Emphasizing data minimization among protocol participants

draft-arkko-iab-data-minimization-principle-05

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.

Copyright Notice

Copyright (c) 2023 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.
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.lab-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 Partitioning [I-D.lab-privacy-partitioning]. This may involve designs where no single party has all information such as with Oblivious DNS [I-D.annee-dprime-oblivious-dns], [I-D.pauly-dprime-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.lab-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 participant (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 Küehlewind [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 Küehlewind, 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]

[Confidentiality]

[Fingerprinting]

[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-
[I-D.arkko-arch-internet-threat-model-guidance]

[I-D.farrell-etm]

[I-D.iab-path-signals-collaboration]

[I-D.iab-privacy-partitioning]

[I-D.iab-use-it-or-lose-it]

[I-D.ietf-ohai-ohttp]

[I-D.ietf-ppm-dap]
Geoghegan, T., Patton, C., Rescorla, E., and C. A. Wood, "Distributed Aggregation Protocol for Privacy Preserving

[I-D.lazanski-smart-users-internet]

[I-D.pauly-dprive-oblivious-doh]

[I-D.thomson-tmi]

[I-D.wood-pearg-website-fingerprinting]

[PoLP]


Author’s Address

Jari Arkko
Ericsson
Valitie 1B
FI- Kauniainen
Finland
Email: jari.arkko@piuha.net
Report from the IAB workshop on Management Techniques in Encrypted Networks (M-TEN)  
draft-iab-m-ten-workshop-02

Abstract

The Management Techniques in Encrypted Networks (M-TEN) workshop was convened by the Internet Architecture Board (IAB) from 17 October 2022 to 19 October 2022 as a three-day online meeting. The workshop was organized in three parts to discuss ways to improve network management techniques in support of even broader adoption of encryption on the Internet. This report summarizes the workshop’s discussion and identifies topics that warrant future work and consideration.

Note that this document is a report on the proceedings of the workshop. The views and positions documented in this report are those of the expressed during the workshop by participants and do not necessarily reflect IAB views and positions.

About This Document

This note is to be removed before publishing as an RFC.

The latest revision of this draft can be found at https://intarchboard.github.io/m-ten-workshop-public/draft-iab-m-ten-workshop.html. Status information for this document may be found at https://datatracker.ietf.org/doc/draft-iab-m-ten-workshop/.

Source for this draft and an issue tracker can be found at https://github.com/intarchboard/m-ten-workshop-public.

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 15 February 2024.

Copyright Notice

Copyright (c) 2023 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.

Table of Contents

1. Introduction ........................................... 3

1.1. About this workshop report content ................. 3

2. Workshop Scope and Discussion .......................... 4

  2.1. Where we are - Requirements and Passive Observations . 4

  2.1.1. Traffic classification and network management .... 4

  2.1.2. Preventing traffic analysis ...................... 5

  2.1.3. Users and privacy .............................. 6

  2.1.4. Discussion ..................................... 6

  2.2. Where we want to go - Collaboration Principles .... 7

  2.2.1. First party collaboration for network management .. 7

  2.2.2. Second and third party collaboration for network management ........................................... 8

  2.2.3. Visible, optional network management ............... 9

  2.2.4. Discussion ..................................... 9

  2.3. How we get there - Collaboration Use cases .......... 10

  2.3.1. Establishing expected contracts to enable security management ........................................... 10

  2.3.2. Zero Knowledge Middleboxes ........................ 11

  2.3.3. Red Rover - A collaborative approach to content filtering ........................................... 12

3. Conclusions ........................................... 12

4. Informative References ................................. 13

Appendix A. Position Papers ............................... 15
1. Introduction

User privacy and security are constantly being improved by increasingly strong and more widely deployed encryption. This workshop aims to discuss ways to improve network management techniques in support of even broader adoption of encryption on the Internet.

Network management techniques need to evolve to work effectively and reliably in the presence of ubiquitous traffic encryption and support techniques that enhance user privacy. In an all-encrypted network, it is not viable to rely on unencrypted metadata for network monitoring and security functions, troubleshooting devices, and passive traffic measurements. New approaches are needed to track network behaviors, e.g., by directly cooperating with endpoints and applications, increasing use of in-band telemetry, increasing use of active measurement approaches, and privacy-preserving inference techniques.

The aim of this IAB online workshop from October 17-19, 2022, has been to provide a platform to explore the interaction between network management and traffic encryption and initiate new work on collaborative approaches that promote security and user privacy while supporting operational requirements. As such the workshop addressed the following questions:

* What are actionable network management requirements?
* Who is willing to work on collaborative solutions?
* What are the starting points for collaborative solutions?

1.1. About this workshop report content

This document is a report on the proceedings of the workshop. The views and positions documented in this report are those of the expressed during the workshop by participants and do not necessarily reflect IAB views and positions.
Furthermore, the content of the report comes from presentations given by workshop participants and notes taken during the discussions, without interpretation or validation. Thus, the content of this report follows the flow and dialog of the workshop but does not attempt to capture a consensus.

2. Workshop Scope and Discussion

The workshop was organized across three days with all-group discussion slots, one per day. The following topic areas were identified and the program committee organized paper submissions into three main themes for each of the three discussion slots. During each discussion, those papers were presented sequentially with open discussion held at the end of each day.

2.1. Where we are - Requirements and Passive Observations

The first day of the workshop agenda focused on the existing state of the relationship between network management and encrypted traffic from various angles. Presentations ranged from discussing classifiers using machine-learning to recognize traffic, to advanced techniques for evading traffic analysis, to user privacy considerations.

After an introduction that covered the goals of the workshop and the starting questions (as described in Section 1), there were four presentations, followed by open discussion.

2.1.1. Traffic classification and network management

Many existing network management techniques are passive in nature: they don’t rely on an explicit signals from end hosts to negotiate with network middleboxes, but instead rely on inspecting packets to recognize traffic and apply various policies. Traffic classification, as a passive technique, is being challenged by increasing encryption.

Traffic classification is commonly performed by networks to infer what applications and services are being used. This information is in turn used for capacity and resource planning, Quality-of-Service (QoS) monitoring, traffic prioritization, network access control, identity management, and malware detection. However, since classification traditionally relies on recognizing unencrypted properties of packets in a flow, increasing encryption of traffic can decrease the effectiveness of classification.
The amount of classification that can be performed on traffic also provides a useful insight onto how "leaky" the protocols used by applications are, and points to areas where information is visible to any observer, which may be malicious or not.

Traditionally, classification has been based on experts crafting specific rules, but there is also a move toward using machine learning to recognize patterns. "Deep learning" machine learning models generally rely on analyzing a large set of traffic over time, and have trouble reacting quickly to changes in traffic patterns.

Models that are based on closed-world data sets also become less useful over time, as traffic changes. [JIANG] describes experiments that showed that a model that performs with high accuracy on an initial data set became severely degraded when running on a newer data set that contained traffic from the same applications. Even in as little time as one week, the traffic classification would become degraded. However, the set of features in packets and flows that were useful for models stayed mostly consistent, even if the models themselves needed to be updated. Models where the feature space is reduced to fewer features showed better resiliency, and could be retrained more quickly. Based on this, [JIANG] recommends more work and research on determining which set of features in IP packets are most useful for focused machine learning analysis. [WU] also recommends further research investment in Artificial Intelligence (AI) analysis for network management.

2.1.2. Preventing traffic analysis

Just as traffic classification is continually adapting, techniques to prevent traffic analysis and obfuscate application and user traffic are continually evolving. An invited talk from the authors of [DITTO] shared a novel approach with the workshop for how to build a very robust system to prevent unwanted traffic analysis.

Usually traffic obfuscation is performed by changing the timing of packets or adding padding data. The practices can be costly and negatively impact performance. DITTO demonstrated the feasibility of applying traffic obfuscation on aggregated traffic in the network with minimal overhead and in line speed.

While traffic obfuscation techniques are today not widely deployed, this study underlines, together with the need for continuous effort to keep traffic models updated over time, the challenges of classification of encrypted traffic as well as opportunities to further enhance user privacy.
2.1.3. Users and privacy

The Privacy Enhancements and Assessments Research Group is working on a document to discuss guidelines for how to measure traffic on the Internet in a safe and privacy-friendly way ([I-D.irtf-pearg-safe-internet-measurement]). These guidelines and principles provide another angle onto the discussion of passive classification and analysis of traffic.

Consent for collection and measurement of metadata is an important consideration in deploying network measurement techniques. This consent can be explicitly given as informed consent, or can be given by proxy or be only implied. For example, a user of a network might need to consent to certain measurement and traffic treatment when joining a network.

Various techniques for data collection can also improve user privacy, such as discarding data after a short period of time, masking out aspects of data that contain user-identifying information, reducing the accuracy of collected data, and aggregating data.

2.1.4. Discussion

The intents and goals of users, application developers, and network operators align in some cases, but not others. One of the recurring challenges that came up was not having a clear way to understand or communicate intents and requirements. Both traffic classification and traffic obfuscation attempt to change the visibility of traffic without cooperation of other parties: traffic classification is a network attempting to inspect application traffic without coordination from applications, and traffic obfuscation is an attempt to hide that same traffic as it transits a network.

Traffic adaptation and prioritization is one dimension in which the incentives for cooperation seem most clear. Even if an application is trying to prevent leaking metadata, it could benefit from signals from network about sudden capacity changes that can help it adapt its application quality, such as bitrates and codecs. Such signalling may not be appropriate for the most privacy-sensitive applications, like Tor, but could be applicable for many others. There are existing protocols that involve explicit signaling between applications and networks, such as Explicit Congestion Notification (ECN) [RFC3168], but that has yet to see wide adoption.
Managed networks (such as private corporate networks) were brought up in several comments as a particularly challenging area for being able to meet management requirements while maintaining encryption and privacy. These networks can have legal and regulated requirements for the detection of specific fraudulent or malicious traffic.

Personal networks that enable managed parental controls have similar complications with encrypted traffic and user privacy. In these scenarios, the parental controls being operated by the network may be as simple as a DNS filter, and can be made ineffective by a device routing traffic to an alternate DNS resolver.

2.2. Where we want to go – Collaboration Principles

The second day of the workshop agenda focused on the emerging techniques for analysing, managing, or monitoring encrypted traffic. Presentations ranged from discussing advanced classification and identification, including machine-learning techniques, for the purposes of managing network flows, monitoring, or monetising usage.

After an introduction that covered the goals of the workshop and the starting questions (as described in Section 1), there were three presentations, followed by open discussion.

2.2.1. First party collaboration for network management

It is the intention of encryption to create a barrier between entities inside the communication channel and everyone else, including network operators, considering end-to-end encryption of traffic. Any attempt, therefore, to overcome that intentional barrier requires an intent to collaborate between the inside and outside entities. Those entities must, at a minimum, agree on the benefits to overcoming the barrier (or solving the problem), that costs are proportional to the benefits, and to additional limitations, or safeguards, against bad behaviour by collaborators including the inclusion of other non-insiders [BARNES].

The Internet is designed interoperably, which means an outside entity wishing to collaborate with the inside might be any number of intermediaries and not, say, a specific person that can be trusted in the human sense. Additionally, the use of encryption, especially network-layer or transport-layer encryption, introduces dynamic or opportunistic or perfunctory discoverability. These realities both point to a need to interrogate the reason why any outside entity might make an engineering case to collaborate with the user of a network with encrypted traffic, and whether the tradeoffs and potential risks are worth it to the user.
However, the answers cannot be specific and the determinations or
guidance need to be general as the encryption boundary is inevitably
an application used by many people. Tradeoffs must make sense to
users who are unlikely to be thinking about network management
considerations. Harms need to be preemptively reduced because in
general terms few users would choose network management benefits over
their own privacy if given the choice.

Some have found that there appears to be little if any actual
evidence that encryption is causing user-meaningful network problems.
Since alignment on problem-solving is a prerequisite to collaboration
on a solution it does not seem that collaboration across the
encryption boundary is called for.

2.2.2. Second and third party collaboration for network management

Even with the wide-scale deployment of encryption in new protocols
and techniques that prevent passive observers of network traffic from
knowing the content of exchanged communications, important
information such as which parties communicate and sometimes even
which services have been requested may still be able to be deduced.
The future is to conceal more data and metadata from passive
observers and also to minimize information exposure to second parties
(where the user is the first party) by, maybe counterintuitively,
introducing third-party relay services to intermediate
communications. As discussed in [KUEHLEWIND], the relay is a
mechanism to separate (using additional levels of encryption) two
important pieces of information: knowledge of the identity of the
person accessing a service is separated from knowledge about the
service being accessed. By contrast a VPN uses only one level of
encryption and does not separate identity (first party) and service
(second party) metadata.

Relay mechanisms are termed "oblivious", there is a future for
specifications in privacy-preserving measurement (PPM), and protocols
like Multiplexed Application Substrate over QUIC Encryption (MASQUE)
are discussed in the IETF. In various schemes, users are ideally
able to share their identity only with the entity they have
identified as a trusted one. That data is not shared with the
service provider. However this is more complicated for network
management, but there may be opportunities for better collaboration
between the network and, say, the application or service at the
endpoint.

A queriable relay mechanism could preserve network management
functions that are disrupted by encryption, such as TCP optimisation,
quality of service, zero-rating, parental controls, access control,
redirection, content enhancement, analytics and fraud prevention.
Instead of encrypted communication between only two ends and passive observation by all on-path elements, intermediate relays could be trusted parties with limited information for the purposes of collaboration between in-network intermediary services’ support.

2.2.3. Visible, optional network management

In encrypted communications, out of all of the possible network management functions that might be ameliorated by proxying, the ability to control congestion has been researched in depth. These techniques are realized based on TCP performance enhancing proxies (PEP) that either entirely intercept a TCP connection or interfere with the transport information in the TCP header. However, despite the challenge that the new encrypted protocol will limit any such in-network interference, these techniques can also have a negative impact on the evolvability of these protocols. Therefore, instead of manipulating existing information, a new approach was presented where additional information is send using a so-called side-car protocol independent of the main transport protocol that is used end-to-end [WELZL]. E.g. side car information can contain additional acknowledgements to enable in-network local retransmission faster end-to-end retransmission by reducing the signaling round trip time.

Taking user privacy benefits for granted, there is a need to investigate the comparable performance outputs of various encrypted traffic configurations such as use of an additional "side-car" protocol, or explicit encrypted and trusted network communication using MASQUE in relation to existing techniques such as TCP performance enhancing proxies (PEP), etc.

2.2.4. Discussion

One size fits all? On the issue of trust, different networks or devices are going to have different requirements for the level of trust that they have in devices, users or each other, and vice versa. For example, imagine networks with really different security requirements, like protecting children in a home versus a national security institution. How could one network architecture solve the needs of all use cases?

Does our destination have consequences? It seems sometimes that there may be consequences many years down the line of ubiquitous, strong encryption of network traffic because it will cause a reaction by intermediaries to find ways to poke holes in what are supposed to be long-term solutions for user privacy and security.
Can we bring the user along? While there has been a focus on the good reasons for why people might collaborate across the encryption barrier, there will always be others who want to disrupt that because they are motivated to exploit the data for their own gain, and sometimes this is called innovation. What high-level policy mitigations have done is to expose how powerless end users are to corporate practices of data harvesting. And yet interfaces to help users understand these lower layer traffic flows to protect their financial transactions or privacy haven’t been achieved yet. That means that engineers are having to make inferences about what users want. Instead we should be making these relationships and tradeoffs more visible.

2.3. How we get there - Collaboration Use cases

The third day focused on techniques that could actually be used to improve management of encrypted networks. A central theme of all of the presentations about potential proposed paths forward included some element of collaboration between networks and subscribing clients that simultaneously want both privacy and protection. Thus, the central theme in the third day became negotiation and collaboration.

2.3.1. Establishing expected contracts to enable security management

When thinking about enterprise networks where client behavior is potentially managed, [COLLINS] proposes "Improving network monitoring through contracts", where contracts describe different states of network behavior.

Because network operators have a limited amount of time to focus on problems and process alerts, contracts and states let the operator focus on a particular aspect of a current situation or problem. The current estimate for the number of events a Security Operations Center (SOC) operator can handle is about 10 per hour. Operators must work within the limits imposed by their organization, and must pick between options that frequently only frustrate attackers -- entirely preventing attacks is potentially impossible. Finally, operators must prioritize and manage the most events possible.

Validating which alerts are true positives is challenging because lots of weird traffic creates many anomalies and not all anomalies are malicious events. Identifying what anomalous traffic is rooted in malicious activity with any level of certainty is extremely challenging. Unfortunately, applying the latest machine learning techniques has only produced mixed results. To make matters worse, the large amounts of Internet-wide scanning has resulted in endless traffic that is technically malicious but only creates an information
overload and challenges event prioritization. Any path forward must succeed in freeing up analyst time to concentrate on the more challenging events.

The proposed contract solution is to define a collection of acceptable behaviors categorized into an envelope of different states that might include IP addresses, domain names, and indicators of compromise. Deviation from a contract might indicate that a system is acting outside a normal mode of behavior, or even a normal mode of behavior is suddenly missing. An example contract might be "this system is expected to update its base OS once a day", and if this doesn’t occur then this expectation has not been met and the system should be checked as it failed to call home to look for (potentially security related) updates.

Within the IETF, the Manufacturer Usage Description Specification (MUD) (?RFC8520) specification is one subset of contracts. Note that contracts are likely to only succeed in a constrained, expected environment maintained by operational staff, and may not work in an open internet environment where end users are driving all network connections.

2.3.2. Zero Knowledge Middleboxes

The world is not only shifting to increased encrypted traffic but is also encrypting more and more of the metadata (e.g. DNS queries and responses). This makes network policy enforcement by middleboxes significantly more challenging. The result is the creation of a significant tension between security enforcement and privacy protection.

A goal for solving this problem should include not weakening encryption, should enable networks to enforce their policies, and should ideally not require newly deployed server software. Existing solutions fail with at least one of these points.

A cryptographic principle of a "zero-knowledge proof" (ZKP) [GRUBBS] maybe one path forward to consider. A ZKP allows a third party to verify that a statement is true, without revealing what the statement actually is. Applying this to network traffic has been shown to allow a middlebox to verify that traffic to a web server is actually compliant with a policy without revealing the actual contents. This solution meets the above three criteria. Using ZKP within TLS 1.3 traffic turns out to be plausible.

An example engine was built to test ZKP using encrypted DNS. Clients were able to create DNS requests that were not listed within a DNS block list. Middleboxes could verify, without knowing the exact
request, that the client’s DNS request was not in the prohibited list. Although the result was functional, the computational overhead was still too slow and future work will be needed to decrease the ZKP imposed latencies.

2.3.3. Red Rover - A collaborative approach to content filtering

The principle challenge being studied is how to deal with the inherent conflict between filtering and privacy. Network operators need to implement policies and regulations that can originate from many locations (e.g. security, governmental, parental, etc). Conversely, clients need to protect user’s privacy and user security.

Safe browsing, originally created by Google, is one example of a mechanism that tries to meet both sides of this conflict. It would be beneficial to standardize this and other similar mechanisms. Operating systems could continually protect their users by ensuring that malicious destinations are not being reached. This would require some coordination between cooperating clients and servers offering protection services. These collaborative solutions may be the best compromise between the tension of privacy vs protection based services [PAULY].

3. Conclusions

Looking forward, the workshop participants identified that solving the entire problem space with a single approach will be challenging for several reasons:

* The scalability of many solutions will likely be an issue as some solutions are complex or expensive to implement.

* Collaboration between multiple parties will be required for many mechanisms to function, and the sets of parties required for different use cases might be disjoint.

* There is an unanswered question of whether or not network operators be willing to participate and allow technologies into their environment requirements in exchange for technologies that prove their clients are being good net-citizens. If so, some of these solutions might be required to exist before networks allow a certain type of increased encryption; consider the example of TLS Encrypted Client Hello being blocked by some network operators.

The breadth of the problem space itself is another complicating factor. There is a wide variety of network architectures, and each has different requirements for both data encryption and network management. Each problem space will have different encumbrances of
multiple types; for example, technical, legal, data ownership, and regulatory concerns. New network architectures might be needed to solve this problem at a larger scope, which would in turn require interoperability support from network product vendors. Education about various solutions will be required in order to ensure regulation and policy organizations can understand and thus support the deployment of developed solutions.

After new technologies and related standards are developed and deployed, unintended consequences can emerge that weren’t considered during the design of the protocol. These lead to effects in multiple directions: on one hand, exposed protocol values not intended for network management might be used by networks to differentiate traffic; on the other hand, changes to a protocol might have impact on private network deployments that break existing use cases. While making decisions on technology direction and protocol design, it is important to consider the impact on various kinds of network deployments and their unique requirements. When protocols change to make different network management functions easier or harder, the impact on various deployment models ought to be considered and documented.

4. Informative References


Interested participants were openly invited to submit position papers on the workshop topics, including Internet-Drafts, relevant academic papers, or short abstracts explaining their interest. The papers below constitute the inputs that were considered relevant for workshop attendees and that focused the discussions themselves. The program committee grouped the papers by theme as such.

A.1. Motivations and principles

Richard Barnes. Whats In It For Me? Revisiting the reasons people collaborate. [BARNES]

Iain R. Learmonth, Gurshabad Grover, Mallory Knodel. Guidelines for Performing Safe Measurement on the Internet. (Additional rationale) [KNODEL]

Qin Wu, Jun Wu, Qiufang Ma. Network Management of Encrypted Traffic: Detect it dont decrypt it. [WU]
A.2. Classification and identification of encrypted traffic

Luca Deri. nDPI Research Proposal. [DERI]


Xi Jiang, Shinan Liu, Saloua Naama, Francesco Bronzino, Paul Schmitt, Nick Feamster. Towards Designing Robust and Efficient Classifiers for Encrypted Traffic in the Modern Internet. [JIANG]

Yupeng Lei, Jun Wu, Xudong Sun, Liang Zhang, Qin Wu. Encrypted Traffic Classification Through Deep Learning. [LEI]

A.3. Ideas for collaboration and coordination between devices and networks

Michael Collins. Improving Network Monitoring Through Contracts. [COLLINS]

Paul Grubbs, Arasu Arun, Ye Zhang, Joseph Bonneau, Michael Walfish. Zero-Knowledge Middleboxes. [GRUBBS]

Mirja Kühlewind, Magnus Westerlund, Zaheduzzaman Sarker, Marcus Ihlar. Relying on Relays: The future of secure communication. [KUEHLEWIND]

Tommy Pauly, Richard Barnes. Red Rover: A collaborative approach to content filtering. [PAULY]

Michael Welzl. The Sidecar: Opting in to PEP Functions. [WELZL]

A.4. Other background material

Pedro Casas. Monitoring User-Perceived Quality in an Encrypted Internet AI to the Rescue. [CASAS]

Nalini Elkins, Mike Ackermann, Mohit P. Tahliani, Dhruv Dhody, Prof. Tommaso Pecorella. Performance Monitoring in Encrypted Networks: PDMv2. [ELKINS]

Appendix B. Workshop participants

The workshop participants were Cindy Morgan, Colin Perkins, Cullen Jennings, Deborah Brungard, Dhruv Dhody, Eric Vyncke, Georg Carle, Ivan Nardi, Jari Arkko, Jason Livingood, Jiankang Yao, Karen O’Donoghue, Keith Winstein, Lars Eggert, Laurent Vanbever, Luca Deri, Mallory Knodel, Marcus Ihlar, Matteo, Michael Ackermann, Michael Collins, Michael Richardson, Michael Welzl, Mike Ackermann, Mirja Knodel, et al.
Appendix C. Program Committee

The workshop program committee members were Wes Hardaker (IAB, USC/ISI), Mallory Knodel (IAB, Center for Democracy and Technology), Mirja Kühlewind (IAB, Ericsson), Tommy Pauly (IAB, Apple), Russ White (IAB, Juniper), Qin Wu (IAB, Huawei).

Acknowledgments

TODO acknowledge.

Authors’ Addresses

Mallory Knodel
Center for Democracy & Technology
Email: mknodel@cdt.org

Wes Hardaker
Email: ietc@hardakers.net

Tommy Pauly
Email: tpauly@apple.com
Partitioning as an Architecture for Privacy
draft-iab-privacy-partitioning-03

Abstract

This document describes the principle of privacy partitioning, which selectively spreads data and communication across multiple parties as a means to improve privacy by separating user identity from user data. This document describes emerging patterns in protocols to partition what data and metadata is revealed through protocol interactions, provides common terminology, and discusses how to analyze such models.

Discussion Venues

This note is to be removed before publishing as an RFC.

Discussion of this document takes place on the Internet Architecture Board Internet Engineering Task Force mailing list (iab@iab.org), which is archived at .

Source for this draft and an issue tracker can be found at https://github.com/intarchboard/draft-obliviousness.

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."
1. Introduction

Protocols such as TLS and IPsec provide a secure (authenticated and encrypted) channel between two endpoints over which endpoints transfer information. Encryption and authentication of data in transit are necessary to protect information from being seen or modified by parties other than the intended protocol participants. As such, this kind of security is necessary for ensuring that information transferred over these channels remains private.
However, a secure channel between two endpoints is insufficient for the privacy of the endpoints themselves. In recent years, privacy requirements have expanded beyond the need to protect data in transit between two endpoints. Some examples of this expansion include:

* A user accessing a service on a website might not consent to reveal their location, but if that service is able to observe the client’s IP address, it can learn something about the user’s location. This is problematic for privacy since the service can link user data to the user’s location.

* A user might want to be able to access content for which they are authorized, such as a news article, without needing to have which specific articles they read on their account being recorded. This is problematic for privacy since the service can link user activity to the user’s account.

* A client device that needs to upload metrics to an aggregation service might want to be able to contribute data to the system without having their specific contributions attributed to them. This is problematic for privacy since the service can link client contributions to the specific client.

The commonality in these examples is that clients want to interact with or use a service without exposing too much user-specific or identifying information to that service. In particular, separating the user-specific identity information from user-specific data is necessary for privacy. Thus, in order to protect user privacy, it is important to keep identity (who) and data (what) separate.

This document defines "privacy partitioning," sometimes also referred to as the "decoupling principle" [DECOUPLING], as the general technique used to separate the data and metadata visible to various parties in network communication, with the aim of improving user privacy. Partitioning is a spectrum and not a panacea. It is difficult to guarantee there is no link between user-specific identity and user-specific data. However, when applied properly, privacy partitioning helps ensure that user privacy violations become more technically difficult to achieve over time.

Several IETF working groups are working on protocols or systems that adhere to the principle of privacy partitioning, including Oblivious HTTP Application Intermediation (OHAI), Multiplexed Application Substrate over QUIC Encryption (MASQUE), Privacy Pass, and Privacy Preserving Measurement (PPM). This document summarizes work in those groups and describes a framework for reasoning about the resulting privacy posture of different endpoints in practice.
Privacy partitioning is particularly relevant as a tool for data minimization, which is described in Section 6.1 of [RFC6973]. [RFC6973] provides guidance for privacy considerations in Internet protocols, along with a set of questions on how to evaluate the data minimization of a protocol in Section 7.1 of [RFC6973]. Protocols that employ privacy partitioning ought to consider the questions in that section when evaluating their design, particularly with regard to how identifiers and data can be correlated by protocol participants and observers in each context that has been partitioned. Privacy partitioning can also be used as a way to separate identity providers from relying parties (see Section 6.1.4 of [RFC6973]), as in the case of Privacy Pass (see Section Section 3.3).

2. Privacy Partitioning

For the purposes of user privacy, this document focuses on user-specific information. This might include any identifying information that is specific to a user, such as their email address or IP address, or data about the user, such as their date of birth. Informally, the goal of privacy partitioning is to ensure that each party in a system beyond the user themselves only have access to one type of user-specific information.

This is a simple application of the principle of least privilege, wherein every party in a system only has access to the minimum amount of information needed to fulfill their function. Privacy partitioning advocates for this minimization by ensuring that protocols, applications, and systems only reveal user-specific information to parties that need access to the information for their intended purpose.

Put simply, privacy partitioning aims to separate _who_ someone is from _what_ they do. In the rest of this section, we describe how privacy partitioning can be used to achieve this goal.

2.1. Privacy Contexts

Each piece of user-specific information exists within some context, where a context is abstractly defined as a set of data, metadata, and the entities that share access to that information. In order to prevent the correlation of user-specific information across contexts, partitions need to ensure that any single entity (other than the client itself) does not participate in more than one context where the information is visible.

[RFC6973] discusses the importance of identifiers in reducing correlation as a way of improving privacy:
Correlation is the combination of various pieces of information related to an individual or that obtain that characteristic when combined... Correlation is closely related to identification.

Internet protocols can facilitate correlation by allowing individuals’ activities to be tracked and combined over time.

Pseudonymity is strengthened when less personal data can be linked to the pseudonym; when the same pseudonym is used less often and across fewer contexts; and when independently chosen pseudonyms are more frequently used for new actions (making them, from an observer’s or attacker’s perspective, unlinkable).

Context separation is foundational to privacy partitioning and reducing correlation. As an example, consider an unencrypted HTTP session over TCP, wherein the context includes both the content of the transaction as well as any metadata from the transport and IP headers; and the participants include the client, routers, other network middleboxes, intermediaries, and server.

```
+-------------------------------------------------------------------+
| Context A                                                         |
|  +--------+                +-----------+              +--------+  |
|  |        +------HTTP------+           +--------------+        |  |
|  | Client |                | Middlebox |              | Server |  |
|  |        +------TCP-------+           +--------------+        |  |
|  +--------+      flow      +-----------+              +--------+  |
+-------------------------------------------------------------------+
```

Figure 1: Diagram of a basic unencrypted client-to-server connection with middleboxes

Adding TLS encryption to the HTTP session is a simple partitioning technique that splits the previous context into two separate contexts: the content of the transaction is now only visible to the client, TLS-terminating intermediaries, and server; while the metadata in transport and IP headers remain in the original context. In this scenario, without any further partitioning, the entities that participate in both contexts can allow the data in both contexts to be correlated.
Another way to create a partition is to simply use separate connections. For example, to split two separate HTTP requests from one another, a client could issue the requests on separate TCP connections, each on a different network, and at different times; and avoid including obvious identifiers like HTTP cookies across the requests.
Using separate connections to create separate contexts can reduce or eliminate the ability of specific parties to correlate activity across contexts. However, any identifier at any layer that is common across different contexts can be used as a way to correlate activity. Beyond IP addresses, many other factors can be used together to create a fingerprint of a specific device (such as MAC addresses, device properties, software properties and behavior, application state, etc).

2.2. Context Separation

In order to define and analyze how various partitioning techniques work, the boundaries of what is being partitioned need to be established. This is the role of context separation. In particular, in order to prevent the correlation of user-specific information across contexts, partitions need to ensure that any single entity (other than the client itself) does not participate in contexts where both identities are visible.

Context separation can be achieved in different ways, for example, over time, across network paths, based on (en)coding, etc. The privacy-oriented protocols described in this document generally involve more complex partitioning, but the techniques to partition communication contexts still employ the same techniques:

1. Encryption allows partitioning of contexts within a given network path.

2. Using separate connections across time or space allows partitioning of contexts for different application transactions.

These techniques are frequently used in conjunction for context separation. For example, encrypting an HTTP exchange might prevent a network middlebox that sees a client IP address from seeing the user account identity, but it doesn’t prevent the TLS-terminating server from observing both identities and correlating them. As such, preventing correlation requires separating contexts, such as by using proxying to conceal a client’s IP address that would otherwise be used as an identifier.

2.3. Approaches to Partitioning

While all of the partitioning protocols described in this document create separate contexts using encryption and/or connection separation, each one has a unique approach that results in different sets of contexts. Since many of these protocols are new, it is yet to be seen how each approach will be used at scale across the Internet, and what new models will emerge in the future.
There are multiple factors that lead to a diversity in approaches to partitioning, including:

* Adding privacy partitioning to existing protocol ecosystems places requirements and constraints on how contexts are constructed. CONNECT-style proxying is intended to work with servers that are unaware of privacy contexts, requiring more intermediaries to provide strong separation guarantees. Oblivious HTTP, on the other hand, assumes servers that cooperate with context separation, and thus reduces the overall number of elements in the solution.

* Whether or not information exchange needs to happen bidirectionally in an interactive fashion determines how contexts can be separated. Some use cases, like metrics collection for PPM, can occur with information flowing only from clients to servers, and can function even when clients are no longer connected. Privacy Pass is an example of a case that can be either interactive or not, depending on whether tokens can be cached and reused. CONNECT-style proxying and Oblivious HTTP often requires bidirectional and interactive communication.

* The degree to which contexts need to be partitioned depends in part on the client’s threat models and level of trust in various protocol participants. For example, in Oblivious HTTP, clients allow relays to learn that clients are accessing a particular application-specific gateway. If clients do not trust relays with this information, they can instead use a multi-hop CONNECT-style proxy approach wherein no single party learns whether specific clients are accessing a specific application. This is the default trust model for systems like Tor, where multiple hops are used to drive down the probability of privacy violations due to collusion or related attacks.

3. A Survey of Protocols using Partitioning

The following section discusses current on-going work in the IETF that is applying privacy partitioning.

3.1. CONNECT Proxying and MASQUE

HTTP forward proxies, when using encryption on the connection between the client and the proxy, provide privacy partitioning by separating a connection into multiple segments. When connections to targets via the proxy themselves are encrypted, the proxy cannot see the end-to-end content. HTTP has historically supported forward proxying for TCP-like streams via the CONNECT method. More recently, the Multiplexed Application Substrate over QUIC Encryption (MASQUE)
working group has developed protocols to similarly proxy UDP [CONNECT-UDP] and IP packets [CONNECT-IP] based on tunneling.

In a single-proxy setup, there is a tunnel connection between the client and proxy and an end-to-end connection that is tunnelled between the client and target. This setup, as shown in the figure below, partitions communication into:

* a Client-to-Target encrypted context, which contains the end-to-end content within the TLS session to the target, such as HTTP content;

* a Client-to-Target proxied context, which is the end-to-end data to the target that is also visible to the proxy, such as a TLS session;

* a Client-to-Proxy context, which contains the transport metadata between the client and the target, and the request to the proxy to open a connection to the target;

* and a Proxy-to-Target context, which for TCP and UDP proxying contains any packet header information that is added or modified by the proxy, e.g., the IP and TCP/UDP headers.
Using two (or more) proxies provides better privacy partitioning. In particular, with two proxies, each proxy sees the Client metadata, but not the Target; the Target, but not the Client metadata; or neither.

In addition to the contexts described above for the single proxy case, the two-hop proxy case shown in the figure below changes the contexts in several ways:

* the Client-to-Target proxied context only includes the second proxy (referred to here as "Proxy B");
* a new Client-to-Proxy B context is added, which is the TLS session from the client to Proxy B that is also visible to the first proxy (referred to here as "Proxy A");

* the contexts that see transport data only (TCP or UDP over IP) are separated out into three separate contexts, a Client-to-Proxy A context, a Proxy A-to-Proxy B context, and a Proxy B-to-Target context.
Forward proxying, such as the protocols developed in MASQUE, uses both encryption (via TLS) and separation of connections (via proxy hops that see only the next hop) to achieve privacy partitioning.

### 3.2. Oblivious HTTP and DNS

Oblivious HTTP [OHTTP], developed in the Oblivious HTTP Application Intermediation (OHAI) working group, adds per-message encryption to HTTP exchanges through a relay system. Clients send requests through an Oblivious Relay, which cannot read message contents, to an Oblivious Gateway, which can decrypt the messages but cannot communicate directly with the client or observe client metadata like IP address. Oblivious HTTP relies on Hybrid Public Key Encryption [HPKE] to perform encryption.

Oblivious HTTP uses both encryption and separation of connections to achieve privacy partitioning.

* End-to-end messages are encrypted between the Client and Gateway. The content of these inner messages are visible to the Client, Gateway, and Target. This is the Client-to-Target context.

* The encrypted messages exchanged between the Client and Gateway are visible to the Relay, but the Relay cannot decrypt the messages. This is the Client-to-Gateway context.

* The transport (such as TCP and TLS) connections between the Client, Relay, and Gateway form two separate contexts: a Client-to-Relay context and a Relay-to-Gateway context. It is important to note that the Relay-to-Gateway connection can be a single connection, even if the Relay has many separate Clients. This provides better anonymity by making the pseudonym presented by the Relay to be shared across many Clients.
Oblivious DNS over HTTPS [ODOH] applies the same principle as Oblivious HTTP, but operates on DNS messages only. As a precursor to the more generalized Oblivious HTTP, it relies on the same HPKE cryptographic primitives, and can be analyzed in the same way.
3.3. Privacy Pass

Privacy Pass is an architecture [PRIVACYPASS] and a set of protocols being developed in the Privacy Pass working group that allows clients to present proof of verification in an anonymous and unlinkable fashion, via tokens. These tokens originally were designed as a way to prove that a client had solved a CAPTCHA, but can be applied to other types of user or device attestation checks as well. In Privacy Pass, clients interact with an attester and issuer for the purposes of issuing a token, and clients then interact with an origin server to redeem said token.

In Privacy Pass, privacy partitioning is achieved with cryptographic protection (in the form of blind signature protocols or similar) and separation of connections across two contexts: a "redemption context" between clients and origins (servers that request and receive tokens), and an "issuance context" between clients, attestation servers, and token issuance servers. The cryptographic protection ensures that information revealed during the issuance context is separated from information revealed during the redemption context.

![Diagram of contexts in Privacy Pass](image-url)

Since the redemption context and issuance context are separate connections that involve separate entities, they can also be further decoupled by running those parts of the protocols at different times. Clients can fetch tokens through the issuance context early, and cache the tokens to later use in redemption contexts. This can aid in partitioning identifiers and data.
[PRIVACYPASS] describes different deployment models for which entities operate origins, attesters, and issuers; in some models, they are all separate entities, but in others, they can be operated by the same entity. The model impacts the effectiveness of partitioning, and some models (such as when all three are operated by the same entity) only provide effective partitioning when the timing of connections on the two contexts are not correlated, and when the client uses different identifiers (such as different IP addresses) on each context.

3.4. Privacy Preserving Measurement

The Privacy Preserving Measurement (PPM) working group is chartered to develop protocols and systems that help a data aggregation or collection server (or multiple, non-colluding servers) compute aggregate values without learning the value of any one client’s individual measurement. Distributed Aggregation Protocol (DAP) is the primary working item of the group.

At a high level, DAP uses a combination of cryptographic protection (in the form of secret sharing amongst non-colluding servers) to establish two contexts: an "upload context" between clients and non-colluding aggregation servers (in which the servers are separated into "Helper" and "Leader" roles) wherein aggregation servers possibly learn client identity but nothing about their individual measurement reports, and a "collect context" wherein a collector learns aggregate measurement results and nothing about individual client data.

![Diagram of contexts in DAP](image)

Figure 8: Diagram of contexts in DAP

4. Applying Privacy Partitioning

Applying privacy partitioning to an existing or new system or protocol requires the following steps:
1. Identify the types of information used or exposed in a system or protocol, some of which can be used to identify a user or correlate to other contexts.

2. Partition data to minimize the amount of user-identifying or correlatable information in any given context to only include what is necessary for that context, and prevent the sharing of data across contexts wherever possible.

The most impactful types of information to partition are (a) user-identifying information (including account names or IP addresses) that can be linked and (b) non-user-identifying information (including content a user generates or accesses), which can be often sensitive when combined with a user identity.

In this section, we discuss considerations for partitioning these types of information.

4.1. User-Identifying Information

User data can itself be user-identifying, in which case it should be treated as an identifier. For example, Oblivious DoH and Oblivious HTTP partition the client IP address and client request data into separate contexts, thereby ensuring that no entity beyond the client can observe both. Collusion across contexts could reverse this partitioning, but can also promote non-user-identifying information to user-identifying. For example, in CONNECT proxy systems that use QUIC, the QUIC connection ID is inherently non-user-identifying since it is generated randomly ([QUIC], Section 5.1). However, if combined with another context that has user-identifying information such as the client IP address, the QUIC connection ID can become user-identifying information.

Some information is innate to client user-agents, including details of implementation of protocols in hardware and software, and network location. This information can be used to construct user-identifying information, which is a process sometimes referred to as fingerprinting. Depending on the application and system constraints, users may not be able to prevent fingerprinting in privacy contexts. As a result, fingerprinting information, when combined with non-user-identifying user data, could promote user data to user-identifying information.
4.2. Incorrect or Incomplete Partitioning

Privacy partitioning can be applied incorrectly or incompletely. Contexts may contain more user-identifying information than desired, or some information in a context may be more user-identifying than intended. Moreover, splitting user-identifying information over multiple contexts has to be done with care, as creating more contexts can increase the number of entities that need to be trusted to not collude. Nevertheless, partitions can help improve the client’s privacy posture when applied carefully.

Evaluating and qualifying the resulting privacy of a system or protocol that applies privacy partitioning depends on the contexts that exist and the types of user-identifying information in each context. Such evaluation is helpful for identifying ways in which systems or protocols can improve their privacy posture. For example, consider DNS-over-HTTPS [DOH], which produces a single context which contains both the client IP address and client query. One application of privacy partitioning results in ODoH, which produces two contexts, one with the client IP address and the other with the client query.

4.3. Identifying Information for Partitioning

Recognizing potential applications of privacy partitioning requires identifying the contexts in use, the information exposed in a context, and the intent of information exposed in a context. Unfortunately, determining what information to include in a given context is a nontrivial task. In principle, the information contained in a context should be fit for purpose. As such, new systems or protocols developed should aim to ensure that all information exposed in a context serves as few purposes as possible. Designing with this principle from the start helps mitigate issues that arise if users of the system or protocol inadvertently ossify on the information available in contexts. Legacy systems that have ossified on information available in contexts may be difficult to change in practice. As an example, many existing anti-abuse systems depend on some notion of client identity such as client IP address, coupled with client data, to provide value. Partitioning contexts in these systems such that they no longer see the client’s identity requires new solutions to the anti-abuse problem.

5. Limits of Privacy Partitioning

Privacy Partitioning aims to increase user privacy, though as stated, it is not a panacea. The privacy properties depend on numerous factors, including, though not limited to:
* Non-collusion across contexts; and

* The type of information exposed in each context.

We elaborate on each below.

5.1. Violations by Collusion

Privacy partitions ensure that only the client, i.e., the entity which is responsible for partitioning, can link all user-specific information together up to collusion. No other entity individually knows how to link all the user-specific information as long as they do not collude with each other across contexts. This is why non-collusion is a fundamental requirement for privacy partitioning to offer meaningful privacy for end-users. In particular, the trust relationships that users have with different parties affect the resulting impact on the user’s privacy.

As an example, consider OHTTP, wherein the Oblivious Relay knows the Client identity but not the Client data, and the Oblivious Gateway knows the Client data but not the Client identity. If the Oblivious Relay and Gateway collude, they can link Client identity and data together for each request and response transaction by simply observing requests in transit.

It is not currently possible to guarantee with technical protocol measures that two entities are not colluding. Even if two entities do not collude directly, if both entities reveal information to other parties, it will not be possible to guarantee that the information won’t be combined. However, there are some mitigations that can be applied to reduce the risk of collusion happening in practice:

* Policy and contractual agreements between entities involved in partitioning to disallow logging or sharing of data, along with auditing to validate that the policies are being followed. For cases where logging is required (such as for service operation), such logged data should be minimized and anonymized to prevent it from being useful for collusion.

* Protocol requirements to make collusion or data sharing more difficult.

* Adding more partitions and contexts, to make it increasingly difficult to collude with enough parties to recover identities.
5.2. Violations by Insufficient Partitioning

It is possible to define contexts that contain more than one type of user-specific information, despite efforts to do otherwise. As an example, consider OHTTP used for the purposes of hiding client-identifying information for a browser telemetry system. It is entirely possible for reports in such a telemetry system to contain both client-specific telemetry data, such as information about their specific browser instance, as well as client-identifying information, such as the client’s location or IP address. Even though OHTTP separates the client IP address from the server via a relay, the server still learns this directly from the client.

Other relevant examples of insufficient partitioning include TLS Encrypted Client Hello (ECH) [I-D.ietf-tls-esni] and VPNs. ECH uses cryptographic protection ( encryption) to hide information from unauthorized parties, but both clients and servers (two entities) can link user-specific data to user-specific identity (IP address). Similarly, while VPNs hide identity from end servers, the VPN server has still can see the identity of both the client and server. Applying privacy partitioning would advocate for at least two additional entities to avoid revealing both (identity (who) and user actions (what)) from each involved party.

While straightforward violations of user privacy like this may seem straightforward to mitigate, it remains an open problem to determine whether a certain set of information reveals "too much" about a specific user. There is ample evidence of data being assumed "private" or "anonymous" but, in hindsight, winds up revealing too much information such that it allows one to link back to individual clients; see [DataSetReconstruction] and [CensusReconstruction] for more examples of this in the real world.

Beyond the information that is intentionally revealed by applying privacy partitioning, it is also possible for the information to be unintentionally revealed through side channels. For example, in the two-hop proxy arrangement described in Section 3.1, Proxy A sees and proxies TLS data between the client and Proxy B. While it does not directly learn information that Proxy B sees, it does learn information through metadata, such as the timing and size of encrypted data being proxied. Traffic analysis could be exploited to learn more information from such metadata, including, in some cases, application data that Proxy A was never meant to see. Although privacy partitioning does not obviate such attacks, it does increase the cost necessary to carry them out in practice. See Section 7 for more discussion on this topic.
6. Partitioning Impacts

Applying privacy partitioning to communication protocols leads to a substantial change in communication patterns. For example, instead of sending traffic directly to a service, essentially all user traffic is routed through a set of intermediaries, possibly adding more end-to-end round trips in the process (depending on the system and protocol). This has a number of practical implications, described below.

1. Service operational or management challenges. Information that is traditionally passively observed in the network or metadata that has been unintentionally revealed to the service provider cannot be used anymore for e.g., existing security procedures such as application rate limiting or DDoS mitigation. However, network management techniques deployed at present often rely on information that is exposed by most traffic but without any guarantees that the information is accurate.

Privacy partitioning provides an opportunity for improvements in these management techniques with opportunities to actively exchange information with each entity in a privacy-preserving way and requesting exactly the information needed for a specific task or function rather than relying on the assumption that are derived from a limited set of unintentionally revealed information which cannot be guaranteed to be present and may disappear at any time in future.

2. Varying performance effects and costs. Depending on how context separation is done, privacy partitioning may affect application performance. As an example, Privacy Pass introduces an entire end-to-end round trip to issue a token before it can be redeemed, thereby decreasing performance. In contrast, while systems like CONNECT proxying may seem like they would regress performance, oftentimes the highly optimized nature of proxy-to-proxy paths leads to improved performance.

Performance may also push back against the desire to apply privacy partitioning. For example, HTTPS connection reuse [HTTP2], Section 9.1.1 allows clients to use an existing HTTPS session created for one origin to interact with different origins (provided the original origin is authoritative for these alternative origins). Reusing connections saves the cost of connection establishment, but means that the server can now link the client’s activity with these two or more origins together. Applying privacy partitioning would prevent this, while typically at the cost of less performance.
In general, while performance and privacy tradeoffs are often cast as a zero-sum game, in practice this is often not the case. The relationship between privacy and performance varies depending on a number of related factors, such as application characteristics, network path properties, and so on.

3. Increased attack surface. Even in the event that information is adequately partitioned across non-colluding parties, the resulting effects on the end-user may not always be positive. For example, using OHTTP as a basis for illustration, consider a hypothetical scenario where the Oblivious Gateway has an implementation flaw that causes all of its decrypt requests to be inappropriately logged to a public or otherwise compromised location. Moreover, assume that the Target Resource for which these requests are destined does not have such an implementation flaw. Applications which use OHTTP with this flawed Oblivious Gateway to interact with the Target Resource risk their user request information being made public, albeit in a way that is decoupled from user identifying information, whereas applications that do not use OHTTP to interact with the Target Resource do not risk this type of disclosure.

4. Centralization. Depending on the protocol and system, as well as the desired privacy properties, the use of partitioning may inherently force centralization to a select set of trusted participants. As an example, the impact of OHTTP on end-user privacy generally increases proportionally to the number of users that exist behind a given Oblivious Relay. That is, the probability of an Oblivious Gateway determining the client associated with a request forwarded through an Oblivious Relay decreases as the number of possible clients behind the Oblivious Relay increases. This tradeoff encourages the centralization of the Oblivious Relays.

7. Security Considerations

Section 5 discusses some of the limitations of privacy partitioning in practice. In general, privacy is best viewed as a spectrum and not a binary state (private or not). Applied correctly, partitioning helps improve an end-user’s privacy posture, thereby making violations harder to do via technical, social, or policy means. For example, side channels such as traffic analysis [I-D.irtf-pearg-website-fingerprinting] or timing analysis are still possible and can allow an unauthorized entity to learn information about a context they are not a participant of. Proposed mitigations for these types of attacks, e.g., padding application traffic or generating fake traffic, can be very expensive and are therefore not typically applied in practice. Nevertheless, privacy partitioning
moves the threat vector from one that has direct access to user-specific information to one which requires more effort, e.g., computational resources, to violate end-user privacy.

8. IANA Considerations

This document has no IANA actions.

9. Informative References


Acknowledgments

TODO acknowledge.

Authors’ Addresses

Mirja Kühlewind
Ericsson Research
Email: mirja.kuehlewind@ericsson.com

Tommy Pauly
Apple
Email: tpauly@apple.com

Christopher A. Wood
Cloudflare
Email: caw@heapingbits.net
Abstract

Internet communications and applications have both environmental costs and benefits. The IAB ran an online workshop in December 2022 on exploring and understanding these impacts.

The role of the workshop was to discuss the impacts, discuss the evolving needs from industry, and to identify areas for improvements and future work. A key goal of the workshop was to call further attention to the topic and to bring together a diverse stakeholder community to discuss these issues.

This report summarises the workshop inputs and discussions.

About This Document

This note is to be removed before publishing as an RFC.

Status information for this document may be found at https://datatracker.ietf.org/doc/draft-arkko-iab-ws-environmental-impacts-report/.

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/.
1. Introduction

The IAB ran an online workshop in December 2022 on exploring and understanding the environmental impacts of the Internet.
The background for the workshop was that Internet communications and applications have both environmental costs and benefits. In the positive direction, they can reduce the environmental impact of our society, for instance, by allowing virtual interaction to replace physical travel. Of course, the Internet can equally well act as an enabler for increasing physical goods consumption, for instance, through easing commerce.

Beyond the effects associated with its use, Internet applications do not come for free either. The Internet runs on systems that require energy and raw materials to manufacture and operate. While the environmental benefits of the Internet may certainly outweigh this use of resources in many cases, it is incumbent on the Internet industry to ensure that this use of resources is minimized and optimized. In many cases, this is already an economic necessity due to operational costs. And because many consumers, businesses, and civil societies care deeply about the environmental impact of the services and technologies they use, there is also a clear demand for providing Internet services with minimal environmental impact.

The role of the workshop was to discuss the Internet’s environmental impact, discuss the evolving needs from industry, and to identify areas for improvements and future work. A key goal of the workshop was to call further attention to the topic and to bring together a diverse stakeholder community to discuss these issues. This report summarises the workshop inputs and discussions.

The workshop drew many position paper submissions. Of these, 26 were accepted and published to stimulate discussion. There were active discussions both in the meeting and on the workshop mailing list with altogether 73 participants.

Perhaps the main overriding observation is how much there is interest and urgency on this topic, among engineers, researchers, and businesses.

The workshop discussions and conclusions are covered in Section 3. The position papers, and links to recordings of workshop sessions, can be found at https://www.iab.org/activities/workshops/e-impact/. Presentations held during the discussions can be found from the IETF Datatracker at https://datatracker.ietf.org/group/eimpactws/meetings/.

The discussion at the IETF will continue after the workshop, both around specific proposals as well as general discussion on a new mailing list, the e-impact list (e-impact@ietf.org). You can subscribe to this list at https://www.ietf.org/mailman/listinfo/e-impact.
Some improvements addressing specific situations are being discussed at the IETF, such as the Time Variant Routing (TVR) proposal that can help optimize connectivity with systems that are periodically on or reachable (such as satellites). We expect more proposals in the future.

2. Scope

Environmental impact assessment and improvements are broad topics, ranging from technical questions to economics, business decisions, and policies.

The technical, standards, and research communities can help ensure that we have a sufficient understanding of the environmental impact of the Internet and its applications. They can also help to design the right tools to continue to build and improve all aspects of the Internet, such as addressing new functional needs, easing of operations, improving performance and/or efficiency, or reducing environmental impacts in other ways.

The workshop was expected to discuss:

* The direct environmental impacts of the Internet, including but not limited to energy usage by Internet systems themselves (the network equipment along with the associated power and cooling infrastructure), energy usage of the relevant end-user devices, resources needed for manufacturing the associated devices, or the environmental impacts throughout the life-cycle of Internet systems. This included discussion about the breakdown of those impacts across different system components and operations, and predictions about the potential future trends for these impacts based on changed usage patterns and emerging technologies.

* Discussion of the indirect environmental impacts of the Internet, i.e., its effects on society through enabling communications, virtual services, or global commerce.

* Sharing information about relevant measurement metrics and data, and identify the need for additional metric or measurements.

* Discussing about the need for improvements or associated new functionality.

* Sharing information about the societal, business, and regulatory situation, to help identify areas of opportunity.

* Identifying areas where further technical work would be most impactful.
* Discussing specific improvement proposals.

* Discussion of past work in the IETF, IRTF, and IAB in this area and the status of such work.

* Discussion of observed user behaviours as they relate to environmental impacts.

We expected that the workshop discussions connect analysis of the issues (e.g., scale of energy consumption or carbon footprint) to industry needs (e.g., deployment opportunities) and solutions.

Business and societal policy questions were in scope only insofar as they informed the workshop participants about the context we are in, but what those policies should be was not for the workshop to decide or even extensively discuss. The scope excluded also how the technical community works and meets, such as the question of in-person or hybrid meetings (although it should be noted that the workshop itself was run as an on-line meeting).

2.1. Practical Arrangements

The IAB discussed a potential workshop in this area during its May 2022 retreat. A call for position papers went out in August 2022. Position papers were to be submitted by end of October, a deadline which was later extended by one week.

As noted, the workshop itself was run as an on-line meeting, with four half-day long sessions complemented by email discussions and the position papers submitted by the participants.

All in all, 73 people participated in at least one session in the workshop. Participation was by invitation only, based on the position paper submissions.

Every submission was read by at least three members of the program committee, and acceptance decisions were communicated back to the authors. Review comments were provided for authors for information, and some of the papers were revised before the workshop.

The program committee decided that due to interest and differing areas of expertise, all co-authors were to be invited, and most of them did attend. The program committee also invited a handful of additional participants, where they were seen as providing valuable input. Similarly, as is traditional in IAB workshops, the program committee members and members of the IAB and IESG were offered an opportunity to participate even in cases where they did not submit a position paper.
The IETF secretariat and communications staff provided practical support during the process, sending announcements, maintaining the workshop web page with position papers, setting up mailing lists, tracking submissions, helping with blog article submissions, and so on.

3. Workshop Topics and Discussion

The meeting part of the workshop was divided into four sessions:

* The first session was about the big picture and relationships between different aspects of sustainability (see Section 3.1).
* The second session focused on what we know and do not know, and how we can measure environmental impacts (see Section 3.2).
* The third session was about potential improvements (see Section 3.3).
* The final fourth session was about conclusions and next steps (see Section 3.4).

3.1. The Big Picture

This session was about the big picture and how the Internet influences the rest of the society. We also spoke about the goals of the workshop.

The session began with a discussion about what is overall involved in this topic. We also looked at how the IETF has approached this topic in the past.

The discussions also expressed the urgency of action and the importance of continuous improvement: an incremental change every year is needed for larger savings at the end of the decade. We continued to talk about the need to recognize how climate changes impact different communities in the world, often unfairly. Finally, we focused on the need to be aware of carbon footprint rather than pure energy consumption - carbon intensity of energy sources varies.
The starting observation from this session was that the issue is much bigger than Internet technology alone. The issue influences all parts of society, and even matters such as (in)equality, externalized costs, and justice. Another key observation was that improvements come in many forms; there is no silver bullet. The opportunity to bring together people with different backgrounds helped us see how we approach the topic from different angles - none of them wrong, but also none of the sole angle to focus on either. Only the combined effects of complementary efforts can provide the required level of changes.

Some of the useful tools for approaching the issue included of course technical solutions, but also solidarity, aiming for sufficiency, and awareness. It is important to not stand still waiting for the perfect solution. Renewable energy and carbon awareness were seen as a part of the solution, but not, however, sufficient by themselves.

As an example demonstration of the diversity of angles and improvements relating to environmental issues, the figure below classifies the areas that workshop position papers fell on:

```
+----- Actors & organizations
|      +----- Avoidance
+----- Benefits to other fields
|      +----- User behaviour
+----- Society, awareness, &
|    justice
|    +----- Implementation

Workshop ++ Improvements ------------------+
|    +----- Understanding &
|    Measurements
|      +----- Dataplane
|      Protocols +----- Routing
|         +----- Edge cloud
|         +----- Mobile
|         +----- Metrics
|         +----- Other
```

Figure 1: Position paper submission topics

Some of the goals for the IETF should include:
* Connecting the IETF with others. Given that the issue is broad, it is difficult for one standards organisation alone to make a significant impact, or even have the full picture. Working in collaboration with others is necessary. And understanding the situation beyond technology will be needed.

* Continuous improvement. It is important that the IETF (among others) sets itself on a continuous improvement cycle. No single improvement will change the overall situation sufficiently, but over a longer period of time, even smaller changes every year will result in larger improvements.

* Finding the right targets for improvements in the Internet. These should perhaps not be solely defined by larger speeds or bigger capacity, but rather increased usefulness to society and declining emissions from the information and communications technology (ICT) sector.

* Specifying what research needs to be done, i.e., where additional knowledge would allow us to find better improvements. For instance, not enough is known about environmental impacts beyond energy, such as natural resources used for manufacturing, or the use of water. Carbon-awareness and measurements across domains is also poorly understood today. And business model impacts -- such as the role of advertising on Internet’s carbon footprint -- deserve more study.

3.2. Understanding the Impacts

The second session focused on what we know and do not know, and how we can measure environmental impacts.

The initial presentation focused on narrowing down the lower and upper limits of the energy use of the Internet and putting some common but erroneous claims into context. There was also discussion regarding the energy consumption of the ICT sector and how it compares to some other selected industries such as aviation.

Dwelling deeper into the energy consumption and the carbon footprint of the ICT sector there was discussion regarding how the impact was split amongst the networks, data centres and user devices (with the user devices appearing to contribute to the largest fraction of the impact). Also, while lot of the energy consumption related studies and discussions have been focused on data centers, some studies suggested that data center energy usage is still a small fraction of energy use as compared to residential and commercial buildings.
There were also further discussions both during the presentations and in the hallway chats regarding the press and media coverage of the potential environment technologies. The overall sense of the participants seemed to be that there was a lot of sensational headlines, but they were not really backed by measurements done by the industry and academia, and were fraught with errors. Some of these media reports were off by quite a bit, sometimes even by an order of magnitude (e.g., confusing MBps vs Mbps in calculations). The potential harm is having widely circulating misinformation was noted; it can hinder realistic efforts to reduce carbon emissions.

In the rest of the session we looked at both additional data collected from the operators as well as factors that - depending on circumstances - may drive energy consumption. These include for instance peak capacity and energy proportionality.

If energy consumption is little affected by offered load, the ratio of peak capacity to typical usage becomes a critical factor in energy consumption. On the other hand, systems with energy proportionality scale their resource and energy consumption more dynamically based on offered load. The lack of energy proportionality in many parts of the network infrastructure was noted, along with the potential gains if it can be improved.

There were also observations that showed that the energy consumption grew as a step function when the peak capacity was reached (even instantaneously) and additional capacity was built up by performing network upgrades to handle these new peaks. This resulted in a overall higher baseline energy consumption even when the average demand did not change that much. Thus, the ability to shift load to reduce peak demand was highlighted as a potential way to delay increases in consumption when energy proportionality is lacking.

3.3. Improvements

The third session was about potential improvements.

As noted earlier, there are many different types of improvements. In the discussion we focused mostly on protocol aspects, and looked at metrics, telemetry, routing, multicast, and data encoding formats.

The initial two presentations focused on metrics and telemetry with the premise that visibility is a very important first step (paraphrasing Peter Drucker’s mantra of "You cannot improve what you don’t measure"). There was a discussion of the scopes of emissions and it seemed that from a networking vendor perspective, while directly controlled emissions and emissions from purchased energy are easily measurable, emissions from across the entire value chain can
be much larger. Thus it seemed important that the networking vendors had to put in effort into helping their customers measure and mitigate their environmental impact as well. The need for standardized metrics was very clear as it helps avoid proprietary, redundant and even contradictory metrics across vendors.

The initial and the near-term focus was related to metrics and techniques related to energy consumption of the networking devices themselves while the longer term focus can go into topics much further removed from the IETF such as packaging, circular design in order to form a more holistic picture. The overall feeling was that the topic of metrics, telemetry, and management are quite specific and could be targets to be worked on in the IETF in the near term.

The next part of the discussion highlighted the need to understand the trade-offs involved in changing forwarding decisions - such as increased jitter and stretch. Jitter is about delay fluctuation between packets in a stream [RFC4689]. Stretch is defined as the difference between the absolute shortest path traffic could take through the network and the path the traffic actually takes [RFC7980]. Impacts on jitter and stretch point to the need for careful design and analysis of improvements from a system perspective, to ensure that the intended effect is indeed reached across the entire system, and is not only a local optimum.

We also talked about the potentially significant impact, provided the network exhibits energy proportionality, of using efficient binary formats instead of textual representations when carrying data in protocols. This is something that can be relatively easily adopted in new protocols as they are developed. Indeed, some recently finished protocols such as HTTP/2 have already chosen to use this technique [RFC7540]. General-purpose binary formats such as Concise Binary Object Representation (CBOR) [RFC8949] are also available for use.

There were also some interesting discussions regarding the use of multicast and whether it would help or hurt on the energy efficiency of communications. There were some studies and simulations that showed the potential gains to be had but they were to be balanced against some of the well known barriers to deployment of multicast. We also heard from a leading Content Delivery Network (CDN) operator regarding their views on multicast and how it relates to media usage and consumption models. The hallway conversations also talked about the potential negative effects of multicast in wireless and constrained networks. Overall the conclusion was that the use of multicast can potentially provide some savings but only in some specific scenarios.
For all improvements, the importance of metrics was frequently highlighted to ensure changes lead to a meaningful reduction in overall system carbon footprint.

3.4. Next Steps

The final fourth session was about conclusions and next steps. This section highlights some of these conclusions.

3.4.1. Overall Strategy

While only a few things are easy, the road ahead for making improvements seems clear: we need to continue to improve our understanding of the environmental impact, and have a continuous cycle of improvements that lead not just to better energy efficiency but to reduced overall carbon emissions. The IETF can play an important part in this process, but of course there are other aspects beyond protocols.

On understanding our environmental impact the first step is better awareness of sustainability issues in general, which helps us understand better where our issues are. The second step is willingness to understand in detail what the causes and relationships are within our issues. What parts, components, or behaviours in the network cause what kinds of impacts? An overall drive in the society to report and improve environmental impacts can be helpful in creating a willingness to get to this information.

On establishing a continuous cycle of improvements, the ability to understand where we are, making improvements, and then seeing the impact of those improvements is of course central. But obviously a key question is what are the potential improvements, and how can we accelerate them? It should be noted that quick, large changes are not likely. But a continuous stream of smaller changes can create a large impact over a longer period of time.

One of the key realizations from this workshop was that the problem to be solved is very large, complex and that there is no single solution that fixes everything. There are some solutions that could help in the near term and others that would only show benefits over longer periods, but they are both necessary.

One further challenge is that due to the size and complexity of the problem, it was very likely that there might be varying opinions on what KPIs need to be measured and improved.
3.4.2. Improvements

In looking at potential improvements, it is essential that any associated tradeoffs can be understood (note that not all improvements do indeed entail a tradeoff).

Importantly, the role of the Internet in improving other areas of society must not be diminished. Understanding the costs and benefits requires taking a holistic view of energy consumption, focusing not just on the carbon footprint of the Internet but of the broader systems in which it is used. For instance, discussion in session three revealed how some changes might impact latency and jitter. Given that these characteristics are an important factor how virtual meetings are perceived by potential participants, it is important that the performance of networks satisfies these participants at a level where there’s willingness to use them over other potentially more environmentally harmful methods, such as travel. Focusing solely on the carbon footprint of the Internet, or solely on the carbon footprint of travel, risks missing the bigger picture potential savings.

Note that while virtual meetings are a common example, it is important to consider different use cases, some of which may not be as obvious to us human users as meetings are. Improvements may bring different or even larger impacts in other situations, e.g., Internet connected electronics might benefit from different characteristics than human users, e.g., with regards to support for intermittent connectivity.

The relationships between different system components and the impact of various detailed design choices in networks is not always apparent. A local change in one node may have an impact in other nodes. When considering environmental sustainability, in most cases the overall system impact is what counts more than local impacts. Of course, other factors, such as device battery life and availability of power may result in other preferences, such as optimising for low power usage of end-user devices, even at the cost of increases elsewhere.

In terms of useful tools for building improvements, the following were highlighted in discussions:

* Measures beyond protocol design, such as implementations or renewable energy use. Not everything is about protocols.
* Metrics, measurements, and data are very beneficial. Carbon-aware metrics would in particular be very useful. All additional information makes us more aware of what the environmental impacts are, but also enables optimization, AI-based adjustments, or carbon-directed computing and networking tools, and so on.

* It would be beneficial to be able to provide various systems a more dynamic ability to slow down and sleep. Awareness of energy availability and type would also allow us to employ time and place shifting for reducing carbon impacts.

* When we design systems, paying attention to the used data formats may pay off significantly, as argued in [Moran].

* Possibly there's a new opportunity for deploying multicast as well [Navarre].

* Designing systems for energy constrained situations may actually make the resulting systems work well in several environments.

3.4.3. Actions

The workshop discussed a number of possible actions. These actions are not about how to take specific technical solutions forward, but rather about how to discuss the topic going forward or what technical areas to focus on:

* We need to continue the discussion not all questions are answered. Additional discussion within the IETF will be needed. Continuing to connect the IETF with others in society and other SDOs around this topic is also useful.

* It is useful to find a role and a scope for IETF work in this area. The IETF will not develop alternative energy sources, work on social issues or have detailed discussions about implementation strategies or electronics design. However, the IETF has a role in measurement mechanisms, protocol design and standards — but of course activities in this role need to be aware of other aspects, such as implementation strategies.
* Increase our understanding of the environmental impacts of Internet technologies. One discussion topic during the workshop was also whether each new RFC should dedicate a section to discuss these impacts. No conclusion was drawn about the way to document these in RFCs, but it is clear that the IETF community will need to understand the environmental issues better. (Perhaps in addition to learning about the actual issues, guidelines for analysing protocols with regards to their impacts could be useful.)

* IETF activities on specific technologies are already ongoing or starting, such as metrics discussed, for instance, at the NMRG research group [NMRG] or the OPSAWG working group [OPSAWG], or the new Time Variant Routing (TVR) working group [TVR]. It may be also useful to start from picking the low-hanging fruits, such as:

- Focusing on improving energy proportionality and the consequent use of efficient data formats.

- Avoiding crypto assets - such as Non-Fungible Tokens (NFTs) and cryptocurrencies.

- Being able to carry information that needs to be shared for the purposes of enabling load and time shifting.

* Help initiate research activities that address some of the issues, such as broader gathering and sharing of measurement data, analysis of this data and looking at business related issues such as the impact of peering or advertising impacts sustainability. In addition, there may be a need to look at research for specific areas of improvements that are promising but not ready for standards discussion.

In summary, the goals that the IETF should have include:

* Full understanding of the Internets environmental impact.

* Continuous improvement of our technology.

* Launching research relevant activities.

4. Feedback

The organizers received generally positive feedback about the workshop.
One practical issue from the organizer’s point of view was that due to the extension of the deadline, the final submissions and paper reviews collided in part with the IETF-115 meeting. This led to it being very difficult for the program committee and practical organization staff to find time for the activity. We recommend avoiding such collisions in the future.

5. Security Considerations

The workshop itself did not address specific security topics. Of course, individual changes in Internet technology or operations that influence environmental impacts may also influence security aspects. These need to be looked at for every proposed change.

Such influence on security may come in different forms. For instance:

* A mechanism that makes, for instance, energy consumption information available may be susceptible to tampering or providing false information. For instance, [McDaniel] argues that economics and history shows that different players will attempt to cheat if a benefit can be accrued by doing so, e.g., by misreporting. As a result, sustainability measures and systems must be modeled as systems under threat.

* A mechanism that allows control of network elements for optimization purposes may be misused to cause denial-of-service or other types of attacks.

* Avoiding the use of crypto assets where other mechanisms suffice.

* Streamlining what data is sent may improve privacy if less information is shared.

6. IANA Considerations

This document has no IANA actions.

7. Position Papers

The following position papers were submitted to the workshop:

* Chris Adams, Stefano Salsano, Hesham ElBakoury: "Extending IPv6 to support Carbon Aware Networking" [Adams]

* Per Anderson, Suresh Krishnan, Jan Lindblad, Snezana Mitrovic, Marisol Palmero, Esther Roure, Gonzalo Salgueiro: "Sustainability Telemetry" [Anderson]
* Jari Arkko, Nina Lövehagen, Pernilla Bergmark: "Environmental Impacts of the Internet: Scope, Improvements, and Challenges" [Arkko]


* Alexander Clemm, Lijun Dong, Greg Mirsky, Laurent Ciavaglia, Jeff Tantsura, Marie-Paule Odini: "Green Networking Metrics" [ClemmA]

* Alexander Clemm, Cedric Westphal, Jeff Tantsura, Laurent Ciavaglia, Marie-Paule Odini: "Challenges and Opportunities in Green Networking" [ClemmB]

* Toerless Eckert, Mohamed Boucadair, Pascal Thubert, Jeff Tantsura: "IETF and Energy - An Overview" [Eckert]

* Greening of Streaming: "Tune In. Turn On. Cut Back. Finding the optimal streaming 'default' mode to increase energy efficiency, shift consumer expectations, and safeguard choice" [GOS]

* Romain Jacob: "Towards a power-proportional Internet" [Jacob]

* Fieke Jansen and Maya Richman: "Environment, internet infrastructure, and digital rights" [Jansen]

* Michael King, Suresh Krishnan, Carlos Pignataro, Pascal Thubert, Eric Voit: "On Principles for a Sustainability Stack" [King]

* Suresh Krishnan, Carlos Pignataro: "Sustainability considerations for networking equipment" [Krishnan]

* Jukka Manner: "Sustainability Considerations" [Manner]

* Vesna Manojlovic: "Internet Infrastructure and Climate Justice" [Manojlovic]

* Mike Mattera: "Understanding the Full Emissions Impact from Internet Traffic" [Mattera]

* John Preuss Mattsson: "Environmental Impact of Crypto-Assets" [Mattsson]

* Brendan Moran, Henk Birkholz, Carsten Bormann: "CBOR is Greener than JSON" [Moran]
* Louis Navarre, François Michel, Olivier Bonaventure: "It is time to reconsider multicast" [Navarre]

* Bruce Nordman: "Applying Internet Architecture to Energy Systems" [Nordman]

* Alvaro Retana, Russ White, Manuel Paul: "A Framework and Requirements for Energy Aware Control Planes" [Retana]

* Shayna Robinson, Remy Hellstern, Mariana Diaz: "Sea Change: Prioritizing the Environment in Internet Architecture" [Robinson]

* Daniel Schien, Paul Shabajee, Chris Preist: "Rethinking Allocation in High-Baseload Systems: A Demand-Proportional Network Electricity Intensity Metric" [Schien]


* Selome Kostentinos Tesfatsion, Xuejun Cai, Arif Ahmed: "End-to-end Energy Efficiency at Service-level in Edge Cloud" [Kostentinos]

* Pascal Thubert: "Digital Twin and Automation" [Thubert]

* Wim Vanderbauwhede: "Frugal Computing" [Vanderbauwhede]

* Michael Welzl, Ozgu Alay, Peyman Teymoori, Safiqul Islam: "Reducing Green House Gas Emissions With Congestion Control" [Welzl]

8. Program Committee

The program committee members were:

* Jari Arkko, Ericsson (program committee co-chair)
* Lars Eggert, Netapp (program committee co-chair)
* Colin Perkins, University of Glasgow (program committee co-chair)
* Luis M. Contreras, Telefónica
* Toerless Eckert, Futurewei
* Martin Flack, Akamai
* Mike Mattera, Akamai
9. Workshop Participants

The participants who attended at least one of the four sessions were:

* Alex Clemm
* Ali Rezaki
* Arif Ahmed
* Beatrice Siccardi
* Brendan Moran
* Bruce Nordman
* Carlos Pignataro
* Carsten Bormann
* Cedric Westphal
* Chiara Lombardo
* Chris Adams
* Colin Perkins
* Daniel Schien
* Dawn Nafus
* Dom Robinson
* Eric Voit
* Eric Vyncke
* Esther Roure Vila
* Eve M. Schooler
* Fieke Jansen
* Franco Davoli
* Gonzalo Salgueiro
* Greg Mirsky
* Henk Birkholz
* Hesham ElBakoury
* Hosein Badran
* Iankang Yao
* Jan Lindblad
* Jari Arkko
* Jens Malmolin
* Jiankang Yao
* John Preuß Mattsson
* Jukka Manner
* Julien Maisonneuve
* Kristin Moyer
* Lars Eggert
* Laurent Ciavaglia
* Lijun Dong
* Louis Navarre
* Louise Krug
* Luis M. Contreras
* Marisol Palmero Amador
10. Informative References


[Manojlovic]


[Vanderbauwhede]  


Appendix A. IAB Members at the Time of Approval
* Jari Arkko, Ericsson
* Deborah Brungard, AT&T
* Lars Eggert, NetApp
* Wes Hardaker, USC/ISI
* Cullen Jennings, Cisco Systems
* Mallory Knodel, Center for Democracy and Technology
* Mirja Kühlewind, Ericsson
* Zhenbin Li, Huawei
* Tommy Pauly, Apple
* David Schinazi, Google
* Russ White, Akamai
* Qin Wu, Huawei Technologies
* Jiankang Yao, CNNIC China Internet Network Information Center

Appendix B. Acknowledgments

Naturally, most of the credit goes to the workshop participants.

The organizers wish to thank Cindy Morgan and Greg Wood for their work on the practical arrangements and communications relating to the workshop. This report was greatly enhanced by the feedback provided on it, thanks to Michael Welzl in particular for his detailed review.
Authors’ Addresses

Jari Arkko
Ericsson
Email: jari.arkko@ericsson.com

Colin Perkins
University of Glasgow
Email: csp@csperkins.org

Suresh Krishnan
Cisco
Email: suresh.krishnan@gmail.com