Network Working Group J. Arkko
Internet-Draft Ericsson
Intended status: Informational March 14, 2023
Expires: September 15, 2023

Data minimization among protocol participants
draft-arkko-iab-data-minimization-principle-04

Abstract

Communications security has been at the center of many security improvements in the Internet. The goal has been to ensure that communications are protected against outside observers and attackers. Privacy has also been a key focus area, and understanding the privacy implications of new Internet technology is an important factor when IETF works on such technologies. One key aspect of privacy is minimization of data disclosed to other parties.

This document highlights the need for a particular focus with respect to data minimization. Avoiding data leakage to outside parties is of course important, but it can also be necessary to limit it among the primary protocol participants.

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.

Copyright Notice

Table of Contents

1. Introduction

Communications security has been at the center of many security improvements on the Internet. The goal has been to ensure that communications are protected against outside observers and attackers.

This has been exemplified in many aspects of IETF efforts, in the threat models [RFC3552], concerns about surveillance [RFC7258], IAB statements [Confidentiality], and the introduction of encryption in many protocols [RFC9000], [RFC7858], [RFC8484]. This has been very successful. Advances in protecting most of our communications with strong cryptographic has resulted in much improved security. Work on these advances continues to be a key part of IETF's security effort.

Privacy has also been at the center of many activities in the IETF. Improvements in communications security obviously have improved privacy as well, but the concept is broader. 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.

One key aspect of privacy is minimization of data disclosed to other parties.

This document highlights the need for a particular focus with respect to data minimization. Avoiding data leakage to outside parties is of course important, but it can also be necessary to limit it among the primary protocol participants (and not just observers/attackers). 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 and motivation 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. This document also suggests that what information is given to any other participant should depend on the role of that participant.

The reason why this is important is that it is possible that endpoints are compromised, malicious, or have interests that do not align with the interests of users. 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.

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

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

2.1. 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 ([RFC6973]). Information may also be combined from multiple sources, e.g., websites and social media systems collaborating to identify visiting users [WP2021].

2.2. Dealing with compromise

Even with careful exposure of information, compromises may occur. It is important to build defenses to protect information, even when some component in a system becomes compromised. 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], service side encryption of data at rest, confidential computing, and other mechanisms.

Protocols can ensure that forward secrecy is provided, so that the damage resulting from a compromise of keying material has limited impact.

On the client side, the client may trust that another party handles information appropriately, but take steps to ensure or verify that this is the case. For instance, as discussed above, the client can encrypt a message only to the actual final recipient, even if the server holds the message before it is delivered.

A corollary of the recommendation is 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.

For instance, 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. For instance, 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.

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.

2.3. Related work

Cooper et al. [RFC6973] discuss the general concept of privacy, including data minimization. Section 6.1 in RFC 6973 is about data minimization, and there they provide the general statement quoted in Section 1. That is exactly about what the present document is also about. Section 7 in RFC 6973 is about guidelines to authors of specifications. And Section 7.1 in turn asks the a number of questions that protocol designers can use to further analyse 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. For the

present document this is the particularly important question in RFC 6973:

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?

These questions are in line with avoiding sharing information to a protocol participant unless it is needed for its role. Therefore, this too is exactly what the present document is about. It is clear from both the text and interviewing the authors of RFC 6973 that they believed in the importance of limiting data disclosure across all possible parties, from outsiders to the primary protocol participants, just as the author of the current document does.

However, if there's something to be wished, RFC 6973 could perhaps been more explicit:

o It should be stated that information should not even be shared with a participant if it is not necessary for the expected role of that participant. Yet, RFC 6973 says "Limiting the data collected by protocol elements to only what is necessary (collection limitation)". But when there are potentially conflicting interests among the protocol participants, expecting a participant to limit its data collection itself seems insufficient. What we need to do is to not even give that that participant the information. Interviewing the authors of the RFC, this what the intent of the text was, but the text isn't explicit about it.

o Similarly, the Section 7.1 guidance is merely questions, not recommendations about limiting to the necessary information.

o The examples in Section 6 are largely related to cases where some information is relayed to some parties but not others. For instance, the anonymity and identity confidentiality examples are about withholding identity from some parties such as the other endpoint of a call or outsiders observing an authentication exchange. But they still disclose an identity to a party running a service. This is of course necessary in many situations, but it would be useful to provide examples where information is withheld entirely.

These are of course nuances that may change in a future revision of RFC 6973.

In the years after publishing [RFC6973] there has also been a number of documents on specific issues, 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] discusses 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. The document recommends a principle that implicit signals should be avoided and that explicit signals be used instead, and only when the signal's originator intends that it be used by the network elements on the path.

Arkko, Kuhlewind, Pauly, and Hardie [I-D.iab-path-signals-collaboration] discuss the same topic, and extend the RFC 8558 principles with recommendations to ensure minimum set of parties, minimum information, and consent.

Kuehlewind, 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. This is of course a highly relevant technique, and should be a part of the data minimization toolkit.

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, e.g., deliberate selection of protocol participants or limiting the capabilities or information exposure related to the intermediaries.

Trammel and Kuehlewind [RFC8546] discuss the concept of a "wire image" of a protocol. This is an abstraction of the information available to an on-path non-participant in a networking protocol. It relates to the topic of non-participants interpreting information that is available to them in the "wire image" (or associated timing and other indirect information). The issues are largely the same even for participants. Even proper protocol participants may start

to use information available to them, regardless of whether it was intended to that participant or simply relayed through them.

3. 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 Kuehlewind, Tommy Pauly, Jaime Jimenez 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],

4. 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", draft-annee-dprive-oblivious-dns-00 (work in progress), 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", draft-arkko-arch-internet-threat-model-guidance-00 (work in progress), 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", draft-farrell-etm-03 (work in progress), 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. Kuehlewind, "Considerations on Application - Network Collaboration Using Path Signals", draft-iab-path-signals-collaboration-03 (work in progress), February 2023, <https://datatracker.ietf.org/doc/html/draft-iab-path-signals-collaboration-03>.

[I-D.iab-privacy-partitioning]
Kuehlewind, M., Pauly, T., and C. Wood, "Partitioning as an Architecture for Privacy", draft-iab-privacy-partitioning-01 (work in progress), 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", draft-iab-use-it-or-lose-it-04 (work in progress), 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. Wood, "Oblivious HTTP", draft-ietf-ohai-ohttp-07 (work in progress), March 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-ohai-ohttp-07>.

[I-D.ietf-ppm-dap]
Geoghegan, T., Patton, C., Rescorla, E., and C. Wood, "Distributed Aggregation Protocol for Privacy Preserving Measurement", draft-ietf-ppm-dap-04 (work in progress), March 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-ppm-dap-04>.

[I-D.lazanski-smart-users-internet]
Lazanski, D., "An Internet for Users Again", draft-lazanski-smart-users-internet-00 (work in progress), 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. Wood, "Oblivious DNS over HTTPS", draft-pauly-dprive-oblivious-doh-11 (work in progress), 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", draft-thomson-tmi-04 (work in progress), September 2022, <https://datatracker.ietf.org/doc/html/draft-thomson-tmi-04>.

[I-D.wood-pearg-website-fingerprinting]
Goldberg, I., Wang, T., and C. Wood, "Network-Based Website Fingerprinting", draft-wood-pearg-website-fingerprinting-00 (work in progress), 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.

[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>.

[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>.

[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., "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
Kauniainen
Finland

Email: jari.arkko@piuha.net