# Understanding the Ethical Frameworks of Internet Measurement Studies

Eric Pauley
University of Wisconsin–Madison
epauley@cs.wisc.edu

Patrick McDaniel
University of Wisconsin–Madison
mcdaniel@cs.wisc.edu

***Abstract*—Measurement of network data received from or transmitted over the public Internet has yielded a myriad of insights towards improving the security and privacy of deployed services. Yet, the collection and analysis of this data necessarily involves the processing of data that could impact human subjects, and anonymization often destroys the very phenomena under study. As a result, Internet measurement faces the unique challenge of studying data from human subjects who could not conceivably consent to its collection, and yet the measurement community has tacitly concluded that such measurement is beneficial and even necessary for its positive impacts. We are thus at an impasse: academics and practitioners routinely collect and analyze sensitive user data, and yet there exists no cohesive set of ethical norms for the community that justifies these studies. In this work, we examine the ethical considerations of Internet traffic measurement and analysis, analyzing the ethical considerations and remediations in prior works and general trends in the community. We further analyze ethical expectations in calls-for-papers, finding a general lack of cohesion across venues. Through our analysis and recommendations, we hope to inform future studies and venue expectations towards maintaining positive impact while respecting and protecting end users.**

## I. Introduction

Works in Internet measurement, specifically those making use of Internet telescopes, make distinct contributions to our understanding of the security of deployed systems. By collecting traffic from malicious or other sources, and aggregating or analyzing this traffic, we can understand trends in emerging threats, user behavior, and network design for performance and security. As a result, Internet measurement using telescopes has enjoyed a long history of well-regarded works.

In recent years, the security community generally, and the measurement community specifically, has increasingly discussed the ethical implications of their work. Indeed, previous works [13], [15], [23], [26] have highlighted several unique challenges that exist within the measurement space, and recommended that measurement researchers take care to discuss the ethical implications of their work during submission. While measurement venues have responded through the expectation of an ethics section on such works, there is as of yet no cohesive understanding of what constitutes ethical behavior in the measurement space.

Internet measurement for security faces issues under existing governance structures due to its scope and data diversity: disclosure of measurement to all parties is not practical, and the the free-form nature of network data means that even identifying involved parties is often intractable. At the same time, data from benign end-users is inevitably included in this analysis, and it has been increasingly shown that this end-user data can contain sensitive information, even when the telescope does not solicit it. As a result, existing ethical frameworks such as that of informed consent [22] cannot be directly applied to Internet measurement.

The measurement community has largely relied on existing governance structures (such as institutional review boards (IRB) under United States law) to exempt studies, either as not dealing with human subjects or as mitigating potential harms [26]. While this fulfills ethical requirements of institutions and some venues, IRB alone is specifically ill-equipped to establish the harm done by measurement studies for two reasons: (1) IRB may fail to identify harms to humans that are a result of emergent properties of the network under study, and (2) the risk of such harms may be acceptable due to the overall benefits of the work, even though these harms would disqualify the study under IRB rules.

We are thus faced with an existential challenge for Internet measurement: collection methodology, if interpreted in a more complete and pessimistic way, may not meet the requirements of existing ethical practices codified at institutions. Yet, these measurement studies have undoubtedly benefited society in their positive impacts on security of deployed systems. There is increasing need for an ethical framework for these studies that establishes bounds on data collection and analysis that are sensitive to the needs of the measurement community, while maximizing benefits and minimizing harms to end users.

In this work, we examine the ethical implications of Internet measurement for security, specifically the collection and analysis of Internet traffic by telescopes and related means. We first begin by analyzing existing works on measurement ethics, developing a space of *parties*, *benefits*, and *harms* that apply to Internet measurement research. We then examine 10 published Internet measurement papers with ethical considerations, and seek to establish a consensus on norms with respect to these benefits and harms.

In tandem with our analysis of published works, we also characterize the ethical expectations of the community. We examine the ethics sections of calls-for-papers in 8 measurement and security venues. We find highly disparate expectations, differing levels of sophistication and clarity, and an overall lack of cohesive standards for the community. We expect this lack of cohesive clarity across venues may be contributing to the varied behavior of accepted submissions. From these observations,

Workshop on Ethics in Computer Security (EthiCS) 2023
27 February 2023, San Diego, CA, USA
ISBN 1-891562-85-1
https://dx.doi.org/10.14722/ethics.2023.239547
www.ndss-symposium.org

we discuss trends towards an ethical framework for Internet measurement, and make recommendations for authors and venues to improve their postures.

We anticipate that, through feedback and discussion on our recommendations, the community might achieve a cohesive vision for future ethical Internet measurement.

## II. Background: Ethical Measurement

Because of past negative outcomes, the security community has increasingly focused on the ethical implications of their works. Documents such as the Menlo report [22] (based on the original Belmont report [28]) form a basis of this analysis for security research in general, focusing on four key priciples: respect for persons (consent), beneficence (benefits and harms), justice (equity), and accountability. Works in security generally cite these principles, and can also operate under their respective Institutional Review Board (IRB) to ensure compliance with human subjects research requirements. For instance, a user study on end-user or system administrator security postures readily fits within this framework.

### A. Existing Structures: IRB

Requirements for Institutional Review Boards (IRB) are stipulated by the US federal government as a requirement for various funded medical research [1]. However, universities and professional groups have (nearly universally) extended this federal mandate to a general expectation in other fields, and similar governance structures exist in other geographies (for convenience this paper uses IRB to refer to institutional ethics boards generally). IRB is designed to regulate the studying of human subjects, and research can generally fall into two groups: non-exempt (generally research that directly measures human subjects in a medical setting) and exempt. Studies can broadly be exempted either because, in the board's view, the study does not collect data from human subjects, or it falls into a set of pre-determined exemption categories [1]. These exemptions require, for instance, that data be anonymized or that non-anonymized data could have no conceivable means of harming the human subjects. Because IRB members are generally not experts in network measurement, several scenarios can cause studies to be approved despite potentially not meeting exemption requirements:

- *Failing to identify human subjects.* Studied network endpoints may be representative of the non-public behavior of a human subject, including subjects only incidentally included in the study (U.S. law makes no exemption for incidental participants). For instance, measurement of Internet scans could also contain legitimate and sensitive traffic from end users due to misconfigurations.
- *Incomplete or missing anonymization.* Collected data may be traceable back to individuals because of the structure of the Internet or data collection, in ways not apparent to reviewers. In one case, for instance, Narayanan et al. [25] showed that statistical techniques could be used to infer personal information from ostensibly anonymized datasets. Web browsers also pose a challenge for anonymization, as metadata can often be used to fingerprint individual users [19].
- *Hidden harms.* Disclosure of collected data could cause harm to individuals in non-obvious ways. For instance, disclosure of scanning IP addresses could reveal human-owned devices with security vulnerabilities.

**Other exemptions to IRB.** Several other routes can also be taken towards an IRB exemption, as is often done in medical fields. For instance, study of dead human subjects is explicitly exempted, as are measurements taken from the publicly-visible behavior of subjects, so long as this behavior is anonymized or disclosure could not harm individuals. While these exemptions can enable a variety of studies that would otherwise not be possible without informed consent, they are not especially compelling for network measurement, as live users' engagement through the network is non-public.

### B. Ethical Challenges in Internet Measurement

While the principles from the Menlo report and IRB readily apply to much of security research, network measurement poses unique challenges with respect to consent and beneficence. Measurement studies can include many thousands or millions of users [26], and may cause unmitigatable (if only remotely possible) harms due to the structure of networked systems. Faced with these challenges, works have considered how measurement methodologies can comply with the spirit of ethical principles, especially with respect to data collection, impact, and anonymization [15].

Ethical measurement largely hinges on the types of data being collected (beneficence), and parties involved (consent). Khan et al. [23] discuss a variety of data types acquired towards measurement endpoints, and note that the sensitivity of these can vary. Partridge and Allman [26] also discuss this phenomenon, concluding that measurement papers should tune their ethics discussions to the nature and sensitivity of their data collection (e.g., low-sensitvity non-anonymous data is acceptable if properly protected). Specific phenomena can also limit the ability of measurement studies to inform end-users of collection: for instance, disclosing measurement of illegal activity could cause users to obfuscate their network traffic [32].

When possible, anonymization before analysis [23] or prior to publication and data sharing [13] can mitigate harms to measured parties. When anonymization is performed at collection-time, the data cleanly complies with existing ethical principles laid out in IRB guidelines, effectively being the equivalent of commercially-obtained biological specimens. In this way, it can be argued academics are shielded from responsibility for harms due to disclosure of individual's data, though larger-scale harms can still occur (for instance, adversarial exploitation of discovered vulnerabilities). Similarly to biological studies, however, anonymization can hide the very phenomena under study [15], and research on these must inherently analyze non-anonymous data. When anonymization is not possible, researchers can take steps to protect sensitive data access during analysis and anonymize results for publication.

Based on these insights from prior works, conferences have increasingly expected authors to enumerate and justify their ethical decisions when submitting works involving or impacting human subjects (Table I). While such considerations allow reviewers and readers to evaluate each work's individual decision, these decisions are still made in an ad-hoc manner. By considering the overall space of ethical decisions made by authors, we can more accurately understand the views of the community on measurement ethics.

TABLE I. MAJOR SECURITY AND MEASUREMENT VENUES AND CONFERENCE YEARS WHEN ETHICAL CONSIDERATIONS WERE FIRST MENTIONED IN CALLS-FOR-PAPERS. ADDITIONAL COLUMNS DESCRIBE ATTRIBUTES OF THE MOST RECENTLY-PUBLISHED CFP.

| Conference | Ethics in CFP since | Latest CFP[8] | IRB[1] | Impact[3] | Disclosure[4] | Legal[5] | REC[6] | Framework[7] |
|---|---|---|---|---|---|---|---|---|
| ACM IMC | 2009 [6] | 2022 | ● | ● | ○ | ○ | ○ | Belmont [28] (B/C) |
| USENIX Security | 2013 [7] | 2023 | ●[2] | ● | ● | ○ | ● | Menlo [22] (B) |
| NDSS | 2015 [8] | 2023 | ● | ○ | ● | ● | ○ | |
| ACM CCS | 2017 [10] | 2022 | ● | ○ | ● | ○ | ○ | |
| ACM ASIACCS | 2017 [9] | 2023 | ○ | ○ | ● | ● | ○ | |
| IEEE S&P | 2017 [11] | 2023 | ●[2] | ○ | ○ | ● | ● | |
| IEEE EuroS&P | 2017 [5] | 2023 | ●[2] | ○ | ● | ○ | ○ | Menlo [22] (B) |
| ACM SIGMETRICS | 2018 [12] | 2023 | ●[2] | ○ | ○ | ○ | ○ | Menlo [22] (B/C) |
| ACSAC | 2021 [2] | 2022 | ● | ● | ● | ○ | ○ | |

[1] Require IRB or equivalent when potentially relevant [2] Emphasize that IRB is necessary but not sufficient [3] Discuss possibility of unforeseen impacts [4] Discuss disclosure of vulnerabilties [5] Discuss legal issues [6] Research Ethics Committee [7] Cites an ethical framework (B=beneficence, C=consent) [8] Analyzed text is provided in the appendix.

## III. STUDYING MEASUREMENT ETHICS IN PRACTICE

In response to community discussions on ethics in measurement and security broadly, major venues and other have instituted requirements to add discussion of ethical considerations to paper submissions, and program committees (PCs) evaluate papers on their ethical as well as technical merits. As papers have been submitted and accepted by PCs in subsequent years, reviewing these papers provides insight into community norms on ethical measurement. To this end, we collect 10 papers in Internet measurement (outlined in Table II) from the past 5 years that were submitted to and accepted to conferences that specifically highlighted ethical considerations in the call-for-papers. We consider the ethical models of each paper with respect to parties studied, consent received, and data collection/analysis.

For each paper, we determine the set of parties measured, both intentionally and as a byproduct of the measurement. We examine the types of data collected, anonymization techniques (both during analysis and for publication). Next, we study how measurement techniques can impact end-users during collection, and the extent to which end-users can opt in/out of the study. Our examined works are distributed broadly across these dimensions, yet demonstrate ethical boundaries in the community that may be fruitful for discussion.

*Note:* While our work interprets the ethical decisions made by each work, we do not wish to pass judgment on the resulting decisions of authors or reviewers. Rather, this work aims to identify *de facto* ethical norms in in the field and make recommendations towards adopting or improving on these norms. While individual papers are not anonymized, the conclusions of this would should not be taken as evaluations of individual author choices. To this end, we refer to analyzed papers by their reference numbers without author names.

### A. Measured Parties

While each studied work generally targets measurement of a specific phenomenon, this can often span across behaviors of multiple parties. Further, collection methodologies that target one party (such as scanners) can unintentionally collect data sent by end users.

*1) Measuring scanners:* Of the studied papers, seven studied the activities of Internet scanners to some end. Each achieved this by exposing collection endpoints on some publicly-routable IP address, and monitored/responded to incoming traffic. Traffic from end-users (i.e. networks operated by users that do not themselves intend to scan) can be received by these endpoints for a variety of reasons:

1) *Configuration*. Clients could be configured to connect due to services (a) deployed, or (b) previously deployed at the IP address. For instance, [29] measures traffic at IP addresses that are also used for legitimate purposes, but filters out these ports from analysis to avoid collecting end-user data. [18], [20], [21] can also hypothetically receive traffic due to previously-deployed services. [27] explicitly seeks to understand the effects of previously deployed services, and so receiving this traffic is by-design. When configuration causes end-user connections, it is often not possible to soundly distinguish from scanner traffic.
2) *Client infection*. Clients may be inadvertently co-opted as scanners through the use of malware, such as Mirai [14]. In these cases, scan traffic is often sourced from residential IP addresses with infected devices. While all examined papers studying scanners could also receive this data, [18] explicitly isolates and analyzes these end-user IPs. Without anonymization, sharing these addresses could leave vulnerable systems subject to targeted attacks.

When measuring scanners, traffic from end-users can inevitably be collected. Papers collecting this data generally focus on the impact of this collection (subsection III-D), rather than the incidental collection of information. From this, we conclude that studies of scanners can acceptably focus on their main study goals, so long as legitimate traffic is not purposefully elicited and reasonable effort is taken to protect data.

In these works, we also see a tacit assumption that measuring the scanners themselves is not an ethical issue. Ostensibly, these scanners are an aggregate and unavoidable phenomenon on the public Internet. However, scanners are inevitably being designed and operated by individuals and organizations, many of which would likely explicitly not consent to measurement. Further, disclosing personal details on scanner operators would expose those users to personal or legal harm, which expressly violates IRB exemptions.

The legal basis of such work may provide some insights. Consider the regulatory definition of private information [1]:

> Private information includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation

TABLE II. EXAMINED WORKS IN INTERNET MEASUREMENT.
EACH PAPER COLLECTS AND ANALYZES CERTAIN DATA COLLECTED FROM USERS THROUGH THE INTERNET, SUBJECT TO ETHICS RESTRICTIONS OF MAJOR VENUES.

| Ref | Venue | Vantage Point | Data Collected | Target Parties | Incidental Parties | Ethics Sec. | Anon.[1] | Impact[2] |
|---|---|---|---|---|---|---|---|---|
| [20] | ASIACCS '18 | Campus Net | Transport-Layer | Scanners | End-Users | ○ | ◐ | ○ |
| [17] | IMC '19 | DNS Resolver | DNS Queries | Recursive Resolvers | End-Users | ● | ● | ○ |
| [29] | IMC '19 | CDN IPs | Transport Layer | Scanners | | ○ | ◐ | ○ |
| [18] | CCS '21 | Cloud IPs | DDoS Traffic | Scanners | End-Users | ● | ◐ | ● |
| [21] | SEC '21 | Cloud IPs | Application Layer | Scanners | End-Users | ● | ◐ | ◐ |
| [16] | EuroS&PW '22 | Campus Net | Application Layer | Scanners | | ○ | ◐ | ◐ |
| [24] | SEC '22 | Container Registries | Download counts | End-Users | | ● | ● | ◐ |
| [27] | S&P '22 | Cloud IPs | Application Layer | Scanners, End-Users | | ● | ◐ | ●[3] |
| [30] | IMC '22 | Web Browser | Aggregate Browsing Behavior | End-Users | | ● | ● | ○ |
| [31] | IMC '22 | Darknet | Passive IP + DNS | Scanners, DNS Servers | | ● | ◐ | ●[3] |

[1] ◐ anonymized for publication ● anonymized at collection [2] ◐ reactive to inbound traffic ● probing/outbound [3] Outbound DNS queries

> or recording is taking place, and information that has been provided for specific purposes by an individual and that the individual can reasonably expect will not be made public (e.g., a medical record).

Based on these requirements, we're faced with a seemingly-simple question: *is measuring scanners without consent acceptable?* One might argue that malicious scanners are targeting deployed systems, and would therefore reasonably expect that their activity would be recorded and analyzed for security purposes. As such, personal details inferred from scanning activity are arguably not considered private information and the scanner does not represent a human subject for purposes of IRB. Additionally, an official interpretation [4] suggests that a bot itself is not a human subject, though data provided by a bot could easily contain personal information from other subjects. That being said, because scanners can also be deployed on end-user systems, there is an open question of whether the parties scanning are actually malicious and have an expectation of recording. Through this, we observe that an act as seemingly harmless as measuring scanning traffic can have ethical pitfalls when working under existing frameworks. The community must reach consensus on an ethical basis for this measurement, as existing works do not address these issues.

*2) Targeting end-users:* Internet measurement papers also aim to characterize the behaviors of end users. By measuring end users, studies can better understand user behavior, but also infer the performance of Internet resources and resulting user experiences. Of our studied works, three targeted end-users for study. Here, we see more involved efforts within the papers themselves to characterize ethical implications: all end-user-targeting works had ethics discussion, and two of those [27], [30] included concrete discussions of data management practices for collection, storage, anonymization, and analysis of data (the third by construction did not have access to non-anonymized information [24]).

**Takeaways & Recommendations:** From these works, we can see that end-user involvement in measurement is often an inevitable or even desirable phenomenon. Protections for these parties should be far more strict than would be applied to bot or scanner traffic, especially when receiving this data is explicitly part of the measurement design. Even when scanners are measured, it should be assumed that end-user data could be received unless countermeasures are taken to prevent this, and best practices should be applied to protect this collected data.

Multiple techniques can be employed to protect end-users from undue involvement in measurement studies. Telescopes can be deployed in IP ranges less likely to receive legitimate traffic (e.g., data received by cloud telescopes is inherently more sensitive due to latent configuration [27]). Telescopes deployed on configured IPs can take advantage of that to filter legitimate traffic [29], or signatures of residential IPs or known botnets could be used to drop human subject data.

When end-user data is received, encryption and access control techniques can be enumerated in the paper to demonstrate protection of user data. Additionally, steps can be taken post-collection to filter traffic that is likely end-users. For instance, [27] identifies individual users based partially on overall IPs or ports contacted. While in this case the technique is used to isolate and study misconfigurations, it could be applied in reverse to analyze only likely scanner traffic.

### B. Types of data collected

As measurement studies vary in their endpoints, we likewise see variation in the types of data collected: layer 7 (application) traffic and IP- or transport-layer metadata are the most commonly collected (9 of 10 papers in some form), with some targeting specific subsets of these (e.g., DNS queries).

*1) Application-layer data:* Works have collected application-layer data to measure misconfigurations [27], targeted vulnerabilities [16], [21], and core Internet performance [17]. While collection of such data at major venues has recently required ethical discussion, collection that includes end-user data is acceptable so long as steps are taken to reduce impact (subsection III-D). For instance, [27] describes processes for encrypted data storage and access control to protect sensitive data, and [21] only incidentally receives end-user application data so follows standard best practices for data protection.

*2) Metadata:* In some cases, only metadata about network flows are collected, with expectations for controls on data collection and management being lower. For instance, [20] records and analyzed TCP flow tuples of inbound and outbound traffic on a campus network. While this study is intended to monitor scanning behavior, the collection also incidentally measures large amounts of end-user metadata. Another work, [29], collects flow tuples of scanner traffic on IPs that also receive and process legitimate traffic, though steps are taken to ensure that only illegitimate scanner traffic is collected. This paper does not contain an ethics section, highlighting that metadata collection of scanner traffic with only incidental user metadata does not constitute an ethical concern to the authors.

**Takeaways & Recommendations:** Ethical expectations differ depending on the type of data collected, though some key overall trends emerge. First, collected data should be the minimal required: works that collect application-layer data present justifications for why this data is needed, and others tailor their analysis to function with limited data. Works can also take steps to filter out data that is unrelated to their study goals, for instance by extracting features from application data that are pertinent to the study instead of storing raw data.

### C. Anonymization

When collecting network data, there is an expectation that authors protect parties by anonymizing data. In every studied work, some level of anonymization was used in the published work. For instance, non-public vulnerable IPs are not disclosed, and data is presented in aggregate. However, there is a precedent for not fully anonymizing data when references are not to human subjects. For instance, [17], [27], [31] disclose vulnerable domain names and companies, though not specific vulnerable IP addresses unless otherwise publicly known. In other cases [17], [30], where the parties performing collection have some duty of care to parties, data is fully anonymized and aggregated during collection. In two cases [24], [30], data is indirectly collected and anonymized by a third-party before being received by researchers, ensuring the anonymity of involved parties.

**Takeaways & Recommendations:** From these works, we can see trends in acceptable anonymization of published results, as well as more stringent requirements for anonymization as the sensitivity of data or privileged collection access increase. Works should maximally anonymize their data as early as possible while maintaining study endpoints and staying within the bounds of accepted practice. As with the types of data collected, anonymization can be an opportunity for technical contribution of a work in addition to an ethical requirement. For instance, binning data during collection without reducing study accuracy other collection-time anonymization techniques could provide fruitful technical insights.

At the same time, anonymization techniques can have limitations [19], [25]. When anonymizing data, and especially when anonymizing for publication or dissemination, authors can fail to account for the limitations of their techniques. Here, program committees are faced with the challenge of evaluating the technical soundness of ethical approaches, a task that could potentially go overlooked and risk sensitive subject data.

### D. Impact on End-Users

Measurement studies that interact with or otherwise impact their surroundings can have negative effects on the subjects they measure, or other related parties. However, often a passive approach is not sufficient to fully characterize a phenomenon (for instance, application-layer TCP payloads cannot be collected non-interactively). Further, some works employ outbound traffic to measure additional data or to properly elicit adversarial behavior.

*1) Reactive measurement:* Reactive (i.e., interactive) measurement involves the use of vantage points that respond to inbound traffic, usually to elicit behavior that could not be measured otherwise. Reactive approaches have seen increasing use in recent years, with 5 studied works including some reactive component. Of these, [16], [21], [27] use this reactivity to collect application-layer payloads, with [16] additionally employing honeypot-type responses to elicit further behavior. Notably, [16]'s vantage point reduces the likelihood of interaction with end-users, and the protocols used would likely not cause users to submit sensitive data. In contrast, [21], [27] are deployed to cloud IP addresses, and so further interactivity could elicit legitimate end-users to submit sensitive information.

When end-users are incidental parties in reactive measurement, there could also be negative impacts on clients. For instance, an erroneous HTTP response could cause application errors. Both reactive studies on cloud IPs [21], [27], appear to cap interactivity at TCP session establishment, reducing the likelihood that clients could process erroneous data.

In one other case [24], clients receive responses not from a research apparatus, but from a public container registry. In this case, the authors seek to measure container registry typo-squatting, and so end-users download containers as a byproduct of the measurement. Here, the authors took steps to ensure the downloaded containers would not directly harm client systems, though they note that developer confusion and frustration are likely experimental outcomes. The authors deleted all deployed containers after study completion to prevent long-term impacts.

*2) Outbound traffic:* In some cases, works also used outbound traffic (other than in response to inbound traffic). In two cases [27], [31], this took the form of DNS lookups from public zones, with authors either not discussing ethical implications or concluding no likely harms. In the case of [18], outbound traffic was used the allow the telescope to realistically resemble a target for DDoS amplification, and so careful steps were taken to ensure use of the telescope by an adversary would not actually cause traffic amplification in practice.

**Takeaways & Recommendations:** By looking at examples of reactive and outbound Internet measurement in literature, we can see an important trend of carefully minimizing the negative impacts of the measurement. In the case of reactive measurement, reasonable care should be taken to minimize negative impacts on client systems, though there is precedent for not entirely eliminating this risk if potential harms to legitimate clients are low. In the case of outbound traffic to non-involved end-users, the standard is much more strict, with a studied work in this space introducing novel techniques to safeguard recipients of traffic from undue harm such as DDoS attacks.

One way that measurement studies can limit their impact is by comparing their study apparatus with a distribution of benign Internet participants. If telescope behavior is a subset of what one would expect from a legitimate service (even under unusual circumstances), it is less likely to cause harm to clients that are functional under this legitimate service. For instance, negotiating TCP sessions without sending application-layer replies is semantically equivalent to a frozen server application or network failure, scenarios that are unlikely to cause more harm to clients than a non-responsive service. When initiating outbound connections, rate-limiting should be discussed and mentioned as part of ethical considerations [18].

### E. Consent & Opt-Out

Because most of the studied works passively collect data from sending parties, there is no opportunity to inform clients

or provide opt-out. In the case of targeted collection on specific networks, such as [16], [20], data collection consistent is ostensibly received from the network operator, rather than from users. In one work in particular [30], which measures browsing behavior, the authors mention receiving opt-in consent from users of the Chrome browser. The Chrome User Experience Report (CrUX) [3] notes that this data collection requires an opt in to generalized data collection and history syncing by the end user, though this option is by default opted-in on new installations. Trends across these works demonstrate that collection of received traffic generally does not require consent or opt-out capability, but collection of traffic from otherwise non-participating end-users should be accompanied by some form of opt-out, and ideally explicitly informed consent.

**Takeaways & Recommendations:** Informed consent can be a complex process, especially when academics are not in full control of data collection. When working with third-party data, authors should be aware of the circumstances under which this data is collected, as this could have ethical pitfalls for the analysis work as well. Additional considerations for third-party data are well-discussed in literature [13], [26].

## IV. Ethical Expectations of Measurement Venues

Ultimately, the ethical postures of published papers are influenced by community norms, with authors largely being informed of these through the ethical requirements in calls-for-papers. To better understand this, we also look at language used in the ethical considerations sections at 8 major venues, finding large variation in stated expectations. Table I displays several attributes of ethics sections at the most recent CFPs for major conferences. While all imply some ethical duty of care (with all but ASIACCS explicitly calling out IRB as one such structure), the factors that are considered relevant for consideration vary. For instance, few venues emphasize the risks of indirect or unforeseen impacts from work. In contrast, more general concepts such as vulnerability disclosure are more comprehensively discussed in CFPs. In some cases, venues also explicitly reference a set of ethical principles such as the Menlo Report, incorporating those expectations by reference.

**Recommendations:** Disparate venue recommendations suggest a lack of cohesive expectations in security research generally, and especially in Internet measurement. One potential avenue here is to encourage sharing of best practices between PCs of major conferences. Indeed, several historical CFPs acknowledge other venues for adapted concepts. Alternately, this diversity may be evidence that a new document is needed establishing concrete norms for the field, based on experience from recent years. Increased cohesion and specificity of ethics sections may give researchers confidence in designing ethical measurement studies, and security works more broadly.

## V. Discussion

From our analysis of accepted Internet measurement papers with ethical implications, we can see a variety of trends. Generally speaking, data collection from scanners is acceptable, and data collection from end-users is acceptable if it is incidental, or if the data is anonymous or otherwise securely managed. Impacts on end-users are discouraged, but works that take appropriate measures to minimize these can still be acceptable if the impact is unavoidable. Yet, we also observe that works can be accepted despite not absolutely minimizing harm, for instance by not anonymizing before analysis when this is possible. We hypothesize that this may be due to constraints in the ability of ethics committees to give constructive feedback, or reviewers not being focused on ethical issues.

### A. Towards Improving Ethics Reviews

Because few works in this niche have been published under the scrutiny of modern standards at conferences, it is difficult to draw broad conclusions on the stance of the community from these works alone. Further, these works are only positive examples of methods acceptable to a set of reviewers, not negative examples of works deemed unacceptable. We foresee two key thrusts towards addressing these limitations: principled feedback from conference reviewers and the community at large, and analysis of works that did not pass ethical scrutiny.

Broader and more directed feedback from the community would improve our understanding of ethical expectations from Internet measurement. To this end, a survey of community and PC members about hypothetical Internet measurement scenarios could provide broader datapoints on when and why measurement is ethical. Such a survey might be composed of scenarios generated by practitioners that are plausible but incomplete studies.

Analysis of extant works that did not pass ethical review may also prove fruitful as negative examples towards understanding ethical norms. In most respects, the publication process encourages academics to continually push the boundaries of prior works. The exact opposite is true for ethics, where there is a need for concrete norms that can be consistently observed by the community. For this reason, we would also encourage conferences to publish anonymized information on papers rejected for ethical reasons when possible. Committees could request consent from authors for this or additionally request that authors write brief anonymous retrospectives on ethical failures. While some such examples on unacceptable work are publicly visible due to retractions, collaboration with ethics committees at conferences to more closely analyze trends in these could also identify new guidance for the community.

## VI. Conclusions

Internet measurement poses unique ethical challenges across data collection, processing, and user consent. That being said, this work also identifies inconsistencies in the ethical expectations of security venues generally. Here, we see an opportunity for venues to come to more cohesive and concrete expectations for authors. We also anticipate that this work, along with subsequent discussions, might lead to the development of a concrete ethical playbook for Internet measurement research. By improving ethical guidance, members of the community may be empowered to answer new measurement questions while protecting Internet users.

## References


[1] "45 CFR Part 46 (2018-07-19) – Protection of Human Subjects." [Online]. Available: https://www.ecfr.gov/on/2018-07-19/title-45/subtitle-A/subchapter-A/part-46

[2] "Call for Papers - Annual Computer Security Applications Conference (ACSAC)." [Online]. Available: https://www.acsac.org/2021/submissions/papers/

[3] "Chrome UX Report." [Online]. Available: https://developer.chrome.com/docs/crux/

[4] "Considerations for Research with the Internet." [Online]. Available: https://www.hhs.gov/ohrp/sachrp-committee/recommendations/considerations-for-research-with-the-internet/index.html

[5] "IEEE European Symposium on Security and Privacy 2017." [Online]. Available: https://www.ieee-security.org/TC/EuroSP2017/cfp.php

[6] "Internet Measurement Conference 2009," 2009. [Online]. Available: http://conferences.sigcomm.org/imc/2009/cfp.html

[7] "USENIX Security 2013 Call for Papers," 2013. [Online]. Available: https://www.usenix.org/sites/default/files/sec13cfp_111912.pdf

[8] "NDSS 2015 Call for Papers | Internet Society," 2015. [Online]. Available: http://www.internetsociety.org/events/ndss-symposium-2015/ndss-2015-call-papers

[9] "Call for Papers – ACM Asia Conference on Computer and Communications Security (ASIACCS) 2017," Nov. 2016. [Online]. Available: https://web.archive.org/web/20161119170123/http://asiaccs2017.com/calls/call-for-papers

[10] "ACM CCS 2017 | ACM CCS 2017 Website," 2017. [Online]. Available: https://ccs2017.sigsac.org/papers.html

[11] "IEEE Symposium on Security and Privacy 2017," 2017. [Online]. Available: https://www.ieee-security.org/TC/SP2017/cfpapers.html

[12] "Call for Papers - ACM SIGMETRICS 2018," 2018. [Online]. Available: https://www.sigmetrics.org/sigmetrics2018/call_for_papers.html

[13] M. Alllman and V. Paxson, "Issues and etiquette concerning use of shared measurement data," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement - IMC '07*. San Diego, California, USA: ACM Press, 2007, p. 135. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1298306.1298327

[14] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, and M. Kallitsis, "Understanding the mirai botnet," in *26th USENIX security symposium (USENIX Security 17)*, 2017, pp. 1093–1110.

[15] J. R. Crandall, M. Crete-Nishihata, and J. Knockel, "Forgive Us our SYNs: Technical and Ethical Considerations for Measuring Internet Filtering," in *Proceedings of the 2015 ACM SIGCOMM Workshop on Ethics in Networked Systems Research*. London United Kingdom: ACM, Aug. 2015, pp. 3–3. [Online]. Available: https://dl.acm.org/doi/10.1145/2793013.2793021

[16] T. Favale, D. Giordano, I. Drago, and M. Mellia, "What Scanners do at L7? Exploring Horizontal Honeypots for Security Monitoring," in *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, Jun. 2022, pp. 307–313, iSSN: 2768-0657.

[17] P. Foremski, O. Gasser, and G. C. M. Moura, "DNS Observatory: The Big Picture of the DNS," in *Proceedings of the Internet Measurement Conference*. Amsterdam Netherlands: ACM, Oct. 2019, pp. 87–100. [Online]. Available: https://dl.acm.org/doi/10.1145/3355369.3355566

[18] H. Griffioen, K. Oosthoek, P. van der Knaap, and C. Doerr, "Scan, Test, Execute: Adversarial Tactics in Amplification DDoS Attacks," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. Virtual Event Republic of Korea: ACM, Nov. 2021, pp. 940–954. [Online]. Available: https://dl.acm.org/doi/10.1145/3460120.3484747

[19] A. Gómez-Boix, P. Laperdrix, and B. Baudry, "Hiding in the crowd: an analysis of the effectiveness of browser fingerprinting at large scale," in *Proceedings of the 2018 world wide web conference*, 2018, pp. 309–318.

[20] H. Heo and S. Shin, "Who is knocking on the Telnet Port: A Large-Scale Empirical Study of Network Scanning," in *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, ser. ASIACCS '18. New York, NY, USA: Association for Computing Machinery, May 2018, pp. 625–636. [Online]. Available: https://doi.org/10.1145/3196494.3196537

[21] R. Hiesgen, M. Nawrocki, A. King, A. Dainotti, T. C. Schmidt, and M. Wählisch, "Spoki: Unveiling a New Wave of Scanners through a Reactive Network Telescope," in *31st USENIX Security Symposium (USENIX Security 22)*. Boston, MA: USENIX Association, Aug. 2022, pp. 431–448. [Online]. Available: https://www.usenix.org/conference/usenixsecurity22/presentation/hiesgen

[22] E. Kenneally and D. Dittrich, "The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research," *SSRN Electronic Journal*, 2012. [Online]. Available: http://www.ssrn.com/abstract=2445102

[23] M. T. Khan and C. Kanich, "High Fidelity, High Risk, High Reward: Using High-Fidelity Networking Data in Ethically Sound Research," in *Proceedings of the 2015 ACM SIGCOMM Workshop on Ethics in Networked Systems Research*, ser. NS Ethics '15. New York, NY, USA: Association for Computing Machinery, Aug. 2015, pp. 23–26. [Online]. Available: https://doi.org/10.1145/2793013.2793024

[24] G. Liu, X. Gao, H. Wang, and K. Sun, "Exploring the Unchartered Space of Container Registry Typosquatting," 2022, pp. 35–51. [Online]. Available: https://www.usenix.org/conference/usenixsecurity22/presentation/liu-guannan

[25] A. Narayanan and V. Shmatikov, "How To Break Anonymity of the Netflix Prize Dataset," Nov. 2007, arXiv:cs/0610105.

[26] C. Partridge and M. Allman, "Ethical considerations in network measurement papers," *Communications of the ACM*, vol. 59, no. 10, pp. 58–64, Sep. 2016. [Online]. Available: https://doi.org/10.1145/2896816

[27] E. Pauley, R. Sheatsley, B. Hoak, Q. Burke, Y. Beugin, and P. McDaniel, "Measuring and Mitigating the Risk of IP Reuse on Public Clouds," in *2022 IEEE Symposium on Security and Privacy (SP)*, May 2022, pp. 558–575, arXiv:2204.05122 [cs].

[28] O. f. H. R. Protections (OHRP), "The Belmont Report," Jan. 2010, last Modified: 2022-10-17T14:50:16-0400. [Online]. Available: https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html

[29] P. Richter and A. Berger, "Scanning the Scanners: Sensing the Internet from a Massively Distributed Network Telescope," in *Proceedings of the Internet Measurement Conference*, ser. IMC '19. New York, NY, USA: Association for Computing Machinery, Oct. 2019, pp. 144–157. [Online]. Available: https://doi.org/10.1145/3355369.3355595

[30] K. Ruth, A. Fass, J. Azose, M. Pearson, E. Thomas, C. Sadowski, and Z. Durumeric, "A world wide view of browsing the world wide web," in *Proceedings of the 22nd ACM Internet Measurement Conference*. Nice France: ACM, Oct. 2022, pp. 317–336. [Online]. Available: https://dl.acm.org/doi/10.1145/3517745.3561418

[31] R. Sommese, K. Claffy, R. van Rijswijk-Deij, A. Chattopadhyay, A. Dainotti, A. Sperotto, and M. Jonker, "Investigating the impact of DDoS attacks on DNS infrastructure," in *Proceedings of the 22nd ACM Internet Measurement Conference*. Nice France: ACM, Oct. 2022, pp. 51–64. [Online]. Available: https://dl.acm.org/doi/10.1145/3517745.3561458

[32] J. Van Der Ham, "Ethics and Internet Measurements," in *2017 IEEE Security and Privacy Workshops (SPW)*, May 2017, pp. 247–251.

## ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE1255832. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## APPENDIX

## CFP ETHICS REQUIREMENTS

For reference, the following sections are taken verbatim from the calls-for-papers of the associated conferences.

**ACM IMC 2022.** The program committee may raise concerns around the ethics of the work, even if it does not involve human subjects. All papers must include, in a clearly marked appendix section with the heading "Ethics", a statement about ethical issues; papers that do not include such a statement may be rejected. This could be, if appropriate for the paper, simply the sentence "This work does not raise any ethical issues.". If the work involves human subjects or potentially sensitive data (e.g., user traffic or social network information, evaluation of censorship, etc.), the paper should clearly discuss these issues, perhaps in a separate subsection.

Research that entails experiments involving human subjects or user data (e.g., network traffic, passwords, social network information) should adhere to community norms. Any work that raises potential ethics considerations should indicate this on the submission form. The basic principles of ethical research are outlined in the Belmont Report: (1) respect for persons (which may involve obtaining consent); (2) beneficence (a careful consideration of risks and benefits); and (3) justice (ensuring that parts of the population that bear the risks of the research also are poised to obtain some benefit from it). Authors should further consult the ACM policy on research involving human subjects for further information on ethical principles that apply to this conference.

Research involving human subjects must be approved by the researchers' respective Institutional Review Boards before the research takes place. Authors should indicate on the submission form whether the work involves human subjects. If so, the authors must indicate whether an IRB protocol has been approved for the research, or if the research has been determined exempt (either self-determination or IRB determination). We expect that any research follows the practices and procedures of the institution(s) where the work

is being carried out; for example, some universities require separate approval for the use of campus data. We expect researchers to abide by these protocols.

We recognize that different IRBs follow different procedures for determining the status of human subject research, and approval or exempt status from a single institution may not align with community norms. To help the Ethics Committee review cases of concern, there is a need for more information about the research protocol. To this end, if the work involves human subjects, the authors must include with their submission a copy of the form that was used to determine IRB status (approved or exempt), sufficiently anonymized to preserve double-blind review.

If the submission describes research involving human subjects and none of the authors are at an institution with an IRB (or equivalent), the authors are nonetheless expected to follow a research protocol that adheres to ethical principles, as stated in the ACM policy on research involving human subjects. In such cases, the authors must use the Ethics section of their appendix to explain how their research protocol satisfies the principles of ethical research.

Some research does not involve human subjects yet nonetheless raises questions of ethics, which may be wide-ranging and not necessarily limited to direct effects. We encourage authors to be mindful of the ethics of the research that they undertake; these considerations are often not clear-cut, but often warrant thoughtful consideration. Discussions of these issues should be placed in the "Ethics" appendix section mentioned above, or in the main body of the paper where appropriate.

Additionally, the program committee reserves the right to conduct additional evaluations and reviews of research ethics and reserves the right to independent judgment concerning the ethics of the conducted research.

**USENIX Security 2023.** We expect authors to carefully consider and address the potential harms associated with carrying out their research, as well as the potential negative consequences that could stem from publishing their work. Failure to do so may result in rejection of a submission regardless of its quality and scientific value.

Although causing harm is sometimes a necessary and legitimate aspect of scientific research in computer security and privacy, authors are expected to document how they have addressed and mitigated the risks. This includes, but is not limited to, considering the impact of your research on deployed systems, understanding the costs your research imposes on others, safely and appropriately collecting data, and following responsible disclosure. In particular, if the submission deals with vulnerabilities (e.g., software vulnerabilities in a given program or design weaknesses in a hardware system), the authors need to discuss in detail the steps they have already taken or plan to take to address these vulnerabilities (e.g., by disclosing vulnerabilities to the vendors).

Papers should include a clear statement about why the benefit of the research outweighs the harms, and how the authors have taken measures and followed best practices to ensure safety and minimize the potential harms caused by their research.

Due to the complexity of today's computing systems, humans can be harmed directly or indirectly in unexpected ways (see The Menlo Report at [URL]). If the submitted research has potential to cause harm, and authors have access to an Institutional Review Board (IRB), we encourage authors to consult this IRB and document its response and recommendations in the paper. We note, however, that IRBs are not expected to understand computer security research well or to know about best practices and community norms in our field, so IRB approval does not absolve researchers from considering ethical aspects of their work. In particular, IRB approval is not sufficient to guarantee that the PC will not have additional concerns with respect to harms associated with the research.

**NDSS 2023.** If a paper relates to human subjects, analyzes data derived from human subjects, may put humans at risk, or might have other ethical implications or introduce legal issues of potential concern to the NDSS community, authors should disclose if an ethics review (e.g., IRB approval) was conducted, and discuss in the paper how ethical and legal concerns were addressed. If the paper reports a potentially high-impact vulnerability the authors should discuss their plan for responsible disclosure. The chairs will contact the authors in case of concerns. The Program Committee reserves the right to reject a submission if insufficient evidence was presented that ethical or relevant legal concerns were appropriately addressed.

**ACM CCS 2022.** For papers that might raise ethical concerns, authors are expected to convince reviewers that proper procedures (such as IRB approval or responsible disclosure) have been followed, and due diligence has been made to minimize potential harm.

**ACM ASIACCS 2023.** The authors should take care of clarifying any potential ethical and legal concerns to their results, highly critical vulnerabilities or exploits, etc. The authors should provide evidence that they have thoroughly considered such issues.The Program Committee reserves the right to reject a submission if insufficient evidence was presented that ethical or relevant legal concerns were appropriately addressed.

**IEEE S&P 2023; Ethical Considerations for Vulnerability Disclosure.** Where research identifies a vulnerability (e.g., software vulnerabilities in a given program, design weaknesses in a hardware system, or any other kind of vulnerability in deployed systems), we expect that researchers act in a way that avoids gratuitous harm to affected users and, where possible, affirmatively protects those users. In nearly every case, disclosing the vulnerability to vendors of affected systems, and other stakeholders, will help protect users. It is the committee's sense that a disclosure window of 45 days to 90 days ahead of publication is consistent with authors' ethical obligations.

Longer disclosure windows (which may keep vulnerabilities from the public for extended periods of time) should only be considered in exceptional situations, e.g., if the affected parties have provided convincing evidence the vulnerabilities were previously unknown and the full rollout of mitigations requires additional time. The authors are encouraged to consult with the PC chairs in case of questions or concerns.

The version of the paper submitted for review must discuss in detail the steps the authors have taken or plan to take to address these vulnerabilities; but, consistent with the timelines above, the authors do not have to disclose vulnerabilities ahead of submission. If a paper raises significant ethical and/or legal concerns, it will be checked by the REC and it might be rejected based on these concerns. The PC chairs will be happy to consult with authors about how this policy applies to their submissions.

**Ethical Considerations for Human Subjects Research.** Submissions that describe experiments that could be viewed as involving human subjects, that analyze data derived from human subjects (even anonymized data), or that otherwise may put humans at risk should:

Disclose whether the research received an approval or waiver from each of the authors' institutional ethics review boards (IRB) if applicable. Discuss steps taken to ensure that participants and others who might have been affected by an experiment were treated ethically and with respect.

If a submission deals with any kind of personal identifiable information (PII) or other kinds of sensitive data, the version of the paper submitted for review must discuss in detail the steps the authors have taken to mitigate harms to the persons identified. If a paper raises significant ethical and/or legal concerns, it will be checked by the REC and it might be rejected based on these concerns. The PC chairs will be happy to consult with authors about how this policy applies to their submissions.

**IEEE EuroS&P 2023.** We expect authors to carefully consider and address the potential harms associated with carrying out their research, as well as the potential negative consequences that could stem from publishing their work. Failure to adequately discuss such potential harms within the body of the submission may result in rejection of a submission, regardless of its quality and scientific value.

Although risking to cause harm is sometimes a necessary and legitimate aspect of scientific research in computer security and privacy, authors are expected to document how they addressed and mitigated such risks. This includes, but is not limited to, considering the impact of the research on deployed systems, understanding the costs the research imposes on others, safely and appropriately collecting data, and following responsible disclosure practices. Papers should include a clear statement as to how the benefit of the research outweighs the potential harms, and how the authors have taken measures and followed best practices to ensure safety and minimize the potential harms caused by their research.

If the submitted research has potential to cause harm, and authors have access to an Institutional Review Board (IRB), we expect that this IRB was consulted appropriately and that its approval and recommendations are documented in the paper. We note that IRBs are not necessarily well-versed in computer security research and may not know the best practices and community norms in our field, so IRB approval does not absolve researchers from considering ethical aspects of their work. In particular, IRB approval is not sufficient to guarantee that the PC will not have additional concerns with respect to harms associated with the research.

We encourage authors to consult existing documentation, e.g., Common Pitfalls in Writing about Security and Privacy Human Subjects Experiments, and How to Avoid Them or the Menlo Report and existing Safety consultation entities, e.g., the Tor Safety Research Board. These can help in thinking about potential harms, and in designing the safest experiments and disclosure processes.

**ACM SIGMETRICS 2023.** Papers describing experiments with users or user data (e.g., network traffic, passwords, social network information), should follow the basic principles of ethical research, e.g., beneficence (maximizing the benefits to an individual or to society while minimizing harm to the individual), minimal risk (appropriateness of the risk versus benefit ratio), voluntary consent, respect for privacy, and limited deception. When appropriate, authors are encouraged to include a subsection describing these issues. Authors may want to consult the Menlo Report for further information on ethical principles, or the Allman/Paxson IMC '07 paper for guidance on ethical data sharing.

Authors must, as part of the submission process, attest that their work complies with all applicable ethical standards of their home institution(s), including but not limited to privacy policies and policies on experiments involving humans. Note that submitting research for approval by one's institution's ethics review body is necessary, but not sufficient—in cases where the PC has concerns about the ethics of the work in a submission, the PC will have its own discussion of the ethics of that work. The PC's review process may examine the ethical soundness of the paper just as it examines the technical soundness.

As a published ACM author, you and your co-authors are subject to all ACM Publications Policies, including ACM's new Publications Policy on Research Involving Human Participants and Subjects. In particular, authors must follow the basic research and publication principle outlined by the ACM Publication Board's Policies and Procedures and the ACM Code of Ethics and Professional Conduct. Relevant policies regarding the publication and review processes include ACM's policies (i) on the roles and responsibilities in ACM publishing, (ii) on the coercion and abuse in the ACM publications process, and on plagiarism, misrepresentation, and falsification.

**ACSAC 2022.** Papers that might raise ethical concerns (e.g., papers that use human subjects or describe experiments related to vulnerabilities in software or systems) must include an Ethical Considerations section that properly describes what procedures have been followed to minimize potential harm. Such papers should discuss the steps taken to avoid negatively affecting any third-parties, whether an institutional ethics committee reviewed the research, or how the authors plan to responsibly disclose the vulnerabilities to the appropriate software/system vendors or owners before publication.