Network Working Group
Internet-Draft
Intended status: Informational
Expires: 8 July 2022

V. Bertola
Open-Xchange

4 January 2022

Everything But The User Is An Intermediary
draft-bertola-everything-but-the-user-00

Abstract

This document provides the author's perspective on the shortcomings of the Internet threat model defined in RFC 3552 and currently in use at the IETF. It then proposes the basic conceptual framework for an additional model, the "holistic threat model", describing how it could be used to broaden the analysis of non-technical protocol impacts in the design phase.

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 8 July 2022.

Copyright Notice

Table of Contents

1. Introduction and Background

1.1. The End-to-End Principle

The architecture and design of Internet protocols traditionally stems from the "end-to-end principle" - the choice of concentrating the intelligence and the complexity of the services of a network at its edges, reducing to the minimum possible level the functions and the prerogatives of the inner nodes of the network. When the implementation of a network's functions is built as a set of layers relying on each other, the intelligence and the complexity are pushed upwards, to the layers nearer to the end-user, minimizing the activities of the lower layers [Salzer].

Under this principle, application-level protocols are designed as a communication between two endpoints sitting at the edge of the network, directly interacting with the respective end-users, or, for human-to-server communication, with the physical or juridical entity providing the service at the endpoint. While in some cases communications can be directed to more than one endpoint, in-network intermediaries to these communications simply do not exist; all parties involved in the communication sit at the network's edge.

1.2. Intermediaries in Today's Internet

Over time, throughout the technical development of the Internet, it has not always been possible to upkeep the end-to-end principle in

full. There now are services and functions that involve intermediaries, that is, entities other than the endpoints, sitting on the network path between them and performing higher-level functions.

For example, the relative scarcity of IPv4 addresses prompted the introduction of nodes that would encapsulate the end-user's traffic and pretend to be the actual endpoint of the Internet connection, then passing back the packets received in response. Security checks have been implemented by having intermediate network nodes examine all the traffic, sometimes up to content at the highest application level, and determine whether such traffic should be allowed to reach the other endpoint or not.

Some functionalities also require the involvement of third parties acting as "side intermediaries", that is, additional parties to the communication that sit elsewhere than on the network path between the endpoints, and are brought into the communication as an endpoint to a separate network connection. An example is the DNS resolution process, as the network stack setting up a connection to the endpoint must establish a separate connection to a separate service, the DNS resolver, to discover the IP address of the endpoint from its name.

Given the general recognition of the end-to-end principle, the role of intermediaries in the design of Internet protocols has generally been controversial. As added parties into a communication, intermediaries can be used as a point of surveillance and control over traffic that the end-user did not mean them to see. On the other hand, intermediaries can also be used to provide valuable services, such as enabling the user to identify network destinations by name rather than by IP address, or such as increased security for users that would not be able to identify threats directly (e.g. against phishing websites).

1.3. The Current Internet Threat Model

The threat model used when evaluating the safety of Internet protocols has generally assumed that endpoint behaviour and risks are out of scope, only focusing on the communication itself. Thus, in the last decade, the increased attention to the privacy and security of the end-user's network activities led to the progressive

elimination of intermediaries, especially those that were deployed by network operators halfway on the path of Internet communications.

Frequently, the instrument used to this purpose has been the encryption and disguise of communications between the two network endpoints; sometimes, protocol changes, including high-level routing changes and traffic aggregation, have also been conceived to reduce opportunities for discriminating or tracking traffic.

In some cases, however, these instruments have been claimed to produce effects that actually endanger the user's privacy and security - among these, the aggregation and centralisation of traffic

into the infrastructure of a limited number of operators, and the elimination of any possibility for positive intermediation, such as in-network security checks.

As the overall privacy of the end-user's communication is determined both by the privacy during the transport of the information and by the privacy of the processing that happens within the endpoints, an increase in the former could be compensated by a decrease in the latter, with a negative net effect. At the IETF and elsewhere, discussions are ongoing on how Internet protocol design could also take into account this risk.

Since many of these threats are related to endpoint behaviour, or to a combination of protocol features and non-technical endpoint decisions such as the model, timeframe and policies for their deployment, they cannot be caught by a threat model that declares endpoints to be out of scope.

As a contribution to this discussion, this document proposes a different and more general approach to the assessment of potential risks, considering both in-network and endpoint actors as intermediaries, and extending the analysis to non-technical issues. This model could be used whenever a broader impact analysis is considered useful.

2. Shortcomings of the Current Threat Model

The current Internet threat model is formalized in RFC 3552 [RFC3552]. While the goals described in section 2 of that document

are still generally valid, in today's Internet environment the threat model described in section 3 sometimes fails to capture all the relevant threats that could affect the accomplishment of those goals.

2.1. Shortcomings in the Scope of Attackers

The first shortcoming of the current Internet threat model is that it attempts to evaluate the impact on the properties of a communication between human beings (or between a human and a computer service) by only assessing the properties of the communication between the protocol endpoints, assuming for the purpose of protocol design that no attacks can happen during the processing of communications between the protocol endpoint and the end-user, or that these attacks can be effectively pre-empted and countered by the end-user.

Section 3 of RFC 3552 supplies a justification for this assumption; it states that "Protecting against an attack when one of the end-systems has been compromised is extraordinarily difficult." Indeed, under this assumption the protocol designer can conflate two separate

communication layers, the protocol one - an exchange of information between two software or hardware elements acting as protocol endpoints at the edges of the network - and the end-user one - an exchange of information between human beings. This assumption is quite helpful for protocol design, as it reduces the quantity and quality of threats that have to be examined. However, it inevitably excludes other relevant threats from the analysis, giving users no protection from them.

This assumption could have been justified under a number of external conditions that were mostly true when the original Internet threat model was conceived, such as:

1. The end-user has full control over the device that they are using.

2. The end-user is freely able to choose the application and the operating system that they are using, picking some that they can trust.

3. The end-user is sufficiently competent in technology to be able to assess the risks of misbehaviour by the endpoint elements of

their communication, and to judge whether to entrust them with data and information.

However, in today's Internet, these conditions are generally false. Internet end-users are in average much less tech-literate than they were in 2003, when RFC 3552 was written; even more so if we consider the Internet of 1984, when the end-to-end principle as we know it today was formalized.

Over the last decade, network services based on open, federated protocols - like e-mail or the Web - have increasingly been complemented or replaced by newer services - like instant messaging or social media - based on a single-operator, single-implementation "walled garden" model, thus giving the user much more limited choices, or no choice at all, over which application to use. Most users now access the Internet from smartphones, almost entirely running on one of two dominant operating systems. To account for the more limited technical competence of end-users, newer devices and applications often give them limited configuration choices. The advent of the cloud computing model has brought forth a vast range of home appliances and software applications that only communicate with the servers of their own makers, often with no opportunity for the user to verify or alter their network communications.

While useful in technical terms, this assumption in today's scenario - at least in some cases - leaves too many threats out of scope to be still acceptable.

2.2. Shortcomings in the Scope of Threats

Another shortcoming of the current Internet threat model derives from its limited technical focus. When it was designed, the threat model focused only on the security of the communication; from the end-user's viewpoint, communication security was defined in section 2.1 of RFC 3552 as the union of three more specific goals - confidentiality, data integrity and peer entity authentication.

In 2013, the IETF deemed it necessary to add a separate, specific focus on potential threats to the user's privacy, releasing RFC 6973

[RFC6973]. While building on the original goal of confidentiality, this document moved the analysis to a higher, less technical level, bringing into scope the non-technical consequences of technical communication breaches. To protect the user's privacy effectively, it was deemed necessary to consider the non-technical purposes and motivations that could prompt an attacker to exploit any flaws and weaknesses in communication protocols. Moreover, in section 4 RFC 6973 already questioned the validity of RFC 3552's assumption of non-compromised endpoints.

In 2017, the scope of the suggested analysis was further broadened by RFC 8280 [RFC8280]. While still being a proposal subject to further research, this document codified an even broader range of non-technical threats to social, economical and political rights, such as freedom of expression, non-discrimination, freedom of assembly and so on; it then suggested how to consider potential adverse effects to these rights when designing protocols.

In the last twenty years, the IETF appears to have followed the societal trend that, in parallel with the increased usage and pervasiveness of the Internet in the everyday life of all human beings, brought the non-technical consequences of technical design choices by the Internet's technical community and industry to the attention of the public opinion and of the policymakers. It is now expected that a sense of social and corporate responsibility, and policy objectives defined outside of the technical community, contribute to shaping the choices that will determine the future evolution of the Internet.

However, the analysis of non-technical consequences of technical design choices does not just require technical skills, but also competence in fields like business administration, economy, policy, law and social sciences. The efforts of protocol engineers to take

into account the dynamics and the effects in these other fields are commendable, but there is the need to bring the appropriate competences into the discussion, while avoiding mission creep for technical standards organizations. Any attempt to address this shortcoming should also consider this issue.

3. The Holistic Threat Model

As a consequence of the shortcomings that have been identified above, to perform a complete evaluation of the risks to the security and privacy of end-users and to their rights in general, it would be necessary to expand the scope of the threat analysis in two directions.

First, it would be necessary to consider all elements that have access to the end-user's information, data, metadata and content, independently from whether they sit within the network or within the endpoint devices; any element located anywhere between the fingers (or other human-machine interface) of the sending end-user and the eyes of the receiving end-user (or the operator of the application-level Internet service) is in scope. In short, "everything but the user is an intermediary"; since both the endpoint devices and any in-network intermediaries have access to user information, all of them should be considered as third parties that could potentially attack the end-user.

Secondly, to understand how these elements might realistically behave, it would be necessary to consider not just the technical building blocks, but the physical or juridical entities that operate or control them, and the non-technical motivations that could push them to attack or constrain the user, such as economic advantages, the accumulation and preservation of power, the desire to surveil or stalk another person and many others.

Under this extended threat model, no claim should be made over the privacy, security or any other property of Internet communications from the end-user's perspective, unless all parties different from the user(s) that take part in the communication, and all their possible motivations, have been considered: hence the "holistic" threat model.

3.1. Applicability of the Holistic Threat Model

In theory, the holistic threat model is a necessary response to the shortcomings identified above. The more limited analysis performed under the RFC 3552 threat model could not only lead to incomplete results, failing to identify significant threats, but could even lead to counterproductive results. Architectural choices that are assessed as increasing the user's privacy could, under a broader analysis of likely non-technical dynamics, actually cause new threats to the user's privacy, for example by promoting business models based on data monetization or the centralization of user information into fewer hands.

However, the skillsets and research necessary to conduct such a multidisciplinary analysis, and the sheer amount of actors and possible behaviours, could make the effort too heavy and practically unfeasible, or simply excessive for specifications of limited impact.

As an initial experiment, the holistic threat model could be applied as a second step for selected documents, after performing a security and privacy analysis under the current threat model, whenever the initial examination of the properties of a new specification raises concerns on potential broader impacts.

It is important to note that even the assessment on whether a specification could have impacts on non-technical issues and stakeholders requires in principle the broader skillset necessary to perform the full analysis. In other words, to determine whether a holistic assessment would be necessary, it is necessary to consult the non-technical stakeholders that could be affected by the new specification, letting them do a preliminary analysis and raise any potential concerns before the specification is finalized.

In the end, the need identified by this document calls for a stronger and more effective interaction between the IETF and other parts of the Internet's industry, policy and user communities on a global scale. Efforts to establish such an interaction should be made jointly by the IETF and by the other relevant stakeholders; how to do that in practice could be the subject of a separate discussion.

4. Security Considerations

This document discusses the assumptions under which security considerations should be developed in the future. It is meant to contribute to the ongoing discussion over a possible revision of RFC 3552 [RFC3552], which specifies how to write security considerations.

5. IANA Considerations

This document has no IANA actions.

6. Informative References

[RFC3552] Rescorla, E. and B. Korver, "Guidelines for Writing RFC Text on Security Considerations", BCP 72, RFC 3552, DOI 10.17487/RFC3552, July 2003, <https://www.rfc-editor.org/info/rfc3552>.

[RFC6973] Cooper, A., Tschofenig, H., Aboba, B., Peterson, J., Morris, J., Hansen, M., and R. Smith, "Privacy Considerations for Internet Protocols", RFC 6973, DOI 10.17487/RFC6973, July 2013, <https://www.rfc-editor.org/info/rfc6973>.

[RFC8280] ten Oever, N. and C. Cath, "Research into Human Rights Protocol Considerations", RFC 8280, DOI 10.17487/RFC8280, October 2017, <https://www.rfc-editor.org/info/rfc8280>.

[Salzer] Salzer, J. H., Reed, D. P., and D. D. Clark, "End-to-end arguments in system design", 1 November 1984, <https://dl.acm.org/doi/10.1145/357401.357402>.

Author's Address

Vittorio Bertola
Open-Xchange Srl
Via Treviso 12
10144 Torino
Italy

Email: vittorio.bertola@open-xchange.com
URI: https://www.open-xchange.com/