ï»¿

Internet Engineering Task Force (IETF)
Request for Comments: 8932
BCP: 232
Category: Best Current Practice
ISSN: 2070-1721

S. Dickinson
Sinodun IT
B. Overeinder
R. van Rijswijk-Deij
NLnet Labs
A. Mankin
Salesforce

October 2020

Recommendations for DNS Privacy Service Operators

Abstract

This document presents operational, policy, and security considerations for DNS recursive resolver operators who choose to offer DNS privacy services. With these recommendations, the operator can make deliberate decisions regarding which services to provide, as well as understanding how those decisions and the alternatives impact the privacy of users.

This document also presents a non-normative framework to assist writers of a Recursive operator Privacy Statement, analogous to DNS Security Extensions (DNSSEC) Policies and DNSSEC Practice Statements described in RFC 6841.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

The Domain Name System (DNS) is at the core of the Internet; almost every activity on the Internet starts with a DNS query (and often several). However, the DNS was not originally designed with strong security or privacy mechanisms. A number of developments have taken place in recent years that aim to increase the privacy of the DNS, and these are now seeing some deployment. This latest evolution of the DNS presents new challenges to operators, and this document attempts to provide an overview of considerations for privacy-focused DNS services.

In recent years, there has also been an increase in the availability of "public resolvers" [RFC8499], which users may prefer to use instead of the default network resolver, either because they offer a specific feature (e.g., good reachability or encrypted transport) or because the network resolver lacks a specific feature (e.g., strong privacy policy or unfiltered responses). These public resolvers have tended to be at the forefront of adoption of privacy-related enhancements, but it is anticipated that operators of other resolver services will follow.

Whilst protocols that encrypt DNS messages on the wire provide protection against certain attacks, the resolver operator still has (in principle) full visibility of the query data and transport identifiers for each user. Therefore, a trust relationship (whether explicit or implicit) is assumed to exist between each user and the operator of the resolver(s) used by that user. The ability of the operator to provide a transparent, well-documented, and secure privacy service will likely serve as a major differentiating factor for privacy-conscious users if they make an active selection of which resolver to use.

It should also be noted that there are both advantages and disadvantages to a user choosing to configure a single resolver (or a fixed set of resolvers) and an encrypted transport to use in all network environments. For example, the user has a clear expectation of which resolvers have visibility of their query data. However, this resolver/transport selection may provide an added mechanism for tracking them as they move across network environments. Commitments from resolver operators to minimize such tracking as users move between networks are also likely to play a role in user selection of resolvers.

More recently, the global legislative landscape with regard to personal data collection, retention, and pseudonymization has seen significant activity. Providing detailed practice advice about these areas to the operator is out of scope, but Section 5.3.3 describes some mitigations of data-sharing risk.

This document has two main goals:

* To provide operational and policy guidance related to DNS over encrypted transports and to outline recommendations for data handling for operators of DNS privacy services.

* To introduce the Recursive operator Privacy Statement (RPS) and present a framework to assist writers of an RPS. An RPS is a document that an operator should publish that outlines their operational practices and commitments with regard to privacy, thereby providing a means for clients to evaluate both the measurable and claimed privacy properties of a given DNS privacy service. The framework identifies a set of elements and specifies an outline order for them. This document does not, however, define a particular privacy statement, nor does it seek to provide legal advice as to the contents of an RPS.

A desired operational impact is that all operators (both those providing resolvers within networks and those operating large public services) can demonstrate their commitment to user privacy, thereby driving all DNS resolution services to a more equitable footing. Choices for users would (in this ideal world) be driven by other factors -- e.g., differing security policies or minor differences in operator policy -- rather than gross disparities in privacy concerns.

Community insight (or judgment?) about operational practices can change quickly, and experience shows that a Best Current Practice (BCP) document about privacy and security is a point-in-time statement. Readers are advised to seek out any updates that apply to this document.

2. Scope

"DNS Privacy Considerations" [RFC7626] describes the general privacy issues and threats associated with the use of the DNS by Internet users; much of the threat analysis here is lifted from that document and [RFC6973]. However, this document is limited in scope to best-practice considerations for the provision of DNS privacy services by servers (recursive resolvers) to clients (stub resolvers or forwarders). Choices that are made exclusively by the end user, or those for operators of authoritative nameservers, are out of scope.

This document includes (but is not limited to) considerations in the following areas:

1. Data "on the wire" between a client and a server.

2. Data "at rest" on a server (e.g., in logs).

3. Data "sent onwards" from the server (either on the wire or shared with a third party).

Whilst the issues raised here are targeted at those operators who choose to offer a DNS privacy service, considerations for areas 2 and 3 could equally apply to operators who only offer DNS over unencrypted transports but who would otherwise like to align with privacy best practice.

3. Privacy-Related Documents

There are various documents that describe protocol changes that have the potential to either increase or decrease the privacy properties of the DNS in various ways. Note that this does not imply that some documents are good or bad, better or worse, just that (for example) some features may bring functional benefits at the price of a reduction in privacy, and conversely some features increase privacy with an accompanying increase in complexity. A selection of the most relevant documents is listed in Appendix A for reference.

4. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

DNS terminology is as described in [RFC8499], except with regard to the definition of privacy-enabling DNS server in Section 6 of [RFC8499]. In this document we use the full definition of a DNS over (D)TLS privacy-enabling DNS server as given in [RFC8310], i.e., that such a server should also offer at least one of the credentials described in Section 8 of [RFC8310] and implement the (D)TLS profile described in Section 9 of [RFC8310].

Other Terms:

RPS: Recursive operator Privacy Statement; see Section 6.

DNS privacy service: The service that is offered via a privacy-enabling DNS server and is documented either in an informal statement of policy and practice with regard to users privacy or a formal RPS.

5. Recommendations for DNS Privacy Services

In the following sections, we first outline the threats relevant to the specific topic and then discuss the potential actions that can be taken to mitigate them.

We describe two classes of threats:

* Threats described in [RFC6973], "Privacy Considerations for Internet Protocols"

  - Privacy terminology, threats to privacy, and mitigations as described in Sections 3, 5, and 6 of [RFC6973].

* DNS Privacy Threats

  - These are threats to the users and operators of DNS privacy services that are not directly covered by [RFC6973]. These may be more operational in nature, such as certificate-management or service-availability issues.

We describe three classes of actions that operators of DNS privacy services can take:

* Threat mitigation for well-understood and documented privacy threats to the users of the service and, in some cases, the operators of the service.

* Optimization of privacy services from an operational or management perspective.

* Additional options that could further enhance the privacy and usability of the service.

This document does not specify policy, only best practice. However, for DNS privacy services to be considered compliant with these best-practice guidelines, they SHOULD implement (where appropriate) all:

* Threat mitigations to be minimally compliant.

* Optimizations to be moderately compliant.

* Additional options to be maximally compliant.

The rest of this document does not use normative language but instead refers only to the three differing classes of action that correspond to the three named levels of compliance stated above. However, compliance (to the indicated level) remains a normative requirement.

5.1. On the Wire between Client and Server

In this section, we consider both data on the wire and the service provided to the client.

5.1.1. Transport Recommendations

Threats described in [RFC6973]:
   Surveillance:
      Passive surveillance of traffic on the wire.

DNS Privacy Threats:
   Active injection of spurious data or traffic.

Mitigations:
   A DNS privacy service can mitigate these threats by providing service over one or more of the following transports:

   * DNS over TLS (DoT) [RFC7858] [RFC8310].

   * DNS over HTTPS (DoH) [RFC8484].

It is noted that a DNS privacy service can also be provided over DNS over DTLS [RFC8094]; however, this is an Experimental specification, and there are no known implementations at the time of writing.

It is also noted that DNS privacy service might be provided over DNSCrypt [DNSCrypt], IPsec, or VPNs. However, there are no specific RFCs that cover the use of these transports for DNS, and any discussion of best practice for providing such a service is out of scope for this document.

Whilst encryption of DNS traffic can protect against active injection on the paths traversed by the encrypted connection, this does not diminish the need for DNSSEC; see Section 5.1.4.

5.1.2. Authentication of DNS Privacy Services

Threats described in [RFC6973]:
   Surveillance:
      Active attacks on client resolver configuration.

Mitigations:

DNS privacy services should ensure clients can authenticate the server. Note that this, in effect, commits the DNS privacy service to a public identity users will trust.

When using DoT, clients that select a "Strict Privacy" usage profile [RFC8310] (to mitigate the threat of active attack on the client) require the ability to authenticate the DNS server. To enable this, DNS privacy services that offer DoT need to provide credentials that will be accepted by the client's trust model, in the form of either X.509 certificates [RFC5280] or Subject Public Key Info (SPKI) pin sets [RFC8310].

When offering DoH [RFC8484], HTTPS requires authentication of the server as part of the protocol.

5.1.2.1. Certificate Management

Anecdotal evidence to date highlights the management of certificates as one of the more challenging aspects for operators of traditional DNS resolvers that choose to additionally provide a DNS privacy service, as management of such credentials is new to those DNS operators.

It is noted that SPKI pin set management is described in [RFC7858] but that key-pinning mechanisms in general have fallen out of favor operationally for various reasons, such as the logistical overhead of rolling keys.

DNS Privacy Threats:

* Invalid certificates, resulting in an unavailable service, which might force a user to fall back to cleartext.

* Misidentification of a server by a client -- e.g., typos in DoH URL templates [RFC8484] or authentication domain names [RFC8310] that accidentally direct clients to attacker-controlled servers.

Mitigations:

It is recommended that operators:

* Follow the guidance in Section 6.5 of [RFC7525] with regard to certificate revocation.

* Automate the generation, publication, and renewal of certificates. For example, Automatic Certificate Management Environment (ACME) [RFC8555] provides a mechanism to actively manage certificates through automation and has been implemented by a number of certificate authorities.

* Monitor certificates to prevent accidental expiration of certificates.

* Choose a short, memorable authentication domain name for the service.

5.1.3. Protocol Recommendations

5.1.3.1. DoT

DNS Privacy Threats:

* Known attacks on TLS, such as those described in [RFC7457].

* Traffic analysis, for example: [Pitfalls-of-DNS-Encryption] (focused on DoT).

* Potential for client tracking via transport identifiers.

* Blocking of well-known ports (e.g., 853 for DoT).

Mitigations:

In the case of DoT, TLS profiles from Section 9 of [RFC8310] and the "Countermeasures to DNS Traffic Analysis" from Section 11.1 of [RFC8310] provide strong mitigations. This includes but is not limited to:

* Adhering to [RFC7525].

* Implementing only (D)TLS 1.2 or later, as specified in [RFC8310].

* Implementing Extension Mechanisms for DNS (EDNS(0)) Padding [RFC7830] using the guidelines in [RFC8467] or a successor specification.

* Servers should not degrade in any way the query service level provided to clients that do not use any form of session resumption mechanism, such as TLS session resumption [RFC5077] with TLS 1.2 (Section 2.2 of [RFC8446]) or Domain Name System (DNS) Cookies [RFC7873].

* A DoT privacy service on both port 853 and 443. If the operator deploys DoH on the same IP address, this requires the use of the "dot" Application-Layer Protocol Negotiation (ALPN) value [dot-ALPN].

Optimizations:

* Concurrent processing of pipelined queries, returning responses as soon as available, potentially out of order, as specified in [RFC7766]. This is often called "OOOR" -- out-of-order responses (providing processing performance similar to HTTP multiplexing).

* Management of TLS connections to optimize performance for clients using [RFC7766] and EDNS(0) Keepalive [RFC7828]

Additional Options:

Management of TLS connections to optimize performance for clients using DNS Stateful Operations [RFC8490].

5.1.3.2. DoH

DNS Privacy Threats:

* Known attacks on TLS, such as those described in [RFC7457].

* Traffic analysis, for example: [DNS-Privacy-not-so-private] (focused on DoH).

* Potential for client tracking via transport identifiers.

Mitigations:

* Clients must be able to forgo the use of HTTP cookies [RFC6265] and still use the service.

* Use of HTTP/2 padding and/or EDNS(0) padding, as described in Section 9 of [RFC8484].

* Clients should not be required to include any headers beyond the absolute minimum to obtain service from a DoH server. (See Section 6.1 of [BUILD-W-HTTP].)

5.1.4. DNSSEC

DNS Privacy Threats:

Users may be directed to bogus IP addresses that, depending on the application, protocol, and authentication method, might lead users to reveal personal information to attackers. One example is a website that doesn't use TLS or whose TLS authentication can somehow be subverted.

Mitigations:

All DNS privacy services must offer a DNS privacy service that performs Domain Name System Security Extensions (DNSSEC) validation. In addition, they must be able to provide the DNSSEC Resource Records (RRs) to the client so that it can perform its own validation.

The addition of encryption to DNS does not remove the need for DNSSEC [RFC4033]; they are independent and fully compatible protocols, each solving different problems. The use of one does not diminish the need nor the usefulness of the other.

While the use of an authenticated and encrypted transport protects origin authentication and data integrity between a client and a DNS privacy service, it provides no proof (for a nonvalidating client) that the data provided by the DNS privacy service was actually DNSSEC authenticated. As with cleartext DNS, the user is still solely trusting the Authentic Data (AD) bit (if present) set by the resolver.

It should also be noted that the use of an encrypted transport for DNS actually solves many of the practical issues encountered by DNS validating clients -- e.g., interference by middleboxes with cleartext DNS payloads is completely avoided. In this sense, a validating client that uses a DNS privacy service that supports DNSSEC has a far simpler task in terms of DNSSEC roadblock avoidance [RFC8027].

5.1.5. Availability

DNS Privacy Threats:
A failing DNS privacy service could force the user to switch providers, fall back to cleartext, or accept no DNS service for the duration of the outage.

Mitigations:
A DNS privacy service should strive to engineer encrypted services to the same availability level as any unencrypted services they provide. Particular care should to be taken to protect DNS privacy services against denial-of-service (DoS) attacks, as experience has shown that unavailability of DNS resolving because of attacks is a significant motivation for users to switch services. See, for example, Section IV-C of [Passive-Observations-of-a-Large-DNS].

Techniques such as those described in Section 10 of [RFC7766] can be of use to operators to defend against such attacks.

5.1.6. Service Options

DNS Privacy Threats:
Unfairly disadvantaging users of the privacy service with respect to the services available. This could force the user to switch providers, fall back to cleartext, or accept no DNS service for the duration of the outage.

Mitigations:
A DNS privacy service should deliver the same level of service as offered on unencrypted channels in terms of options such as filtering (or lack thereof), DNSSEC validation, etc.

5.1.7. Impact of Encryption on Monitoring by DNS Privacy Service Operators

DNS Privacy Threats:
Increased use of encryption can impact a DNS privacy service operator's ability to monitor traffic and therefore manage their DNS servers [RFC8404].

Many monitoring solutions for DNS traffic rely on the plaintext nature of this traffic and work by intercepting traffic on the wire, either using a separate view on the connection between clients and the resolver, or as a separate process on the resolver system that inspects network traffic. Such solutions will no longer function when traffic between clients and resolvers is encrypted. Many DNS privacy service operators still need to inspect DNS traffic -- e.g., to monitor for network security threats. Operators may therefore need to invest in an alternative means of monitoring that relies on either the resolver software directly, or exporting DNS traffic from the resolver using, for example, [dnstap].

Optimization:
When implementing alternative means for traffic monitoring, operators of a DNS privacy service should consider using privacy-conscious means to do so. See Section 5.2 for more details on data handling and the discussion on the use of Bloom Filters in Appendix B.

5.1.8. Limitations of Fronting a DNS Privacy Service with a Pure TLS Proxy

DNS Privacy Threats:

* Limited ability to manage or monitor incoming connections using DNS-specific techniques.

* Misconfiguration (e.g., of the target-server address in the proxy configuration) could lead to data leakage if the proxy-to-target-server path is not encrypted.

Optimization:
Some operators may choose to implement DoT using a TLS proxy (e.g., [nginx], [haproxy], or [stunnel]) in front of a DNS nameserver because of proven robustness and capacity when handling large numbers of client connections, load-balancing capabilities, and good tooling. Currently, however, because such proxies typically have no specific handling of DNS as a protocol over TLS or DTLS, using them can restrict traffic management at the proxy layer and the DNS server. For example, all traffic received by a nameserver behind such a proxy will appear to originate from the proxy, and DNS techniques such as Access Control Lists (ACLs), Response Rate Limiting (RRL), or DNS64 [RFC6147] will be hard or impossible to implement in the nameserver.

Operators may choose to use a DNS-aware proxy, such as [dnsdist], that offers custom options (similar to those proposed in [DNS-XPF]) to add source information to packets to address this shortcoming. It should be noted that such options potentially significantly increase the leaked information in the event of a misconfiguration.

5.2. Data at Rest on the Server

5.2.1. Data Handling

Threats described in [RFC6973]:

* Surveillance.

* Stored-data compromise.

* Correlation.

* Identification.

* Secondary use.

* Disclosure.

Other Threats

* Contravention of legal requirements not to process user data.

Mitigations:
The following are recommendations relating to common activities for DNS service operators; in all cases, data retention should be minimized or completely avoided if possible for DNS privacy

services. If data is retained, it should be encrypted and either aggregated, pseudonymized, or anonymized whenever possible. In general, the principle of data minimization described in [RFC6973] should be applied.

* Transient data (e.g., data used for real-time monitoring and threat analysis, which might be held only in memory) should be retained for the shortest possible period deemed operationally feasible.

* The retention period of DNS traffic logs should be only as long as is required to sustain operation of the service and meet regulatory requirements, to the extent that they exist.

* DNS privacy services should not track users except for the particular purpose of detecting and remedying technically malicious (e.g., DoS) or anomalous use of the service.

* Data access should be minimized to only those personnel who require access to perform operational duties. It should also be limited to anonymized or pseudonymized data where operationally feasible, with access to full logs (if any are held) only permitted when necessary.

Optimizations:

* Consider use of full-disk encryption for logs and data-capture storage.

5.2.2. Data Minimization of Network Traffic

Data minimization refers to collecting, using, disclosing, and storing the minimal data necessary to perform a task, and this can be achieved by removing or obfuscating privacy-sensitive information in network traffic logs. This is typically personal data or data that can be used to link a record to an individual, but it may also include other confidential information -- for example, on the structure of an internal corporate network.

The problem of effectively ensuring that DNS traffic logs contain no or minimal privacy-sensitive information is not one that currently has a generally agreed solution or any standards to inform this discussion. This section presents an overview of current techniques to simply provide reference on the current status of this work.

Research into data minimization techniques (and particularly IP address pseudonymization/anonymization) was sparked in the late 1990s / early 2000s, partly driven by the desire to share significant corpuses of traffic captures for research purposes. Several techniques reflecting different requirements in this area and different performance/resource trade-offs emerged over the course of the decade. Developments over the last decade have been both a blessing and a curse; the large increase in size between an IPv4 and an IPv6 address, for example, renders some techniques impractical, but also makes available a much larger amount of input entropy, the better to resist brute-force re-identification attacks that have grown in practicality over the period.

Techniques employed may be broadly categorized as either anonymization or pseudonymization. The following discussion uses the definitions from [RFC6973], Section 3, with additional observations from [van-Dijkhuizen-et-al].

* Anonymization. To enable anonymity of an individual, there must exist a set of individuals that appear to have the same attribute(s) as the individual. To the attacker or the observer, these individuals must appear indistinguishable from each other.

* Pseudonymization. The true identity is deterministically replaced with an alternate identity (a pseudonym). When the pseudonymization schema is known, the process can be reversed, so the original identity becomes known again.

In practice, there is a fine line between the two; for example, it is difficult to categorize a deterministic algorithm for data minimization of IP addresses that produces a group of pseudonyms for a single given address.

5.2.3. IP Address Pseudonymization and Anonymization Methods

A major privacy risk in DNS is connecting DNS queries to an individual, and the major vector for this in DNS traffic is the client IP address.

There is active discussion in the space of effective pseudonymization of IP addresses in DNS traffic logs; however, there seems to be no single solution that is widely recognized as suitable for all or most use cases. There are also as yet no standards for this that are unencumbered by patents.

Appendix B provides a more detailed survey of various techniques employed or under development in 2020.

5.2.4. Pseudonymization, Anonymization, or Discarding of Other Correlation Data

DNS Privacy Threats:

* Fingerprinting of the client OS via various means, including: IP TTL/Hoplimit, TCP parameters (e.g., window size, Explicit Congestion Notification (ECN) support, selective acknowledgment (SACK)), OS-specific DNS query patterns (e.g., for network connectivity, captive portal detection, or OS-specific updates).

* Fingerprinting of the client application or TLS library by, for example, HTTP headers (e.g., User-Agent, Accept, Accept-Encoding), TLS version/Cipher-suite combinations, or other connection parameters.

* Correlation of queries on multiple TCP sessions originating from the same IP address.

* Correlating of queries on multiple TLS sessions originating from the same client, including via session-resumption mechanisms.

* Resolvers _might_ receive client identifiers -- e.g., Media Access Control (MAC) addresses in EDNS(0) options. Some customer premises equipment (CPE) devices are known to add them [MAC-address-EDNS].

Mitigations:

* Data minimization or discarding of such correlation data.

5.2.5. Cache Snooping

Threats described in [RFC6973]:

Surveillance:

Profiling of client queries by malicious third parties.

Mitigations:

See [ISC-Knowledge-database-on-cache-snooping] for an example discussion on defending against cache snooping. Options proposed include limiting access to a server and limiting nonrecursive queries.

5.3. Data Sent Onwards from the Server

In this section, we consider both data sent on the wire in upstream queries and data shared with third parties.

5.3.1. Protocol Recommendations

Threats described in [RFC6973]:
Surveillance:
Transmission of identifying data upstream.

Mitigations:
The server should:

* implement QNAME minimization [RFC7816].

* honor a SOURCE PREFIX-LENGTH set to 0 in a query containing the EDNS(0) Client Subnet (ECS) option ([RFC7871], Section 7.1.2). This is as specified in [RFC8310] for DoT but applicable to any DNS privacy service.

Optimizations:
As per Section 2 of [RFC7871], the server should either:

* not use the ECS option in upstream queries at all, or

* offer alternative services, one that sends ECS and one that does not.

If operators do offer a service that sends the ECS options upstream, they should use the shortest prefix that is operationally feasible and ideally use a policy of allowlisting upstream servers to which to send ECS in order to reduce data leakage. Operators should make clear in any policy statement what prefix length they actually send and the specific policy used.

Allowlisting has the benefit that not only does the operator know which upstream servers can use ECS, but also the operator can decide which upstream servers apply privacy policies that the operator is happy with. However, some operators consider allowlisting to incur significant operational overhead compared to dynamic detection of ECS support on authoritative servers.

Additional options:

* "Aggressive Use of DNSSEC-Validated Cache" [RFC8198] and "NXDOMAIN: There Really Is Nothing Underneath" [RFC8020] to reduce the number of queries to authoritative servers to increase privacy.

* Run a local copy of the root zone [RFC8806] to avoid making queries to the root servers that might leak information.

5.3.2. Client Query Obfuscation

Additional options:

Since queries from recursive resolvers to authoritative servers are performed using cleartext (at the time of writing), resolver services need to consider the extent to which they may be directly leaking information about their client community via these upstream queries and what they can do to mitigate this further. Note that, even when all the relevant techniques described above are employed, there may still be attacks possible -- e.g., [Pitfalls-of-DNS-Encryption]. For example, a resolver with a very small community of users risks exposing data in this way and ought to obfuscate this traffic by mixing it with "generated" traffic to make client characterization harder. The resolver could also employ aggressive prefetch techniques as a further measure to counter traffic analysis.

At the time of writing, there are no standardized or widely recognized techniques to perform such obfuscation or bulk prefetches.

Another technique that particularly small operators may consider is forwarding local traffic to a larger resolver (with a privacy policy that aligns with their own practices) over an encrypted protocol, so that the upstream queries are obfuscated among those of the large resolver.

5.3.3. Data Sharing

Threats described in [RFC6973]:

* Surveillance.
* Stored-data compromise.
* Correlation.
* Identification.
* Secondary use.
* Disclosure.

DNS Privacy Threats:

Contravention of legal requirements not to process user data.

Mitigations:

Operators should not share identifiable data with third parties.

If operators choose to share identifiable data with third parties in specific circumstances, they should publish the terms under which data is shared.

Operators should consider including specific guidelines for the collection of aggregated and/or anonymized data for research purposes, within or outside of their own organization. This can benefit not only the operator (through inclusion in novel research) but also the wider Internet community. See the policy published by SURFnet [SURFnet-policy] on data sharing for research as an example.

6. Recursive Operator Privacy Statement (RPS)

To be compliant with this Best Current Practice document, a DNS recursive operator SHOULD publish a Recursive operator Privacy Statement (RPS). Adopting the outline, and including the headings in the order provided, is a benefit to persons comparing RPSs from multiple operators.

Appendix C provides a comparison of some existing policy and privacy statements.

6.1. Outline of an RPS

The contents of Sections 6.1.1 and 6.1.2 are non-normative, other than the order of the headings. Material under each topic is present to assist the operator developing their own RPS. This material:

* Relates _only_ to matters around the technical operation of DNS privacy services, and no other matters.
* Does not attempt to offer an exhaustive list for the contents of an RPS.
* Is not intended to form the basis of any legal/compliance documentation.

Appendix D provides an example (also non-normative) of an RPS statement for a specific operator scenario.

6.1.1. Policy

1. Treatment of IP addresses. Make an explicit statement that IP addresses are treated as personal data.

2. Data collection and sharing. Specify clearly what data (including IP addresses) is:

* Collected and retained by the operator, and for what period it is retained.

* Shared with partners.

* Shared, sold, or rented to third parties.

In each case, specify whether data is aggregated, pseudonymized, or anonymized and the conditions of data transfer. Where possible provide details of the techniques used for the above data minimizations.

3. Exceptions. Specify any exceptions to the above -- for example, technically malicious or anomalous behavior.

4. Associated entities. Declare and explicitly enumerate any partners, third-party affiliations, or sources of funding.

5. Correlation. Whether user DNS data is correlated or combined with any other personal information held by the operator.

6. Result filtering. This section should explain whether the operator filters, edits, or alters in any way the replies that it receives from the authoritative servers for each DNS zone before forwarding them to the clients. For each category listed below, the operator should also specify how the filtering lists are created and managed, whether it employs any third-party sources for such lists, and which ones.

   * Specify if any replies are being filtered out or altered for network- and computer-security reasons (e.g., preventing connections to malware-spreading websites or botnet control servers).

   * Specify if any replies are being filtered out or altered for mandatory legal reasons, due to applicable legislation or binding orders by courts and other public authorities.

   * Specify if any replies are being filtered out or altered for voluntary legal reasons, due to an internal policy by the operator aiming at reducing potential legal risks.

   * Specify if any replies are being filtered out or altered for any other reason, including commercial ones.

6.1.2. Practice

Communicate the current operational practices of the service.

1. Deviations. Specify any temporary or permanent deviations from the policy for operational reasons.

2. Client-facing capabilities. With reference to each subsection of Section 5.1, provide specific details of which capabilities (transport, DNSSEC, padding, etc.) are provided on which client-facing addresses/port combination or DoH URI template. For Section 5.1.2, clearly specify which specific authentication mechanisms are supported for each endpoint that offers DoT:

   a. The authentication domain name to be used (if any).

   b. The SPKI pin sets to be used (if any) and policy for rolling keys.

3. Upstream capabilities. With reference to Section 5.3, provide specific details of which capabilities are provided upstream for data sent to authoritative servers.

4. Support. Provide contact/support information for the service.

5. Data Processing. This section can optionally communicate links

to, and the high-level contents of, any separate statements the operator has published that cover applicable data-processing legislation or agreements with regard to the location(s) of service provision.

6.2. Enforcement/Accountability

Transparency reports may help with building user trust that operators adhere to their policies and practices.

Where possible, independent monitoring or analysis could be performed of:

* ECS, QNAME minimization, EDNS(0) padding, etc.

* Filtering.

* Uptime.

This is by analogy with several TLS or website-analysis tools that are currently available -- e.g., [SSL-Labs] or [Internet.nl].

Additionally, operators could choose to engage the services of a third-party auditor to verify their compliance with their published RPS.

7. IANA Considerations

This document has no IANA actions.

8. Security Considerations

Security considerations for DNS over TCP are given in [RFC7766], many of which are generally applicable to session-based DNS. Guidance on operational requirements for DNS over TCP are also available in [DNS-OVER-TCP]. Security considerations for DoT are given in [RFC7858] and [RFC8310], and those for DoH in [RFC8484].

Security considerations for DNSSEC are given in [RFC4033], [RFC4034], and [RFC4035].

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC4033] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "DNS Security Introduction and Requirements", RFC 4033, DOI 10.17487/RFC4033, March 2005, <https://www.rfc-editor.org/info/rfc4033>.

[RFC5280] Cooper, D., Santesson, S., Farrell, S., Boeyen, S., Housley, R., and W. Polk, "Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile", RFC 5280, DOI 10.17487/RFC5280, May 2008, <https://www.rfc-editor.org/info/rfc5280>.

[RFC6973] Cooper, A., Tschofenig, H., Aboba, B., Peterson, J., Morris, J., Hansen, M., and R. Smith, "Privacy Considerations for Internet Protocols", RFC 6973, DOI 10.17487/RFC6973, July 2013, <https://www.rfc-editor.org/info/rfc6973>.

[RFC7457] Sheffer, Y., Holz, R., and P. Saint-Andre, "Summarizing Known Attacks on Transport Layer Security (TLS) and Datagram TLS (DTLS)", RFC 7457, DOI 10.17487/RFC7457, February 2015, <https://www.rfc-editor.org/info/rfc7457>.

[RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <https://www.rfc-editor.org/info/rfc7525>.

[RFC7766] Dickinson, J., Dickinson, S., Bellis, R., Mankin, A., and D. Wessels, "DNS Transport over TCP - Implementation Requirements", RFC 7766, DOI 10.17487/RFC7766, March 2016, <https://www.rfc-editor.org/info/rfc7766>.

[RFC7816] Bortzmeyer, S., "DNS Query Name Minimisation to Improve Privacy", RFC 7816, DOI 10.17487/RFC7816, March 2016, <https://www.rfc-editor.org/info/rfc7816>.

[RFC7828] Wouters, P., Abley, J., Dickinson, S., and R. Bellis, "The edns-tcp-keepalive EDNS0 Option", RFC 7828, DOI 10.17487/RFC7828, April 2016, <https://www.rfc-editor.org/info/rfc7828>.

[RFC7830] Mayrhofer, A., "The EDNS(0) Padding Option", RFC 7830, DOI 10.17487/RFC7830, May 2016, <https://www.rfc-editor.org/info/rfc7830>.

[RFC7858] Hu, Z., Zhu, L., Heidemann, J., Mankin, A., Wessels, D., and P. Hoffman, "Specification for DNS over Transport Layer Security (TLS)", RFC 7858, DOI 10.17487/RFC7858, May 2016, <https://www.rfc-editor.org/info/rfc7858>.

[RFC7871] Contavalli, C., van der Gaast, W., Lawrence, D., and W. Kumari, "Client Subnet in DNS Queries", RFC 7871, DOI 10.17487/RFC7871, May 2016, <https://www.rfc-editor.org/info/rfc7871>.

[RFC8020] Bortzmeyer, S. and S. Huque, "NXDOMAIN: There Really Is Nothing Underneath", RFC 8020, DOI 10.17487/RFC8020, November 2016, <https://www.rfc-editor.org/info/rfc8020>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8198] Fujiwara, K., Kato, A., and W. Kumari, "Aggressive Use of DNSSEC-Validated Cache", RFC 8198, DOI 10.17487/RFC8198, July 2017, <https://www.rfc-editor.org/info/rfc8198>.

[RFC8310] Dickinson, S., Gillmor, D., and T. Reddy, "Usage Profiles for DNS over TLS and DNS over DTLS", RFC 8310, DOI 10.17487/RFC8310, March 2018, <https://www.rfc-editor.org/info/rfc8310>.

[RFC8467] Mayrhofer, A., "Padding Policies for Extension Mechanisms for DNS (EDNS(0))", RFC 8467, DOI 10.17487/RFC8467, October 2018, <https://www.rfc-editor.org/info/rfc8467>.

[RFC8484] Hoffman, P. and P. McManus, "DNS Queries over HTTPS (DoH)", RFC 8484, DOI 10.17487/RFC8484, October 2018, <https://www.rfc-editor.org/info/rfc8484>.

[RFC8490] Bellis, R., Cheshire, S., Dickinson, J., Dickinson, S., Lemon, T., and T. Pusateri, "DNS Stateful Operations", RFC 8490, DOI 10.17487/RFC8490, March 2019, <https://www.rfc-editor.org/info/rfc8490>.

[RFC8499] Hoffman, P., Sullivan, A., and K. Fujiwara, "DNS Terminology", BCP 219, RFC 8499, DOI 10.17487/RFC8499, January 2019, <https://www.rfc-editor.org/info/rfc8499>.

[RFC8806] Kumari, W. and P. Hoffman, "Running a Root Server Local to a Resolver", RFC 8806, DOI 10.17487/RFC8806, June 2020,

<https://www.rfc-editor.org/info/rfc8806>.

9.2. Informative References

[Bloom-filter]
van Rijswijk-Deij, R., Rijnders, G., Bomhoff, M., and L. Allodi, "Privacy-Conscious Threat Intelligence Using DNSBLOOM", IFIP/IEEE International Symposium on Integrated Network Management (IM2019), 2019, <http://dl.ifip.org/db/conf/im/im2019/189282.pdf>.

[Brekne-and-Arnes]
Brekne, T. and A. Ã\205rnes, "Circumventing IP-address pseudonymization", Communications and Computer Networks, 2005, <https://pdfs.semanticscholar.org/7b34/12c951cebe71cd2cddac5fda164fb2138a44.pdf>.

[BUILD-W-HTTP]
Nottingham, M., "Building Protocols with HTTP", Work in Progress, Internet-Draft, draft-ietf-httpbis-bcp56bis-09, 1 November 2019, <https://tools.ietf.org/html/draft-ietf-httpbis-bcp56bis-09>.

[Crypto-PAn]
CESNET, "Crypto-PAn", commit 636b237, March 2015, <https://github.com/CESNET/ipfixcol/tree/master/base/src/intermediate/anonymization/Crypto-PAn>.

[DNS-OVER-TCP]
Kristoff, J. and D. Wessels, "DNS Transport over TCP - Operational Requirements", Work in Progress, Internet-Draft, draft-ietf-dnsop-dns-tcp-requirements-06, 6 May 2020, <https://tools.ietf.org/html/draft-ietf-dnsop-dns-tcp-requirements-06>.

[DNS-Privacy-not-so-private]
Silby, S., Juarez, M., Vallina-Rodriguez, N., and C. Troncoso, "DNS Privacy not so private: the traffic analysis perspective.", Privacy Enhancing Technologies Symposium, 2018, <https://petsymposium.org/2018/files/hotpets/4-siby.pdf>.

[DNS-XPF] Bellis, R., Dijk, P. V., and R. Gacogne, "DNS X-Proxied-For", Work in Progress, Internet-Draft, draft-bellis-dnsop-xpf-04, 5 March 2018, <https://tools.ietf.org/html/draft-bellis-dnsop-xpf-04>.

[DNSCrypt] "DNSCrypt - Official Project Home Page", <https://www.dnscrypt.org>.

[dnsdist] PowerDNS, "dnsdist Overview", <https://dnsdist.org>.

[dnstap] "dnstap", <https://dnstap.info>.

[DoH-resolver-policy]
Mozilla, "Security/DOH-resolver-policy", 2019, <https://wiki.mozilla.org/Security/DOH-resolver-policy>.

[dot-ALPN] IANA, "Transport Layer Security (TLS) Extensions: TLS Application-Layer Protocol Negotiation (ALPN) Protocol IDs", <https://www.iana.org/assignments/tls-extensiontype-values>.

[Geolocation-Impact-Assessment]
Conversion Works, "Anonymize IP Geolocation Accuracy Impact Assessment", 19 May 2017, <https://www.conversionworks.co.uk/blog/2017/05/19/anonymize-ip-geo-impact-test/>.

[haproxy] "HAProxy - The Reliable, High Performance TCP/HTTP Load Balancer", <https://www.haproxy.org/>.

[Harvan] Harvan, M., "Prefix- and Lexicographical-order-preserving IP Address Anonymization", IEEE/IFIP Network Operations and Management Symposium, DOI 10.1109/NOMS.2006.1687580, 2006, <http://mharvan.net/talks/noms-ip_anon.pdf>.

[Internet.nl] Internet.nl, "Internet.nl Is Your Internet Up To Date?", 2019, <https://internet.nl>.

[IP-Anonymization-in-Analytics] Google, "IP Anonymization in Analytics", 2019, <https://support.google.com/analytics/answer/2763052?hl=en>.

[ipcipher1] Hubert, B., "On IP address encryption: security analysis with respect for privacy", Medium, 7 May 2017, <https://medium.com/@bert.hubert/on-ip-address-encryption-security-analysis-with-respect-for-privacy-dabe1201b476>.

[ipcipher2] PowerDNS, "ipcipher", commit fd47abe, 13 February 2018, <https://github.com/PowerDNS/ipcipher>.

[ipcrypt] veorq, "ipcrypt: IP-format-preserving encryption", commit 8cc12f9, 6 July 2015, <https://github.com/veorq/ipcrypt>.

[ipcrypt-analysis] Aumasson, J-P., "Subject: Re: [Cfrg] Analysis of ipcrypt?", message to the Cfrg mailing list, 22 February 2018, <https://mailarchive.ietf.org/arch/msg/cfrg/cFx5WJo48ZEN-a5cj_LlyrdN8-0/>.

[ISC-Knowledge-database-on-cache-snooping] Goldlust, S. and C. Almond, "DNS Cache snooping - should I be concerned?", ISC Knowledge Database, 15 October 2018, <https://kb.isc.org/docs/aa-00482>.

[MAC-address-EDNS] Hubert, B., "Embedding MAC address in DNS requests for selective filtering", DNS-OARC mailing list, 25 January 2016, <https://lists.dns-oarc.net/pipermail/dns-operations/2016-January/014143.html>.

[nginx] nginx.org, "nginx news", 2019, <https://nginx.org/>.

[Passive-Observations-of-a-Large-DNS] de Vries, W. B., van Rijswijk-Deij, R., de Boer, P-T., and A. Pras, "Passive Observations of a Large DNS Service: 2.5 Years in the Life of Google", DOI 10.23919/TMA.2018.8506536, 2018, <http://tma.ifip.org/2018/wp-content/uploads/sites/3/2018/06/tma2018_paper30.pdf>.

[pcap] The Tcpdump Group, "Tcpdump & Libpcap", 2020, <https://www.tcpdump.org/>.

[Pitfalls-of-DNS-Encryption] Shulman, H., "Pretty Bad Privacy: Pitfalls of DNS Encryption", Proceedings of the 13th Workshop on Privacy in the Electronic Society, pp. 191-200, DOI 10.1145/2665943.2665959, November 2014, <https://dl.acm.org/citation.cfm?id=2665959>.

[policy-comparison] Dickinson, S., "Comparison of policy and privacy statements 2019", DNS Privacy Project, 18 December 2019, <https://dnsprivacy.org/wiki/display/DP/Comparison+of+policy+and+privacy+statements+2019>.

[PowerDNS-dnswasher]
PowerDNS, "dnswasher", commit 050e687, 24 April 2020, <https://github.com/PowerDNS/pdns/blob/master/pdns/dnswasher.cc>.

[Ramaswamy-and-Wolf]
Ramaswamy, R. and T. Wolf, "High-Speed Prefix-Preserving IP Address Anonymization for Passive Measurement Systems", DOI 10.1109/TNET.2006.890128, 2007, <http://www.ecs.umass.edu/ece/wolf/pubs/ton2007.pdf>.

[RFC4034] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "Resource Records for the DNS Security Extensions", RFC 4034, DOI 10.17487/RFC4034, March 2005, <https://www.rfc-editor.org/info/rfc4034>.

[RFC4035] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "Protocol Modifications for the DNS Security Extensions", RFC 4035, DOI 10.17487/RFC4035, March 2005, <https://www.rfc-editor.org/info/rfc4035>.

[RFC5077] Salowey, J., Zhou, H., Eronen, P., and H. Tschofenig, "Transport Layer Security (TLS) Session Resumption without Server-Side State", RFC 5077, DOI 10.17487/RFC5077, January 2008, <https://www.rfc-editor.org/info/rfc5077>.

[RFC6147] Bagnulo, M., Sullivan, A., Matthews, P., and I. van Beijnum, "DNS64: DNS Extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", RFC 6147, DOI 10.17487/RFC6147, April 2011, <https://www.rfc-editor.org/info/rfc6147>.

[RFC6235] Boschi, E. and B. Trammell, "IP Flow Anonymization Support", RFC 6235, DOI 10.17487/RFC6235, May 2011, <https://www.rfc-editor.org/info/rfc6235>.

[RFC6265] Barth, A., "HTTP State Management Mechanism", RFC 6265, DOI 10.17487/RFC6265, April 2011, <https://www.rfc-editor.org/info/rfc6265>.

[RFC7626] Bortzmeyer, S., "DNS Privacy Considerations", RFC 7626, DOI 10.17487/RFC7626, August 2015, <https://www.rfc-editor.org/info/rfc7626>.

[RFC7873] Eastlake 3rd, D. and M. Andrews, "Domain Name System (DNS) Cookies", RFC 7873, DOI 10.17487/RFC7873, May 2016, <https://www.rfc-editor.org/info/rfc7873>.

[RFC8027] Hardaker, W., Gudmundsson, O., and S. Krishnaswamy, "DNSSEC Roadblock Avoidance", BCP 207, RFC 8027, DOI 10.17487/RFC8027, November 2016, <https://www.rfc-editor.org/info/rfc8027>.

[RFC8094] Reddy, T., Wing, D., and P. Patil, "DNS over Datagram Transport Layer Security (DTLS)", RFC 8094, DOI 10.17487/RFC8094, February 2017, <https://www.rfc-editor.org/info/rfc8094>.

[RFC8404] Moriarty, K., Ed. and A. Morton, Ed., "Effects of Pervasive Encryption on Operators", RFC 8404, DOI 10.17487/RFC8404, July 2018, <https://www.rfc-editor.org/info/rfc8404>.

[RFC8446] Rescorla, E., "The Transport Layer Security (TLS) Protocol Version 1.3", RFC 8446, DOI 10.17487/RFC8446, August 2018, <https://www.rfc-editor.org/info/rfc8446>.

[RFC8555] Barnes, R., Hoffman-Andrews, J., McCarney, D., and J. Kasten, "Automatic Certificate Management Environment (ACME)", RFC 8555, DOI 10.17487/RFC8555, March 2019,

<https://www.rfc-editor.org/info/rfc8555>.

[RFC8618] Dickinson, J., Hague, J., Dickinson, S., Manderson, T., and J. Bond, "Compacted-DNS (C-DNS): A Format for DNS Packet Capture", RFC 8618, DOI 10.17487/RFC8618, September 2019, <https://www.rfc-editor.org/info/rfc8618>.

[SSL-Labs] SSL Labs, "SSL Server Test", 2019, <https://www.ssllabs.com/ssltest/>.

[stunnel] Goldlust, S., Almond, C., and F. Dupont, "DNS over TLS", ISC Knowledge Database", 1 November 2018, <https://kb.isc.org/article/AA-01386/0/DNS-over-TLS.html>.

[SURFnet-policy] Baartmans, C., van Wynsberghe, A., van Rijswijk-Deij, R., and F. Jorna, "SURFnet Data Sharing Policy", June 2016, <https://surf.nl/datasharing>.

[tcpdpriv] Ipsilon Networks, Inc., "TCPDRIV - Program for Eliminating Confidential Information from Traces", 2004, <http://fly.isti.cnr.it/software/tcpdpriv/>.

[van-Dijkhuizen-et-al] Van Dijkhuizen, N. and J. Van Der Ham, "A Survey of Network Traffic Anonymisation Techniques and Implementations", ACM Computing Surveys, DOI 10.1145/3182660, May 2018, <https://doi.org/10.1145/3182660>.

[Xu-et-al] Fan, J., Xu, J., Ammar, M.H., and S.B. Moon, "Prefix-preserving IP address anonymization: measurement-based security evaluation and a new cryptography-based scheme", DOI 10.1016/j.comnet.2004.03.033, 2004, <http://an.kaist.ac.kr/~sbmoon/paper/intl-journal/2004-cn-anon.pdf>.

Appendix A. Documents

This section provides an overview of some DNS privacy-related documents. However, this is neither an exhaustive list nor a definitive statement on the characteristics of any document with regard to potential increases or decreases in DNS privacy.

A.1. Potential Increases in DNS Privacy

These documents are limited in scope to communications between stub clients and recursive resolvers:

* "Specification for DNS over Transport Layer Security (TLS)" [RFC7858].

* "DNS over Datagram Transport Layer Security (DTLS)" [RFC8094]. Note that this document has the category of Experimental.

* "DNS Queries over HTTPS (DoH)" [RFC8484].

* "Usage Profiles for DNS over TLS and DNS over DTLS" [RFC8310].

* "The EDNS(0) Padding Option" [RFC7830] and "Padding Policies for Extension Mechanisms for DNS (EDNS(0))" [RFC8467].

These documents apply to recursive and authoritative DNS but are relevant when considering the operation of a recursive server:

* "DNS Query Name Minimisation to Improve Privacy" [RFC7816].

A.2. Potential Decreases in DNS Privacy

These documents relate to functionality that could provide increased tracking of user activity as a side effect:

* "Client Subnet in DNS Queries" [RFC7871].

* "Domain Name System (DNS) Cookies" [RFC7873]).

* "Transport Layer Security (TLS) Session Resumption without Server-Side State" [RFC5077], referred to here as simply TLS session resumption.

* [RFC8446], Appendix C.4 describes client tracking prevention in TLS 1.3

* "Compacted-DNS (C-DNS): A Format for DNS Packet Capture" [RFC8618].

* Passive DNS [RFC8499].

* Section 8 of [RFC8484] outlines the privacy considerations of DoH. Note that (while that document advises exposing the minimal set of data needed to achieve the desired feature set), depending on the specifics of a DoH implementation, there may be increased identification and tracking compared to other DNS transports.

A.3. Related Operational Documents

* "DNS Transport over TCP - Implementation Requirements" [RFC7766].

* "DNS Transport over TCP - Operational Requirements" [DNS-OVER-TCP].

* "The edns-tcp-keepalive EDNS0 Option" [RFC7828].

* "DNS Stateful Operations" [RFC8490].

Appendix B. IP Address Techniques

The following table presents a high-level comparison of various techniques employed or under development in 2019 and classifies them according to categorization of technique and other properties. Both the specific techniques and the categorizations are described in more detail in the following sections. The list of techniques includes the main techniques in current use but does not claim to be comprehensive.

Categorization/Property	GA	d	TC	C	TS	i	B
Anonymization	X	X	X				X
Pseudonymization				X	X	X	
Format preserving	X	X	X	X	X	X	
Prefix preserving			X	X	X		
Replacement			X				
Filtering	X						
Generalization							X
Enumeration		X					
Reordering/Shuffling			X				
Random substitution			X				
Cryptographic permutation				X	X	X	
IPv6 issues					X		

CPU intensive				X			
Memory intensive			X				
Security concerns						X	

Table 1: Classification of Techniques

Legend of techniques:

GA = Google Analytics
d = dnswasher
TC = TCPdpriv
C = CryptoPAn
TS = TSA
i = ipcipher
B = Bloom filter

The choice of which method to use for a particular application will depend on the requirements of that application and consideration of the threat analysis of the particular situation.

For example, a common goal is that distributed packet captures must be in an existing data format, such as PCAP [pcap] or Compacted-DNS (C-DNS) [RFC8618], that can be used as input to existing analysis tools. In that case, use of a format-preserving technique is essential. This, though, is not cost free; several authors (e.g., [Brekne-and-Arnes]) have observed that, as the entropy in an IPv4 address is limited, if an attacker can

* ensure packets are captured by the target and

* send forged traffic with arbitrary source and destination addresses to that target and

* obtain a de-identified log of said traffic from that target,

any format-preserving pseudonymization is vulnerable to an attack along the lines of a cryptographic chosen-plaintext attack.

B.1. Categorization of Techniques

Data minimization methods may be categorized by the processing used and the properties of their outputs. The following builds on the categorization employed in [RFC6235]:

Format-preserving. Normally, when encrypting, the original data length and patterns in the data should be hidden from an attacker. Some applications of de-identification, such as network capture de-identification, require that the de-identified data is of the same form as the original data, to allow the data to be parsed in the same way as the original.

Prefix preservation. Values such as IP addresses and MAC addresses contain prefix information that can be valuable in analysis -- e.g., manufacturer ID in MAC addresses, or subnet in IP addresses. Prefix preservation ensures that prefixes are de-identified consistently; for example, if two IP addresses are from the same subnet, a prefix preserving de-identification will ensure that their de-identified counterparts will also share a subnet. Prefix preservation may be fixed (i.e., based on a user-selected prefix length identified in advance to be preserved) or general.

Replacement. A one-to-one replacement of a field to a new value of the same type -- for example, using a regular expression.

Filtering. Removing or replacing data in a field. Field data can be overwritten, often with zeros, either partially (truncation or reverse truncation) or completely (black-marker anonymization).

Generalization. Data is replaced by more general data with reduced specificity. One example would be to replace all TCP/UDP port numbers with one of two fixed values indicating whether the original port was ephemeral (>=1024) or nonephemeral (>1024). Another example, precision degradation, reduces the accuracy of, for example, a numeric value or a timestamp.

Enumeration. With data from a well-ordered set, replace the first data item's data using a random initial value and then allocate ordered values for subsequent data items. When used with timestamp data, this preserves ordering but loses precision and distance.

Reordering/shuffling. Preserving the original data, but rearranging its order, often in a random manner.

Random substitution. As replacement, but using randomly generated replacement values.

Cryptographic permutation. Using a permutation function, such as a hash function or cryptographic block cipher, to generate a replacement de-identified value.

B.2. Specific Techniques

B.2.1. Google Analytics Non-Prefix Filtering

Since May 2010, Google Analytics has provided a facility [IP-Anonymization-in-Analytics] that allows website owners to request that all their users' IP addresses are anonymized within Google Analytics processing. This very basic anonymization simply sets to zero the least significant 8 bits of IPv4 addresses, and the least significant 80 bits of IPv6 addresses. The level of anonymization this produces is perhaps questionable. There are some analysis results [Geolocation-Impact-Assessment] that suggest that the impact of this on reducing the accuracy of determining the user's location from their IP address is less than might be hoped; the average discrepancy in identification of the user city for UK users is no more than 17%.

Anonymization: Format-preserving, Filtering (truncation).

B.2.2. dnswasher

Since 2006, PowerDNS has included a de-identification tool, dnswasher [PowerDNS-dnswasher], with their PowerDNS product. This is a PCAP filter that performs a one-to-one mapping of end-user IP addresses with an anonymized address. A table of user IP addresses and their de-identified counterparts is kept; the first IPv4 user addresses is translated to 0.0.0.1, the second to 0.0.0.2, and so on. The de-identified address therefore depends on the order that addresses arrive in the input, and when running over a large amount of data, the address translation tables can grow to a significant size.

Anonymization: Format-preserving, Enumeration.

B.2.3. Prefix-Preserving Map

Used in [tcpdpriv], this algorithm stores a set of original and anonymized IP address pairs. When a new IP address arrives, it is compared with previous addresses to determine the longest prefix match. The new address is anonymized by using the same prefix, with the remainder of the address anonymized with a random value. The use of a random value means that TCPdpriv is not deterministic; different anonymized values will be generated on each run. The need to store previous addresses means that TCPdpriv has significant and unbounded memory requirements. The need to allocate anonymized addresses sequentially means that TCPdpriv cannot be used in parallel processing.

Anonymization: Format-preserving, prefix preservation (general).

B.2.4. Cryptographic Prefix-Preserving Pseudonymization

Cryptographic prefix-preserving pseudonymization was originally proposed as an improvement to the prefix-preserving map implemented in TCPdpriv, described in [Xu-et-al] and implemented in the [Crypto-PAn] tool. Crypto-PAn is now frequently used as an acronym for the algorithm. Initially, it was described for IPv4 addresses only; extension for IPv6 addresses was proposed in [Harvan]. This uses a cryptographic algorithm rather than a random value, and thus pseudonymity is determined uniquely by the encryption key, and is deterministic. It requires a separate AES encryption for each output bit and so has a nontrivial calculation overhead. This can be mitigated to some extent (for IPv4, at least) by precalculating results for some number of prefix bits.

Pseudonymization: Format-preserving, prefix preservation (general).

B.2.5. Top-Hash Subtree-Replicated Anonymization

Proposed in [Ramaswamy-and-Wolf], Top-hash Subtree-replicated Anonymization (TSA) originated in response to the requirement for faster processing than Crypto-PAn. It used hashing for the most significant byte of an IPv4 address and a precalculated binary-tree structure for the remainder of the address. To save memory space, replication is used within the tree structure, reducing the size of the precalculated structures to a few megabytes for IPv4 addresses. Address pseudonymization is done via hash and table lookup and so requires minimal computation. However, due to the much-increased address space for IPv6, TSA is not memory efficient for IPv6.

Pseudonymization: Format-preserving, prefix preservation (general).

B.2.6. ipcipher

A recently released proposal from PowerDNS, ipcipher [ipcipher1] [ipcipher2], is a simple pseudonymization technique for IPv4 and IPv6 addresses. IPv6 addresses are encrypted directly with AES-128 using a key (which may be derived from a passphrase). IPv4 addresses are similarly encrypted, but using a recently proposed encryption [ipcrypt] suitable for 32-bit block lengths. However, the author of ipcrypt has since indicated [ipcrypt-analysis] that it has low security, and further analysis has revealed it is vulnerable to attack.

Pseudonymization: Format-preserving, cryptographic permutation.

B.2.7. Bloom Filters

van Rijswijk-Deij et al. have recently described work using Bloom Filters [Bloom-filter] to categorize query traffic and record the traffic as the state of multiple filters. The goal of this work is to allow operators to identify so-called Indicators of Compromise (IOCs) originating from specific subnets without storing information about, or being able to monitor, the DNS queries of an individual user. By using a Bloom Filter, it is possible to determine with a high probability if, for example, a particular query was made, but the set of queries made cannot be recovered from the filter. Similarly, by mixing queries from a sufficient number of users in a single filter, it becomes practically impossible to determine if a particular user performed a particular query. Large numbers of queries can be tracked in a memory-efficient way. As filter status is stored, this approach cannot be used to regenerate traffic and so cannot be used with tools used to process live traffic.

Anonymized: Generalization.

Appendix C. Current Policy and Privacy Statements

A tabular comparison of policy and privacy statements from various DNS privacy service operators based loosely on the proposed RPS structure can be found at [policy-comparison]. The analysis is based on the data available in December 2019.

We note that the existing policies vary widely in style, content, and detail, and it is not uncommon for the full text for a given operator to equate to more than 10 pages (A4 size) of text in a moderate-sized font. It is a nontrivial task today for a user to extract a meaningful overview of the different services on offer.

It is also noted that Mozilla has published a DoH resolver policy [DoH-resolver-policy] that describes the minimum set of policy requirements that a party must satisfy to be considered as a potential partner for Mozilla's Trusted Recursive Resolver (TRR) program.

Appendix D. Example RPS

The following example RPS is very loosely based on some elements of published privacy statements for some public resolvers, with additional fields populated to illustrate what the full contents of an RPS might look like. This should not be interpreted as

* having been reviewed or approved by any operator in any way

* having any legal standing or validity at all

* being complete or exhaustive

This is a purely hypothetical example of an RPS to outline example contents -- in this case, for a public resolver operator providing a basic DNS Privacy service via one IP address and one DoH URI with security-based filtering. It does aim to meet minimal compliance as specified in Section 5.

D.1. Policy

1. Treatment of IP addresses. Many nations classify IP addresses as personal data, and we take a conservative approach in treating IP addresses as personal data in all jurisdictions in which our systems reside.

2. Data collection and sharing.

    a. IP addresses. Our normal course of data management does not have any IP address information or other personal data logged to disk or transmitted out of the location in which the query was received. We may aggregate certain counters to larger network block levels for statistical collection purposes, but those counters do not maintain specific IP address data, nor is the format or model of data stored capable of being reverse-engineered to ascertain what specific IP addresses made what queries.

    b. Data collected in logs. We do keep some generalized location information (at the city / metropolitan-area level) so that we can conduct debugging and analyze abuse phenomena. We also use the collected information for the creation and sharing of telemetry (timestamp, geolocation, number of hits, first seen, last seen) for contributors, public publishing of general statistics of system use (protections, threat types, counts, etc.). When you use our DNS services, here is the full list of items that are included in our logs:

        * Requested domain name -- e.g., example.net

        * Record type of requested domain -- e.g., A, AAAA, NS, MX, TXT, etc.

        * Transport protocol on which the request arrived -- i.e., UDP, TCP, DoT, DoH

* Origin IP general geolocation information -- i.e., geocode, region ID, city ID, and metro code

* IP protocol version -- IPv4 or IPv6

* Response code sent -- e.g., SUCCESS, SERVFAIL, NXDOMAIN, etc.

* Absolute arrival time using a precision in ms

* Name of the specific instance that processed this request

* IP address of the specific instance to which this request was addressed (no relation to the requestor's IP address)

We may keep the following data as summary information, including all the above EXCEPT for data about the DNS record requested:

* Currently advertised BGP-summarized IP prefix/netmask of apparent client origin

* Autonomous system number (BGP ASN) of apparent client origin

All the above data may be kept in full or partial form in permanent archives.

c. Sharing of data. Except as described in this document, we do not intentionally share, sell, or rent individual personal information associated with the requestor (i.e., source IP address or any other information that can positively identify the client using our infrastructure) with anyone without your consent. We generate and share high-level anonymized aggregate statistics, including threat metrics on threat type, geolocation, and if available, sector, as well as other vertical metrics, including performance metrics on our DNS Services (i.e., number of threats blocked, infrastructure uptime) when available with our Threat Intelligence (TI) partners, academic researchers, or the public. Our DNS services share anonymized data on specific domains queried (records such as domain, timestamp, geolocation, number of hits, first seen, last seen) with our Threat Intelligence partners. Our DNS service also builds, stores, and may share certain DNS data streams which store high level information about domain resolved, query types, result codes, and timestamp. These streams do not contain the IP address information of the requestor and cannot be correlated to IP address or other personal data. We do not and never will share any of the requestor's data with marketers, nor will we use this data for demographic analysis.

3. Exceptions. There are exceptions to this storage model: In the event of actions or observed behaviors that we deem malicious or anomalous, we may utilize more detailed logging to collect more specific IP address data in the process of normal network defense and mitigation. This collection and transmission off-site will be limited to IP addresses that we determine are involved in the event.

4. Associated entities. Details of our Threat Intelligence partners can be found at our website page (insert link).

5. Correlation of Data. We do not correlate or combine information from our logs with any personal information that you have provided us for other services, or with your specific IP address.

6. Result filtering.

   a. Filtering. We utilize cyber-threat intelligence about malicious domains from a variety of public and private sources and block access to those malicious domains when your system attempts to contact them. An NXDOMAIN is returned for blocked sites.

i. Censorship. We will not provide a censoring component and will limit our actions solely to the blocking of malicious domains around phishing, malware, and exploit-kit domains.

ii. Accidental blocking. We implement allowlisting algorithms to make sure legitimate domains are not blocked by accident. However, in the rare case of blocking a legitimate domain, we work with the users to quickly allowlist that domain. Please use our support form (insert link) if you believe we are blocking a domain in error.

D.2. Practice

1. Deviations from Policy. None in place since (insert date).

2. Client-facing capabilities.

   a. We offer UDP and TCP DNS on port 53 on (insert IP address)

   b. We offer DNS over TLS as specified in RFC 7858 on (insert IP address). It is available on port 853 and port 443. We also implement RFC 7766.

      i. The DoT authentication domain name used is (insert domain name).

      ii. We do not publish SPKI pin sets.

   c. We offer DNS over HTTPS as specified in RFC 8484 on (insert URI template).

   d. Both services offer TLS 1.2 and TLS 1.3.

   e. Both services pad DNS responses according to RFC 8467.

   f. Both services provide DNSSEC validation.

3. Upstream capabilities.

   a. Our servers implement QNAME minimization.

   b. Our servers do not send ECS upstream.

4. Support. Support information for this service is available at (insert link).

5. Data Processing. We operate as the legal entity (insert entity) registered in (insert country); as such, we operate under (insert country/region) law. Our separate statement regarding the specifics of our data processing policy, practice, and agreements can be found here (insert link).

Acknowledgements

Many thanks to Amelia Andersdotter for a very thorough review of the first draft of this document and Stephen Farrell for a thorough review at Working Group Last Call and for suggesting the inclusion of an example RPS. Thanks to John Todd for discussions on this topic, and to StÃ©phane Bortzmeyer, Puneet Sood, and Vittorio Bertola for review. Thanks to Daniel Kahn Gillmor, Barry Green, Paul Hoffman, Dan York, Jon Reed, and Lorenzo Colitti for comments at the mic. Thanks to Loganaden Velvindron for useful updates to the text.

Sara Dickinson thanks the Open Technology Fund for a grant to support the work on this document.

Contributors

The below individuals contributed significantly to the document:

John Dickinson
Sinodun IT
Magdalen Centre
Oxford Science Park
Oxford
OX4 4GA
United Kingdom

Jim Hague
Sinodun IT
Magdalen Centre
Oxford Science Park
Oxford
OX4 4GA
United Kingdom

Authors' Addresses

Sara Dickinson
Sinodun IT
Magdalen Centre
Oxford Science Park
Oxford
OX4 4GA
United Kingdom

Email: sara@sinodun.com

Benno J. Overeinder
NLnet Labs
Science Park 400
1098 XH Amsterdam
Netherlands

Email: benno@nlnetLabs.nl

Roland M. van Rijswijk-Deij
NLnet Labs
Science Park 400
1098 XH Amsterdam
Netherlands

Email: roland@nlnetLabs.nl

Allison Mankin
Salesforce.com, Inc.
Salesforce Tower
415 Mission Street, 3rd Floor
San Francisco, CA 94105
United States of America

Email: allison.mankin@gmail.com