dprive
Internet-Draft
Intended status: Best Current Practice
Expires: May 21, 2020

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

November 18, 2019

Recommendations for DNS Privacy Service Operators draft-ietf-dprive-bcp-op-06

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, and how the decisions and alternatives impact the privacy of users.

This document also presents a framework to assist writers of a DNS Recursive Operator Privacy Statement (analogous to DNS Security Extensions (DNSSEC) Policies and DNSSEC Practice Statements described in RFC6841).

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on May 21, 2020.

Copyright Notice

Copyright (c) 2019 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction 3
2. Scope 5
3. Privacy related documents 5
4. Terminology 6
5. Recommendations for DNS privacy services 6
5.1. On the wire between client and server 7
5.1.1. Transport recommendations 7
5.1.2. Authentication of DNS privacy services 8
5.1.3. Protocol recommendations 9
5.1.4. DNSSEC 11
5.1.5. Availability 11
5.1.6. Service options 12
5.1.7. Impact on DNS Privacy Service Operators 12
5.1.8. Limitations of using a pure TLS proxy 13
5.2. Data at rest on the server 13
5.2.1. Data handling 13
5.2.2. Data minimization of network traffic 14
5.2.3. IP address pseudonymization and anonymization methods 15
5.2.4. Pseudonymization, anonymization or discarding of other correlation data 17
5.2.5. Cache snooping 17
5.3. Data sent onwards from the server 18
5.3.1. Protocol recommendations 18
5.3.2. Client query obfuscation 19
5.3.3. Data sharing 19
6. DNS Recursive Operator Privacy (DROP) statement 20
6.1. Recommended contents of a DROP statement 20
6.1.1. Policy 20
6.1.2. Practice 21
6.2. Current policy and privacy statements 22
6.3. Enforcement/accountability 23
7. IANA considerations 23
8. Security considerations 23
9. Acknowledgements 23
10. Contributors 24
11. Changelog 24

12. References 26
12.1. Normative References 26
12.2. Informative References 28
12.3. URIs 30
Appendix A. Documents 31
A.1. Potential increases in DNS privacy 31
A.2. Potential decreases in DNS privacy 32
A.3. Related operational documents 32
Appendix B. IP address techniques 32
B.1. Google Analytics non-prefix filtering 33
B.2. dnswasher 34
B.3. Prefix-preserving map 34
B.4. Cryptographic Prefix-Preserving Pseudonymisation 34
B.5. Top-hash Subtree-replicated Anonymisation 35
B.6. ipcipher 35
B.7. Bloom filters 35
Appendix C. Example DROP statement 36
C.1. Policy 36
C.2. Practice 39
Authors' Addresses 40

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 which aim to increase the privacy of the DNS system 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 because they offer a specific feature (e.g. good reachability, encrypted transport, strong privacy policy, filtering (or lack of), etc.). These open 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 exists. 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 the choice of a user to configure a single resolver (or a fixed set of resolvers) and an encrypted transport to use in all network environments has both advantages and disadvantages. 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 to track them as they move across network environments. Commitments from operators to minimize such tracking 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. It is an untested area that simply using a DNS resolution service constitutes consent from the user for the operator to process their query data. The impact of recent legislative changes on data pertaining to the users of both Internet Service Providers and public DNS resolvers is not fully understood at the time of writing.

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 DNS Recursive Operator Privacy (DROP) statement and present a framework to assist writers of this document. A DROP statement is a document that an operator can publish outlining their operational practices and commitments with regard to privacy thereby providing a means for clients to evaluate the privacy properties of a given DNS privacy service. In particular, the framework identifies the elements that should be considered in formulating a DROP statement. This document does not, however, define a particular Privacy statement, nor does it seek to provide legal advice or recommendations as to the contents.

A desired operational impact is that all operators (both those providing resolvers within networks and those operating large anycast 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 difference 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 errata or updates that apply to this document.

2. Scope

"DNS Privacy Considerations" [I-D.ietf-dprive-rfc7626-bis] describes the general privacy issues and threats associated with the use of the DNS by Internet users and much of the threat analysis here is lifted from that document and from [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). Privacy considerations specifically from the perspective of an 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 (taken from [I-D.ietf-dprive-rfc7626-bis]):

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 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 of the DNS. Note 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 are 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] with one modification: we restate the clause in the original definition of Privacy-enabling DNS server in [RFC8310] to include the requirement that a DNS over (D)TLS server should also offer at least one of the credentials described in Section 8 and implement the (D)TLS profile described in Section 9 of [RFC8310].

Other Terms:

- DROP: DNS 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 DROP statement.

5. Recommendations for DNS privacy services

We describe two classes of threats:

- 'Privacy Considerations for Internet Protocols' [RFC6973] Threats
  - 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 to 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

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

[RFC6973] Threats:

- Surveillance:
  - Passive surveillance of traffic on the wire [I-D.ietf-dprive-rfc7626-bis] Section 2.4.2.

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 [RFC7858] and [RFC8310]
- 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 IPSec, DNSCrypt or VPNs. However, use of these transports for DNS are not standardized 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 this does not diminish the need for DNSSEC, see Section 5.1.4.

5.1.2. Authentication of DNS privacy services

[RFC6973] Threats:

- Surveillance:
  - Active attacks that can redirect traffic to rogue servers [I-D.ietf-dprive-rfc7626-bis] Section 2.5.3.

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 DNS-over-TLS 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 DNS-over-TLS should provide credentials in the form of either X.509 certificates [RFC5280] or 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.
- Mis-identification of a server by a client e.g. typos in URLs or authentication domain names

Mitigations:

It is recommended that operators:

- Follow the guidance in Section 6.5 of [RFC7525] with regards to certificate revocation
- Choose a short, memorable authentication name for the service
- Automate the generation, publication and renewal of certificates. For example, 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

5.1.3. Protocol recommendations

5.1.3.1. DNS-over-TLS

DNS Privacy Threats:

- Known attacks on TLS such as those described in [RFC7457]
- Traffic analysis, for example: [Pitfalls-of-DNS-Encryption]
- Potential for client tracking via transport identifiers
- Blocking of well known ports (e.g. 853 for DNS-over-TLS)

Mitigations:

In the case of DNS-over-TLS, TLS profiles from Section 9 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 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 DNS-over-TLS privacy service on both port 853 and 443. This practice may not be possible if e.g. the operator deploys DoH on the same IP address.

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 either

  * [RFC7766] and EDNS(0) Keepalive [RFC7828] and/or

  * 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: the traffic analysis perspective [1]

- Potential for client tracking via transport identifiers

Mitigations:

- Clients must be able to forego the use of HTTP Cookies [RFC6265] and still use the service

- 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 [I-D.ietf-httpbis-bcp56bis].)

5.1.4. DNSSEC

DNS Privacy Threats:

- Users may be directed to bogus IP addresses for e.g. websites where they might reveal personal information to attackers.

Mitigations:

- All DNS privacy services must offer a DNS privacy service that performs DNSSEC validation. In addition they must be able to provide the DNSSEC 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 non-validating 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 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 which supports DNSSEC has a far simpler task in terms of DNS Roadblock avoidance.

5.1.5. Availability

DNS Privacy Threats:

- A failed DNS privacy service could force the user to switch providers, fallback to cleartext or accept no DNS service for 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 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 Service: 2.5 Years in the Life of Google [2].

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, fallback to cleartext or accept no DNS service for the outage.

Mitigations:

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

5.1.7. Impact on DNS Privacy Service Operators

DNS Privacy Threats:

- Increased use of encryption impacts operator ability to manage their network [RFC8404]

Many monitoring solutions for DNS traffic rely on the plain text 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. There are, however, legitimate reasons for operators to inspect DNS traffic, e.g. to monitor for network security threats. Operators may therefore need to invest in alternative means of monitoring that relies on either the resolver software directly, or exporting DNS traffic from the resolver using e.g. dnstap [3].

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 Section 5.2 for more details on data handling and also the discussion on the use of Bloom Filters in Appendix A.

5.1.8. Limitations of using a pure TLS proxy

DNS Privacy Threats:

- Limited ability to manage or monitor incoming connections using DNS specific techniques
- Misconfiguration of the target server could lead to data leakage if the proxy to target server path is not encrypted.

Optimization:

Some operators may choose to implement DNS-over-TLS using a TLS proxy (e.g. nginx [4], haproxy [5] or stunnel [6]) 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 at 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 ACLs, RRL or DNS64 will be hard or impossible to implement in the nameserver.

Operators may choose to use a DNS aware proxy such as dnsdist [7] which offer custom options (similar to that proposed in [I-D.bellis-dnsop-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

[RFC6973] Threats:

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

Other Threats

- Contravention of legal requirements not to process user data?

Mitigations:

The following are common activities for DNS service operators and in all cases 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. that is 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 those required to sustain operation of the service and, to the extent that such exists, meet regulatory requirements.

- 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 were 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 may also include revealing 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 and 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 tradeoffs 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. [8]

- 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, how 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

As [I-D.ietf-dprive-rfc7626-bis] makes clear, the big 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.

The following table presents a high level comparison of various techniques employed or under development today and classifies them according to categorization of technique and other properties. Appendix B provides a more detailed survey of these techniques and definitions for the categories and properties listed below. The list of techniques includes the main techniques in current use, but does not claim to be comprehensive.

Categorisation/Property	GA	d	TC	C	TS	i	B
Anonymisation	X	X	X				X
Pseudoanonymisation				X	X	X	
Format preserving	X	X	X	X	X	X	
Prefix preserving			X	X	X		
Replacement			X				
Filtering	X						
Generalisation							X
Enumeration		X					
Reordering/Shuffling			X				
Random substitution			X				
Crytpographic permutation				X	X	X	
IPv6 issues					X		
CPU intensive				X			
Memory intensive			X				
Security concerns						X	

Table 1: Classification 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 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. Brenker & Arnes [9]) have observed that, as the entropy in an IPv4 address is limited, given a de-identified log from a target, if an attacker is capable of ensuring packets are captured by the target and the attacker can send

forged traffic with arbitrary source and destination addresses to that target, any format-preserving pseudonymization is vulnerable to an attack along the lines of a cryptographic chosen plaintext attack.

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, ECN support, 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 e.g. TLS version/Cipher suite combinations or other connection parameters.
- Correlation of queries on multiple TCP session 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. MAC addresses in EDNS(0) options - some CPE devices are known to add them.
- HTTP headers (e.g., User-Agent, Accept, Accept-Encoding)

Mitigations:

- Data minimization or discarding of such correlation data.

5.2.5. Cache snooping

[RFC6973] Threats:

- Surveillance:
  - Profiling of client queries by malicious third parties

Mitigations:

- See ISC Knowledge database on cache snooping [10] for an example discussion on defending against cache snooping

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

[RFC6973] Threats:

- Surveillance:
  - Transmission of identifying data upstream.

Mitigations:

As specified in [RFC8310] for DNS-over-TLS but applicable to any DNS Privacy services 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 and not send an ECS option in upstream queries.

Optimizations:

- 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 whitelisting upstream servers to send ECS to in order to minimize data leakage. Operators should make clear in any policy statement what prefix length they actually send and the specific policy used.

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

Additional options:

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

- Run a copy of the root zone on loopback [RFC7706] 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 obfuscate this traffic by mixing it with 'generated' traffic to make client characterization harder. The resolver could also employ aggressive pre-fetch 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 pre-fetches.

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

[RFC6973] Threats:

- Surveillance

- Stored data compromise

- Correlation

- Identification

- Secondary use

- Disclosure

DNS Privacy Threats:

o Contravention of legal requirements not to process user data

Mitigations:

Operators should not provide identifiable data to third-parties without explicit consent from clients (we take the stance here that simply using the resolution service itself does not constitute consent).

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 SURFnet's policy [11] on data sharing for research as an example.

6. DNS Recursive Operator Privacy (DROP) statement

The following section outlines the recommended contents of a DROP statement an operator might choose to publish. An example statement for a specific scenario is provided for guidance only in Appendix C.

6.1. Recommended contents of a DROP statement

6.1.1. Policy

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

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

   and in each case whether it is aggregated, pseudonymized or anonymized and the conditions of data transfer.

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

4. Associated entities. Declare 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

This section should explain 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 section Section 5 provide specific details of which capabilities are provided on which client facing addresses and ports:

   1. For DoT, specify the authentication name to be used (if any)

   2. For DoT, specify the SPKI pin sets to be used (if any) and policy for rolling keys

3. Upstream capabilities. With reference to section 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. Jurisdiction. This section should communicate the applicable jurisdictions and law enforcement regimes under which the service is being provided.

    1. Specify the operator entity or entities that will control the data and be responsible for their treatment, and their legal place of business.

    2. Specify, either directly or by pointing to the applicable privacy policy, the relevant privacy laws that apply to the treatment of the data, the rights that users enjoy in regard to their own personal information that is treated by the service, and how they can contact the operator to enforce them.

    3. Additionally specify the countries in which the servers handling the DNS requests and the data are located (if the operator applies a geolocation policy so that requests from certain countries are only served by certain servers, this should be specified as well).

    4. Specify whether the operator has any agreement in place with law enforcement agencies, or other public and private parties dealing with security and intelligence, to give them access to the servers and/or to the data.

6.2. Current policy and privacy statements

A tabular comparison of existing policy and privacy statements from various DNS Privacy service operators based loosely on the proposed DROP structure can be found on dnsprivacy.org [12].

We note that the existing set of 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 of moderate font sized A4 text. It is a non-trivial task today for a user to extract a meaningful overview of the different services on offer.

It is also noted that Mozilla have published a Security/DoH-resolver policy [13], which 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.

6.3. Enforcement/accountability

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

Independent monitoring or analysis could be performed where possible of:

- ECS, QNAME minimization, EDNS(0) padding, etc.
- Filtering
- Uptime

This is by analogy with e.g. several TLS or website analysis tools that are currently available e.g. SSL Labs [14] or Internet.nl [15].

Additionally operators could choose to engage the services of a third party auditor to verify their compliance with their published DROP statement.

7. IANA considerations

None

8. Security considerations

Security considerations for DNS-over-TCP are given in [RFC7766], many of which are generally applicable to session based DNS.

9. 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 WGLC and for suggesting the inclusion of an example DROP statement. Thanks to John Todd for discussions on this topic, and to Stephane Bortzmeyer, Puneet Sood and Vittorio Bertola for review. Thanks to Daniel Kahn Gillmor, Barry Green, Paul Hoffman, Dan York, John Reed, 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.

10. Contributors

The below individuals contributed significantly to the document:

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

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

11. Changelog

draft-ietf-dprive-bcp-op-05

- Final minor changes from second WGLC.

draft-ietf-dprive-bcp-op-05

- Remove some text on consent:
  - Paragraph 2 in section 5.3.3
  - Item 6 in the DROP Practice statement (and example)
- Remove .onion and TLSA options
- Include ACME as a reference for certificate management
- Update text on session resumption usage
- Update section 5.2.4 on client fingerprinting

draft-ietf-dprive-bcp-op-04

- Change DPPPS to DROP (DNS Recursive Operator Privacy) statement
- Update structure of DROP slightly
- Add example DROP statement

- Add text about restricting access to full logs
- Move table in section 5.2.3 from SVG to inline table
- Fix many editorial and reference nits

draft-ietf-dprive-bcp-op-03

- Add paragraph about operational impact
- Move DNSSEC requirement out of the Appendix into main text as a privacy threat that should be mitigated
- Add TLS version/Cipher suite as tracking threat
- Add reference to Mozilla TRR policy
- Remove several TODOs and QUESTIONS.

draft-ietf-dprive-bcp-op-02

- Change 'open resolver' for 'public resolver'
- Minor editorial changes
- Remove recommendation to run a separate TLS 1.3 service
- Move TLSA to purely a optimisation in Section 5.2.1
- Update reference on minimal DoH headers.
- Add reference on user switching provider after service issues in Section 5.1.4
- Add text in Section 5.1.6 on impact on operators.
- Add text on additional threat to TLS proxy use (Section 5.1.7)
- Add reference in Section 5.3.1 on example policies.

draft-ietf-dprive-bcp-op-01

- Many minor editorial fixes
- Update DoH reference to RFC8484 and add more text on DoH
- Split threat descriptions into ones directly referencing RFC6973 and other DNS Privacy threats

- Improve threat descriptions throughout
- Remove reference to the DNSSEC TLS Chain Extension draft until new version submitted.
- Clarify use of whitelisting for ECS
- Re-structure the DPPPS, add Result filtering section.
- Remove the direct inclusion of privacy policy comparison, now just reference dnsprivacy.org and an example of such work.
- Add an appendix briefly discussing DNSSEC
- Update affiliation of 1 author

draft-ietf-dprive-bcp-op-00

- Initial commit of re-named document after adoption to replace draft-dickinson-dprive-bcp-op-01

12. References

12.1. Normative References

[I-D.ietf-dprive-rfc7626-bis]
Bortzmeyer, S. and S. Dickinson, "DNS Privacy Considerations", draft-ietf-dprive-rfc7626-bis-02 (work in progress), October 2019.

[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>.

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

[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>.

[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>.

[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>.

[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>.

[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>.

[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>.

[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>.

12.2. Informative References

[I-D.bellis-dnsop-xpf]
Bellis, R., Dijk, P., and R. Gacogne, "DNS X-Proxied-For", draft-bellis-dnsop-xpf-04 (work in progress), March 2018.

[I-D.ietf-dnsop-dns-tcp-requirements]
Kristoff, J. and D. Wessels, "DNS Transport over TCP - Operational Requirements", draft-ietf-dnsop-dns-tcp-requirements-05 (work in progress), November 2019.

[I-D.ietf-httpbis-bcp56bis]
Nottingham, M., "Building Protocols with HTTP", draft-ietf-httpbis-bcp56bis-09 (work in progress), November 2019.

[pcap] tcpdump.org, "PCAP", 2016, <http://www.tcpdump.org/>.

[Pitfalls-of-DNS-Encryption]
Shulman, H., "Pretty Bad Privacy: Pitfalls of DNS Encryption", 2014, <https://dl.acm.org/citation.cfm?id=2665959>.

[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>.

[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>.

[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>.

[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>.

[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>.

[RFC7706] Kumari, W. and P. Hoffman, "Decreasing Access Time to Root Servers by Running One on Loopback", RFC 7706, DOI 10.17487/RFC7706, November 2015, <https://www.rfc-editor.org/info/rfc7706>.

[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>.

[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>.

[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>.

[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>.

[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>.

12.3. URIs

[1] https://petsymposium.org/2018/files/hotpets/4-siby.pdf

[2] http://tma.ifip.org/2018/wp-content/uploads/sites/3/2018/06/tma2018_paper30.pdf

[3] http://dnstap.info

[4] https://nginx.org/

[5] https://www.haproxy.org/

[6] https://kb.isc.org/article/AA-01386/0/DNS-over-TLS.html

[7] https://dnsdist.org

[8] https://doi.org/10.1145/3182660

[9] https://pdfs.semanticscholar.org/7b34/12c951cebe71cd2cddac5fda164fb2138a44.pdf

[10] https://kb.isc.org/docs/aa-00482

[11] https://surf.nl/datasharing

[12] https://dnsprivacy.org/wiki/display/DP/Comparison+of+policy+and+privacy+statements

[13] https://wiki.mozilla.org/Security/DOH-resolver-policy

[14] https://www.ssllabs.com/ssltest/

[15] https://internet.nl

[16] https://support.google.com/analytics/answer/2763052?hl=en

[17] https://www.conversionworks.co.uk/blog/2017/05/19/anonymize-ip-geo-impact-test/

[18] https://github.com/edmonds/pdns/blob/master/pdns/dnswasher.cc

[19] http://ita.ee.lbl.gov/html/contrib/tcpdpriv.html

[20] http://an.kaist.ac.kr/~sbmoon/paper/intl-journal/2004-cn-anon.pdf

[21] https://www.cc.gatech.edu/computing/Telecomm/projects/cryptopan/

[22] http://mharvan.net/talks/noms-ip_anon.pdf

[23] http://www.ecs.umass.edu/ece/wolf/pubs/ton2007.pdf

[24] https://medium.com/@bert.hubert/on-ip-address-encryption-security-analysis-with-respect-for-privacy-dabe1201b476

[25] https://github.com/PowerDNS/ipcipher

[26] https://github.com/veorq/ipcrypt

[27] https://www.ietf.org/mail-archive/web/cfrg/current/msg09494.html

[28] http://dl.ifip.org/db/conf/im/im2019/189282.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 characteristic of the document.

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], referred to here as 'DNS-over-TLS'.
- 'DNS over Datagram Transport Layer Security (DTLS)' [RFC8094], referred to here as 'DNS-over-DTLS'. Note that this document has the Category of Experimental.
- 'DNS Queries over HTTPS (DoH)' [RFC8484] referred to here as DoH.
- 'Usage Profiles for DNS over TLS and DNS over DTLS' [RFC8310]
- 'The EDNS(0) Padding Option' [RFC7830] and 'Padding Policy for EDNS(0)' [RFC8467]

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

- 'DNS Query Name minimization to Improve Privacy' [RFC7816] referred to here as 'QNAME minimization'

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.
- 'A DNS Packet Capture Format' [RFC8618]
- Passive DNS [RFC8499]

Note that depending on the specifics of the implementation [RFC8484] may also provide increased tracking.

A.3. Related operational documents

- 'DNS Transport over TCP - Implementation Requirements' [RFC7766]
- 'Operational requirements for DNS-over-TCP' [I-D.ietf-dnsop-dns-tcp-requirements]
- 'The edns-tcp-keepalive EDNS0 Option' [RFC7828]
- 'DNS Stateful Operations' [RFC8490]

Appendix B. IP address 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, subnet in IP addresses. Prefix preservation ensures that prefixes are de-identified consistently; e.g. 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 (and thus truncating) or replacing data in a field. Field data can be overwritten, often with zeros, either partially (grey marking) or completely (black marking).

- 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 non-ephemeral (>1024). Another example, precision degradation, reduces the accuracy of e.g. a numeric value or a timestamp.

- Enumeration. With data from a well-ordered set, replace the first data item 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.1. Google Analytics non-prefix filtering

Since May 2010, Google Analytics has provided a facility [16] 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 [17] which 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 (grey marking).

B.2. dnswasher

Since 2006, PowerDNS have included a de-identification tool dnswasher [18] 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 running over a large amount of data the address translation tables can grow to a significant size.

Anonymization: Format-preserving, Enumeration.

B.3. Prefix-preserving map

Used in TCPdpriv [19], this algorithm stores a set of original and anonymised 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 TCPdrpiv 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, and because of the need to allocated anonymized addresses sequentially cannot be used in parallel processing.

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

B.4. Cryptographic Prefix-Preserving Pseudonymisation

Cryptographic prefix-preserving pseudonymisation was originally proposed as an improvement to the prefix-preserving map implemented in TCPdpriv, described in Xu et al. [20] and implemented in the Crypto-PAn tool [21]. 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 & Schoenwaelder [22] and implemented in snmpdump. 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, so has a non-trivial calculation overhead. This can be mitigated to some extent (for IPv4, at least) by pre-calculating results for some number of prefix bits.

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

B.5. Top-hash Subtree-replicated Anonymisation

Proposed in Ramaswamy & Wolf [23], Top-hash Subtree-replicated Anonymisation (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 pre-calculated 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 pre-calculated structures to a few Mb 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.6. ipcipher

A recently-released proposal from PowerDNS [24], ipcipher [25] 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 [26] suitable for 32bit block lengths. However, the author of ipcrypt has since indicated [27] that it has low security, and further analysis has revealed it is vulnerable to attack.

Pseudonymization: Format-preserving, cryptographic permutation.

B.7. Bloom filters

van Rijswijk-Deij et al. [28] have recently described work using Bloom filters 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 be 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. Example DROP statement

The following example DROP statement is very loosely based on some elements of published privacy statements for some public resolvers, with additional fields populated to illustrate the what the full contents of a DROP statement 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 a DROP statement 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.

C.1. Policy

1. Treatment of IP addresses. Many nations classify IP addresses as Personally-Identifiable Information (PII), and we take a conservative approach in treating IP addresses as PII in all jurisdictions in which our systems reside.

2. Data collection and sharing.

    1. IP addresses. Our normal course of data management does not have any IP address information or other PII 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.

    2. 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 use of system (protections, threat types, counts, etc.) When you use our DNS Services, here is the full list of items that are
included in our logs:

+ Request 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

+ 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.

3. 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 the 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 its threat intelligence partners. Our DNS Services 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 IP address information of requestor and cannot be correlated to IP address or other PII. We do not and never will share any of its data with marketers, nor will it use this data for demographic analysis.

3. Exceptions. There are exceptions to this storage model: In the event of events or observed behaviors which we deem malicious or anomalous, we may utilize more detailed logging to collect more specific IP address data in the process of normal network defence 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.

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

        1. 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.

2. Accidental blocking. We implement whitelisting 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 whitelist that domain. Please use our support form (insert link) if you believe we are blocking a domain in error.

C.2. Practice

1. Deviations from Policy. None currently in place.

2. Client facing capabilities.

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

    2. We offer DNS-over-TLS as specified in RFC7858 on (insert IP address). It is available on port 853 and port 443. We also implement RFC7766.

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

        2. We do not publish SPKI pin sets.

    3. We offer DNS-over-HTTPS as specified in RFC8484 on (insert URI template). Both POST and GET are supported.

    4. Both services offer TLS 1.2 and TLS 1.3.

    5. Both services pad DNS responses according to RFC8467.

    6. Both services provide DNSSEC validation.

3. Upstream capabilities.

    1. Our servers implement QNAME minimisation.

    2. Our servers do not send ECS upstream.

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

5. Jurisdiction.

    1. We operate as the legal entity (insert entity) registered in (insert country) as (insert company identifier e.g Company Number). Our Headquarters are located at (insert address).

2. As such we operate under (insert country) law. For details of our company privacy policy see (insert link). For questions on this policy and enforcement contact our Data Protection Officer on (insert email address).

3. We operate servers in the following countries (insert list).

4. We have no agreements in place with law enforcement agencies to give them access to the data. Apart from as stated in the Policy section of this document with regard to cyber threat intelligence, we have no agreements in place with other public and private parties dealing with security and intelligence, to give them access to the servers and/or to the data.

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
Amsterdam 1098 XH
The Netherlands

Email: benno@nlnetLabs.nl

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

Email: roland@nlnetLabs.nl

Allison Mankin
Salesforce

Email: allison.mankin@gmail.com