Workgroup: Network Working Group
Internet-Draft: draft-thomson-tmi-00
Published: 13 July 2020
Intended Status: Informational
Expires: 14 January 2021
Authors: M. Thomson
Mozilla

Principles for the Involvement of Intermediaries in Internet Protocols

Abstract

This document proposes a set of principles for designing protocols with rules for intermediaries. The goal of these principles is to limit the ways in which intermediaries can produce undesirable effects and to protect the useful functions that intermediaries legitimately provide.

Discussion Venues

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

Discussion of this document takes place on the IAB Model-T list (modelt@iab.org), which is archived at https://mailarchive.ietf.org/arch/browse/model-t/.

Source for this draft and an issue tracker can be found at https://github.com/martinthomson/tmi.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

The Internet owes much of its success to its application of the end-to-end principle [E2E]. The realization that efficiency is best served by moving higher-level functions to endpoints is a key insight in system design, but also a key element of the success of the Internet.

This does not mean that the Internet avoids a relying on functions provided by entities in the network. While the principle establishes that some functions are best provided by endsystems, this does not exclude all intermediary functions. Some level of function in the network is necessary, or else there would be no network. The ways in which intermediaries can assist protocol endpoints are numerous and constantly evolving.

This document explores some of the ways in which intermediaries make both essential and valuable contributions to the function of the system. Problems arise when the interests of intermediaries are poorly aligned with those of endpoints. This can result in systemic costs and tension. Addressing those issues can be difficult.

This document proposes the following design principles for the protocols that might involve the participation of intermediaries:

* Avoid intermediation (Section 9.1)
* Limit the entities that can intermediate (Section 9.2)
* Limit what intermediaries can do (Section 9.3)

These principles aim to provide clarity about the roles and responsibilities of protocol participants. These principles produce more robust protocols with better privacy and security properties. These also limit the secondary costs associated with intermediation.

2. What is Meant by Intermediary

A protocol intermediary is an element that participates in communications. An intermediary is not the primary initiator or recipient of communications, but instead acts to facilitate communications.

An intermediary need not be explicitly present at the request of a participant.

Intermediaries exist at all layers of the stack. A router is an intermediary that acts at the network layer to forward packets. A TURN relay [RFC8155] provides similar forwarding capability for UDP in the presence of a network address translator (NAT) - a different type of intermediary that provides the ability to share a limited supply of addresses. At higher layers of the stack, group messaging servers intermediate the exchange of messages within groups of people; a conference focus aids the sending of media group real-time communications; and a social network intermediates communication and information sharing through the exchange of messages and formation of groups.

It is possible to facilitate communication without being an intermediary. The DNS provides information that is critical to locating and communicating with other Internet hosts, but it does so without intermediating those communications. Thus, this definition of intermediary does not include services like the DNS. That said, though the DNS as a service does not result in intermediation of other activities, there are roles for intermediaries within the DNS that fit this definition, such as recursive resolvers.

3. Intermediation Is Essential

Intermediaries are essential to scalable communications. The service an intermediary provides usually involves access to resources that would not otherwise be available. For instance, the Internet does not function without routers that enable packets to reach other networks.

Thus, there is some level of intermediation that is essential for the proper functioning of the Internet.

Scalable solutions to the introduction problem often depend on services that provide access to information and capabilities. As it is with the network layer of the Internet, the use of an intermediary can be absolutely essential. For example, a social networking application acts as an intermediary that provides a communications medium, content discovery and publication, and related services. Video conferencing applications often depend on an intermediary that mixes audio and selectively forwards video so that bandwidth requirements don't increase beyond what is available for participants as conferences grow in size.

4. Intermediation Is Useful

That intermediaries provide access to valuable resources does not imply that all intermediaries have exclusive control over access to resources. A router might provide access to other networks, but similar access might be obtained via a different route. The same web content might be provided by multiple CDNs. Multiple DNS resolvers can provide answers to the same queries. The ability to access the same capabilities from multiple entities contributes greatly to the robustness of a system.

Intermediaries often provide capabilities that benefit from economies of scale by providing a service that aggregates demand from multiple individuals. For instance, individuals are unlikely to be in a position to negotiate connections to multiple networks, but an ISP can. Similarly, an individual might find it difficult to acquire the capacity necessary to withstand a DDoS attack, but the scale at which a CDN operates means that this capacity is likely available to it. Or the value of a social network is in part due to the existing participation of other people.

Aggregation also provides other potential benefits. For instance, caching of shared information can allow for performance advantages. From an efficiency perspective, the use of shared resources might allow load to be more evenly distributed over time. For privacy, individual activity might be mixed with the activity of many others, thereby making it difficult to isolate that activity.

The ability of an intermediary to operate at scale can therefore provide a number of different benefits to performance, scalability, privacy, and other areas.

5. Intermediation Enables Scaling Of Control

An action by an intermediary can affect all who communicate using that intermediary. For an intermediary that operates at scale, this means it can be seen as an effective control point.

The goal of some intermediary deployments is to effect a policy, relying on the ability of a well-placed intermediary to affect multiple protocol interactions and participants.

The ability of an intermediary to affect a large number of network users can be an advantage or vulnerability, depending on perspective. For instance, network intermediaries have been used to distribute warnings of impending natural disasters like fire, flood, or earthquake, which save lives and property. In contrast, control over large-scale communications can enable censorship [RFC7754], misinformation, or pervasive monitoring [RFC7258].

Intermediaries that can affect many people can therefore be powerful agents for control. Though it is clear that the morality of actions taken can be subjective, network users have to consider the potential for the power they vest in intermediaries to be abused or subverted.

6. Incentive Misalignment at Scale

A dependency on an intermediary can represent a risk to those that take the dependency. The incentives and motives of intermediaries can be important to consider.

For instance, the information necessary for an intermediary to performs its function can often be used (or abused) for other purposes. Even the simple function of forwarding necessarily involves information about who was communicating, when, and the size of messages. This can reveal more than is obvious [CLINIC].

As uses of networks become more diverse, the extent that incentives for intermediaries and network users align reduce. In particular, acceptance of the costs and risks associated with intermediation by a majority of network users does not mean that all users have the same expectations and requirements. This can be a significant problem if it becomes difficult to avoid or refuse participation by a particular intermediary; see (TODO CHOKEPOINTS=I-D.iab-chokepoints).

7. Forced and Unwanted Intermediation

The ability to act as intermediary can offer more options than a service that is called upon to provide information. Sometimes those advantages are enough to justify the use of intermediation over alternative designs. However, the use of an intermediary also introduces costs.

The use of transparent or interception proxies in HTTP [HTTP] is an example of a practice that has fallen out of common usage due to increased use of HTTPS. Use of transparent proxies was once widespread with a wide variety of reasons for their deployment. However, transparent proxies were involved in many abuses, such as unwanted transcoding of content and insertion of identifiers to the detriment of individual privacy.

Introducing intermediaries is often done with the intent of avoiding disruption to protocols that operate a higher layer of the stack. However, network layering abstractions often leak, meaning that the effects of the intermediation can be observed. Where those effects cause problems, it can be difficult to detect and fix those problems.

The insertion of an intermediary in a protocol imposes other costs on other protocol participants; see [EROSION] or [MIDDLEBOX]. In particular, poor implementations of intermediaries can adversely affect protocol operation.

As an intermediary is another participant in a protocol, they can make interactions less robust. Intermediaries can also be responsible for ossification, or the inability to deploy new protocol mechanisms; see Section 2.3 of [USE-IT]. For example, measurement of TCP showed that the protocol has poor prospects for extensibility due to widespread use - and poor implementation - of intermediaries [TCP-EXTEND].

8. Contention over Intermediation

The IETF has a long history of dealing with different forms of intermediation poorly.

Early use of NAT was loudly decried by some in the IETF community. Indeed, the use of NAT was regarded as an unwanted intrusion by intermediaries. The eventual recognition - not endorsement - of the existence of NAT ([MIDDLEBOX], [NAT-ARCH]) allowed the community to engage in the design protocols that properly handled NAT devices

([UNSAF], [STUN]) and to make recommendations for best practices [BEHAVE].

Like HTTP, SIP [RFC3261] defines a role for a proxy, which is a form of intermediary with limited ability to interact with the session that it facilitates. In practice, many deployments instead choose to deploy some form of Back-to-Back UA (B2BUA; [RFC7092]) for reasons that effectively reduce to greater ability to implement control functions.

There are several ongoing debates in the IETF that are rooted in disagreement about the rule of intermediaries. The interests of network-based devices - which are sometimes intermediaries - is fiercely debated in the context of TLS 1.3 [TLS], where the design renders certain practices obsolete. Proposed uses of IPv6 header extensions in [SRv6NP] called into question the extent to which header extensions are the exclusive domain of endpoints as opposed to being available to intermediaries.

It could be that the circumstances in each of these debates is different enough that there is no singular outcome. The complications resulting from large-scale deployments of great diversity might render a single clear outcome impossible for an established protocol.

9. Proposed Principles

Many problems caused by intermediation are the result of intermediaries that are introduced without the involvement of protocol endpoints. Limiting the extent to which protocol designs depend on intermediaries makes the resulting system more robust.

These principles are set out in three stages:

1. Prefer designs without intermediaries (Section 9.1);

2. Failing that, control which entities can intermediate the protocol (Section 9.2); and

3. Limit actions and information that are available to intermediaries (Section 9.3).

The use of technical mechanisms to ensure that these principles are enforced is necessary. It is expected that protocols will need to use cryptography for this.

New protocols should identify what intermediation is anticipated and provide technical mechanisms to guarantee conformance. Modifying existing protocols to follow these principles could be difficult, but worthwhile.

9.1. Prefer Services to Intermediaries

Protocols should prefer designs that do not involve additional participants, such as intermediaries.

Designing protocols to use services rather than intermediaries ensures that responsibilities of protocol participants are clearly defined. Where functions can provided by means other than intermediation, the design should prefer that alternative.

If there is a need for information, defining a means for querying a service for that information is preferable to adding an intermediary. Similarly, direct invocation of service to perform an action is better than involving that service as a participant in the protocol.

Involving an entity as an intermediary can greatly increase the degree to which that entity becomes a dependency. For example, it might be necessary to negotiate the use of new capabilities with all protocol participants, including the intermediary, even when the functions for which the intermediary was added are not affected. It is also more difficult to limit the extent to which a protocol participant can be involved than a service that is invoked for a specific task.

Using discrete services is not always the most performant architecture as additional network interactions can add to overheads. The cost of these overheads need to be weighed against the recurrent costs from involving intermediaries.

Preferring services is analogous to the software design principle that recommends a preference for composition over inheritance [PATTERNS].

9.2. Deliberately Select Protocol Participants

Protocol participants should know what other participants they might be interacting with, including intermediaries.

Protocols that permit the involvement of an intermediary need to do so intentionally and provide measures that prevent the addition of unwanted intermediaries. Ideally, all protocol participants are identified and known to other protocol participants.

The addition of an unwanted protocol participant is an attack on the protocol.

This is an extension of the conclusion of [PATH-SIGNALS], which:

> recommends that implicit signals should be avoided and that an implicit signal should be replaced with an explicit signal only when the signal's originator intends that it be used by the network elements on the path.

Applying principle likely requires the use of authentication and encryption.

9.3. Limit Capabilities of Intermediaries

Protocol participants should be able to limit the capabilities conferred to other protocol participants.

Where the potential for intermediation already exists, or intermediaries provide essential functions, protocol designs should limit the capabilities and information that protocol participants are required to grant others.

Limiting the information that participants are required to provide to other participants has benefits for privacy or to limit the potential for misuse of information; see Section 9.3.1. Where confidentiality is impossible or impractical, integrity protection can be used to ensure that data origin authentication is preserved; see Section 9.3.2.

9.3.1. Limit Information Exposure

Protocol participants should only have access to the information they need to perform their designated function.

Protocol designs based on a principle of providing the minimum information necessary have several benefits. In addition to requiring smaller messages, or fewer exchanges, reducing information provides greater control over exposure of information. This has privacy benefits.

Where an intermediary needs to carry information that it has no need to access, protocols should use encryption to ensure that the intermediary cannot access that information.

Providing information for intermediaries using signals that are separate from other protocol signaling is preferable [RFC8558]. In addition, integrity protection should be applied to these signals to prevent modification.

9.3.2. Limit Permitted Interactions

An action should only be taken based on signals from protocol participants that are authorized to request that action.

Where an intermediary needs to communicate with other protocol participants, ensure that these signals are attributed to an intermediary. Authentication is the best means of ensuring signals generated by protocol participants are correctly attributed. Authentication informs decisions protocol participants make about actions they take.

In some cases, particularly protocols that are primarily two-party protocols, it might be sufficient to allow the signal to be attributed to any intermediary. This is the case in QUIC [QUIC] for ECN [ECN] and ICMP [ICMP], both of which are assumed to be provided by elements on the network path. Limited mechanisms exist to authenticate these as signals that originate from path elements, informing actions taken by endpoints.

9.3.3. Costs of Technical Constraints

Moving from a protocol in which there are two participants (such as [TLS]) to more than two participants can be more complex and expensive to implement and deploy.

More generally, the application of technical measures to control how intermediaries participate in a protocol incur costs that manifest in several ways. Protocols are more difficult to design; implementations are larger and more complex; and deployments might suffer from added operational costs, higher computation loads, and more bandwidth consumption. These costs are reflective of the true cost of involving additional entities in protocols. In protocols without technical measures to limit participation, these costs have historically been borne by other protocol participants.

10. Applying Non-Technical Constraints

Not all intermediary functions can be tightly constrained. For instance, as described in Section 6, some functions involve granting intermediaries access to information that can be used for more than its intended purpose. Applying strong technical constraints on how that information is used might be infeasible or impossible.

The use of authentication allows for other forms of control on intermediaries. Auditing systems or other mechanisms for ensuring accountability can use authentication information. Authentication can also enable the use of legal, social, or other types of control that might cover any shortfall in technical measures.

11. The Effect on Existing Practices

The application of these principles can have an effect on existing operational practices, particularly where they rely on protocols not limiting intermediary access. Several documents have explored aspects of this in detail:

* [RFC8404] describes effects of encryption on practices performed by intermediaries;

* [RFC8517] describes a broader set of practices;

* [TSV-ENC] explores the effect on transport-layer intermediaries in more detail; and

* [NS-IMPACT] examines the effect of TLS on operational network security practices.

In all these documents, the defining characteristic is the move from a system that lacked controls on participation to one in which technical controls are deployed. In each case the protocols in question provided no technical controls or only limited technical controls that prevent the addition of intermediaries. This allowed the deployment of techniques that involved the insertion of intermediaries into sessions without permission or knowledge of other protocol participants. By adding controls like encryption, these practices are disrupted. Overall, the advantages derived from having greater control and knowledge of other protocol participants outweighs these costs.

The process of identifying critical functions for intermediaries is ongoing. There are three potential classes of outcome of these discussion:

* Practices might be deemed valuable and methods that allow limited participation by intermediaries will be added to protocols.

* The use case supported by the practice might be deemed valuable, but alternative methods that address the use case without the use of an intermediary will be sought.

* Practices might be deemed harmful and no replacement mechanism will be sought.

Many factors could influence the outcome of this analysis. For instance, deployment of alternative methods or limited roles for intermediaries could be relatively simple for new protocol deployments; whereas it might be challenging to retrofit controls on existing protocol deployments.

12. Security Considerations

Controlling the level of participation and access intermediaries have is a security question. The principles in Section 9 are fundamentally an application of a security principle: namely the principle of least privilege [LEAST-PRIVILEGE].

Lack of proper controls on intermediaries protocols has been the source of significant security problems.

13. IANA Considerations

This document has no IANA actions.

14. Informative References

[BEHAVE] Audet, F., Ed. and C. Jennings, "Network Address Translation (NAT) Behavioral Requirements for Unicast UDP", BCP 127, RFC 4787, DOI 10.17487/RFC4787, January 2007, <https://www.rfc-editor.org/info/rfc4787>.

[CLINIC] Miller, B., Huang, L., Joseph, A., and J. Tygar, "I Know Why You Went to the Clinic: Risks and Realization of HTTPS Traffic Analysis", DOI 10.1007/978-3-319-08506-7_8, Privacy Enhancing Technologies pp. 143-163, 2014, <https://doi.org/10.1007/978-3-319-08506-7_8>.

[E2E] Saltzer, J., Reed, D., and D. Clark, "End-to-end arguments in system design", DOI 10.1145/357401.357402, ACM Transactions on Computer Systems (TOCS) Vol. 2, pp. 277-288, November 1984, <https://doi.org/10.1145/357401.357402>.

[ECN] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, DOI 10.17487/RFC3168, September 2001, <https://www.rfc-editor.org/info/rfc3168>.

[EROSION] Hildebrand, J. and P. McManus, "Erosion of the moral authority of transparent middleboxes", Work in Progress, Internet-Draft, draft-hildebrand-middlebox-erosion-01, 10 November 2014, <http://www.ietf.org/internet-drafts/draft-hildebrand-middlebox-erosion-01.txt>.

[HTTP] Fielding, R., Nottingham, M., and J. Reschke, "HTTP Semantics", Work in Progress, Internet-Draft, draft-ietf-httpbis-semantics-10, 12 July 2020, <http://www.ietf.org/internet-drafts/draft-ietf-httpbis-semantics-10.txt>.

[ICMP] Postel, J., "Internet Control Message Protocol", STD 5, RFC 792, DOI 10.17487/RFC0792, September 1981, <https://www.rfc-editor.org/info/rfc792>.

[LEAST-PRIVILEGE] Saltzer, J., "Protection and the control of information sharing in multics", DOI 10.1145/361011.361067, Communications of the ACM Vol. 17, pp. 388-402, July 1974, <https://doi.org/10.1145/361011.361067>.

[MIDDLEBOX] Carpenter, B. and S. Brim, "Middleboxes: Taxonomy and Issues", RFC 3234, DOI 10.17487/RFC3234, February 2002, <https://www.rfc-editor.org/info/rfc3234>.

[NAT-ARCH] Hain, T., "Architectural Implications of NAT", RFC 2993, DOI 10.17487/RFC2993, November 2000, <https://www.rfc-editor.org/info/rfc2993>.

[NS-IMPACT] Cam-Winget, N., Wang, E., Danyliw, R., and R. DuToit, "Impact of TLS 1.3 to Operational Network Security Practices", Work in Progress, Internet-Draft, draft-ietf-opsec-ns-impact-00, 23 June 2020, <http://www.ietf.org/internet-drafts/draft-ietf-opsec-ns-impact-00.txt>.

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

[PATTERNS] Gamma, E., Helm, R., Johnson, R., and J. Vlissides, "Design Patterns: Elements of Reusable Object-Oriented Software", 1994.

[QUIC] Iyengar, J. and M. Thomson, "QUIC: A UDP-Based Multiplexed and Secure Transport", Work in Progress, Internet-Draft, draft-ietf-quic-transport-29, 9 June 2020, <http://www.ietf.org/internet-drafts/draft-ietf-quic-transport-29.txt>.

[RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, DOI 10.17487/RFC3261, June 2002, <https://www.rfc-editor.org/info/rfc3261>.

[RFC3552] Rescorla, E. and B. Korver, "Guidelines for Writing RFC Text on Security Considerations", BCP 72, RFC 3552, DOI

10.17487/RFC3552, July 2003, <https://www.rfc-editor.org/info/rfc3552>.

[RFC3724] Kempf, J., Ed., Austein, R., Ed., and IAB, "The Rise of the Middle and the Future of End-to-End: Reflections on the Evolution of the Internet Architecture", RFC 3724, DOI 10.17487/RFC3724, March 2004, <https://www.rfc-editor.org/info/rfc3724>.

[RFC7092] Kaplan, H. and V. Pascual, "A Taxonomy of Session Initiation Protocol (SIP) Back-to-Back User Agents", RFC 7092, DOI 10.17487/RFC7092, December 2013, <https://www.rfc-editor.org/info/rfc7092>.

[RFC7258] Farrell, S. and H. Tschofenig, "Pervasive Monitoring Is an Attack", BCP 188, RFC 7258, DOI 10.17487/RFC7258, May 2014, <https://www.rfc-editor.org/info/rfc7258>.

[RFC7754] Barnes, R., Cooper, A., Kolkman, O., Thaler, D., and E. Nordmark, "Technical Considerations for Internet Service Blocking and Filtering", RFC 7754, DOI 10.17487/RFC7754, March 2016, <https://www.rfc-editor.org/info/rfc7754>.

[RFC8155] Patil, P., Reddy, T., and D. Wing, "Traversal Using Relays around NAT (TURN) Server Auto Discovery", RFC 8155, DOI 10.17487/RFC8155, April 2017, <https://www.rfc-editor.org/info/rfc8155>.

[RFC8404] Moriarty, K., Ed. and A. Morton, Ed., "Effects of Pervasive Encryption on Operators", RFC 8404, DOI 10.17487/RFC8404, July 2018, <https://www.rfc-editor.org/info/rfc8404>.

[RFC8517] Dolson, D., Ed., Snellman, J., Boucadair, M., Ed., and C. Jacquenet, "An Inventory of Transport-Centric Functions Provided by Middleboxes: An Operator Perspective", RFC 8517, DOI 10.17487/RFC8517, February 2019, <https://www.rfc-editor.org/info/rfc8517>.

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

[SRv6NP] Filsfils, C., Camarillo, P., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "SRv6 Network Programming", Work in Progress, Internet-Draft, draft-ietf-spring-srv6-network-programming-16, 27 June 2020, <http://www.ietf.org/internet-drafts/draft-ietf-spring-srv6-network-programming-16.txt>.

[STUN]
Petit-Huguenin, M., Salgueiro, G., Rosenberg, J., Wing, D., Mahy, R., and P. Matthews, "Session Traversal Utilities for NAT (STUN)", Work in Progress, Internet-Draft, draft-ietf-tram-stunbis-21, 22 March 2019, <http://www.ietf.org/internet-drafts/draft-ietf-tram-stunbis-21.txt>.

[TCP-EXTEND] Honda, M., Nishida, Y., Raiciu, C., Greenhalgh, A., Handley, M., and H. Tokuda, "Is it still possible to extend TCP?", DOI 10.1145/2068816.2068834, Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference - IMC '11, 2011, <https://doi.org/10.1145/2068816.2068834>.

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

[TSV-ENC] Fairhurst, G. and C. Perkins, "Considerations around Transport Header Confidentiality, Network Operations, and the Evolution of Internet Transport Protocols", Work in Progress, Internet-Draft, draft-ietf-tsvwg-transport-encrypt-15, 1 May 2020, <http://www.ietf.org/internet-drafts/draft-ietf-tsvwg-transport-encrypt-15.txt>.

[UNSAF] Daigle, L., Ed. and IAB, "IAB Considerations for UNilateral Self-Address Fixing (UNSAF) Across Network Address Translation", RFC 3424, DOI 10.17487/RFC3424, November 2002, <https://www.rfc-editor.org/info/rfc3424>.

[USE-IT] Thomson, M., "Long-term Viability of Protocol Extension Mechanisms", Work in Progress, Internet-Draft, draft-iab-use-it-or-lose-it-00, 7 August 2019, <http://www.ietf.org/internet-drafts/draft-iab-use-it-or-lose-it-00.txt>.

Appendix A. Acknowledgments

This document is merely an attempt to codify existing practice. Practice that is inspired, at least in part, by prior work, including [RFC3552] and [RFC3724] which both advocate for clearer articulation of trust boundaries.

Eric Rescorla and David Schinazi are acknowledged for their contributions of thought, review, and text.

Author's Address

Martin Thomson

Mozilla

Email: mt@lowentropy.net