Workgroup: Network Working Group
Internet-Draft:
draft-arkko-iab-data-minimization-principle-05
Published: July 2023
Intended Status: Informational
Expires: 11 January 2024
Authors: J. Arkko
Ericsson

Emphasizing data minimization among protocol participants

Abstract

Data minimization is an important privacy technique, as it can reduce the amount information exposed about a user. This document emphasizes the need for data minimization among primary protocol participants, such as between clients and servers. Avoiding data leakage to outside parties is of course very important as well, but both need to be considered in minimization.

This is because is necessary to protect against endpoints that are compromised, malicious, or whose interests simply do not align with the interests of users. It is important to consider the role of a participant and limit any data provided to it according to that role.

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 2 January 2024.

Table of Contents

1. Introduction
2. Recommendations
3. Discussion
  3.1. Types of Protocol Exchanges
  3.2. Types of information
  3.3. Different Ways of Avoiding Information Sharing
  3.4. Role of Trust
  3.5. Evolvability and Fingerprinting
  3.6. Related work
4. Acknowledgements
5. Informative References
Author's Address

1. Introduction

Privacy been at the center of many activities in the IETF. Privacy and its impact on protocol development activities at IETF is discussed in [RFC6973], covering a number of topics, from understanding privacy threats to threat mitigation, including data minimization.

This document emphasizes the need for data minimization among primary protocol participants, such as between clients and servers. Avoiding data leakage to outside parties such as observers or attackers is of course very important as well, but minimization needs to consider both.

As RFC 6973 states:

"Limiting the data collected by protocol elements to
 only what is necessary (collection limitation) is
 the most straightforward way to help reduce privacy
 risks associated with the use of the protocol."

This document offers some further discussion, recommendations, and clarifications for this. This document suggests that limiting the sharing of data to the protocol participants is a key technique in limiting the data collection mentioned above. It is important that

minimization happens prior to disclosing information to another party, rather than relying on the good will of the other party to avoid storing the information.

This is because is necessary to protect against endpoints that are compromised, malicious, or whose interests simply do not align with the interests of users. It is important to consider the role of a participant and limit any data provided to it according to that role.

Even closed, managed networks may have compromised nodes, justifying careful consideration of what information is provided to different nodes in the network. And in all networks, increased use of communication security means adversaries may resort to new avenues of attack. New adversaries and risks have also arisen, e.g., due to increasing amount of information stored in various Internet services. And in situations where interests do not align across the protocol participants, limiting data collection by a protocol participant itself - who is interested in data collection - may not be sufficient.

Careful control of information is also useful for technology evolution. For instance, allowing a party to unnecessarily collect or receive information may lead to a similar effect as described in [RFC8546] for protocols: regardless of initial expectations, over time unnecessary information will get used, leading to, for instance, ossification. Systems end up depend on having access to exactly the same information as they had access to previously. This makes it hard to change what information is provided or how it is provided.

2. Recommendations

The Principle of Least Privilege [PoLP] is applicable:

> "Every program and every user of the system should operate using the least set of privileges necessary to complete the job."

In this context, it is recommended that the protocol participants minimize the information they share. I.e., they should provide only the information to each other that is necessary for the function that is expected to be performed by the other party.

3. Discussion

3.1. Types of Protocol Exchanges

Information sharing may relate to different types of protocol exchanges, e.g., interaction of an endpoint with outsiders, the network, or intermediaries.

Other documents address aspects related to networks ([RFC8546], [RFC8558], [I-D.iab-path-signals-collaboration]). Thomson [I-D.thomson-tmi] discusses the role intermediaries. Communications security largely addresses observers and outsider adversaries, see for instance [Confidentiality], [RFC7858], [RFC8446], [RFC8484], [RFC9000]. And [RFC6973] discusses associated traffic analysis threats.

The focus in this document is on the primary protocol participants, such as a server in a client-server architecture or a service enables some kind of interaction among groups of users.

As with communication security, we try to avoid providing too much information as it may be misused or leak through attacks. The same principle applies not just to routers and potential attackers on path, but also many other services in the Internet, including servers that provide some function.

3.2. Types of information

The use of identifiers has been extensively discussed in [RFC6973],

Note that indirectly inferred information can also end up being shared, such as message arrival times or patterns in the traffic flow ([RFC6973]). Information may also be obtained from fingerprinting the protocol participants, in an effort to identify unique endpoints or users. Information may also be combined from multiple sources, e.g., websites and social media systems collaborating to identify visiting users [WP2021].

3.3. Different Ways of Avoiding Information Sharing

The most straightforward approach is of course to avoid sending a particular piece of information at all.

Or the information needs to be encrypted to very specific recipients, even if the encrypted message is shared with a broader set of protocol participants. For instance, a client can encrypt a message only to the actual final recipient, even if the server holds the message before it is delivered.

> Architectural note: A transport connection between two components of a system is not an end-to-end connection even if it encompasses all the protocol layers up to the application layer. It is not end-to-end, if the information or control function it carries extends beyond those components. Just because an e-mail server can read the contents of an e-mail message do not make it a legitimate recipient of the e-mail.

This document recommends that information should not be disclosed, stored, or routed in cleartext through services that do not need to have that information for the function they perform.

Where the above methods are not possible due to the information being necessary for a function that the user wishes to be performed, there are still methods to set limits on the information sharing.

Kühlewind et al discuss the concept of Privacy Partititioning [I-D.iab-privacy-partitioning]. This may involve designs where no single party has all information such as with Oblivious DNS [I-D.annee-dprive-oblivious-dns], [I-D.pauly-dprive-oblivious-doh] or HTTP [I-D.ietf-ohai-ohttp], cryptographic designs where a service such as with the recent IETF PPM effort [I-D.ietf-ppm-dap], and so on.

3.4. Role of Trust

Of course, participants may provide more information to each other after careful consideration, e.g., information provided in exchange of some benefit, or to parties that are trusted by the participant.

3.5. Evolvability and Fingerprinting

The general topic of ensuring that protocol mechanisms stays evolvable and workable is covered in [I-D.iab-use-it-or-lose-it]. But the associated methods for reducing fingerprinting possibilities probably deserve further study [Fingerprinting] [AmIUnique]. [I-D.wood-pearg-website-fingerprinting] discusses one aspect of this.

3.6. Related work

Cooper et al. [RFC6973] discuss the general concept of privacy, including data minimization. Among other things, it provides the general statement quoted in Section 1. It also provides guidelines to authors of specifications, a number of questions that protocol designers can use to further analyze the impact of their design. For data minimization the questions relate to identifiers, data, observers, and fingerprinting. This includes, for instance, asking

what information is exposed to which protocol entities, and if there are ways to limit such exposure:

> Observers. Which information discussed in (a) and (b) is exposed to each other protocol entity (i.e., recipients, intermediaries, and enablers)? Are there ways for protocol implementers to choose to limit the information shared with each entity? Are there operational controls available to limit the information shared with each entity?

This is very much in line with this document, although today it would be desirable to have recommendation as well as questions. For instance, recommending against sharing information with a participant if it is not necessary for the expected role of that participant. And, as discussed earlier, it is important to distinguish between the choices of a sender not sharing information and a receiver choosing to not collect information. Trusting an entity to not collect is not sufficient.

There has also been a number of documents that address data minimization for specific situations, such as one DNS Query Name Minimization [RFC7816], general DNS privacy advice including data minimization [RFC9076], advice for DHCP clients for minimizing information in requests they send to DHCP servers [RFC7844] (along with general privacy considerations of DHCP [RFC7819] [RFC7824]). These are on the topic of limiting information sent by one primary protocol particiant (client) to another (server).

Hardie [RFC8558] and Arkko et al. [I-D.iab-path-signals-collaboration] discuss path signals, i.e., messages to or from on-path elements to endpoints. In the past, path signals were often implicit, e.g., network nodes interpreting in a particular way transport protocol headers originally intended for end-to-end consumption. Implicit signals should be avoided and that explicit signals be used instead.

Kühlewind, Pauly, and Wood [I-D.iab-privacy-partitioning] discuss the concept of privacy partitioning: how information can be split and carefully shared in ways where no individual party beyond the client requesting a service has full picture of who is asking and what is being asked.

Thomson [I-D.thomson-tmi] discusses the role intermediaries in the Internet architecture, at different layers of the stack. For instance, a router is an intermediary, some parts of DNS infrastructure can be intermediaries, messaging gateways are intermediaries. Thomson discusses when intermediaries are or are not an appropriate tool and presents a number of principles relating to the use of intermediaries.

Trammel and Kühlewind [RFC8546] discuss the concept of a “wire image” of a protocol, and how network elements may start to rely on information in the image, even if it was not originally intended for them. The issues are largely the same even for participants.

4. Acknowledgements

The author would like to thank the participants of various IAB workshops and programs, and IETF discussion list contributors for interesting discussions in this area. The author would in particular like to acknowledge the significant contributions of Martin Thomson, Nick Doty, Alissa Cooper, Stephen Farrell, Mark McFadden, John Mattsson, Chris Wood, Dominique Lazanski, Eric Rescorla, Russ Housley, Robin Wilton, Mirja Kühlewind, Tommy Pauly, Jaime Jiménez and Christian Huitema.

This work has been influenced by [RFC6973], [RFC8980], [I-D.farrell-etm] [I-D.arkko-arch-internet-threat-model-guidance], [I-D.lazanski-smart-users-internet],

5. Informative References

[AmIUnique] INRIA, ., "Am I Unique?", https://amiunique.org , 2020.

[Confidentiality] The Internet Architecture Board, ., "IAB Statement on Internet Confidentiality", https://www.iab.org/2014/11/14/iab-statement-on-internet-confidentiality/ , November 2014.

[Fingerprinting] Laperdrix, P., Bielova, N., Baudry, B., and G. Avoine, "Browser Fingerprinting: A survey", arXiv: 1905.01051v2 [cs.CR] 4 Nov 2019 , November 2019.

[I-D.annee-dprive-oblivious-dns] Edmundson, A., Schmitt, P., Feamster, N., and A. Mankin, "Oblivious DNS - Strong Privacy for DNS Queries", Work in Progress, Internet-Draft, draft-annee-dprive-oblivious-dns-00, 2 July 2018, <https://datatracker.ietf.org/doc/html/draft-annee-dprive-oblivious-dns-00>.

[I-D.arkko-arch-internet-threat-model-guidance] Arkko, J. and S. Farrell, "Internet Threat Model Guidance", Work in Progress, Internet-Draft, draft-arkko-arch-internet-threat-model-guidance-00, 12 July 2021, <https://datatracker.ietf.org/doc/html/draft-arkko-arch-internet-threat-model-guidance-00>.

[I-D.farrell-etm] Farrell, S., "We're gonna need a bigger threat model", Work in Progress, Internet-Draft, draft-farrell-

etm-03, 6 July 2019, <https://datatracker.ietf.org/doc/html/draft-farrell-etm-03>.

[I-D.iab-path-signals-collaboration] Arkko, J., Hardie, T., Pauly, T., and M. Kühlewind, "Considerations on Application - Network Collaboration Using Path Signals", Work in Progress, Internet-Draft, draft-iab-path-signals-collaboration-03, 3 February 2023, <https://datatracker.ietf.org/doc/html/draft-iab-path-signals-collaboration-03>.

[I-D.iab-privacy-partitioning] Kühlewind, M., Pauly, T., and C. A. Wood, "Partitioning as an Architecture for Privacy", Work in Progress, Internet-Draft, draft-iab-privacy-partitioning-01, 13 March 2023, <https://datatracker.ietf.org/doc/html/draft-iab-privacy-partitioning-01>.

[I-D.iab-use-it-or-lose-it] Thomson, M. and T. Pauly, "Long-Term Viability of Protocol Extension Mechanisms", Work in Progress, Internet-Draft, draft-iab-use-it-or-lose-it-04, 12 October 2021, <https://datatracker.ietf.org/doc/html/draft-iab-use-it-or-lose-it-04>.

[I-D.ietf-ohai-ohttp] Thomson, M. and C. A. Wood, "Oblivious HTTP", Work in Progress, Internet-Draft, draft-ietf-ohai-ohttp-08, 15 March 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-ohai-ohttp-08>.

[I-D.ietf-ppm-dap] Geoghegan, T., Patton, C., Rescorla, E., and C. A. Wood, "Distributed Aggregation Protocol for Privacy Preserving Measurement", Work in Progress, Internet-Draft, draft-ietf-ppm-dap-05, 10 July 2023, <https://datatracker.ietf.org/api/v1/doc/document/draft-ietf-ppm-dap/>.

[I-D.lazanski-smart-users-internet] Lazanski, D., "An Internet for Users Again", Work in Progress, Internet-Draft, draft-lazanski-smart-users-internet-00, 8 July 2019, <https://datatracker.ietf.org/doc/html/draft-lazanski-smart-users-internet-00>.

[I-D.pauly-dprive-oblivious-doh] Kinnear, E., McManus, P., Pauly, T., Verma, T., and C. A. Wood, "Oblivious DNS over HTTPS", Work in Progress, Internet-Draft, draft-pauly-dprive-oblivious-doh-11, 17 February 2022, <https://

datatracker.ietf.org/doc/html/draft-pauly-dprive-oblivious-doh-11>.

[I-D.thomson-tmi] Thomson, M., "Principles for the Involvement of Intermediaries in Internet Protocols", Work in Progress, Internet-Draft, draft-thomson-tmi-04, 8 September 2022, <https://datatracker.ietf.org/doc/html/draft-thomson-tmi-04>.

[I-D.wood-pearg-website-fingerprinting] Goldberg, I., Wang, T., and C. A. Wood, "Network-Based Website Fingerprinting", Work in Progress, Internet-Draft, draft-wood-pearg-website-fingerprinting-00, 4 November 2019, <https://datatracker.ietf.org/doc/html/draft-wood-pearg-website-fingerprinting-00>.

[PoLP] Saltzer, J. and M. Schroader, "The Protection of Information in Computer Systems", Fourth ACM Symposium on Operating System Principles , October 1975.

[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>.

[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>.

[RFC7819] Jiang, S., Krishnan, S., and T. Mrugalski, "Privacy Considerations for DHCP", RFC 7819, DOI 10.17487/RFC7819, April 2016, <https://www.rfc-editor.org/info/rfc7819>.

[RFC7824] Krishnan, S., Mrugalski, T., and S. Jiang, "Privacy Considerations for DHCPv6", RFC 7824, DOI 10.17487/RFC7824, May 2016, <https://www.rfc-editor.org/info/rfc7824>.

[RFC7844] Huitema, C., Mrugalski, T., and S. Krishnan, "Anonymity Profiles for DHCP Clients", RFC 7844, DOI 10.17487/RFC7844, May 2016, <https://www.rfc-editor.org/info/rfc7844>.

[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>.

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

[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>.

[RFC8546] Trammell, B. and M. Kuehlewind, "The Wire Image of a Network Protocol", RFC 8546, DOI 10.17487/RFC8546, April 2019, <https://www.rfc-editor.org/info/rfc8546>.

[RFC8558] Hardie, T., Ed., "Transport Protocol Path Signals", RFC 8558, DOI 10.17487/RFC8558, April 2019, <https://www.rfc-editor.org/info/rfc8558>.

[RFC8980] Arkko, J. and T. Hardie, "Report from the IAB Workshop on Design Expectations vs. Deployment Reality in Protocol Development", RFC 8980, DOI 10.17487/RFC8980, February 2021, <https://www.rfc-editor.org/info/rfc8980>.

[RFC9000] Iyengar, J., Ed. and M. Thomson, Ed., "QUIC: A UDP-Based Multiplexed and Secure Transport", RFC 9000, DOI 10.17487/RFC9000, May 2021, <https://www.rfc-editor.org/info/rfc9000>.

[RFC9076] Wicinski, T., Ed., "DNS Privacy Considerations", RFC 9076, DOI 10.17487/RFC9076, July 2021, <https://www.rfc-editor.org/info/rfc9076>.

[WP2021] Fowler, Geoffrey A., "There’s no escape from Facebook, even if you don’t use it", Washington Post , August 2021.

Author's Address

Jari Arkko
Ericsson
Valitie 1B
FI- Kauniainen
Finland

Email: jari.arkko@piuha.net