979
CONTABO	772	267	SANREN DATA LIMITED	8,065	1,605
Guangdong Mobile Communication Co.Ltd.	608	267	DataWeb Global Group B.V.	2,740	1,488
Netmihan Communication Company Ltd	592	203	TIER-NET	2,577	1,331
HOSTWINDS	979	191	SERVER-MANIA	2,133	1,312
Host-Africa-AS	219	169	H4Y-TECHNOLOGIES	1,332	1,275

(10k, 100k)

Botnet C&C			Malware		
AS	# Domains	Rate	AS	# Domains	Rate
STEADFAST	1,490	117	AS-COLOCROSSING	112	10
Private Layer INC	57	4	PUBLIC-DOMAIN-REGISTRY	424	8
LeaseWeb Netherlands B.V.	50	2	NOCIX	207	7
CHINA UNICOM China169 Backbone	23	2	DIMENOC	124	6
Gigabit Hosting Sdn Bhd	16	2	Contabo GmbH	146	5
Tencent Building, Kejizhongyi Avenue	47	1	xneelo	74	4
CNSERVERS	42	1	GO-DADDY-COM-LLC	94	3
MICROSOFT-CORP-MSN-AS-BLOCK	73	1	SERVERCENTRAL	54	3
Chinanet	16	1	SINGLEHOP-LLC	64	2
Shenzhen Tencent Computer Systems Company Limited	21	1	STEADFAST	28	2

Phishing			Spam		
AS	# Domains	Rate	AS	# Domains	Rate
AS-COLOCROSSING	2,319	198	Clayer Limited	75,874	1,992
ASN-QUADRANET-GLOBAL	4,399	65	PUBLIC-DOMAIN-REGISTRY	49,259	919
Contabo GmbH	1,670	53	HENGTONG-IDC-LLC	10,029	694
DEDIPATH-LLC	618	53	acens Technologies, S.L.	8,784	684
JSC The First	716	46	PEGTECHINC	28,749	561
BGPNET Global ASN	644	41	LEASEWEB-USA-LAX-11	14,805	427
MICROSOFT-CORP-MSN-AS-BLOCK	2,530	40	LEASEWEB-USA-SFO-12	5,819	347
Private Layer INC	597	38	POWER LINE DATACENTER	22,405	310
AS-CHOOPA	1,575	32	ABCDE GROUP COMPANY LIMITED	6,615	310
Online S.a.s.	629	27	AS-COLOCROSSING	3,051	261

(100k, 1M]

Botnet C&C			Malware		
AS	# Domains	Rate	AS	# Domains	Rate
Alibaba (US) Technology Co., Ltd.	119	1	UNIFIEDLAYER-AS-1	1,369	2
AMAZON-AES	149	1	CLOUDFLARENET	3,131	2
Hangzhou Alibaba Advertising Co.,Ltd.	58	1	NAMECHEAP-NET	644	1
Linode, LLC	39	0	Alibaba (US) Technology Co., Ltd.	92	1
DIGITALOCEAN-ASN	51	0	AMAZON-AES	201	1
Strato AG	62	0	Hetzner Online GmbH	192	1
1&1 Ionos Se	147	0	DIGITALOCEAN-ASN	102	1
EGIHOSTING	21	0	OVH SAS	231	1
CLOUDFLARENET	241	0	Host Europe GmbH	66	0
GOOGLE	390	0	AS-26496-GO-DADDY-COM-LLC	314	0

Phishing			Spam		
AS	# Domains	Rate	AS	# Domains	Rate
Alibaba (US) Technology Co., Ltd.	4,895	47	DXTL Tseung Kwan O Service	37,214	329
NAMECHEAP-NET	19,649	40	EGIHOSTING	20,592	186
UNIFIEDLAYER-AS-1	15,947	29	Alibaba (US) Technology Co., Ltd.	6,584	63
DIGITALOCEAN-ASN	4,564	23	AMAZON-AES	15,180	56
CLOUDFLARENET	15,217	12	AMAZON-02	35,236	49
Hetzner Online GmbH	2,894	11	OVH SAS	21,839	49
OVH SAS	2,755	6	DIGITALOCEAN-ASN	8,275	42
Verotel International B.V.	1,602	6	GMO Internet,Inc	3,822	34
Linode, LLC	603	5	Linode, LLC	3,784	32
AMAZON-02	3,381	5	CLOUDFLARENET	40,367	31

(1M, ∞)

Table 13: Top 10 AS with the highest absolute (# Domains) relative concentrations (Rate) of blacklisted domains grouped by their corresponding AS size and abuse type.

(a) Occurrence of abuse

(b) Rate of abuse
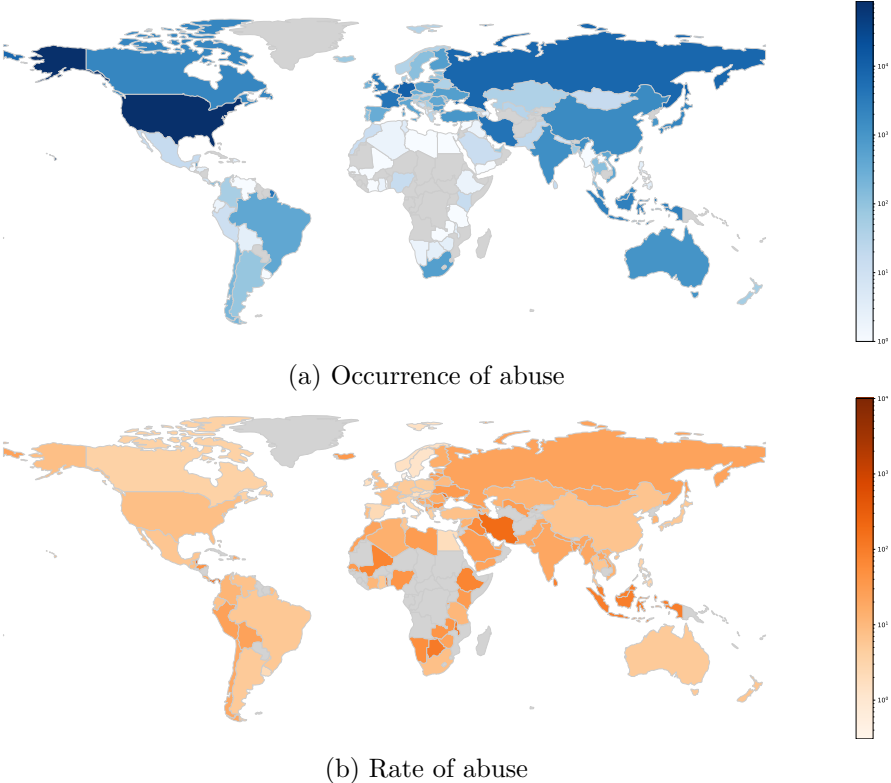
Figure 9: Distribution and rates of IP addresses related to phishing.

(a) Occurrence of abuse

(b) Rate of abuse
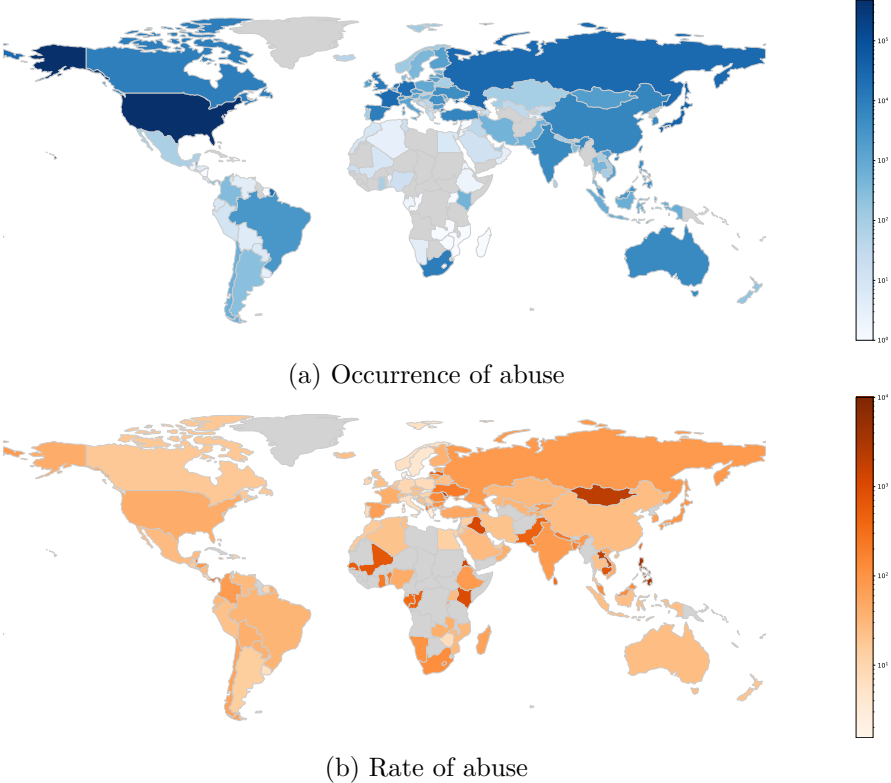
Figure 10: Distribution and rates of IP addresses related to spam.

(a) Occurrence of abuse

(b) Rate of abuse
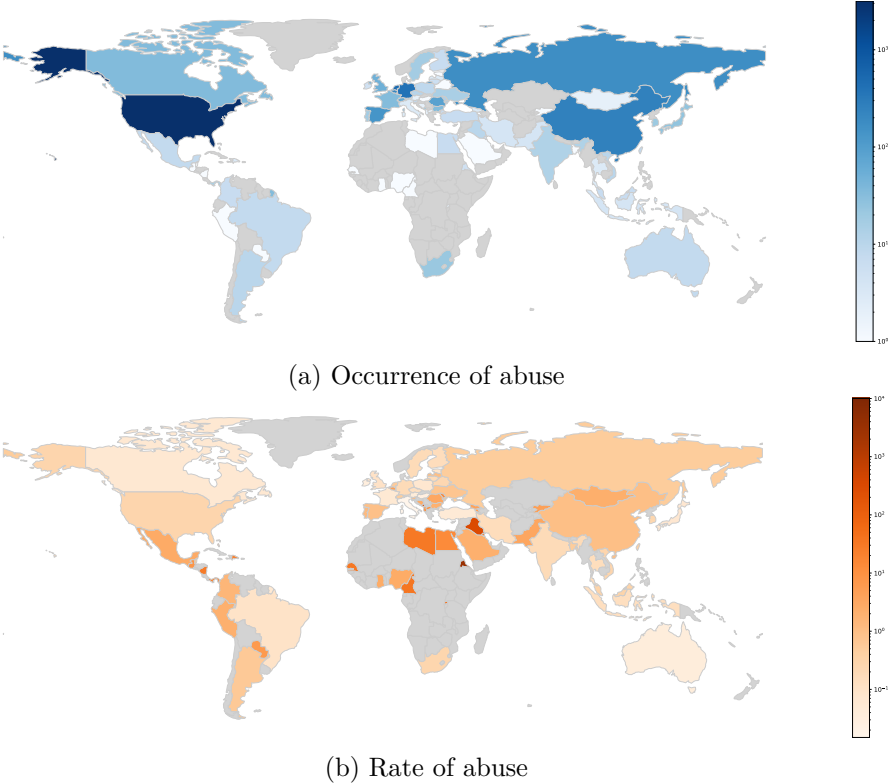
Figure 11: Distribution and rates of IP addresses related to botnet C&C.

(a) Occurrence of abuse

(b) Rate of abuse
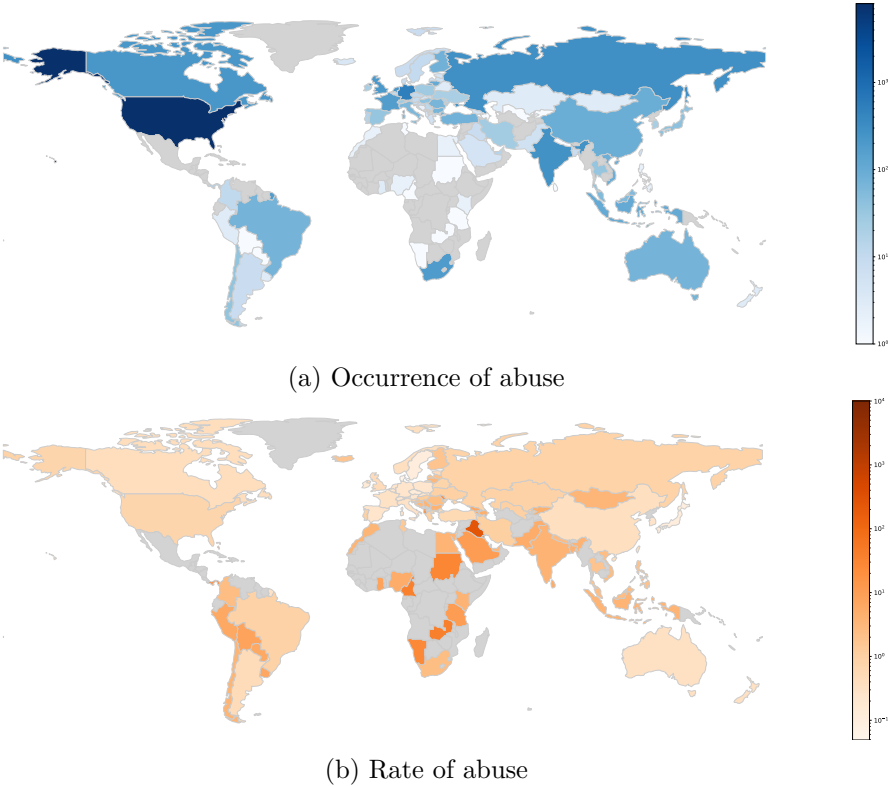
Figure 12: Distribution and rates of IP addresses related to malware distribution.

offshore providers (hosting resellers) claim to use the hosting infrastructure of some of these providers.[42] Note that abused hosting providers often offer very cheap servers, such as Contabo, and therefore attract legitimate and malicious users.

Among the largest and most abused hosting providers, we observe companies that also offer cheap domain registration, such as NAMECHEAP-NET, Alibaba (US) Technology Co., Ltd, or OVH SAS. Similarly, competitive pricing may be one of the main reasons why these companies suffer from abuse. CLOUDFLARENET is among the top ten most abused providers in all four abuse categories (malware, spam, phishing, and botnet C&C). Note that Cloudflare provides a reverse proxy service that is widely used by cybercriminals. When a criminal (or legitimate) site is behind a reverse proxy, the IP address of the backend server hosting, for example, a phishing site, is not visible. Instead, the domain resolves to the Cloudflare IP address, which affects its reputation.

The IP address distributions for each type of abuse and each country (phishing, spam, botnet C&C, and malware) are shown in Figures 9, 10, 11, 12, respectively. While in absolute terms, the hosting infrastructure in the US, the Russian Federation, and Germany are generally among the countries with the highest number of abuses, in terms of rates, these countries are relatively less affected.

> **Recommendation**: In a similar manner with respect to the TLD registries and the registrars, the abuse rates of hosting providers should be monitored on an ongoing basis by independent researchers in cooperation with institutions and regulatory bodies (e.g., European Commission, European Union Agency for Cybersecurity – ENISA or national authorities). Abuse rates should not exceed predetermined thresholds. Incentive structures should be studied to induce hosting providers to develop technical solutions that effectively curb hosting and content abuse.

---

[42]e.g., http://bbar.sottoglialberi.it/free-vps-for-ddosing.html

AS	count	mean	median
UNIFIEDLAYER-AS-1	3,853	3 days 09:19:31	1 days 00:00:00
CLOUDFLARENET	3,291	2 days 04:16:19	0 days 01:00:00
Asiatech Data Transmission company	1,933	0 days 03:14:37	0 days 00:00:00
NAMECHEAP-NET	1,671	2 days 02:01:42	0 days 06:00:00
Hetzner Online GmbH	1,633	0 days 11:49:26	0 days 00:00:00
OVH SAS	802	2 days 01:17:45	0 days 00:30:00
Netmihan Communication Company Ltd	577	0 days 00:26:55	0 days 00:00:00
AS-26496-GO-DADDY-COM-LLC	426	3 days 09:42:22	1 days 00:00:00
AMAZON-02	392	2 days 12:01:41	0 days 12:00:00
DIGITALOCEAN-ASN	387	3 days 05:09:37	1 days 00:00:00
Contabo GmbH	369	3 days 04:47:39	1 days 00:00:00
PUBLIC-DOMAIN-REGISTRY	365	3 days 21:38:50	1 days 00:00:00
ASN-QUADRANET-GLOBAL	257	3 days 18:07:48	1 days 00:00:00
INMOTION	251	3 days 07:41:06	1 days 00:00:00
DDOS-GUARD CORP.	226	0 days 11:33:01	0 days 00:00:00
Faraso Samaneh Pasargad Ltd.	214	0 days 00:22:27	0 days 00:00:00
GOOGLE	211	3 days 15:05:42	0 days 06:00:00
AS-COLOCROSSING	207	2 days 06:37:49	0 days 12:00:00
LIQUIDWEB	206	2 days 04:41:44	0 days 09:00:00
IMH-IAD	204	2 days 04:36:19	0 days 06:00:00

Table 14: Uptimes of the phishing URLs for top 20 most abused ASes.

Table 14, shows the uptime analysis (mean and median) of phishing URLs for the 20 most abused ASes. 'UNIFIEDLAYER-AS-1' suffers from the highest number of phishing domains (3,853 domains), followed by 'CLOUDFLARENET' (3,291 domains; as mentioned earlier, Cloudflare provides a proxy service, not hosting) and 'Asiatech Data Transmission company' (1,933 domains). Figure 13 shows the survival analysis of the ten most abused hosting providers shown in Table 14. Among these ten, 'Netmihan communication Company Ltd' demonstrated the fastest remediation in Q2 2021 with a median of zero and an average of 26 minutes, followed closely by 'Asiatech Data Transmission company. Results indicate that 'UNIFIEDLAYER-AS-1', which suffered from the highest number of phishing attacks, seems to be among the slowest in terms of remediation efforts.

12 Special Domains

Special domains are those whose original intent of registration is not malicious. They are registered by companies to provide legitimate services. Examples of such services include dynamic DNS services, free subdomains, and URL shorter systems. Due to the
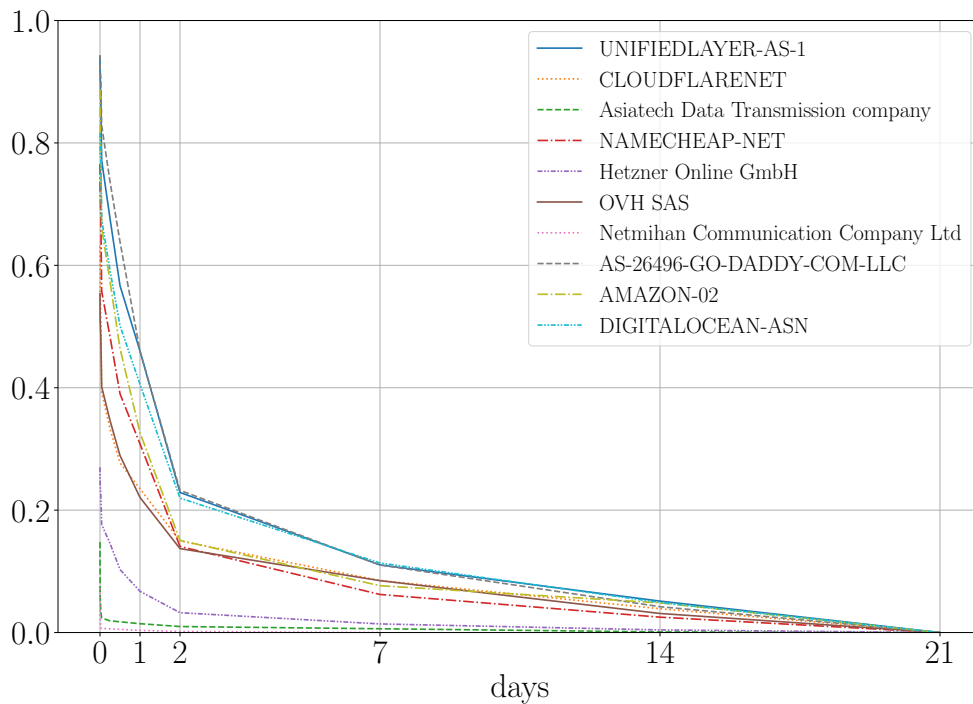
Figure 13: Survival analysis of the 10 most commonly abused ASes for phishing.

nature of these services, they are very enticing to cybercriminals. For example, free service operators, such as 000webhost[43], provide free subdomains and free hosting, which is very conducive especially to phishers.

Table 15 shows the top 10 service providers with the most blacklisted fully qualified domain names (FQDNs) by abuse type. ngrok[44] – the most abused service in phishing – is a tunneling reverse proxy that establishes secure tunnels from a public endpoint to a locally running web service. A malicious user can hide their backend server (e.g., a compromised end-host with a local IP address) and can use a subdomain under ngrok.io to launch a phishing campaign (e.g., user-account-verification.ngrok.io). As we can see, phishers make heavy use of such services because, in most cases, they incur no (or meager) cost, which makes them practical for serving malicious content. However, these services are less suitable for spam attacks because criminals use domains to send fake/deceptive emails rather than directly deliver content. Similarly, for botnet C&C communications, attackers generally need to control domains to ensure communication between compromised end-hosts and malicious servers.

Subdomains typically contain brand names to lure their victims into entering, for

---

[43]https://000webhost.com
[44]https://ngrok.com

Botnet C&C		Malware		Phishing		Spam	
Provider	# Domains	Provider	# Domains	Provider	# Domains	Provider	# Domains
Duck DNS	9	dns.army	208	ngrok	23,531	Google Cloud	118
ChangeiP	3	NoIP	92	000webhost	16,867	Google Firebase	30
000webhost	2	000webhost	41	Google Firebase	13,371	NoIP	14
		Duck DNS	32	Duck DNS	7,252	amazonaws.com	12
		amazonaws.com	23	Google Cloud	5,440	wixsite.com	11
		soundcast.me	14	NoIP	4,004	blogspot.com	6
		DynuDNS	10	weebly.com	3,853	IBM cloud	6
		tmweb.ru	4	ChangeiP	3,340	glitch.me	5
		weebly.com	3	tmweb.ru	3,125	Duck DNS	4
		blogspot.com	2	yolasite.com	1,952	netlify.app	4

Table 15: Top 10 special service providers with the highest occurrence of blacklisted FQDNs per abuse type.

example, credentials (in the case of phishing attacks). Since such services are commonly abused, providers should employ advanced preventive and reactive solutions to curb subdomain name abuse and hosting infrastructure. They should proactively detect suspicious domain names containing brand keywords most commonly targeted by attacks. They should also work closely with the most attacked companies and develop trusted notifier programs to remove subdomain names with offensive and abusive content effectively and timely.

name	mean	median
appspot.com	5 days 07:46:21	2 days 00:00:00
blogspot.com	3 days 02:19:00	0 days 12:00:00
IBM cloud	1 days 19:38:31	0 days 12:00:00
tmweb.ru	1 days 10:26:32	0 days 12:00:00
glitch.me	1 days 02:13:37	1 days 00:00:00
Google Firebase	1 days 01:31:41	0 days 12:00:00
000webhostapp.com	0 days 21:20:56	0 days 06:00:00
yolasite.com	0 days 20:56:32	0 days 12:00:00
weebly.com	0 days 14:05:16	0 days 01:00:00
ngrok.io	0 days 00:19:15	0 days 00:05:00

Table 16: Mean and median uptime of top 10 most abused subdomain service providers.

Table 16 and Figures 14, 15, and 16, show mean, median, and the distribution of uptimes of the 10 most abused subdomain providers. While the mean uptime is the most intuitive metric, it comes with its limitation: it may be skewed by long-lived malicious domains. To overcome this limitation, we also consider the use of median uptime of abused subdomains. The median uptime (Table 16) may indicate the policies imposed by different providers. ngrok.io suspends the majority of malicious subdomains within first 5 minutes after blacklisting, 000webhostapp.com within 6 hours, blogspot.com,

IBM Cloud (appdomain.cloud), tmweb.ru, Google Firebase (web.app, firebaseapp.com), and yolasite.com within 12h, glitch.me within one day, whereas appspot.com within two days after blacklisting. Figure 14 shows that ngrok.io (the most abused service according to the collected data) is also the fastest to remediate: 99% of abused domains are suspended within an hour of being blacklisted. It may indicate that they actively detect and suspend abusive domains and/or collaborate with trusted notifiers.

(a) ngrok.io

(b) IBM cloud (appdomain.cloud)

(c) 000webhostapp.com
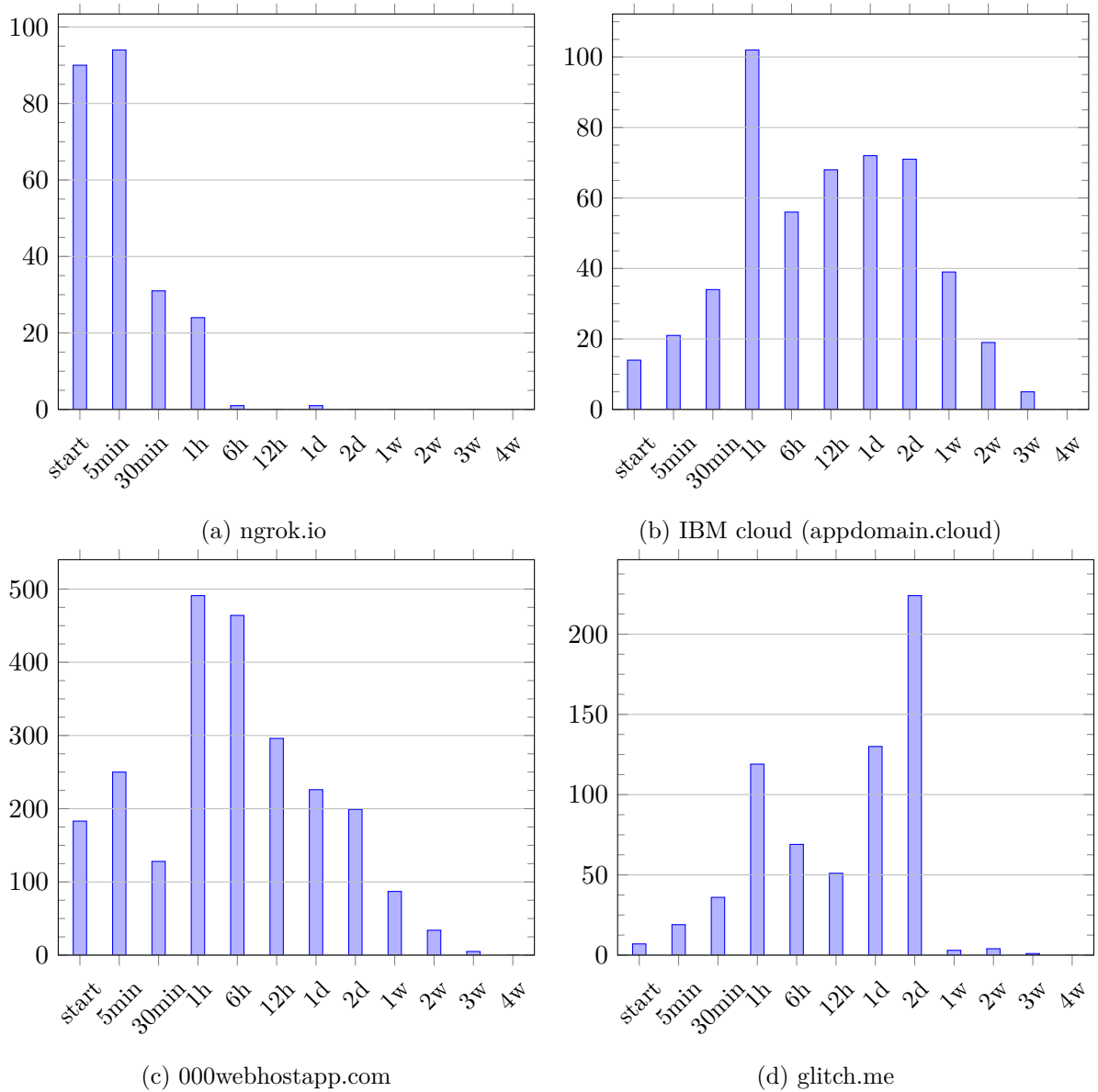
(d) glitch.me

Figure 14: Uptimes subdomain providers (ngrok.io, IBM cloud, 000webhostapp.com, glitch.me).
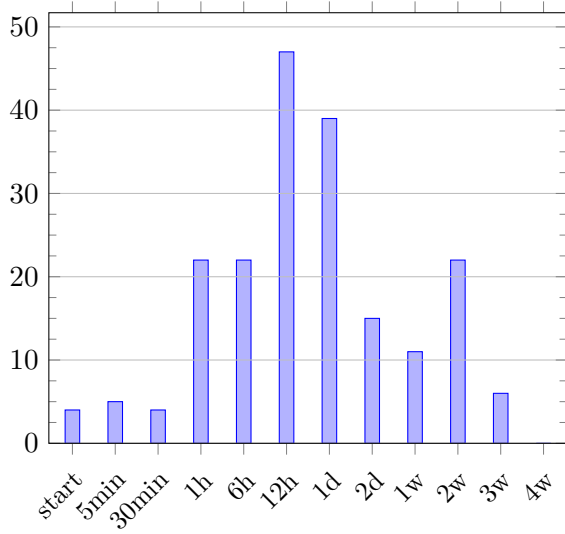
(a) blogspot.com
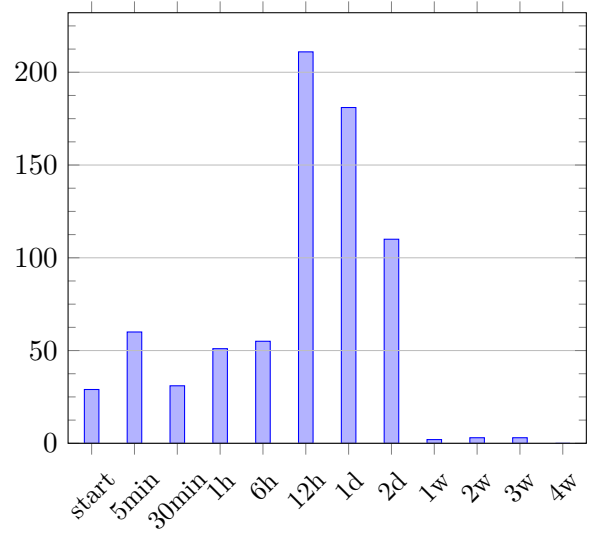
(b) yolasite.com
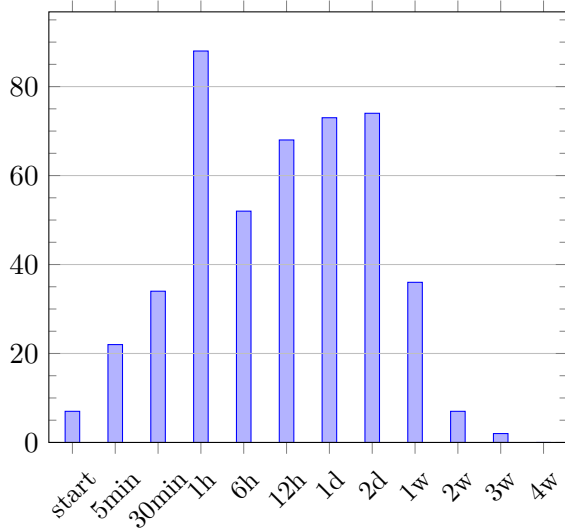
(c) tmweb.ru
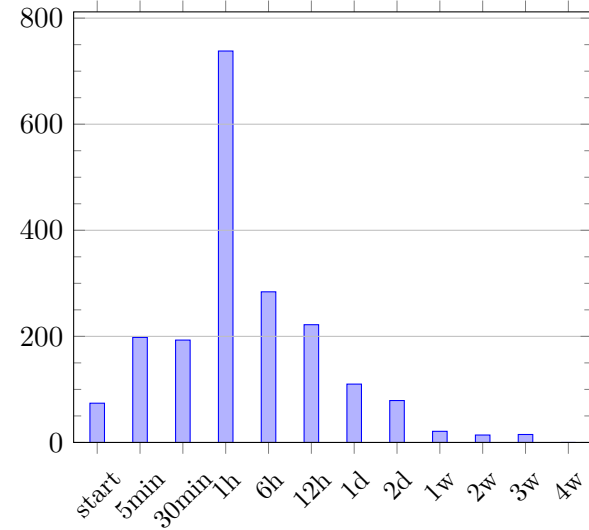
(d) weebly.com

Figure 15: Uptimes subdomain providers (blogspot.com, yolasite.com, tmweb.ru, eebly.com).

Recommendation: Since free services (e.g., free hosting and subdomains) are commonly exploited in phishing attacks, their operators should employ advanced prevention and remediation solutions to quickly curb abuses of subdomain names and hosting infrastructure. They should proactively detect suspicious domain names containing keywords of the most frequently targeted brands and names and work closely with the most heavily attacked companies and develop trusted notifier programs.

(a) appspot.com

(b) Firebase (web.app, firebaseapp.com)
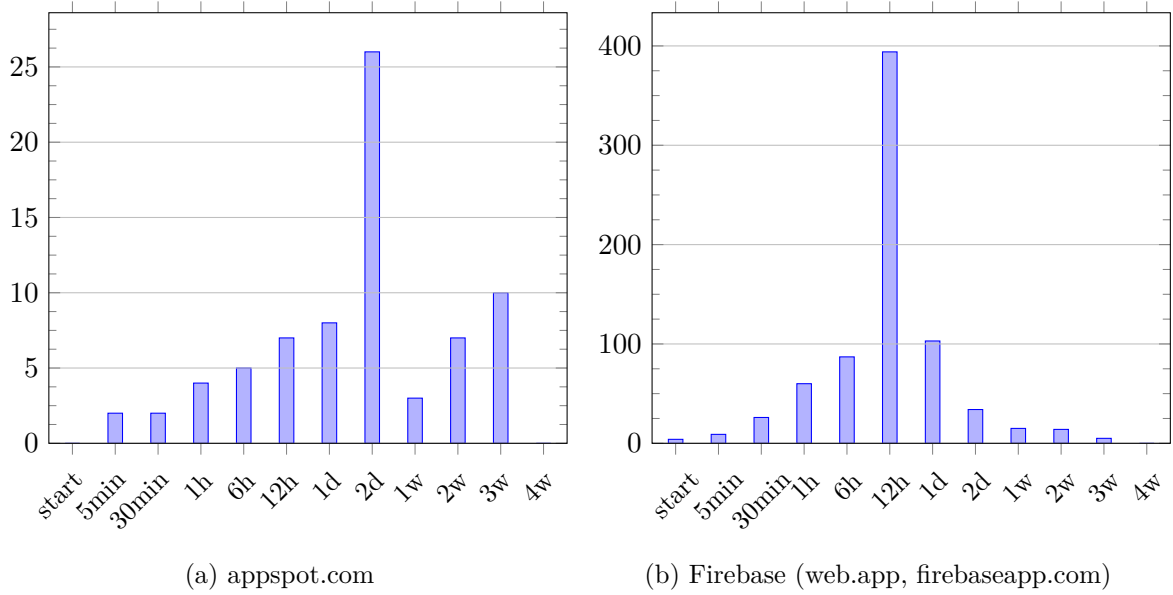
Figure 16: Uptimes subdomain providers (appspot.com, Firebase).

13 Targeted Brands
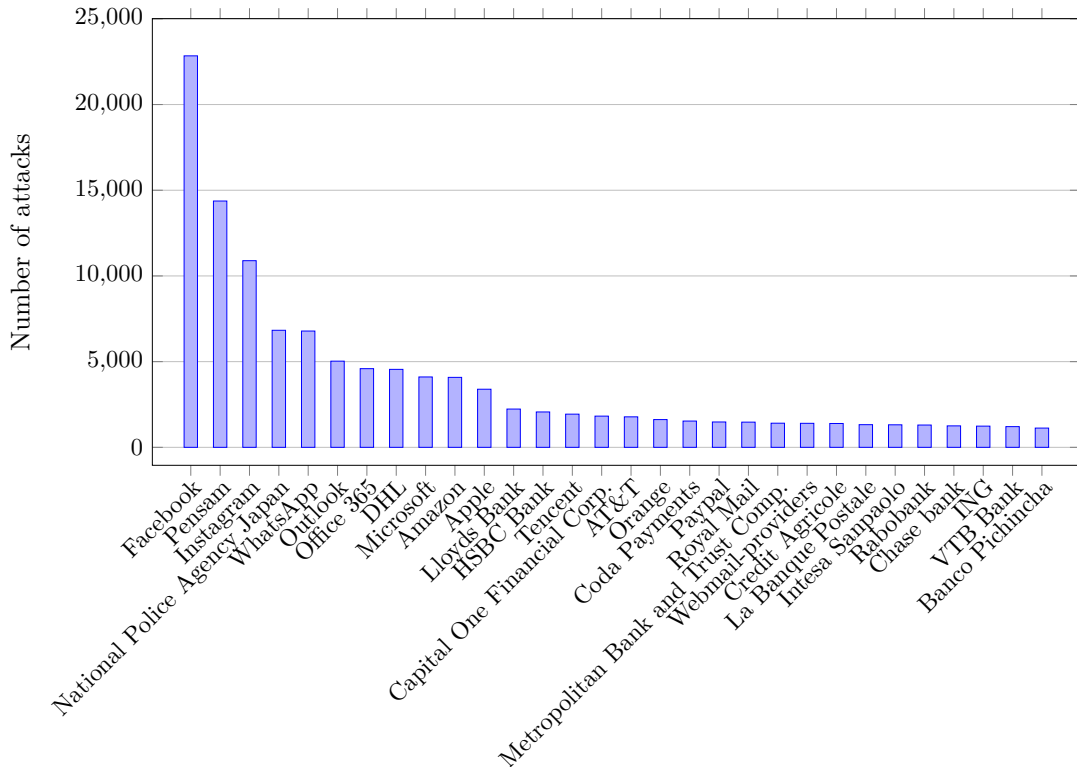
Figure 17: Top 30 most targeted brands.

We also studied 495,085 URLs that APWG, PhishTank, and OpenPhish identified as phishing and containing information about the targeted brands. The brand name

in the provided data appears in different variants (e.g. `Facebook` or `Facebook Inc.`) that point to the same entity (`Facebook Inc.`). Therefore, we manually tagged 1,424 variations for 1,076 target brand names.

Figure 17 shows the 30 most frequent brands in 405,431 URLs (81.89%) that were reported together with a specific target brand. As can be seen, Facebook Inc. is the most targeted brand, followed by Pensam (danish labour market pension fund[45]) and Instagram. Among the most targeted brands there are multiple banks such as Lloyds, HSBC, Crédit Agricole, Chase, or Rabobank. The results obtained should be considered indicative, as they may be driven (biased) by individual phishing campaigns, or methods of identifying and reporting abuse.

---

[45]https://www.pensam.dk/in-english

Part IV

DNS Infrastructure Abuse

In this part, we consider the issue of how the technical DNS infrastructure can be abused to perform different types of illegal activities. For example, domain names not deploying Domain Name System Security Extensions (DNSSEC) can be hijacked (cache poisoning [33] or zone poisoning attacks [2]) and used in phishing attacks, whereas domains that do not deploy DNS-based email anti-abuse security measures, such as SPF or DMARC, are forged to distribute phishing emails. Many administrators do not take the precaution of configuring their DNS resolvers to process recursive queries only from internal IP addresses. As a consequence, such open DNS resolvers are increasingly abused to amplify Distributed Denial-of-Service (DDoS) attacks.

In this part, we first discuss the deployment of DNSSEC, followed by DNS resolver vulnerabilities, and the study on SPF and DMARC adoption.

14 DNSSEC deployment

14.1 Motivation

The domain name system (DNS) was originally designed more than 30 years ago. At the time, security was not the primary concern on the Internet. That is why the early DNS standard was found vulnerable to many classes of attacks [34]. One of those is *cache poisoning.* When a malicious actor sends a forged reply to a recursive resolver before the genuine reply from an authoritative resolver arrives, it stays in the recursive resolver cache. Such a specifically crafted packet can redirect genuine clients to bogus websites, mail- or name servers.

DNS security extensions (DNSSEC) solve the problem by introducing origin authentication and data integrity [35–37] using a public key infrastructure. DNSSEC is only effective when deployed universally. We analyzed 251 million domain names and found that a small fraction of those attempted to deploy DNSSEC. Even less were correctly signed. We further show that while DNSSEC helps secure certain aspects of DNS, it is also prone to new types of attacks and should be implemented with great caution.

14.2 Background

DNSSEC modifies the normal DNS operation by introducing two new concepts - zone signing and response validation. Zone owners generate public/private key pairs. Private keys are used to sign resource record sets (RRsets) and produce `RRSIG` signatures. The corresponding `DNSKEY` public keys verify the signatures. Although not required by the DNSSEC standard, there are usually two key pairs - key signing key pair (KSK) and zone signing key pair (ZSK). KSK only signs `DNSKEY` RRset and its digest is published in the parent zone as `DS` resource record. ZSK signs the remaining RRsets in the zone.

Zone signing does not protect from manipulation if the keys and signatures are not cryptographically verified. DNSSEC-validating recursive resolvers are pre-configured with one trust anchor, usually the root zone public key (or its digest). The validator follows the chain of `DS`-`DNSKEY` resource records from the root zone down the domain name tree to the requested domain name. It ensures that the digests correspond to public keys and public keys verify the signatures. If all the checks are successful, it returns the response with `NOERROR` status code and `SERVFAIL` otherwise.

14.3 DNSSEC Measurements

We analyze DNSSEC deployment at two different levels. We first show that the majority of TLDs are signed and can be used to publish `DS` records of their children. We then switch to second-level domain names and observe that DNSSEC suffers low deployment rates.

Top-level domains

Operators of DNSSEC-signed zones assume that validating recursive resolvers will be able to establish a chain of trust from the trust anchor down to the zone. Since 2010, such a universally accepted trust anchor is the root zone KSK [38]. Once the root zone was signed, TLD operators had an opportunity to sign their zones and upload `DS` records to the root zone. ICANN publishes a daily report on DNSSEC adoption at the TLD level. As of July 2021, 1,372 out of 1,498 TLDs are signed and publish a key hash at the root [39]. The last generic top-level domain was signed in December 2020 [40] and all the 126 unsigned TLDs are country-code. Note that to implement DNSSEC, the TLD operator must sign the TLD zone. It is the first and most critical step in implementing DNSSEC. As one of the safeguards proposed by ICANN, all operators of

70

new gTLDs are required to sign the TLD zone [41].

> **Recommendation**: Similarly to gTLD registries, the registry operators of ccTLDs should be required to sign TLD zone files with DNS security extensions (DNSSEC) and facilitate its deployment according to good practices.

Second-level domains

DNSSEC-signed zones are different from unsigned ones as they publish additional resource records: `DS`, `DNSKEY`, `RRSIG`, and `NSEC(3)` that can be queried by recursive resolvers as any other regular resource records, such as `A`, `NS` etc. We rely on this fact to enumerate second-level domains that attempted to deploy DNSSEC. We use ZDNS[46] scanner to send `DS` and `DNSKEY` requests efficiently at scale. We operate it in the nameserver mode so that it forwards all the requests to the recursive resolver of our choice. We then set up a resolver using BIND9.[47] By default, it performs validation of all the received responses. However, we disable this functionality so that we receive the responses even if they are bogus. While scanning for `DNSKEY`s, we capture all the incoming traffic and extract `RRSIG` signatures returned along with `DNSKEY`s. At this stage, we only check for the presence of resource records and not their validity.

Table 17: Top 20 TLDs with the highest number of second-level domain in our input list.

TLD	Count	Type
com	145,475,053	generic
net	12,213,558	generic
de	9,601,890	country-code
org	9,540,343	generic
uk	4,263,606	country-code
info	3,492,481	generic
ru	3,473,332	country-code
nl	2,741,787	country-code
xyz	2,516,448	generic
br	2,309,677	country-code
tk	2,298,943	country-code
ga	2,249,643	country-code
fr	2,098,489	country-code
cn	1,949,840	country-code
it	1,758,075	country-code
ml	1,657,468	country-code
eu	1,559,517	country-code
au	1,557,872	country-code
cf	1,487,356	country-code
online	1,443,770	generic

We analyzed DNSSEC deployment of more than 251 million second-level domain

---

[46]https://github.com/zmap/zdns
[47]https://www.isc.org/bind/

names, representing 1,376 TLDs (top 20 TLDs by the number of domains are shown in Table 17). Note that for some TLDs for which we have access to their zone files, we evaluate the DNSSEC deployment for all domain names. However, for most ccTLDs, we assess the deployment based on all enumerated domains rather than all registered domain names (e.g., 9.6 million .de domain names, 3.5 million .ru domain names, or 2.7 .nl domain names). Therefore, the results represent approximate rates of domain names correctly signed with DNSSEC.

Overall, 227 million domain names returned `NOERROR` responses to our scanner for both queries. We refer to these as *responsive* domains. We exclude from the further analysis the remaining 24 million domains, as we cannot determine whether those are missing some of the resource records or we could not retrieve them for other reasons (temporary network failures etc.).

We first check how many responsive domains contain one or more DNSSEC resource records: `DNSKEY`, `RRSIG`, and/or `DS`. The presence of such records does not necessarily mean that domains are *correctly* signed, but rather signifies that domain owners attempted to do so. Only 6.7% (15.2 million) of responsive domains publish at least one DNSSEC resource record. Half of those fail to provide all three RRs. Such misconfigurations have different consequences:

- `DNSKEY-RRSIG`, `DNSKEY`, `RRSIG`: lack of `DS` is a common misconfiguration, as this record needs to be manually added to the parent zone (through the registrar control panel). It was previously shown that around 30% of domains that publish `DNSKEY` do not have an associated `DS` [42]. The responses from these domains are considered *insecure* by the DNSSEC standard [35]. They will not fail the validation check by recursive resolvers, but without a complete chain of trust, we cannot conclude whether the domain is correctly signed. Such DNS zones are referred to as islands of security and can only be used to validate their child zones (if recursive resolvers trust their keys). There are 5.7 million second-level domains from 748 TLDs that fail to provide the `DS` record while providing the two others (`DNSKEY` and/or `RRSIG`).

- `DNSKEY-DS`, `RRSIG-DS`, `DS`: domains with `DS` records at the delegation point have the complete chain of trust and will be verified by validating recursive resolvers. Because of the missing signatures (`RRSIG`) and/or public keys (`DNSKEY`), the vali-

dation will fail (responses from such domains are called *bogus*) and the end clients will receive `SERVFAIL` in response to their requests. Such misconfigurations, combined with using validating resolvers, effectively make these domains *unreachable*. There are 112,648 second-level domains from 422 TLDs that fail to provide `DNSKEY` and/or `RRSIG` while providing `DS` record.

These preliminary findings are alarming. The great majority of tested domains do not contain any resource records that would signal the willingness of domain owners to deploy DNSSEC. Only 15.2 million domains contain one or more DNSSEC-related resource records (`DNSKEY`, `DS`, `RRSIG`). However, we see straight away that 37.6% of them are in the best case, islands of security (because of missing `DS`) and 0.7% of them will fail the validation (because of missing public keys and/or signatures).

Note that in addition to TLD registries, registrars also play a key role in the implementation of DNSSEC, as they must add the `DS` record to the parent zone maintained by the TLD registry. The lack of support from registrars means that all domain names managed by these registrars cannot be signed. The Danish Ministry of Business has implemented a law requiring the .dk registry to ensure that all registrars who offer domain names in the .dk domain support DNSSEC no later than January 1, 2021 and offer DNSSEC signing to registrants.[48] Some registrars not only facilitate the addition of a `DS` record to a master zone but provide "one-click" DNSSEC deployment as a paid option (e.g., GoDaddy) or even at no cost (e.g., OVH SAS). The second option is one of the best ways to increase DNSSEC deployment on a massive scale.

> **Recommendation**: To facilitate the implementation of DNSSEC, domain administrators (registrants) should have easy access to DNSSEC signing of domain names within the TLD. TLD registries should require all registrars that offer domain names in the TLD to support DNSSEC signing for registrants.

Domains that do provide all the three resource records (9.4 million) are likely to be correctly signed but need further validation. We switch our BIND9 recursive resolver into validating mode and query these domains for `SOA` and `DNSKEY` records. The validating recursive resolver retrieves the requested resource records, performs additional queries to establish the chain of trust and validates the signatures. The results are re-

---

[48]https://www.retsinformation.dk/eli/lta/2020/44 (in danish)

assuring: 98.1% of domains publishing all the three resource records correctly sign both `DNSKEY` and `SOA` resource records. Thus, we can conclude that the presence of all the necessary DNSSEC resource records results in a high chance that the zone is *correctly* signed.

Based on our measurements, we categorize all the responsive domains (227 million) into three groups:

- Unsigned (212 million): domains that do not publish any DNSSEC resource records (`DNSKEY`, `DS` and `RRSIG`)

- Incorrectly signed (6 million): domains that either publish some of DNSSEC resource records or all of them, but the validation fails.

- Correctly signed (9 million): domains that publish all the DNSSEC records and when queried by a validating resolver provide correctly signed responses.

Tables 18, 19, 20, and 21 provide information on what TLDs have the highest numbers of second-level domains falling into each category. Table 18 displays top 20 TLDs of unsigned, incorrectly signed, and correctly signed domains. Tables 19 and 20 show similar ranking among generic and country-code TLDs. Table 21 shows the number and DNSSEC-deployment rate of European Union TLDs in each category.

As mentioned earlier, the rates for most ccTLDs were calculated based on a large sample of identified domain names because we do not have access to the zone files and the complete list of domain names. Therefore, the presented rates provide an approximation of the actual adoption. The DNSSEC adoption rates are not different from the general population and are rather modest—21 out of 34 TLDs consist of more than 90% of unsigned domains. On the contrary, the `.cz` TLD exhibits the highest proportion of correctly signed second-level domains. The `cz.nic` domain registry achieved it thanks to incentivizing registrars and ISPs economically and supporting them technically [43]. Moreover, DNSSEC is a part of the governmental initiative called "Digital Czech Republic v. 2.0" [44]. Swedish country-code TLD comes second with the majority (54.84%) of correctly signed domains. The `.se` registry provides guidance on DNSSEC deployment [45] and price incentives. The Dutch TLD `.nl` has high DNSSEC adoption rates (51.43%) thanks to the support from both the government and SIDN – the registry of .nl domains [46]. Registrars are charged lower fees for DNSSEC-signed domain names than for unsigned domain names.

The examples of the .cz, .se, .no, or .nl TLDs show that price incentives are the main driving factor behind the implementation of DNSSEC. All of these registry operators are among those that have used such schemes.

> **Recommendation**: As an incentive to the deployment of DNSSEC, TLD registries might offer discounts for DNSSEC-signed domain names.

Table 18: Top 20 TLDs with the highest number of second-level domains falling into each category. The ratio is computed from all the *responsive* domains.

Unsigned			Incorrectly Signed			Correctly Signed		
TLD	Count	Ratio (%)	TLD	Count	Ratio (%)	TLD	Count	Ratio (%)
com	122,236,139	93.85	com	4,905,793	3.77	com	3,105,826	2.38
net	10,403,214	96.30	ru	203,715	6.43	nl	1,367,067	51.43
de	9,230,789	97.84	nl	162,992	6.13	se	676,318	54.84
org	8,309,362	96.33	net	95,869	0.89	cz	418,299	59.23
uk	3,939,182	96.88	org	52,810	0.61	net	303,482	2.81
info	2,970,680	97.20	se	43,019	3.49	fr	292,072	14.45
ru	2,959,669	93.44	eu	41,999	2.81	pl	279,901	21.31
ga	2,213,920	99.95	fr	35,661	1.76	br	265,991	12.42
tk	2,196,953	99.88	de	31,482	0.33	org	263,955	3.06
br	1,866,791	87.20	cz	30,120	4.26	eu	237,625	15.87
xyz	1,844,580	97.92	be	28,377	3.47	be	208,268	25.47
fr	1,693,420	83.78	pl	24,541	1.87	dk	200,016	29.70
it	1,685,931	99.10	uk	21,705	0.53	de	172,621	1.83
ml	1,644,070	99.92	xn-p1ai	21,630	6.94	no	151,435	48.82
cn	1,636,199	99.91	co	17,798	1.60	sk	105,044	47.74
au	1,499,598	99.56	info	15,855	0.52	uk	104,962	2.58
cf	1,480,148	99.95	nu	12,690	7.94	ch	93,150	7.42
gq	1,272,683	99.98	no	12,614	4.07	nu	81,041	50.68
ca	1,269,112	98.33	hu	11,903	3.00	hu	71,959	18.13
eu	1,217,662	81.32	it	11,828	0.70	info	69,699	2.28

14.4 Challenges

DNSSEC has technically solved the problem of forged DNS replies. However, administrators of signed zones face additional maintenance issues, such as key management and signature expiration. We discuss DNSSEC challenges in this section.

Amplification of DDoS Attacks

DNS has long been known as one of the most used protocols to launch reflection and, especially, amplification DDoS attacks [13,14,47]. DNSSEC introduced a non-negligible overhead to the normal DNS operation because signed responses are larger in size. Van Rijswijk-Deij *et al.* [48] analyzed 2.5 million signed domains and a sample of unsigned domains across 6 TLDs and their amplification factors. While regular queries (`A`, `AAAA`, `DNSKEY`, `NSEC3`, `MX`, `NS`, `TXT`) do increase the amplification factor compared to normal

Table 19: Top 20 generic TLDs with the highest number of second-level domains falling into each category. The ratio is computed for all the *responsive* domains.

	Unsigned			Incorrectly Signed			Correctly Signed	
TLD	Count	Ratio (%)	TLD	Count	Ratio (%)	TLD	Count	Ratio (%)
com	122,236,139	93.85	com	4,905,793	3.77	com	3,105,826	2.38
net	10,403,214	96.30	net	95,869	0.89	net	303,482	2.81
org	8,309,362	96.33	org	52,810	0.61	org	263,955	3.06
info	2,970,680	97.20	info	15,855	0.52	info	69,699	2.28
xyz	1,844,580	97.92	online	10,643	0.94	app	58,808	10.45
online	1,101,814	97.30	xyz	8,840	0.47	page	52,368	67.34
club	748,133	98.56	shop	5,274	1.02	dev	48,187	24.27
vip	510,011	99.84	site	5,131	1.06	xyz	30,391	1.61
shop	501,711	97.50	dev	3,719	1.87	online	19,985	1.76
app	500,579	88.94	app	3,434	0.61	ovh	15,413	37.29
site	474,974	98.09	store	2,860	0.91	one	12,742	25.57
top	411,033	99.45	tech	2,568	1.28	realty	9,272	79.63
icu	398,970	99.80	club	2,480	0.33	club	8,453	1.11
store	306,033	97.45	cloud	1,508	1.10	tech	8,048	4.00
live	294,922	98.35	mobi	1,449	0.65	shop	7,611	1.48
work	266,555	99.36	space	1,368	0.82	cloud	6,461	4.70
mobi	220,399	98.79	top	1,040	0.25	store	5,139	1.64
tech	190,506	94.72	website	1,038	0.79	studio	4,460	8.85
space	161,307	97.21	xn-p1acf	941	6.81	live	4,203	1.40
dev	146,628	73.86	fun	886	0.87	site	4,112	0.85

Table 20: Top 20 country-code TLDs with the highest number of second-level domains falling into each category. The ratio is computed from all the *responsive* domains.

	Unsigned			Incorrectly Signed			Correctly Signed	
TLD	Count	Ratio (%)	TLD	Count	Ratio (%)	TLD	Count	Ratio (%)
de	9,230,789	97.84	ru	203,715	6.43	nl	1,367,067	51.43
uk	3,939,182	96.88	nl	162,992	6.13	se	676,318	54.84
ru	2,959,669	93.44	se	43,019	3.49	cz	418,299	59.23
ga	2,213,920	99.95	eu	41,999	2.81	fr	292,072	14.45
tk	2,196,953	99.88	fr	35,661	1.76	pl	279,901	21.31
br	1,866,791	87.20	de	31,482	0.33	br	265,991	12.42
fr	1,693,420	83.78	cz	30,120	4.26	eu	237,625	15.87
it	1,685,931	99.10	be	28,377	3.47	be	208,268	25.47
ml	1,644,070	99.92	pl	24,541	1.87	dk	200,016	29.70
cn	1,636,199	99.91	uk	21,705	0.53	de	172,621	1.83
au	1,499,598	99.56	xn-p1ai	21,630	6.94	no	151,435	48.82
cf	1,480,148	99.95	co	17,798	1.60	sk	105,044	47.74
gq	1,272,683	99.98	nu	12,690	7.94	uk	104,962	2.58
ca	1,269,112	98.33	no	12,614	4.07	ch	93,150	7.42
eu	1,217,662	81.32	hu	11,903	3.00	nu	81,041	50.68
ch	1,158,848	92.27	it	11,828	0.70	hu	71,959	18.13
us	1,142,047	97.32	dk	11,368	1.69	co	46,556	4.20
nl	1,127,965	42.44	us	10,369	0.88	us	21,048	1.79
co	1,044,857	94.20	sk	8,324	3.78	ca	16,747	1.30
pl	1,009,203	76.82	br	8,064	0.38	io	15,646	6.10

DNS, it mostly does not exceed the theoretical upper bound. A more serious amplifier is `ANY` type query, which results in the amplification factor of 47.2 for signed domains versus 5.9 for unsigned. Zone administrators cannot prevent attackers from querying their nameservers. Yet, they can block or provide minimal responses to `ANY` queries [49] and configure the nameservers with response rate/size limiting.

Table 21: Country-code TLDs of European Union members with the highest number of second-level domains falling into each category. The ratio is computed for all the *responsive* domains.

Unsigned			Incorrectly Signed			Correctly Signed		
TLD	Count	Ratio (%)	TLD	Count	Ratio (%)	TLD	Count	Ratio (%)
xn-qxa6a	8	100.00	nl	162,992	6.13	cz	418,299	59.23
xn-qxam	321	99.69	cz	30,120	4.26	se	676,318	54.84
ie	115,817	99.65	xn-e1a4c	4	3.96	nl	1,367,067	51.43
xn-90ae	475	99.58	sk	8,324	3.78	sk	105,044	47.74
hr	46,763	99.54	se	43,019	3.49	dk	200,016	29.70
lt	94,003	99.42	be	28,377	3.47	be	208,268	25.47
mt	4,116	99.18	hu	11,903	3.00	pl	279,901	21.31
si	66,691	99.15	eu	41,999	2.81	hu	71,959	18.13
it	1,685,931	99.10	lv	1,392	2.67	eu	237,625	15.87
gr	211,576	99.10	gl	44	1.90	fr	292,072	14.45
cy	5,508	98.80	pl	24,541	1.87	ee	8,599	13.09
bg	33,108	98.75	fr	35,661	1.76	xn-e1a4c	13	12.87
at	663,527	97.95	dk	11,368	1.69	lv	4,425	8.49
ro	301,433	97.94	fo	30	1.59	lu	1,620	4.31
de	9,230,789	97.84	ro	4,753	1.54	pt	4,692	2.91
es	626,866	97.43	pt	2,069	1.28	fi	6,712	2.60
fo	1,831	97.14	cy	65	1.17	gl	49	2.12
fi	248,835	96.47	fi	2,404	0.93	es	13,302	2.07
gl	2,220	95.98	lu	345	0.92	de	172,621	1.83
pt	154,656	95.81	mt	34	0.82	at	11,218	1.66
lu	35,633	94.77	bg	260	0.78	fo	24	1.27
lv	46,301	88.84	it	11,828	0.70	ro	1,572	0.51
ee	56,769	86.42	gr	1,398	0.65	bg	158	0.47
fr	1,693,420	83.78	si	405	0.60	xn-90ae	2	0.42
xn-e1a4c	84	83.17	es	3,202	0.50	lt	242	0.26
eu	1,217,662	81.32	ee	321	0.49	gr	527	0.25
hu	313,065	78.87	at	2,638	0.39	si	169	0.25
pl	1,009,203	76.82	hr	159	0.34	it	3,527	0.21
be	581,024	71.06	de	31,482	0.33	hr	58	0.12
dk	462,170	68.62	lt	305	0.32	ie	128	0.11
sk	106,687	48.48	xn-qxam	1	0.31	cy	2	0.04
nl	1,127,965	42.44	ie	274	0.24	-	-	-
se	513,880	41.67	-	-	-	-	-	-
cz	257,861	36.51	-	-	-	-	-	-

Zone Walking

DNSSEC guarantees that all the responses returned for signed domains are accompanied by signatures. As zones are usually signed once (statically), it is not feasible to precompute and sign all the possible negative responses (for example, when a subdomain does not exist in the zone). Instead of showing that something does not exist in the zone, DNSSEC explicitly lists everything that *does* exist. The new resource records (`NSEC` and `NSEC3`) list all the resource record types present for a certain name and the *next* name in the zone in the canonical order. The difference between the two is that `NSEC` lists names in plaintext while `NSEC3` hashes them using parameters in `NSEC3PARAM` resource record. Nevertheless, a determined attacker can attempt to precompute the hashes of all the possible names that could appear in the zone.

Although DNS and DNSSEC were never meant to provide confidentiality, there

was no straightforward way to enumerate all the subdomains in the zone. Thus, zone owners could reasonably consider them to be kept private. However, DNSSEC makes it possible to "walk the zone" by following the linked list of plaintext `Next Domain Name` fields in `NSEC(3)` records. We have scanned all our correctly signed second-level domains and found that 2,4 million of them publish `NSEC` resource records. It is not a misconfiguration as such, but zone owners should be aware of the information they expose.

Signature Validity

`RRSIG` signatures introduce the notion of absolute time in DNS. The two fields (`Signature Inception` and `Signature Expiration`) are timestamps that specify the time period during which the signature can be used for validation. Validating recursive resolvers use "their own notion of current time" [37] to check that the signature expiration field is greater than or equal to it. We examined 12.8 million signatures across 10.6 million second-level domains and found that 17,376 of them are expired. Responses with such signatures are *bogus.* Zone administrators should make sure that their signatures are always valid. RFC 6781 lists more time-related considerations in DNSSEC [50]. For example, signed zones are advised to have TTL values smaller than the signature validity period, which will avoid data being flushed from recursive resolvers caches once signature expiration time is reached.

Key Management

For DNSSEC to be cryptographically secure, zone administrators should only sign their zones with recommended algorithms, defined in RFC 8624 [51]. We checked whether domains in our dataset publishing `DNSKEY` records (15.1 million) adhere to this standard. We found that 25.9% of all the `DNSKEY` (25.4 million) implement not recommended algorithms. Very few keys (507) implement algorithms that *must not* be used.

Chung *et al.* [42] closely examined some of the common issues when it comes to key management in DNSSEC. Key reuse occurs when one private key is used to sign multiple domains. Although it was found that only 0.5% of examined keys are shared, one KSK and ZSK were shared among 130,000 domains. If a private key gets compromised, these many domains will be affected at once. Another concern is the key size. The DNSSEC standard does not dictate the key size requirements but the authors refer to NIST recommendations [52]. They found that 91.7% of examined ZSKs were not meeting the

minimal size requirements.

14.5   Discussion

DNSSEC remains the most effective way to fight DNS cache poisoning. Yet, it is only effective when deployed universally. Surprisingly, many 126 TLDs are still not signed. Consequently, their child zones cannot fully deploy DNSSEC because they will not have the complete chain of trust. Out of 227 million active second-level domain names that we analyzed, a tiny fraction (9.2 million) are correctly signed.

DNSSEC operation is complex and involves multiple parties: registrants, zone administrators (if different from registants), registrars, TLDs, and operators of recursive DNS resolvers (discussed in the next section). To increase adoption (and validation) of DNSSEC, everyone needs to participate. The remaining unsigned country-code TLDs should adopt DNSSEC to improve their reputation and enable their customers to sign their domains. They should also incentivize registrars to deploy it. Registrars, on their side, can encourage domain owners to deploy DNSSEC by offering them discounts and facilitating the signing process [53].

15   DNS Resolvers

In addition to proactive and reactive actions taken by TLD registries, registrars, hosting providers, resellers, and free service providers, DNS resolver operators have also an imperative role in securing the DNS infrastructure. Historically, mainly Internet Service Providers (ISPs) were responsible for maintaining DNS resolvers that resolve domain names to their respective IP addresses on behalf of end users. However, several companies such as Google[49], Cloudflare[50], Quad9[51], or OpenDNS[52] have been offering free and public DNS servers as an alternative way to connect to the Internet in recent years. One of the main advantages of using public DNS resolvers is to speed up domain name resolution, thereby improving the quality of experience for end users.

One fundamental problem is DNSSEC validation. In Section 14, we have described the role of registries and registrars in deploying DNSSEC. However, to protect end-users from cache poisoning attacks, local resolvers must verify the chain of trust to ensure

---

[49]https://developers.google.com/speed/public-dns
[50]https://www.cloudflare.com/learning/dns/what-is-1.1.1.1/
[51]https://www.quad9.net/
[52]https://www.opendns.com/setupguide/

the integrity and authenticity of domain name resolutions. Even complete deployment of DNSSEC by TLD registries, registrars, and registrants will not protect end users if DNS resolvers do not perform validation. One of the challenges is to measure whether ISPs perform validation, as it requires performing DNS queries from within the tested networks. In addition, it is challenging to measure the impact of DNSSEC deployment on global security because detection of cache poisoning attacks can generally be done at the ISP level or using passive DNS data.

> **Recommendation**: Internet Service Providers that operate DNS resolvers should configure DNSSEC validation to protect end users from cache poisoning attacks and ensure the integrity and authenticity of domain name resolutions.

Moreover, regardless of whether DNS resolver service is operated by local ISPs or public resolver operators, they should apply certain measures to improve security of end users. Service operators may subscribe to blacklists and should not resolve maliciously registered domain names to their IP addresses. A malicious domain name should resolve with "NXDOMAIN" indicating that the domain name does not exist or should be resolved to the DNS service provider own blocking site instead of the IP address of the requested malicious domain. The Quad9 system uses threat intelligence from more than a dozen leading cybersecurity companies to provide real-time information about which sites contain malware or other threats. If the system detects that a site a user wants to visit is infected, it automatically blocks the user from accessing it.[53] The public resolver operated by Google does not, in principle, perform any blocking.[54] Instead, malicious URLs (and domain names) are blocked by web browsers (e.g., Chrome, Firefox) using Google Safe Browsing.

Another problem is related to open (misconfigured) DNS servers that facilitate amplification reflection Distributed Denial-of-Service (DRDoS) attacks. Note that open DNS resolvers are not critical for launching DDoS attacks. For example, in September and October 2016, cybercriminals launched massive DDoS attacks using the Mirai botnet, which did not use the reflection amplification attack vector.[55] The attack brought down the DNS provider Dyn and several high-profile websites, including Twit-

---

[53]https://quad9.net/
[54]https://developers.google.com/speed/public-dns/faq
[55]Scott Hilton. "Dyn analysis summary of Friday October 21 attack". In: Dyn Blog, Oct (2016). Available at https://perma.cc/YW5C-MDEV

ter, Guardian, Netflix, Reddit, CNN, and many others in the United States and Europe. However, reflection and amplification DDoS attacks are mainly carried out using open UDP-based protocols, often DNS resolvers. Therefore, to increase the barriers to launching modern DDoS attacks, service providers should significantly reduce the number of misconfigured DNS resolvers. In the following sections, we discuss this problem in detail.

15.1 Open DNS Resolvers

Various open services (such as DNS, NTP, SNMP, SSDP, Memcached, etc.) have long been known as efficient DDoS reflectors and powerful amplificators [13, 14]. Among them, DNS resolvers are the most commonly misconfigured and abused. Open DNS resolvers accept DNS requests from any end-host. These can be misused to either target authoritative nameservers by sending the excessive number of incoming requests or, if combined with IP address spoofing, used to redirect responses to victim end-hosts. In this project, we actively scan for open DNS resolvers in IPv4 and IPv6 address spaces and analyze their distribution across organisations and countries.

15.2 Methodology

Scanning for open resolvers requires sending DNS requests to end-hosts and inspecting the received responses. The response codes (`RCODE`), defined in RFC 1035 [15], signal whether the DNS server processes incoming requests. If the query resolution is successful, open resolvers send back the responses to end clients along with `NOERROR` status code.

We use the three following datasets to scan for open resolvers: IPv4 BGP prefixes [16], IPv6 Hitlist Service [17], and IPv6 addresses learned by traversal from IPv4 to resolve IPv6-only domains, as described by Korczyński *et al.* [18]. All the three contain globally reachable IP addresses that may be operational recursive resolvers. Each end-host from the list receives an 'A' request for the unique domain name under our authority. We developed a software tool that allows us efficiently sending DNS packets at large scale [19].

15.3 Analysis

Scan results

We performed IPv4 and IPv6 open resolver scans in March 2021. Having tested more than 2.8 billion routable IPv4 addresses and 3.5 million IPv6 addresses, we discovered 3.4 million IPv4 and 18,843 IPv6 open recursive resolvers.

Response Integrity

Although the above mentioned open resolvers returned the `NOERROR` responses, they are not necessarily *correctly* operating. We closely inspected the answer section of returned packets and found that 18% IPv4 and 15% IPv6 open resolvers returned empty responses. More importantly, 8.4% and 6,6% of resolvers returned *bogus* replies to our 'A' requests. Previous work has shown that this behavior is likely due to censorship, ad redirection and other doubtful activities [54]. As the majority of such recursive resolvers return custom responses without contacting authoritative nameservers, their use in DDoS attacks is limited. Thus, we exclude those from further analysis.

Table 22: Top 20 IPv4 autonomous systems by the number of open resolvers

ASN	Organization	IPv4 Resolvers
4134	China Telecom	260,649
4837	China Unicom	189,714
45090	Tencent-CN	107,769
4766	Korea Telecom	67,557
47331	TTNET A.S.	58,693
5617	Orange Polska	52,568
3462	HiNet	36,868
4812	China Telecom	33,432
9318	SK Broadband	26,903
4808	China Unicom	26,762
12389	Rostelecom	24,989
209	Centurylink	24,979
7713	Telekomunikasi Indonesia	21,475
4538	China Education and Research Network Center	18,866
9808	China Mobile	17,838
58224	Iran Telecommunication Company	16,036
45804	Meghbela Cable & Broadband Services	15,624
32708	Root Networks	15,502
3269	Telecom Italia S.p.A.	12,371
58659	Quest Consultancy	11,918

Autonomous System Distribution

We map the remaining open resolvers to their autonomous system numbers (ASN) using pyasn[56] and check the PeeringDB[57] and AS Rank[58] for the organization names. Table 22 and Table 23 present the numbers of open DNS resolvers by autonomous systems. Top 20 IPv4 organisations are dominated by Asian telecommunication operators, while IPv6 autonomous systems also include transit and hosting providers. In total,

---

[56]https://pypi.org/project/pyasn/
[57]https://www.peeringdb.com
[58]https://asrank.caida.org

Table 23: Top 20 IPv6 autonomous systems by the number of open resolvers.

ASN	Organization	IPv6 Resolvers
6939	Hurricane Electric	4,458
63949	Linode	548
3462	HiNet	415
4837	China Unicom	364
8966	Etisalat-AS	351
12322	Free SAS	332
4812	China Telecom	294
1241	Forthnet	286
51167	Contabo	228
27839	Comteco	184
16276	OVH	179
7922	Comcast	163
4134	China Telecom	159
37564	Wirulink Pty Ltd	153
8100	QuadraNet Enterprises LLC	137
23910	China Next Generation Internet CERNET2	115
3303	Swisscom (Schweiz) AG	110
3356	Level 3 Parent, LLC	104
14061	DigitalOcean, LLC	102
8251	FreeTel, s.r.o.	102

open resolvers are present in 24,087 IPv4 and 1,607 IPv6 autonomous systems (34.2% and 7,4% of all those in the BGP routing table as of beginning of March 2021).

The big absolute number of recursive resolvers may not be surprising if belonging to a big autonomous system. Thus, we compute a ratio of open resolvers to the size of the address space announced by the IPv4 autonomous systems. Table 24 shows the results. None of the organisations from Table 22 and Table 23 is present in Table 24. These small autonomous systems almost entirely consist of open resolvers. In fact, there are 278 IPv4 autonomous systems where more than half of the address space is occupied by open resolvers.

Geographic Distribution

We map all the open resolvers to countries using the MaxMind database.[59] Overall, open resolvers are present in 230 countries/territories. Table 25 shows the top twenty countries by the number of open IPv4 and IPv6 resolvers. Eleven countries dominate both in IPv4 and IPv6 ranking. Importantly, top twenty countries contain the majority of all the open resolvers worldwide: 84.9% in IPv4 and 80.4% in IPv6. Table 26 displays the number of open resolvers in European Union (EU) countries only. The top three countries account for more than 50% IPv4 and 66% IPv6 open resolvers in EU.

Next, we examine the ratio of open resolvers per regions in Table 27. The majority of IPv4 resolvers are located in Asia. IPv6 resolvers are not dominated by a single region,

---

[59]https://dev.maxmind.com/geoip/geoip2/geolite2/

Table 24: Top 20 IPv4 autonomous systems by the ratio of open resolvers.

ASN	Organization	AS Size	Ratio
269113	Uno Telecom LTDA	1,024	99.5 %
268137	Net Sini Fiber Home Telecomunicação LTDA	1,024	99.5 %
136668	Iana Solutions Digital India	512	99.4 %
263108	Opanet Telecomunicacoes LTDA	2,048	99.3 %
267072	Veloz Net Serviços e Comunicações LTDA	768	99.2 %
267007	Turbo Net Telecom Servicos e Vendas de Equipamento	1,024	99.1 %
134929	Orange City Internet Services	2,048	99.0 %
208070	TILYTEL B., S.L.	1,024	99.0 %
270404	Qualidade Digital Internet e Telecomunicações	1,024	99.0 %
134924	Aph Networks	512	99.0 %
269563	MAX3 TELECOM LTDA	1,024	98.93 %
271003	MARILETE PEREIRA DOS SANTOS	1,024	98.83 %
270657	FNET TELECOM	1,024	98.83 %
34939	NextDNS	768	98.83 %
137045	Athoy Cyber Net	512	98.83 %
47849	Global Communication Net Plc	3,072	98.73 %
269012	Click Net Link Informatica e Telecomunicações LTDA	1,024	98.73 %
265276	SPEED_MAAX BANDA LARGA LTDA - ME	1,024	98.73 %
271070	Ailson Tavares	1,024	98.63 %
47275	Torjon Wieslaw Radka	1,024	98.63 %

as more than 60% of those are shared between Asia and European Union. Africa, Oceania, and Europe (outside the European Union) represent the smallest share of open resolvers.

15.4 Discussion

Open resolvers pose a great security threat—they are prone to misuse by attackers and should only be operated when necessary. We discovered more than **2.5 million** correctly resolving IPv4 and IPv6 open resolvers worldwide. We have shown that they are distributed both in terms of organisations and geographic territories. Nevertheless, the majority of all the open resolvers originate from very few autonomous systems and countries.

Kührer *et al.* fingerprinted 5.4 million open resolvers and concluded that more than 60% of those were routers, modems, gateways, and embedded devices [54]. We hypothesize that telecommunication operators do not configure customer equipment correctly. If this is the case, then some national telecommunication operators could eliminate a significant number of open resolvers in their countries, as it is the case with Orange Polska or Telecom Italia.

Note that the problem is not new. Jared Mauch presented at the NANOG meeting the Open Resolver Project [55]. He uncovered 34 Million DNS servers that responded to UDP/53 probe. Despite different initiatives to mitigate the problem, such as Com-

Table 25: Top 20 countries/territories by the number of open resolvers.

Country	IPv4 Resolvers	Country	IPv6 Resolvers
China	758,083	USA	2,500
Brazil	323,263	Germany	1,323
USA	180,328	China	1,258
India	117,363	France	880
Republic of Korea	116,749	Republic of Korea	708
Russia	97,287	Taiwan	583
Turkey	78,982	Russia	494
Indonesia	75,157	Czech Republic	409
Poland	73,189	Japan	395
Taiwan	42,577	UK	376
Bangladesh	38,061	Brazil	367
Argentina	34,858	United Arab Emirates	354
France	31,720	Greece	342
Italy	28,916	Thailand	310
Ukraine	27,348	Canada	307
Iran	27,343	Iran	295
Japan	24,808	Vietnam	252
Thailand	22,520	India	244
Hong Kong	20,765	Switzerland	242
Bulgaria	19,992	South Africa	239

Table 26: Distribution of open resolvers in European Union countries.

Country	IPv4 Resolvers	Country	IPv6 Resolvers
Poland	73,189	Germany	1,323
France	31,720	France	880
Italy	28,916	Czech Republic	409
Bulgaria	19,992	Greece	342
Germany	18,352	Netherlands	181
Spain	12,400	Hungary	119
Hungary	10,221	Italy	76
Romania	7,766	Romania	74
Czech Republic	7,508	AT	72
Netherlands	7,165	Lithuania	64
Sweden	5,945	SE	59
Greece	4,962	PL	53
Austria	3,722	ES	51
Slovakia	3,663	BG	48
Portugal	3,646	Slovenia	43
Latvia	3,394	Finland	30
Croatia	2,547	Belgium	30
Denmark	1,877	Denmark	26
Finland	1,738	PT	18
Belgium	1,734	Ireland	18
Lithuania	1,178	Croatia	18
Ireland	1,145	Cyprus	17
Cyprus	694	LV	13
Slovenia	687	SK	10
Estonia	355	LU	7
Luxembourg	313	EE	6
Malta	250		

puter Emergency Response Team (CERT) alerts [56], research indicating the scale of the problem [13, 14], and notifications to operators by ShadowServer or locally by the national German CERT [57], the issue has still not been resolved.

Table 27: Ratio of open resolvers per region.

Region	Ratio of IPv4 Resolvers	Ratio of IPv6 Resolvers
Africa	2.4 %	2.1 %
Asia	56.4 %	34.9 %
Europe	6.3 %	8.6 %
European Union	10.1 %	26.9 %
North America	7.8 %	19.0 %
Oceania	0.5 %	1.0 %
South America	16.5%	7.5 %

> **Recommendation**: National CERT teams should subscribe to data sources that identify open DNS resolvers. National governments and Computer Emergency Response Team (CERT) teams should intensify notification efforts to reduce the number of open DNS resolvers (and other open services), which are among the root causes of distributed reflective denial-of-service (DRDoS) attacks.

16 SPF and DMARC

16.1 Motivation

Email spoofing is defined as sending emails with a forged sender address in a way that it appears as sent from a legitimate user or on behalf of a company [12]. Business Email Compromise (BEC) is one of the most financially damaging online crimes [58], and email spoofing is one of the most common techniques used in BEC.

The Simple Mail Transfer Protocol (SMTP) does not provide a built-in approach to fight email spoofing. Therefore, the deployment of the email security extensions such as the Sender Policy Framework (SPF) [59], DomainKeys Identified Mail (DKIM) [60], and Domain-based Message Authentication, Reporting, and Conformance (DMARC) [61] is necessary to fight phishing attacks. In this section, we measure the global adoption of email security extensions, namely, SPF and DMARC for all the domain names in our database as described in Section 4. We do not measure DKIM since it needs access to DKIM subdomains (also known as the *selector* tag). They are not publicly available and can only be found in the header of the received emails.

16.2 Methodology

To measure the deployment of SPF and DMARC, we use the following approach:

- Regarding SPF, first, we collect SPF records (as part of DNS TXT resource records) of all 251 Million enumerated domains using the ZDNS tool.

- Then, for those domains with SPF records, we emulate the *check_host()* function as described in RFC 7208 [59], to evaluate the validity and configurations of the records.

- The next step is to collect the DMARC rules, which exist in the TXT resource records of the `_dmarc` subdomains of the registered domains (e.g., `_dmarc.example.com`).

- Finally, we evaluate DMARC rules to check their strictness in accepting (delivering to the end-users) and/or rejecting the incoming forged emails.

The next section shows the results of our scan for SPF and DMARC protocols.

16.3   Results

Table 28 shows the results of the scan for 247,006,422 domain names that returned *NOERROR*. The second column (status) shows the scan status with ZDNS for SPF rules in the TXT resource record of each domain. **60.44%** of the domains have no SPF record, which is a bad practice to protect a domain name from email spoofing since in the absence of the SPF record, for any incoming message, the result of the *check_host()* function is **None** (i.e., no strict decision), leaving the decision to the receiver mailing system. Since most of the domains currently do not have SPF records, the receiving email servers make a 'softer' decision, generally letting the email be delivered so that the users do not lose any email. This behaviour makes it easier for spammers to send forged emails [12, 62, 63].

#	Status	Count	Percentage (%)
1	AUTHFAIL	275,925	0.111
2	ERROR	169,679	0.068
3	NOERROR	77,487,889	31.370
4	NORECORD	149,305,756	60.446
5	NXDOMAIN	2,475,409	1.002
6	REFUSED	5,979,033	2.420
7	SERVFAIL	10,616,307	4.297
8	TEMPORARY	348	0.0001
9	TIMEOUT	696,076	0.281
Total		247,006,422	

Table 28: Scan results of the SPF rules.

Only 31.37% of the domains have SPF records. However, having an SPF record does not necessarily guarantee any protection against email spoofing. Table 29 shows the results of the *check_host()* function emulation for the domains with SPF records (corresponding to the third row of Table 28 with *NOERROR* status). All the domains with SPF *pass* results are open to email spoofing since they let the sender send emails from any IP address. For other SPF results, the decision is made by the receiver with the help of DMARC rules specified in the TXT resource records of the *_dmarc* subdomain. The SPF *permerror* result means that there is a problem in either parsing or recursive querying SPF rules, which usually happens because of setting a syntactically wrong SPF rule or defining too much DNS lookups (recursions) in the SPF rule [63]. Table 30 shows the most common errors related to the domains with the *permerror* result from the *check_host()* function emulation.

#	Status	Count	Percentage (%)
1	None	2,543,870	3.28
2	Neutral	5,866,297	7.58
3	Pass	200,362	0.26
4	Fail	29,049,907	37.5
5	Softfail	35,929,956	46.37
6	Permerror	3,207,817	4.14
7	Temperror	689,680	0.91
	Total	77,487,889	

Table 29: Results of the check_host function emulation

Error type	Example	Correct rule	Frequency
Too many DNS lookups	-	SPF rule must generate less than 10 DNS query	1,638,092
Two or more SPF records found	-	must set one SPF record for each domain	691,746
Void lookup limit of 2 exceeded	-	rules with empty responses must be removed	64,914
More than 10 MX records returned	-	Total number of lookups must be less than 10	27,699
Invalid IP4 address: ip4:	ip4:xxx.xxx.xxx.xx?all	ip4:xxx.xxx.xxx.xx ?all	16,621

Table 30: Most common syntactically wrong rules that lead to the *Permerror* result.

Table 31 shows the scan results of the domain names for DMARC records. The status *NXDOMAIN* means that there is no DMARC subdomain for the domain name, and *NOERROR* means that the DMARC record is present, which is only true for **3% of the domain names**. However, still having DMARC does not necessarily guarantee any protection. The final decision about the incoming email delivery is up to the **p** tag of the DMARC, which specifies the action to do: i) deliver the message, ii) reject the message, or iii) quarantine the message (labeled as spam). Parsing the DMARC record shows that 49.68% of the domain names with the DMARC record has the **p=none**

#	Status	Count	Percentage (%)
1	AUTHFAIL	284,939	0.115
2	ERROR	166,427	0.067
3	NOERROR	8,129,795	3.299
4	NORECORD	70,277,952	28.518
5	NXDOMAIN	150,842,488	61.212
6	REFUSED	6,019,716	2.442
7	SERVFAIL	10,037,232	4.073
8	TEMPORARY	10,795	0.004
9	TIMEOUT	656,653	0.266
Total		246,425,997	

Table 31: Scan results of the DMARC records

rule, which means they specified no strict action against incoming messages sent from unauthorized servers. 11.20% of the domains have **p=quarantine** (i.e., labeling the incoming message as spam), and 37.14% have **p=reject**, which means rejecting the incoming message with unauthorized sender based on SPF rules.

16.4 Discussion

SPF and DMARC protocols are critical for preventing email spoofing and essential in preventing Business Email Compromise (BEC) fraud, which according to the recent FBI report, caused more than US $1.8 billion in losses to businesses and individuals in 2020 [58]. Note that securing domain names with SPF and DMARC does not solve the problem of BEC scams, as criminals can register, e.g., misspelled (e.g., using special characters) or internationalized domain names. However, suppose SPF and DMARC rules are not correctly configured. In that case, a cybercriminal can send emails on behalf of target brand domain names, making recipients unable to distinguish legitimate emails from fraudulent ones. Correctly implemented and strict SPF and DMARC rules can completely mitigate the problem of domain name spoofing, assuming that recipient mail servers verify and filter emails based on SPF and DMARC rules.

A recent study explored the degree of SPF and DMARC deployment for high-profile domains, including banking domains, and identified misconfigured ones [63]. They notified domain owners through local, and national CERT teams, and as many as 23.2% of the domains were reconfigured. While this was a one-time notification campaign, such ongoing efforts to measure the deployment and raise awareness of the problem should be promoted by governments and national CERTs, especially in the light of the recent FBI report of financial losses caused by BEC scams.

Recommendation: Security community should intensify efforts to continuously measure the adoption of the SPF and DMARC protocols, especially for high risk domain names and raise awareness of the domain spoofing problem among domain owners and email service providers. Correct and strict SPF and DMARC rules can mitigate email spoofing and provide the first line of defense against Business Email Compromise (BEC) scams.

17 Analysis of RFC-compliant Email Aliases

In this section, we consider the problem of effective notifications of domain abuse or vulnerabilities to the domain owners, administrators, and webmasters.

17.1 Motivation

Malicious actors compromise thousands of legitimate domains every day by exploiting vulnerable content management systems, frameworks, or libraries used to build websites. The compromised domains are abused to launch Internet-scale phishing, malware drive-by-download, or spam campaigns. To prevent vulnerable resources from being exploited and to remediate already abused domains (URLs), defenders share information about security threats and incidents through collaborative platforms such as Anti-Phishing Working Group (APWG)[60], PhishTank[61], or URLhaus[62].

Some types of domain abuse or vulnerabilities should be directly reported to the domain registrants (owners), administrators, or webmasters (often at scale, so web forms may not be always appropriate). Therefore, the Internet community needs to maintain the ability of large-scale notification mechanisms. An alternative approach is to report abuse through intermediaries such as CERTs. They validate email notifications and further communicate with the actors responsible for the affected systems. However, previous work showed that security notifications directly addressed to the owners of vulnerable resources promote faster remediation than those sent to national CERTs [64]. Therefore, several researchers used the contacts of domain administrators and owners retrieved from the public WHOIS data [64–71] and studied different communications strategies to increase remediation rates.

Retrieving contact information at scale from public WHOIS proved to be highly problematic [65] and became even more difficult (often impossible) with the introduction of the General Data Protection Regulation (GDPR) on May 25, 2018. ICANN adopted the temporary specification on how to publish the registration data of individuals [72] that prohibits domain registrars and registries from storing personal data in the public WHOIS database, in particular, the email addresses of registrants and administrators. In the absence of direct contact with the registrant, it is recommended

---

to contact the relevant registrar who has to provide access to registrant contact information in "a reasonable time" [73]. However, this rule may cause delays in patching vulnerabilities and mitigating abuse, and in addition, it does not scale.

In this report, based on the previous research [74], motivated by the implications of the EU regulation on data protection, we systematically test available direct contacts of domain owners and administrators as defined in RFC 2142 [75]. For a given domain `example.com`, RFC 2142 requires to configure valid email aliases such as `abuse@example.com` or `security@example.com` for incident and vulnerability notifications. Rather than quantifying remediation rates of DNS abuse, we test whether five RFC-specific generic email aliases are correctly configured and whether notifications can be successfully delivered. We also test the reachability of email addresses collected from the DNS Start of Authority (SOA) resource records (RNAME field).

17.2 Methodology

We have developed a scanner to systematically test available direct contacts of domain owners and administrators. We first scan for the DNS MX records of the domain and select a mail server with the highest priority. Afterwards, we establish different connections using the Simple Mail Transfer Protocol (SMTP) [76] to the selected mail server. We do not send emails, but we verify the existence of an email address using the RCPT TO query followed by the destination email address. If the mail server replies with code 250, the recipient address is considered as valid. In a single SMTP session, we only test one email address to avoid triggering mechanisms preventing email address enumeration [76], which may close the SMTP connection and blacklist the IP address of the sender.

Another countermeasure used by mail servers is to accept any given recipient, even a non-existent one, and return code 250. This procedure is called CATCH ALL (or wildcard email address) [77]. Our scanner is designed to detect if a mail server uses the CATCH ALL mechanism by checking for the existence of a randomly generated email. If such a contact is validated, then the mail server is most likely validating all, even non-existent, addresses.

For each sampled domain name, we generate email aliases using the names defined in RFC 2142 [75]: for the domain `example.com`, we test the validity of the following email aliases: `hostmaster@example.com`, `webmaster@example.com` (for DNS and

HTTP issues), `abuse@example.com` (for generic abuse and vulnerability notifications), `noc@example.com`, and `security@example.com` (for network security).

We scan for DNS SOA records, extract the hostmaster contact stored in the RNAME field as defined in RFC 1035 [78], and check whether the syntax of the email address is correct. Note that the domain name of an email gathered from the RNAME field may be different from the tested domain itself implying that we also need to lookup the MX record of the hostmaster domain.

To compare the results of the most popular websites with less known ones, we leverage top 1 M domains from Tranco—a domain ranking list oriented toward research [79].

To measure the reachability rates of the different TLDs, we categorize the domains based on their TLD and we sample domains from each group. We calculate the size of each sample using binomial approximation [80] as explained in [74]. We perform the experiment on generated samples of `.com` (633 K domain names) and `.net` (772 K domains) gTLD names, new gTLD names (232 K), and ccTLD names (967 K). Finally, we remove duplicates present in more than one sample: e.g., `google.com` is present in the Tranco list and has been randomly drawn for the `.com` sample.

17.3 Results of Email Validation Scans

The scan results are shown in Table 32. We consider a domain as reachable if at least one RFC-specific generic email alias has been validated. It is then labeled as CONTACT FOUND. We find that in all the studied categories, domains are more reachable using SOA contacts. For instance, 39.74% of Tranco top 1 M domains are reachable using an email leveraged from the SOA RNAME field while only 24.16% are reachable using RFC-specific contacts. We also find significantly more missing MX records for RFC-specific contacts than for mail servers of SOA contacts because DNS SOA records are often maintained by DNS service operators and less frequently by the domain owners who often do not have enough expertise in configuring DNS servers.

The ccTLD names seem to be the most reachable for both SOA (35.4%) and RFC-specific (12.68%) contacts when compared to `.com` (20.20% and 8.97%) and especially to new gTLDs (21.35% and only 3.97%). New gTLDs names are far less reachable in the sample with 61.63% of domains without MX records, three times more than ccTLDs and almost twice more than `.com` domains.

The results for the Tranco top 100 to 1 M lists (Figure 18) show that the better

Table 32: Results of email validation scan on the selected TLDs and popular domains.

Category		Domains (%)				
		Selected TLDs				Popular
		.com	.net	ccTLD	ngTLD	Tranco 1 M
RFC emails	NO MX RECORD	35.48	35.59	18.43	61.63	21.46
	CONN. ERROR	25.58	27.32	19,11	9.68	13.00
	CATCH ALL	12.33	11.42	18.86	7.68	20.36
	NO CONTACT FOUND	17.63	17.80	30.91	17.03	21.02
	CONTACT FOUND	8.97	7.87	12.68	3.97	24.16
SOA contact	NO SOA RECORD	4.49	4.66	3.63	11.01	2.84
	NO MX FOR SOA	9.01	7.70	8.31	10.47	9.04
	CONN. ERROR	34.68	35.83	15.30	15.86	18,60
	CATCH ALL	18.09	17.54	22.36	12.12	19.38
	NO CONTACT FOUND	13.53	13.28	14.99	29.17	10.40
	CONTACT FOUND	20.20	20.97	35.40	21.35	39.74
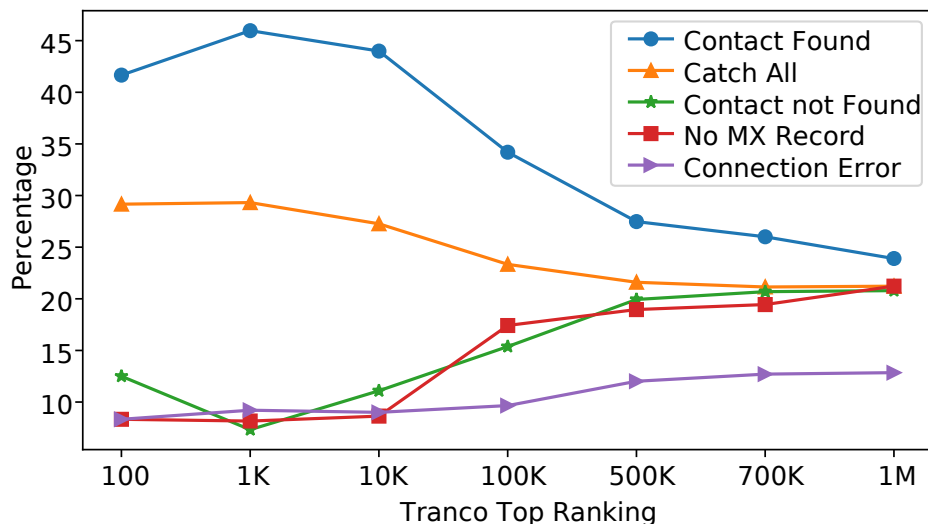
Figure 18: Results of validation scans of RFC-specific email aliases for the Tranco top 100 to Tranco top 1 M lists.

the rank of the domain is, the more likely the domain is reachable and complies with RFC 2142. For the top 1 K domains, at least 43.9% are reachable and the proportion gradually decreases to 22.6% for Tranco top 1 M. Similarly, the rate of CATCH ALL domains decreases from 28% (top 100) to 19% (top 1 M). The results can be explained by the fact that the operators of more popular websites put more emphasis on preventing email address enumeration of their clients.

We also observe an increasing rate of domains with invalid contacts: from 7% (top 1 K) to 19.7% (top 1 M) and domains without MX records: from 8% (top 1 K) to 20% (top 1 M). The connection error rate has also slightly increased from 7% (top 100) to 11.5% (top 1 M).
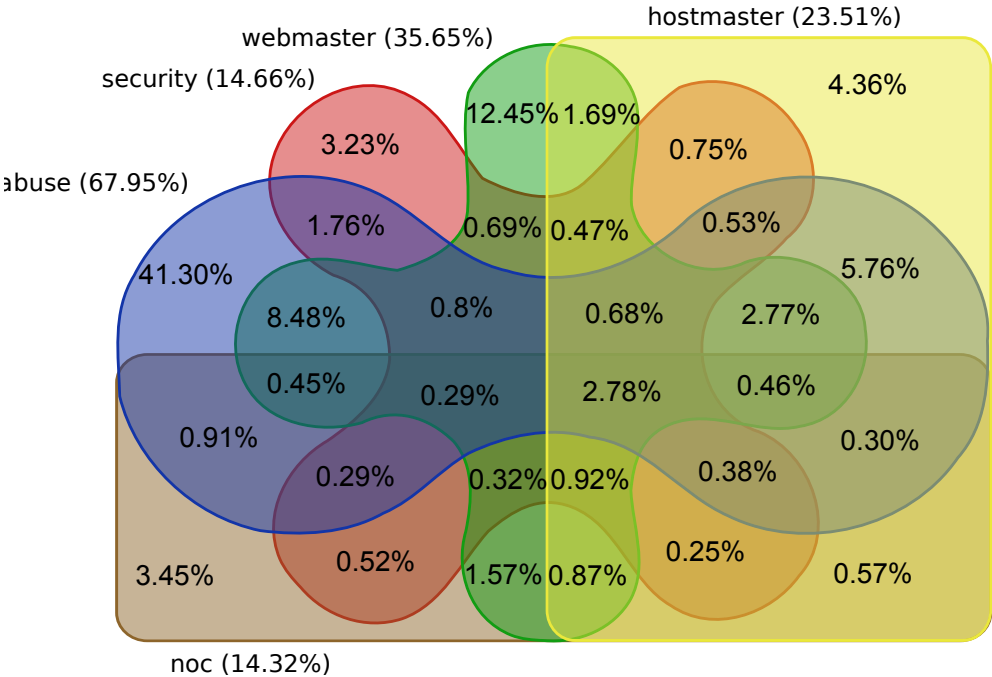
Figure 19: Venn diagram of the used RFC-specific email aliases

We next briefly analyze the most common RFC-specific email aliases. Figure 19 presents the results in the Venn diagram. We observe that the most frequently used contact is *abuse* (67,95%). As many as 64.79% of domains only use one valid contact: 41.3% of domains use the *abuse* contact, while 23.46% use one of the other 4 aliases (half of them use *webmaster*). 35.21% of domains use multiple RFC-specific contacts.

17.4 Discussion

The Internet community needs a large-scale notification system that covers the largest part of Internet domains for direct reporting of threats or abuses.

Recommendation: With no direct contact with domain name registrants and administrators via the public WHOIS database, domain name administrators should also maintain specific email aliases for given domain names (e.g., abuse, hostmaster, webmaster) so that they can be contacted directly in the event of vulnerabilities and domain name abuse.

> **Recommendation**: The email addresses of registrants and domain name administrators that are not visible in the public WHOIS could be displayed as anonymized email addresses to ensure the ability to contact domain owners and administrators directly to notify security vulnerabilities and abuses.

18   Inbound Source Address Validation

The Internet relies on IP packets to enable communication between hosts with the destination and source addresses specified in packet headers. However, there is no packet-level authentication mechanism to ensure that the source address has not been altered [81]. The modification of a source IP address is referred to as "IP spoofing". It results in the anonymity of the sender and prevents a packet from being traced to its origin. Reflection-based Distributed Denial-of-Service (DRDoS) attacks leverage this mechanism and become even more effective using amplification [82–84]. As it is not possible in general to prevent packet header modification, concerted efforts have been undertaken to prevent spoofed packets from reaching potential victims. Filtering packets at the network edge formalized in RFC 2827 and called *Source Address Validation* (SAV) [85, 86] can achieve this goal.

Given the prevalent role of IP spoofing in cyberattacks, there is a need to estimate the level of SAV deployment by network providers. Projects such as Spoofer [87] already enumerate networks that do not implement packet filtering. However, a great majority of this existing work concentrates on *outbound* SAV [81, 87–93] it can prevent DRDoS attacks near their origin [83]. While less obvious, the lack of *inbound* filtering enables an external attacker to masquerade as an internal host of a network, which may reveal valuable information about the network infrastructure usually not seen from the outside. Inbound IP spoofing can serve as a vector for DNS zone poisoning attacks [2] that may lead to domain hijacking or cache poisoning attacks [33] even if the DNS resolver is correctly configured as a closed resolver. A closed resolver only accepts DNS queries from known clients and does so by matching the source IP address of a query against a list of allowed addresses.

The lack of SAV for inbound traffic can also have devastating consequences when combined with the DNS Unchained [84] or the NXDOMAIN attack (also known as

the Water Torture Attack) [94], or the recently discovered NXNSAttack [95]. These attacks result in Denial-of-Service against both recursive resolvers and authoritative servers with a maximum packet amplification factor of 1,620 for the NXNSAttack [95]. IP spoofing is not required for this attack to succeed because any client can attack a resolver if it is allowed to query it. However, IP spoofing can greatly increase the number of affected resolvers by allowing an external attacker to target closed DNS resolvers: the attacker simply needs to masquerade as a legitimate client by spoofing its source IP address. Deploying inbound SAV at the edge of a network is an effective way of protecting closed DNS resolvers from this type of external attacks as well as possible zero-day vulnerabilities that reside within the DNS server software.

The goal of the recently launched Closed Resolver Project [18,96–98] is to enumerate networks vulnerable to inbound spoofing across the Internet as a first step in estimating the scale of the problem. We have presented a novel method to infer the deployment of inbound SAV for the IPv4 and IPv6 address spaces. We have measured the filtering policies of 52% of routable IPv4 autonomous systems (26% for IPv6) and 28% of all the IPv4 BGP prefixes (almost 9% for IPv6). We showed that the vast majority of the networks for which we obtained measurements are (consistently or partially) vulnerable to inbound spoofing and that the vulnerability is not limited to any geographic territory and is spread worldwide.

> **Recommendation**: Network operators should deploy IP Source Address Validation (SAV) not only for outgoing but also for *incoming* traffic at the edge of a network. It provides an effective way of protecting closed DNS resolvers from different external attacks against DNS infrastructure, including possible zero-day vulnerabilities within the DNS server software.

Acknowledgements

References

[1] S. Maroofi, M. Korczyński, C. Hesselman, B. Ampeau, and A. Duda, "COMAR: Classification of Compromised versus Maliciously Registered Domains," in *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*, 2020.

[2] M. Korczyński, M. Król, and M. van Eeten, "Zone Poisoning: The How and Where of Non-Secure DNS Dynamic Updates," in *Proceedings of the 2016 ACM on Internet Measurement Conference*, ser. IMC '16.   ACM, 2016, pp. 271–278.

[3] "The Spamhaus Project," www.spamhaus.org.

[4] "SURBL - URI reputation data," http://www.surbl.org.

[5] "ABUSE CH," https://abuse.ch.

[6] "APWG: Cross-industry Global Group Supporting Tackling the Phishing Menace," http://antiphishing.org.

[7] "PhishTank: A Nonprofit Anti-phishing Organization," http://www.phishtank.com.

[8] "OpenPhish," https://openphish.com.

[9] "The Domain Block List," https://www.spamhaus.org/dbl.

[10] "SURBL Lists," http://www.surbl.org/lists.

[11] P. Nicks, "Uniregistry's registrar and marketplace joining GoDaddy to bring domain investing to the next level," https://www.godaddy.com/garage/uniregistrys-registrar-and-marketplace-joining-godaddy, February 2020.

[12] S. Maroofi, M. Korczynski, and A. Duda, "From Defensive Registration to Subdomain Protection: Evaluation of Email Anti-Spoofing Schemes for High-Profile Domains," in *Proc. Network Traffic Measurement and Analysis Conference (TMA)*, 2020.

[13] C. Rossow, "Amplification Hell: Revisiting Network Protocols for DDoS Abuse," in *Network and Distributed System Security Symposium*, 2014.

[14] M. Kührer, T. Hupperich, C. Rossow, and T. Holz, "Exit from Hell? Reducing the Impact of Amplification DDoS Attacks," in *USENIX Conference on Security Symposium*, 2014.

[15] "Domain names - implementation and specification," RFC 1035, Nov. 1987. [Online]. Available: https://rfc-editor.org/rfc/rfc1035.txt

[16] "University of Oregon Route Views Project," http://www.routeviews.org/routeviews/.

[17] O. Gasser, Q. Scheitle, P. Foremski, Q. Lone, M. Korczynski, S. D. Strowes, L. Hendriks, and G. Carle, "Clusters in the expanse: Understanding and unbiasing ipv6 hitlists," in *Proceedings of the 2018 Internet Measurement Conference.* New York, NY, USA: ACM, 2018.

[18] M. Korczyński, Y. Nosyk, Q. Lone, M. Skwarek, B. Jonglez, and A. Duda, "The Closed Resolver Project: Measuring the Deployment of Source Address Validation of Inbound Traffic," in *arXiv, 2006.05277*, 2020.

[19] M. Skwarek, M. Korczyński, W. Mazurczyk, and A. Duda, "Characterizing Vulnerability of DNS AXFR Transfers with Global-Scale Scanning," in *IEEE Security and Privacy Workshops (SPW)*, 2019.

[20] M. Korczyński and A. Noroozian, "Security reputation metrics," in *Encyclopedia of Cryptography, Security and Privacy.* Springer Berlin Heidelberg, 2021. [Online]. Available: https://doi.org/10.1007/978-3-642-27739-9_1625-1

[21] M. Korczyński, M. Wullink, S. Tajalizadehkhoob, G. 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.* ACM, 2018, pp. 609–623.

[22] M. Korczyński, S. Tajalizadehkhoob, A. Noroozian, M. Wullink, C. Hesselman, and M. van Eeten, "Reputation metrics design to improve intermediary incentives for security of tlds," in *2017 IEEE European Symposium on Security and Privacy (Euro SP)*, April 2017.

[23] A. Noroozian, M. Korczyński, S. Tajalizadehkhoob, and M. van Eeten, "Developing security reputation metrics for hosting providers," in *8th Usenix Workshop on Cyber Security Experimentation and Test (CSET 15)*, 2015.

[24] M. Korczyński, 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 *ACM AsiaCCS*, 2018.

[25] "Competition, Consumer Trust, and Consumer Choice Review Team - Final Report," https://www.icann.org/en/system/files/files/cct-final-08sep18-en.pdf, 2018.

[26] "Second Security, Stability, and Resiliency (SSR2) Review Team Final Report," https://www.icann.org/en/system/files/files/ssr2-review-team-final-report-25jan21-en.pdf, 2021.

[27] (2016) Global phishing survey: Trends and domain name use in 2016. [Online]. Available: https://docs.apwg.org/reports/APWG_Global_Phishing_Report_2015-2016.pdf

[28] M. Korczyński, M. Wullink, S. Tajalizadehkhoob, G. C. Moura, and C. Hesselman, "Statistical Analysis of DNS Abuse in gTLDs Final Report," Tech. Rep., 2017. [Online]. Available: https://www.icann.org/en/system/files/files/sadag-final-09aug17-en.pdf

[29] G. Aaron, L. C. D. Piscitello, and C. Strutt, "Phishing Landscape 2020 A Study of the Scope and Distribution of Phishing," Tech. Rep., 2020. [Online]. Available: interisle.net/PhishingLandscape2020.pdf

[30] T. Halvorson, M. F. Der, I. Foster, S. Savage, L. K. Saul, and G. M. Voelker, "From .Academy to .Zone: An Analysis of the New TLD Land Rush," in *Proc. of ACM IMC*, 2015, pp. 381–394.

[31] A. Noroozian, M. Korczyński, S. Tajalizadehkhoob, and M. van Eeten, "Developing Security Reputation Metrics for Hosting Providers," in *Proc. of the 8th USENIX CSET*, 2015, pp. 1–8.

[32] S. Maroofi, M. Korczyński, and A. Duda, "Are You Human?: Resilience of Phishing Detection to Evasion Techniques Based on Human Verification," in *ACM Internet Measurement Conference, Virtual Event.* ACM, 2020, pp. 78–86.

[33] D. Kaminsky, "It's the End of the Cache as We Know It," https://www.slideshare.net/dakami/dmk-bo2-k8.

[34] D. Atkins and R. Austein, "Threat Analysis of the Domain Name System (DNS)," RFC 3833, Aug. 2004. [Online]. Available: https://rfc-editor.org/rfc/rfc3833.txt

[35] R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose, "DNS Security Introduction and Requirements," Internet Requests for Comments, RFC 4033, March 2005. [Online]. Available: http://www.rfc-editor.org/rfc/rfc4033.txt

[36] S. Rose, M. Larson, D. Massey, R. Austein, and R. Arends, "Protocol Modifications for the DNS Security Extensions," RFC 4035, Mar. 2005. [Online]. Available: https://rfc-editor.org/rfc/rfc4035.txt

[37] S. Rose, M. Larson, D. Massey, R. Austein, and R. Arends, "Resource Records for the DNS Security Extensions," RFC 4034, Mar. 2005. [Online]. Available: https://rfc-editor.org/rfc/rfc4034.txt

[38] IANA, "Root KSK Ceremony 1," 2010, https://www.iana.org/dnssec/ceremonies/1.

[39] I. Research, "TLD DNSSEC Report (2021-05-18 00:05:03)," 2021, http://stats.research.icann.org/dns/tld_report/archive/20210518.000101.html.

[40] ICANN, "Domain Name System Security Extensions Now Deployed in all Generic Top-Level Domains," 12 2020, https://www.icann.org/en/announcements/details/domain-name-system-security-extensions-now-deployed-in-all-generic-top-level-domains-23-12-2020-en.

[41] ICANN, "New gTLD Program Safeguards Against DNS Abuse," https://newgtlds.icann.org/en/reviews/dns-abuse/safeguards-against-dns-abuse-18jul16-en.pdf, July 2016.

[42] T. Chung, R. van Rijswijk-Deij, B. Chandrasekaran, D. Choffnes, D. Levin, B. M. Maggs, A. Mislove, and C. Wilson, "A longitudinal, end-to-end view of the DNSSEC ecosystem," in *26th USENIX Security Symposium (USENIX Security 17).* Vancouver, BC: USENIX Association, Aug. 2017, pp. 1307–1322. [Online]. Available: https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/chung

[43] O. Filip, "DNSSEC.CZ," 10 2012, https://archive.icann.org/en/meetings/toronto2012/bitcache/DNSSEC.CZ-vid=41901&disposition=attachment&op=download.pdf.

[44] "Digital Czech Republic v. 2.0 - The Way to the Digital Economy," 04 2014, https://www.mpo.cz/dokument149132.html.

[45] Internetstiftelsen, "Recommendations for DNSSEC deployment," https://internetstiftelsen.se/en/domains/tech-tools/recommendations-for-dnssec-deployment/.

[46] SIDN, "Frequently Asked Questions," https://www.sidn.nl/en/faq/dnssec.

[47] Y. Afek, A. Bremler-Barr, and L. Shafir, "Nxnsattack: Recursive DNS inefficiencies and vulnerabilities," in *29th USENIX Security Symposium, USENIX Security 2020, August 12-14, 2020*, S. Capkun and F. Roesner, Eds. USENIX Association, 2020, pp. 631–648. [Online]. Available: https://www.usenix.org/conference/usenixsecurity20/presentation/afek

[48] R. van Rijswijk-Deij, A. Sperotto, and A. Pras, "Dnssec and its potential for ddos attacks: A comprehensive measurement study," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, ser. IMC '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 449–460. [Online]. Available: https://doi.org/10.1145/2663716.2663731

[49] J. Abley, Ólafur Guðmundsson, M. Majkowski, and E. Hunt, "Providing Minimal-Sized Responses to DNS Queries That Have QTYPE=ANY," RFC 8482, Jan. 2019. [Online]. Available: https://rfc-editor.org/rfc/rfc8482.txt

[50] O. Kolkman, M. Mekking, and R. M. Gieben, "DNSSEC Operational Practices, Version 2," RFC 6781, Dec. 2012. [Online]. Available: https://rfc-editor.org/rfc/rfc6781.txt

[51] P. Wouters and O. Surý, "Algorithm Implementation Requirements and Usage Guidance for DNSSEC," RFC 8624, Jun. 2019. [Online]. Available: https://rfc-editor.org/rfc/rfc8624.txt

[52] E. Barker and A. Roginsky, "Transitions: Recommendation for Transitioning the Use of Cryptographic Algorithms and Key Lengths," in *Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD*, 01 2011.

[53] T. Chung, R. van Rijswijk-Deij, D. Choffnes, D. Levin, B. M. Maggs, A. Mislove, and C. Wilson, "Understanding the role of registrars in dnssec deployment," in *Proceedings of the 2017 Internet Measurement Conference*, ser. IMC '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 369–383. [Online]. Available: https://doi.org/10.1145/3131365.3131373

[54] M. Kührer, T. Hupperich, J. Bushart, C. Rossow, and T. Holz, "Going Wild: Large-Scale Classification of Open DNS Resolvers," in *Internet Measurement Conference*. ACM, 2015.

[55] J. Mauch, "Spoofing ASNs," http://seclists.org/nanog/2013/Aug/132.

[56] "Alert (TA14-017A) UDP-Based Amplification Attacks," https://us-cert.cisa.gov/ncas/alerts/TA14-017A/.

[57] "Reports on openly accessible services," https://www.bsi.bund.de/EN/Topics/IT-Crisis-Management/CERT-Bund/CERT-Reports/Reports/openly-accessible-services/openly-accessible-services_node.html.

[58] "Business Email Compromise," https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf, 2020.

[59] S. Kitterman, "RFC 7208: Sender Policy Framework (SPF) for Authorizing Use of Domains in Email," Internet Requests for Comments, 2014. [Online]. Available: http://tools.ietf.org/html/rfc7208

[60] D. Crocker, T. Hansen, and M. Kucherawy, "RFC 6376: DomainKeys Identified Mail (DKIM) Signatures," Internet Requests for Comments, 2011. [Online]. Available: http://tools.ietf.org/html/rfc6376

[61] E. Kucherawy, M. Zwicky, and E. Zwicky, "RFC 7489: Domain-Based Message Authentication, Reporting, and Conformance (DMARC)," Internet Requests for Comments, 2015. [Online]. Available: http://tools.ietf.org/html/rfc7489

[62] H. Hu and G. Wang, "End-to-end measurements of email spoofing attacks," in *27th USENIX Security Symposium*, 2018, pp. 1095–1112.

[63] S. Maroofi, M. Korczyński, A. Hölzel, and A. Duda, "Adoption of Email Anti-Spoofing Schemes: A Large Scale Analysis," *IEEE Transactions on Network and Service Management*, 2021.

[64] F. Li, Z. Durumeric, J. Czyz, M. Karami, M. Bailey, D. McCoy, S. Savage, and V. Paxson, "You've Got Vulnerability: Exploring Effective Vulnerability Notifications," in *USENIX Security*, 2016.

[65] O. Cetin, C. Gañán, M. Korczyński, and M. van Eeten, "Make Notifications Great Again: Learning How to Notify in the Age of Large-Scale Vulnerability Scanning," in *WEIS*, 2017.

[66] Z. Durumeric, J. Kasten, D. Adrian, J. A. Halderman, M. Bailey, F. Li, N. Weaver, J. Amann, J. Beekman, M. Payer, and V. Paxson, "The Matter of Heartbleed," in *Proc. IMC*, 2014.

[67] F. Li, G. Ho, E. Kuan, Y. Niu, L. Ballard, K. Thomas, E. Bursztein, and V. Paxson, "Remedying Web Hijacking: Notification Effectiveness and Webmaster Comprehension," in *WWW*, 2016.

[68] B. Stock, G. Pellegrino, C. Rossow, M. Johns, and M. Backes, "Hey, You Have a Problem: On the Feasibility of Large-scale Web Vulnerability Notification," in *USENIX Security*, 2016.

[69] B. Stock, G. Pellegrino, F. Li, M. Backes, and C. Rossow, "Didn't You Hear Me?—Towards More Successful Web Vulnerability Notifications," in *NDSS*, 2018.

[70] E. Zeng, F. Li, E. Stark, A. P. Felt, and P. Tabriz, "Fixing HTTPS Misconfigurations at Scale: An Experiment with Security Notifications," in *WEIS*, 2019.

[71] O. Çetin, M. H. Jhaveri, C. Gañán, M. van Eeten, and T. Moore, "Understanding the Role of Sender Reputation in Abuse Reporting and Cleanup," *J. Cybersecur.*, vol. 2, no. 1, pp. 83–98, 2016.

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

[73] "Advisory Statement: Temporary Specification for gTLD Registration Data," https://www.icann.org/en/system/files/files/advisory-statement-gtld-registration-data-specs-17may18-en.pdf.

[74] W. Soussi, M. Korczynski, S. Maroofi, and A. Duda, "Feasibility of Large-Scale Vulnerability Notifications after GDPR," in *IEEE European Symposium on Security and Privacy Workshops, EuroS&P Workshops 2020.* IEEE, 2020, pp. 532–537.

[75] D. Crocker, "Mailbox names for common services, roles and functions," RFC 2142, 1997.

[76] J. Klensin, "Simple Mail Transfer Protocol," RFC 5321, 2008.

[77] Postfix Virtual Domain Hosting Howto. [Online]. Available: http://www.postfix.org/VIRTUAL_README.html

[78] P. Mockapetris, "Domain Names - Implementation and Specification," RFC 1035, 1987.

[79] V. L. Pochat, T. Van Goethem, S. Tajalizadehkhoob, M. Korczyński, and W. Joosen, "Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation," in *NDSS*, 2019.

[80] S. B. Hulley and S. R. Cummings and W. S. Browner and D. Grady and T. B. Newman, *Designing Clinical Research: an Epidemiologic Approach. 4th ed.* Philadelphia, PA: Lippincott Williams & Wilkins, July 2013.

[81] R. Beverly, A. Berger, Y. Hyun, and k. claffy, "Understanding the Efficacy of Deployed Internet Source Address Validation Filtering," in *Internet Measurement Conference.* ACM, 2009.

[82] C. Rossow, "Amplification Hell: Revisiting Network Protocols for DDoS Abuse," in *Network and Distributed System Security Symposium*, 2014.

[83] M. Kührer, T. Hupperich, C. Rossow, and T. Holz, "Exit from Hell? Reducing the Impact of Amplification DDoS Attacks," in *USENIX Conference on Security Symposium*, 2014.

[84] J. Bushart and C. Rossow, "DNS Unchained: Amplified Application-Layer DoS Attacks Against DNS Authoritatives," in *Research in Attacks, Intrusions, and Defenses*, M. Bailey, T. Holz, M. Stamatogiannakis, and S. Ioannidis, Eds. Cham: Springer International Publishing, 2018, pp. 139–160.

[85] D. Senie and P. Ferguson, "Network Ingress Filtering: Defeating Denial of Service Attacks which Employ IP Source Address Spoofing," RFC 2827, May 2000. [Online]. Available: https://rfc-editor.org/rfc/rfc2827.txt

[86] M. Korczyński and Y. Nosyk, "Source Address Validation," in *Encyclopedia of Cryptography, Security and Privacy.* Springer Berlin Heidelberg, 2021. [Online]. Available: https://doi.org/10.1007/978-3-642-27739-9_1626-1

[87] CAIDA, "The Spoofer Project," https://www.caida.org/projects/spoofer/.

[88] R. Beverly and S. Bauer, "The Spoofer Project: Inferring the Extent of Source Address Filtering on the Internet," in *USENIX Steps to Reducing Unwanted Traffic on the Internet Workshop*, Jul. 2005.

[89] F. Lichtblau, F. Streibelt, T. Krüger, P. Richter, and A. Feldmann, "Detection, Classification, and Analysis of Inter-domain Traffic with Spoofed Source IP Addresses," in *Internet Measurement Conference.* ACM, 2017.

[90] Q. Lone, M. Luckie, M. Korczyński, H. Asghari, M. Javed, and M. van Eeten, "Using Crowdsourcing Marketplaces for Network Measurements: The Case of Spoofer," in *Traffic Monitoring and Analysis Conference*, 2018.

[91] Q. Lone, M. Luckie, M. Korczyński, and M. van Eeten, "Using Loops Observed in Traceroute to Infer the Ability to Spoof," in *Passive and Active Measurement Conference*. Springer International Publishing, 2017.

[92] M. Luckie, R. Beverly, R. Koga, K. Keys, J. Kroll, and k. claffy, "Network Hygiene, Incentives, and Regulation: Deployment of Source Address Validation in the Internet," in *Computer and Communications Security Conference*. ACM, 2019.

[93] Q. Lone, M. Korczyński, C. Gañán, and M. van Eeten, "SAVing the Internet: Explaining the Adoption of Source Address Validation by Internet Service Providers," in *Workshop on the Economics of Information Security*, 2020.

[94] X. Luo, L. Wang, Z. Xu, K. Chen, J. Yang, and T. Tian, "A Large Scale Analysis of DNS Water Torture Attack," in *Conference on Computer Science and Artificial Intelligence*, 2018.

[95] L. Shafir, Y. Afek, and A. Bremler-Barr, "NXNSAttack: Recursive DNS Inefficiencies and Vulnerabilities," in *USENIX Security Symposium*, 2020.

[96] M. Korczyński, Y. Nosyk, Q. Lone, M. Skwarek, B. Jonglez, and A. Duda, "Don't Forget to Lock the Front Door! Inferring the Deployment of Source Address Validation of Inbound Traffic," in *Passive and Active Measurement*. Springer International Publishing, 2020.

[97] "The Closed Resolver Project," https://closedresolver.com.

[98] M. Korczyński, Y. Nosyk, Q. Lone, M. Skwarek, B. Jonglez, and A. Duda, "Inferring the deployment of inbound source address validation using dns resolvers," in *Proceedings of the Applied Networking Research Workshop*, ser. ANRW'20. Association for Computing Machinery, 2020, p. 9–11.