Workgroup: Network Working Group
Internet-Draft:
draft-irtf-pearg-safe-internet-measurement-09
Published: 12 January 2024
Intended Status: Informational
Expires: 15 July 2024
Authors: I. R. Learmonth
HamBSD
G. Grover
Centre for Internet and Society
M. Knodel
Center for Democracy and Technology

Guidelines for Performing Safe Measurement on the Internet

Abstract

Internet measurement is important to researchers from industry, academia and civil society. While measurement of the internet can give insight into the functioning and usage of the internet, it can present risks to user privacy and safety. This document describes briefly those risks and proposes guidelines for ensuring that internet measurements can be carried out safely, with examples.

Note

This document is a draft. It is not an IETF product. It does not propose a standard. Comments are solicited and should be addressed to the research group's mailing list at pearg@irtf.org and/or the author(s).

The sources for this draft are at:

https://github.com/IRTF-PEARG/draft-safe-internet-measurement

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 July 2024.

Copyright Notice

Table of Contents

1. Introduction

Measurement of the internet provides important insights and is a growing area of research. Similarly, the internet plays a role in enhancing research methods of different kinds.

Performing research using the internet, as opposed to an isolated testbed or simulation platform, means that experiments co-exist in a space with other services and end users. Furthermore privacy

considerations are of particular importance in internet measurement research that depends on collaboration and data sharing models between industry and academia[caida].

This document outlines guidelines for academic, industry and civil society researchers who might use the internet as part of scientific experimentation to mitigate risks to the safety of users.

1.1. Scope of this document

These are guidelines for how to measure the internet safely. When performing research on a platform shared with live traffic from other users, that research is considered safe if and only if other users are protected from or unlikely to experience danger, risk, or injury arising due to the research, now or in the future.

Following the guidelines contained within this document is not a substitute for institutional ethics review processes, although these guidelines could help to inform that process. It is particularly important for the growing area of research that includes internet measurement to better equip review boards to evaluate internet measurement methods [SIGCOMM], and we hope that this document is part of that larger effort.

Similarly, these guidelines are not legal advice and local laws must also be considered before starting any experiment that could have adverse impacts on user safety.

The scope of this document is restricted to guidelines that mitigate exposure to risks to user safety when measuring properties of the internet: the network, its constituent hosts and links, or user traffic.

1.2. Terminology

Threat model: A threat is a potential for a security violation, which exists when there is a circumstance, capability, action, or event that could breach security and cause harm [RFC4949].

User: For the purpose of this document, an internet user is an individual or organisation whose data is used in communications over the internet, most broadly, and those who use the internet to communicate or maintain internet infrastructure.

Active measurement: Active measurements generate or modify traffic.

Passive measurement: Passive measurements involve the observation of existing traffic without active intervention.

On/off-path: A measurement that is on-path happens on the network. Off-path indicates activity in a side-channel, end-point or at other points where the user, their connection, or their data can be accessed.

One-/two-ended: A single-ended measurement is like a probe or a trace, whereas a measurement with two-ended control provides more accuracy but requires the cooperation of both endpoints, which might include the network itself if that is the measurement target.

1.3. User impact from measurement studies

Any conceivable internet measurement study might have an impact on an internet user's safety. The measurement of generated traffic may also lead to insights into other users' traffic indirectly as well. It is always necessary to consider the best approach to mitigate the impact of measurements, and to balance the risks of measurements against the benefits to impacted users.

Some possible ways in which users can be affected as a result of an internet measurement study:

Breach of privacy: User privacy can be violated in the context of data collection. This impact also covers the case of an internet user's data being shared beyond that for which a user had given consent. First-order data that distinguishes a person such as name, as well as second-order data that can be used to track behaviour such as IP address, should be considered[Kenneally]

Inadequate data protection: A scenario where data, either in transit or at rest, lacks sufficient protection from disclosure. Failure to meet user expectations for data protection is a concern, even if it does not result in unauthorized access to the data. This includes cases of improper access control (i.e. people having access to user data who do not need it).

Traffic generation: A scenario where undue traffic is generated to traverse the internet.

Traffic modification: A scenario where users' on-path internet traffic is nonconsensually modified.

Impersonation: A scenario where a user is impersonated during a measurement.

Legal: Users and service providers are bound by a wide range of policies from Terms of Service to rule of law, each according to context and jurisdiction. A measurement study may violate these policies, and the consequences of such a violation may be severe.

Unavailability: Users or other entities may rely on the information or systems that are involved in the research and they may be harmed by unexpected or planned unavailability of that information or systems[Menlo].

System or data corruption: A scenario where generated or modified traffic causes the corruption of a system. This covers cases where a user's data may be lost or corrupted, and cases where a user's access to a system may be affected as a result.

Emotional trauma: A scenario where a measurement of or exposure to content or behaviour in an internet measurement study causes a user emotional or psychological harm.

2. Guidelines

2.1. Attribute

Proactively identify your measurement to others on the network. "This allows any party or organization to understand what an unsolicited probe packet is, what its purpose is, and, most importantly, who to contact."[RFC9511]

Example: For a layer 3 IP packet probe you could mark measurements with a probe description URI as defined in RFC9511.

2.2. Obtain consent

Accountability and transparency are fundamentally related to consent. As per the Menlo Report, "Accountability demands that research methodology, ethical evaluations, data collected, and results generated should be documented and made available responsibly in accordance with balancing risks and benefits."[Menlo] A user is best placed to balance the risks and benefits for themselves therefore consent must be obtained. From most transparent to least, there are a few options for obtaining consent.

2.2.1. Informed consent

Informed consent should be collected from all users that may be placed at risk by an experiment.

For consent to be informed, a reasonable coverage of possible risks must be presented to the users. The considerations in this document can be used to provide a starting point although other risks may be present depending on the nature of the measurements to be performed. In addition, it should be clear from the consent language who the asker is, and what the terms of data observation and/or collection are.

Example: A researcher would like to use volunteer-owned mobile devices to collect information about local internet censorship. Connections will be attempted by the volunteer's device with services and content known or suspected to be subject to censorship orders.

This experiment can carry substantial risk for the user depending on their specific circumstances. Trying to access censored material can be seen as (network) policy infringement or breaking laws. Consequences can range from disciplinary action from their employer to arrest or imprisonment by government authorities. If the experimenter wants to expose volunteers to this kind of risk, users must be fully informed, and voluntarily give consent to run the measurement. Even then, experimenters should seriously consider designing their experiment in another way.

Note that informed consent is notoriously tricky to obtain. Conveying all possible risks of a measurement is often simply impractical, depending upon how technical the user audience is, the context of the consent prompt, what the tool is normally used by users for, etc. In addition, consent can have network effects. For example, asking a user to consent to sharing information about their communication with others can have impacts on users who have not personally consented to the study.

2.2.2. Proxy consent

In cases where it is not practical to collect informed consent from all users of a shared network, it may be possible to obtain proxy consent. Proxy consent may be given by a network operator or employer that would be more familiar with the expectations of users of a network than the researcher.

In some cases, a network operator or employer may have terms of service that specifically allow for giving consent to third parties to perform certain experiments.

Example: Some researchers would like to perform a packet capture to determine the TCP options and their values used by all client devices on a corporate wireless network.

The employer may already have terms of service laid out that allow them to provide proxy consent for this experiment on behalf of the employees, in this case the users of the network. The purpose of the experiment may affect whether or not they are able to provide this consent. Say, performing engineering work on the network may be allowed, whereas academic research may not be already covered.

Example: A research project looks at networked "things", yet users' only interface with the network is through a device that does not

provide interaction to the degree that would be sufficient to obtain informed consent at time of use.

However in this case the user can be informed of the use of data for internet measurement research in the device's terms of use and privacy notice, which can be included in a printed, physical manual for the device or accessed at any time via a webpage. These are examples of proxy consent such that the device manufacturer may choose to share data under certain specified conditions, or to conduct their own measurements.

2.2.3. Implied consent

In larger scale measurements, even proxy consent collection may not be practical. In this case, implied consent may be presumed from users for some measurements. Consider that users of a network will have certain expectations of privacy and those expectations may not align with the privacy guarantees offered by the technologies they are using. As a thought experiment, consider how users might respond if asked for their informed consent for the measurements you'd like to perform.

Implied consent should not be considered sufficient for any experiment that may collect sensitive or personally identifying information. If practical, attempt to obtain informed consent or proxy consent from a sample of users to better understand the expectations of other users.

Example: A researcher would like to run a measurement campaign to determine the maximum supported TLS version on popular web servers.

The operator of a web server that is exposed to the internet hosting a popular website would have the expectation that it may be included in surveys that look at supported protocols or extensions but would not expect that attempts be made to degrade the service with large numbers of simultaneous connections.

Example: A researcher would like to perform A/B testing for protocol feature and how it affects web performance. They have created two versions of their software and have instrumented both to report telemetry back. These updates will be pushed to users at random by the software's auto-update framework. The telemetry consists only of performance metrics and does not contain any personally identifying or sensitive information.

As users expect to receive automatic updates, the effect of changing the behaviour of the software is already expected by the user. If users have already been informed that data will be reported back to the developers of the software, then again the addition of new metrics would be expected. Note that the reduced impact of A/B

testing should not be used be an excuse to push updates that might compromise user expectations around security and privacy.

In the event that something does go wrong with the update, it should be easy for users to discover that they have been part of an experiment and roll back the change, allowing for explicit refusal of consent to override the presumed implied consent.

2.3. Share responsibly

Further to use of measurement data, data is often shared with other researchers. Measurement data sharing comes with its own set of expectations and responsibilities of the provider. Likewise there are responsibilities that come with the use of others' measurement data. One obvious expectation is around end-user consent (see "Implied consent" above). Allman and Paxson [Allman] provide "a set of guidelines that aim to aid the process of sharing measurement data... [in] a framework under which providers and users can better attain a mutual understanding about how to treat particular datasets."

Their guidance since 2007 has been for data providers to:

* explicitly indications of the terms of a dataset's acceptable use

* convey what interactions they desire or will accommodate.

Their guidance for researchers is to:

* be thoughtful in the reporting of potentially sensitive information gleaned from providers' data.

* comply with the indications and interactions of the data providers.

Example: Researchers have obtained network measurement data from more than one provider for purposes of conducting analysis of protocol use on both. Where privacy paritioning techniques are used, the researchers' findings may inadvertently collude to uncover private information about users. Once realised, researchers should mitigate this privacy risk to end users as well as disclosing this result to the data providers themselves.

2.4. Isolate risk with a dedicated testbed

Wherever possible, use a testbed. An isolated network means that there are no other users sharing the infrastructure you are using for your experiments.

When measuring performance, competing traffic can have negative effects on the performance of your test traffic and so the testbed approach can also produce more accurate and repeatable results than experiments using the public internet.

Example: WAN link conditions can be emulated through artificial delays and/or packet loss using a tool like [netem]. Competing traffic can also be emulated using traffic generators.

2.5. Be respectful of others' infrastructure

If your experiment is designed to trigger a response from infrastructure that is not your own, consider what the negative consequences of that may be. At the very least your experiment will consume bandwidth that may have to be paid for.

In more extreme circumstances, you could cause traffic to be generated that causes legal trouble for the owner of that infrastructure. The internet is a global network that crosses many legal jurisdictions and so what may be legal for one is not necessarily legal for another.

If you are sending a lot of traffic quickly, or otherwise generally deviating from typical client behaviour, a network may identify this as an attack which means that you will not be collecting results that are representative of what a typical client would see.

One possible way to mitigate this risk is transparency, i.e. mark measurement-related data or activity as such. For example, the popular internet measurement tool ZMap hardcodes its packets to have IP ID 54321 in order to allow identification [ZMap].

2.6. Maintain a "Do Not Scan" list

When performing active measurements on a shared network, maintain a list of hosts that you will never scan regardless of whether they appear in your target lists. When developing tools for performing active measurement, or traffic generation for use in a larger measurement system, ensure that the tool will support the use of a "Do Not Scan" list.

If complaints are made that request you do not generate traffic towards a host or network, you must add that host or network to your "Do Not Scan" list, even if no explanation is given or the request is automated.

You may ask the requester for their reasoning if it would be useful to your experiment. This can also be an opportunity to explain your research and offer to share any results that may be of interest. If you plan to share the reasoning when publishing your measurement

results, e.g. in an academic paper, you must seek consent for this from the requester.

Be aware that in publishing your measurement results, it may be possible to infer your "Do Not Scan" list from those results. For example, if you measured a well-known list of popular websites then it would be possible to correlate the results with that list to determine which are missing. This inference might leak the fact that those websites specifically requested to not be scanned.

2.7. Minimize data

When collecting, using, disclosing, and storing data from a measurement, use only the minimal data necessary to perform a task. Reducing the amount of data reduces the amount of data that can be misused or leaked.

When deciding on the data to collect, assume that any data collected might be disclosed. There are many ways that this could happen, through operational security mistakes or compulsion by a judicial system.

When directly instrumenting a protocol to provide metrics to a passive observer, see section 6.1 of RFC6973[RFC6973] for the data minimization considerations enumerated below that are specific to the use case.

2.7.1. Discard data

Discard data that is not required to perform the task.

When performing active measurements, be sure to only capture traffic that you have generated. Traffic may be identified by IP ranges or by some token that is unlikely to be used by other users.

Again, this can help to improve the accuracy and repeatability of your experiment. For performance benchmarking, [RFC2544] requires that any frames received that were not part of the test traffic are discarded and not counted in the results.

2.7.2. Mask data

Mask data that is not required to perform the task. This technique is particularly useful for content of traffic to indicate that either a particular class of content existed or did not exist, or the length of the content, but not recording the content itself. The content can be replaced with tokens or encrypted.

It is important to note that masking data does not necessarily anonymize it [SurveyNetworkTrafficAnonymisationTech].

2.7.3. Aggregate data

When collecting data, consider if the granularity can be limited by using bins or adding noise. Differential privacy techniques [DifferentialPrivacy] can help with this.

Example: [Tor.2017-04-001] presents a case-study on the in-memory statistics in the software used by the Tor network.

2.8. Reduce accuracy

There are various techniques that can be used to reduce the accuracy of the collected data and make it less identifying.

The use of binning to group numbers of more-or-less continuous values, coarse categorization in modeling, reduction in concentrations of IP address by geography (geoip) or other first- or second-order identifiers, the introduction of noise and all privacy-preserving measurement techniques that allow researchers to safely conduct internet measurement experiments without risking harm to real users[Janson].

2.9. Analyze risk

The benefits of internet measurement should outweigh the risks. Consider auxiliary data (e.g. third-party data sets) when assessing the risks. Consider that while a privacy risk may not be immediately apparent or realisable, in the future increased computing power may then make something possible.

Example: A research project releases encrypted payloads as a method for minimising exposure of sensitive user data. However the encryption could be trivially broken in the future with typical increases in computing power.

3. Security Considerations

This document as a whole addresses user safety considerations for internet measurement studies, and thus discusses security considerations extensively throughout regarding collection and storage of user data.

4. IANA Considerations

This document has no actions for IANA.

5. Acknowledgements

Many of these considerations are based on those from the [TorSafetyBoard] adapted and generalised to be applied to internet research.

Other considerations are taken from the Menlo Report [Menlo] and its companion document [MenloReportCompanion].

Comments of several people on the mailing list was helpful, especially Marwan Fayed and Jeroen van der Ham.

6. Informative References

[netem] Stephen, H., "Network emulation with NetEm", April 2005.

[RFC2544] Bradner, S. and J. McQuaid, "Benchmarking Methodology for Network Interconnect Devices", RFC 2544, DOI 10.17487/RFC2544, March 1999, <https://www.rfc-editor.org/info/rfc2544>.

[TorSafetyBoard] Tor Project, "Tor Research Safety Board", <https://research.torproject.org/safetyboard/>.

[RFC4949] Shirey, R., "Internet Security Glossary, Version 2", August 2007, <https://www.rfc-editor.org/info/rfc4949>.

[Tor.2017-04-001] Herm, K., "Privacy analysis of Tor's in-memory statistics", Tor Tech Report 2017-04-001, April 2017, <https://research.torproject.org/techreports/privacy-in-memory-2017-04-28.pdf>.

[Menlo] Dittrich, D. and E. Kenneally, "The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research", August 2012, <https://www.dhs.gov/sites/default/files/publications/CSD-MenloPrinciplesCORE-20120803_1.pdf>.

[MenloReportCompanion] Bailey, M., Dittrich, D., and E. Kenneally, "Applying Ethical Principles to Information and Communication Technology Research", October 2013, <https://www.impactcybertrust.org/link_docs/Menlo-Report-Companion.pdf>.

[DifferentialPrivacy] Dwork, C., McSherry, F., Nissim, K., and A. Smith, "Calibrating Noise to Sensitivity in Private Data

Analysis", 2006, <https://link.springer.com/chapter/10.1007/11681878_14>.

[SurveyNetworkTrafficAnonymisationTech] Van Dijkhuizen, N. and J. Van Der Ham, "A Survey of Network Traffic Anonymisation Techniques and Implementations", May 2018, <https://dl.acm.org/doi/10.1145/3182660>.

[ZMap] University of Michigan, "ZMap Source Code - packet.c", <https://github.com/zmap/zmap/blob/main/src/probe_modules/packet.c>.

[RFC6973] Cooper, A., Tschofenig, H., Aboba, B., Peterson, J., Morris, J., Hansen, M., and R. Smith, "Privacy Considerations for Internet Protocols", RFC 6973, July 2013, <https://www.rfc-editor.org/info/rfc6937>.

[SIGCOMM] Jones, B., Ensafi, R., Feamster, N., Paxson, V., and N. Weaver, "Ethical Concerns for Censorship Measurement", August 2015, <http://conferences.sigcomm.org/sigcomm/2015/pdf/papers/nsethics/p17.pdf>.

[RFC9511] Vyncke, É., Donnet, B., and J. Iurman, "Attribution of Internet Probes", November 2023, <https://www.rfc-editor.org/info/rfc9511>.

[Allman] Allman, M. and V. Paxson, "Issues and Etiquette Concerning Use of Shared Measurement Data", October 2007, <https://conferences.sigcomm.org/imc/2007/papers/imc80.pdf>.

[caida] CAIDA, "Promotion of Data Sharing", January 2010, <https://www.caida.org/catalog/datasets/sharing>.

[Kenneally] Kenneally, E. and K. Claffy, "Dialing privacy and utility: a proposed data-sharing framework to advance Internet research", 2010, <https://www.caida.org/catalog/papers/2010_dialing_privacy_utility/dialing_privacy_utility.pdf>.

[Janson] Janson, R., Traudt, M., and N. Hopper, "Privacy-Preserving Dynamic Learning of Tor Network Traffic", 2010, <https://dl.acm.org/doi/pdf/10.1145/3243734.3243815>.

Authors' Addresses

Iain R. Learmonth
HamBSD

Email: irl@hambsd.org

Gurshabad Grover
Centre for Internet and Society

Email: gurshabad@cis-india.org

Mallory Knodel
Center for Democracy and Technology

Email: mknodel@cdt.org