

dnsop
Internet-Draft
Intended status: Informational
Expires: May 7, 2020

D. Migault
Ericsson
November 04, 2019

A privacy analysis on DoH deployment
draft-mglt-abcd-doh-privacy-analysis-00

Abstract

This document provides an analysis on DoH impact on privacy

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 <https://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 7, 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 (<https://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.

Internet-Draft A privacy analysis on DoH deployment November 2019

Table of Contents

1.	Requirements Notation	2
2.	Introduction	2
3.	DNS traffic and privacy	3
4.	Privacy impact of DoH	4
4.1.	DNS systems polices: lost of control versus independence	6
5.	Privacy impact related to the choice of the DNS resolver . .	7
6.	Privacy impact of concentration	8
6.1.	Acknowledgment	10
7.	References	10
7.1.	Normative References	10
7.2.	Informative References	11
	Author's Address	11

[1.](#) Requirements Notation

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 [BCP 14](#) [[RFC2119](#)] [[RFC8174](#)] when, and only when, they appear in all capitals, as shown here.

[2.](#) Introduction

DNS Queries over HTTPS (DoH) [[RFC8484](#)] differs from the traditional DNS [[RFC1035](#)] in that DNS exchanges between the DNS client and the resolver are now encrypted and that DNS traffic is not signaled as DNS traffic (with port 53) but instead uses (port 443).

Such approach could enhance end user's privacy by preventing any on-path party to infer any DNS related information from the observed traffic. However, such enhancement may also have counter effects such as the loose of control of the DNS traffic by the end user itself.

This draft aims at providing an analysis on the impact of the deployment of DoH on the current internet.

Section [Section 3](#) details privacy sensitive information carried by the DNS traffic and evaluate how specific this information is specific to DNS or could be inferred from other traffic such as the web traffic depending on Internet concentration.

Section [Section 4](#) exposes the privacy implication of possible usage of DoH and more precisely the ability to circumvent or enforce the end user policies.

While encrypting the DNS traffic enables the section of a DNS resolver, section [Section 5](#) exposes the privacy implications associated to the selection of a resolver and show that choosing a resolver outside the boundaries of an ISP provides in fact limited protection toward that ISP.

Finally, section [Section 6](#) shows that despite the advantages that concentration could provide by obfuscating the IP address, the overall picture of concentration shows that it represents a threat to the end user's privacy.

[3.](#) DNS traffic and privacy

DNS data are public data available to everyone. As a result, the value associated to a DNS exchange are mostly carried by the DNS request that answers to "What this specific end users is interested in?" or "Which end users contact this site?" rather the DNS information provided by the response. Such information is carried by associating the destination IP address (of the IP header) as well as the DNS query field. There are good and bad reasons for monitoring the sites as well as end user that connects to them. Typically, a network administrator may prevent the end user to connect to malicious web sites as well as monitor the sites the end user is connecting to. However, it is out of the control of any protocol to impact the usage of this information.

In most cases, the DNS exchanges are followed by a web connection. If the web session were not encrypted, observation of the web traffic would provide the same information as those carried in the DNS traffic. This information would be richer and more accurate, as web traffic really reflects the web sites the end user is accessing. However, the amount of web traffic is huge compared to the DNS traffic and the DNS traffic was clearly distinct from the HTTP traffic with different port from HTTP(S) with a distinct termination point (DNS resolver).

With an increasing number of encrypted web traffic, analysis of the HTTP traffic is not anymore possible as it is being protected by TLS. However, HTTPS traffic still reveals the destination IP address and the domain name within a TLS field designated as SNI. As mentioned earlier, analysis of the HTTPS traffic, due to the volumes invoked remains a challenge in itself. However, the encryption of the SNI as well as the fact that one IP address provided by cloud provider can be shared by multiple web sites clearly limit the meaning of the information provided by a supposed analysis of the HTTPS traffic. As a result, in addition to be more convenient, the information associated revealed by the DNS traffic may not be inferred from other traffic.

As a result, the information carried by the DNS traffic has the following characteristics:

- o The DNS traffic is a good representation of the web traffic of one end user.
- o When not carried over HTTP, the DNS traffic is by construction logically separate from the web traffic.
- o The DNS traffic is terminated in one point, while web traffic is generally terminated at multiple destinations

The privacy sensitive information carried by the DNS traffic are the IP addresses that "identify" the end user and the content of the DNS query, that reflects the activity of the end user. This information is limited to the administrative domain the DNS traffic is steered to when the DNS traffic is not encrypted. When the DNS traffic is encrypted, this information is limited to the two end points, that is the end user and the DNS resolver.

The same activity can be inferred from the encrypted web traffic unless ESNI together with a high concentration of web sites behind a limited number of IP addresses. In that sense web site concentration and ESNI adds boundaries to the information associated to the DNS traffic, which could enhance the privacy against on-path monitoring. However, concentration of the web traffic transfers the information from the internet providers to large cloud providers. Section [Section 6](#) details furthermore how concentration represents a direct threat to privacy.

As a result privacy sensitive information carried by the DNS traffic is shared between the DNS client, the DNS resolver via DNS traffic. Similar information is provided by the web traffic that is shared between the HTTP client as well as the internet service provider and major cloud providers. The balance between these two depends on the level of concentration.

4. Privacy impact of DoH

The use of DoH to perform DNS exchanges has the following impacts on the DNS traffic:

- o DNS traffic is encrypted
- o DNS traffic is no different from the encrypted web traffic

As mentioned in section [Section 3](#), since DNS traffic is encrypted, the privacy sensitive information of the DNS is exchanged between the

DNS client and the DNS resolver. As per the Internet threat model of [\[RFC3552\]](#), it is expected that "the end-systems engaging in a protocol exchange have not themselves been compromised. Protecting against an attack when one of the end-systems has been compromised is extraordinarily difficult.". The purpose of the protection is to protect against an attacker that may have a complete control of the network. With that threat model in mind, encryption protects the DNS exchanged via DNS exchange between the DNS client and the DNS resolver and as such improves the end user's privacy. In particular it protects against pervasive monitoring attacks [\[RFC7258\]](#).

However, as mentioned in [\[RFC6973\]](#) privacy analysis needs to question the assumption of [\[RFC3552\]](#) on end-systems "since systems are often compromised for the purpose of obtaining personal data". In addition, privacy also includes the ability of the end user to control and protect its information.

The ability to enforce policies for the DNS traffic has been performed until today by having the DNS client centralized in the system of the end user. The configuration at the operating system level ensures that all applications were aligned with the end user policy. A typical policy typically includes the domains that needs

to be resolved, the interface to be used, the DNS resolver to contact...

DoH changes this paradigm in the way that an application can circumvent the policy set by the end user, without the end user being aware of it. Firstly, the encryption is performed by the application and as such does not provide any visibility to the operating system. Second, the use of HTTPS makes DNS traffic indistinguishable from the web traffic. To that extend, DoT would signal the system that some encrypted DNS traffic is being handled by the application. The end user may accept or refuse such traffic depending on its policy. DoH does not provides such capabilities.

Another way to see this issue is to consider that the communication between the DNS client and the DNS resolver is a communication that is secured between the two application end-point. The end resolver policy enforcement is performed on-path inside the end user system, but encryption prevents it to be enforced.

In a nutshell, DoH encrypts and makes DNS traffic undetectable. This provides the ability for an application to circumvent the policies defined by the system and can be seen as a loose of control. The alignment with the policies of the system is enforced by explicit policies from the application and trust the application enforces the claimed policies.

The impact on privacy needs to balance the DNS policies provided by the system versus those provided by the application and more explicitly which of these policies better protects the end user.

[4.1.](#) DNS systems polices: lost of control versus independence

The DNS system policies may or may not reflects the end user's preferences, however, these are part of the configuration parameters of the system and the end user can at least be aware of the policies of its system.

There are cases were the DNS policies in the system expresses the end user's policies. This includes typically the choice of a specific DNS resolver, the subscription to parental control. For such end user, the ability that an application circumvents the policies of the

system represents a threat to their ability to control their DNS traffic.

Similarly, there are cases where the DNS policies are not explicitly specified by the end user, but there is an agreement of the end user to have these policies. This typically includes corporate users that have agreed to comply with the corporate policies with potentially some web sites cannot be accessed. For these end users, the ability that an application can circumvent the DNS policies of the company exposes the end user to risks he may not want to take.

For the two latest category of users the ability for each application to have specific DNS policies present the following drawbacks:

- 1) A per-application control results in defining at multiple places the DNS policies. This at least can create some confusions to the end user, makes configuration prone to errors and eventually debugging harder.
- 2) While some applications may have clear and explicit DNS policies, that the end user could in principle check or configure against the policies he is enforcing, these policies are subject to change over time and without notice, typically during updates. While constantly checking the policies is not something we can rely on, the end user or company may delay the applications to be updated which adds an additional risk to the end user privacy.

There are cases where the DNS policies are imposed to the end user against its will and without agreement from his side. Motivations for such policies could be to enforce surveillance of the end user. In such situation the ability to circumvent the DNS policies by an application improves the end user's privacy. It is also safer that

DNS policies are enforced by the application as the application will be in these situation the trusted system of the end user.

As a result, that an application can enforce there own policies improves or reduce the control of the DNS traffic of the end user depends on what the trust system of the end user is. If the trust system of the end user is the application, this ability clearly improves, otherwise, this may represent a threat. In the later case,

applications should follow the configuration of the system.

5. Privacy impact related to the choice of the DNS resolver

As mentioned in section [Section 3](#) DoH provides end-to-end encryption and as such provides the ability for the end user to chose a specific DNS resolver and share the DNS data only with that resolver. One motivation to chose a specific DNS resolver is to move to a DNS resolver that considers the end user's privacy with more attention. This includes, among other things, not profiling the end users, not selling user's information, or in some places not tracking specific end users. In that sense, the ability for a end user to chose a DNS resolver represents major improvement. When the DNS resolvers are not on-path, and the end user changes from one DNS resolver to the other, encryption does not provide additional protection. In fact, encryption is clearly aiming at protecting against an attacker that would be on-path.

On the other hand, as mentioned in section [Section 3](#), web traffic, unless using more advance IP routing such as with TOR, also leaks similar information. Though gathering the information from the web traffic instead of the DNS traffic raises the bar, it represents a major improvement only if the bar remains sufficiently high. Unfortunately, the bar is nonexistent for user tracking, and remains weak to generalize the tracking to all users. As such the decision is more on the network side to decide the value associated to the DNS traffic or legal requirements to put the necessary infrastructure in place for it. It does not seem to be entirely in the end user's hand. As result, the encryption provides limited protection against on-path parties – such as an ISP. Unless combined wit TOR, moving to a DNS resolver that is not managed by the ISP does not hide much information to the ISP.

The remaining of this section considers that DNS information is not inferred from the web traffic and analyses how moving from a DNS resolver hosted in the ISP network to an DNS resolver outside the ISP network impacts the enduser's privacy. The perspective is considering the information shared with the DNS resolver, and motivations such as bad privacy protection, better latency, DNSSEC resolution, parental filtering are out of scope of this analysis.

provided by the ISP, it could be interpreted that encryption is not necessary. This is not the case, especially as wireless communication might be unprotected or provide the ability to man-in-the middle. As such we assume in this section that the channel between the end user and the DNS resolver is encrypted.

In general a DNS resolver can be seen as an anonymizer. It receives a DNS request from a specific end user, resolves the request under the resolver's identity and sends the response back to the end user. Local ISPs are believed to be fairly close to the end user and as such the IP address of the resolver can be used as a fairly good approximation of the localisation of the end user without revealing information about its identity.

When the end user is using a DNS resolver that is not located into the ISP network, the end user is clearly providing this information to another entity that used not to have this information and that cannot infer it from observing the traffic. Similarly to the ISP, the level of information depends on what is already shared with that entity. If the end user were not sharing any information with that entity, the end user may provide sufficient information to get profiled by an additional entity. If that DNS resolver already got a significant amount of data on that user, that data may fill the little remaining privacy but could have a much smaller impact. For example, in a highly concentrated Internet with one cloud provider for all services, the end user traffic would use one - or a few - destination IP addresses. That cloud provider would have access to all the history of the end user, while the ISP would have little information from the IP addresses or the encrypted SNI. In this specific situation the end user may chose that cloud provider for its DNS resolutions, to minimize the information leakage. Such scenario is currently believed to be purely hypothetical.

As a result, using a public resolver rather than a local resolver can be seen as sharing the web history of the end user. The balance between sharing partially that history versus completely transferring this information depends on the level of concentration of the Internet and the ability of the resolver not to further share that information. Using a public resolver also means that the user has to trust the public resolver handling the information according to the user's wishes.

[6.](#) Privacy impact of concentration

Section [Section 3](#) pointed out that concentration was one factor that could contribute in enforcing boundaries between the information carried through the DNS traffic and the information provided by

observing the web traffic. This section analyses how concentration impacts privacy.

At first sight the ability that concentration has to 'hide' multiple web sites behind a single IP address has to be balanced by the fact that a significant amount of your traffic is going into one place - or at least a single actor. In other words, concentration represent a direct threat to privacy with all you data being provided to one person. The cost associated to hide the signification of the IP address is too high, and even a higher trust in one cloud provider rather than your ISP could hardly justify such an approach.

Firstly, in order to hide the signification of the destination IP address, mechanisms such as TOR should be used instead. Secondly, trust may change over time, but provided data can hardly be retired. As such privacy should be designed in a way that does not depends on one or few players. Ideally, the data should be sufficiently spread among the various players so that none of them could exploit them. This can only be enforced by a *large number of player*. Typically as long as the fraction of data shared with one player is sufficient to start being analysed, privacy is at risk.

To balance that risk, it matters to reduce the amount of data shared as well as to minimize the level of information associated to that data. Typically suppose that a cloud provider proposes both a DNS resolution service as well as hosts the web server `www.example.com`. Performing the DNS resolution over that cloud provider for `www.example.com` will provide limited additional information as the DNS resolution will follow an web connection.

Note that concentration here includes the access to the data. In other words, a cloud provider hosting various web servers without possible access to that data will not fall into the concentration concept described in this section.

While the concentration represents a threat to the privacy, the remaining of this section analysis the impact of a cloud provider providing both a DNS resolution service and hosting service and exposes how this could contribute in balkanizing the internet, or more precisely capturing end users into close to wall gardened networks.

Typically, one cloud provider hosting a DNS resolver is likely to redirect the end user traffic within its data center rather than to the data center of a competitor. Note that such choice may be appropriated according to the localisation of the DNS resolver. The

problem may arise when the end user would benefit of a better connectivity by accessing the web site instantiated in the cloud of

an other cloud provider. In this case, the choice of the DNS resolver may be motivated by its own interest rather than the interest of the end user thus capturing the end user. Furthermore, the former optimization of the data center of the DNS resolver might lead in capturing the end user. Here capturing would mean the cloud provider is keeping the end user - as much as it can - within its borders. Such capture represents a major threat to privacy as the end user is literally kept into one entity, independently of its willingness.

The ability to capture an end user is problematic as it might become a mean to bring the end user into a different jurisdiction as its local jurisdiction. This may represent a direct threat to its private information as some jurisdictions provide little protection regarding to privacy. The comparison of local jurisdiction versus other jurisdictions is not the topic of the document. We do not ignore that certain jurisdictions represent a permanent threat to privacy. However, those jurisdictions put apart, it might also to notice that the local jurisdiction is probably the one best understood by the end user, and that bringing its data into other jurisdiction may goes against its believes. Similarly, some aspects of jurisdictions may also reflect the choice of societies, like the protection of the weakest of their members [IWF].

[6.1.](#) Acknowledgment

We would like to thank the feed backs we received from Bengt Sahlin, Christian Schaefer and Mirja Kuhlewind.

[7.](#) References

[7.1.](#) Normative References

- [RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, [RFC 1035](#), DOI 10.17487/RFC1035, November 1987, <<https://www.rfc-editor.org/info/rfc1035>>.
- [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>>.

- [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>>.

Migault

Expires May 7, 2020

[Page 10]

Internet-Draft A privacy analysis on DoH deployment November 2019

- [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>>.

[7.2](#). Informative References

- [IWF] "DNS over HTTPS Why we're saying DoH could be catastrophic", n.d., <<https://www.iwf.org.uk/news/dns-over-https-why-we%E2%80%99re-saying-doh-could-be-catastrophic>>.
- [RFC3552] Rescorla, E. and B. Korver, "Guidelines for Writing RFC Text on Security Considerations", [BCP 72](#), [RFC 3552](#), DOI 10.17487/RFC3552, July 2003, <<https://www.rfc-editor.org/info/rfc3552>>.
- [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>>.
- [RFC7258] Farrell, S. and H. Tschofenig, "Pervasive Monitoring Is an Attack", [BCP 188](#), [RFC 7258](#), DOI 10.17487/RFC7258, May 2014, <<https://www.rfc-editor.org/info/rfc7258>>.

Author's Address

Daniel Migault
Ericsson
8275 Trans Canada Route
Saint Laurent, QC 4S 0B6

Canada

EMail: mglt.ietf@gmail.com

Migault

Expires May 7, 2020

[Page 11]