Crisis, Ethics, Reliability & a `measurement.network`

Reflections on Active Network Measurements in Academia

Tobias Fiebig
Max-Planck-Institut für Informatik
Saarbrücken, Germany
tfiebig@mpi-inf.mpg.de

ABSTRACT

Network measurements are a necessary component of assessing real-world protocol use to inform the development of new and improvement of old protocols and standards. However, especially active measurements, i.e., measurements in which probes are sent to remote devices to illicit a response, face ethical challenges, are difficult to execute reliably, and may cause unintended harm.

In this paper, we reflect on the connection between the Internet's growing complexity, the practicalities of academic research, and the likelyhood of reliability issues and unintended harm occurring in active measurements. We argue that communal infrastructure providing measurement services to the academic community could be a path forward to improve reliability and accessibility, while reducing the potential for unintended harm, and enabling PhD students to more easily draw from the experience of industry professionals.

CCS CONCEPTS

• **Networks** → **Network measurement**; • **Social and professional topics** → *Codes of ethics*; *Testing, certification and licensing*; • **Security and privacy** → *Network security*.

KEYWORDS

Network measurements, relaibility, reproducibility, open science

ACM Reference Format:
Tobias Fiebig. 2023. Crisis, Ethics, Reliability & a `measurement.network`: Reflections on Active Network Measurements in Academia. In *Applied Networking Research Workshop (ANRW '23), July 22–28, 2023, San Francisco, CA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3606464.3606483

1 INTRODUCTION

Measuring networks and how systems are being used in practice is an integral support in developing *new* and improving old protocols. Network measurements are usually conducted either by industry professionals, often directly targeted at a specific problem, or by researchers in a pursuit of documenting and understanding how the Internet works [21].

In that context, especially active measurements, i.e., network measurements where probes are sent out to illicit a response for the measurements, are often discussed critically, especially if they are run against major parts of the Internet. Despite efforts to establish ethical guidelines for conducting such measurements, e.g., with the Menlo report [2], and best-practices to make probes attributable [26], these may often still be perceived as a nuisance at best. If, due to an implementation error or due to an operator's oversight, measurements with a benign intention cause harm by crashing or overloading systems, the response from the operations community is often–understandably–harsh.

Besides these obvious challenges in network measurements, especially the interaction between academics–usually *not* running systems–and operators–running the systems being measured–can become difficult due to challenges inherent to how academic research tends to work in practice. Together, these interaction effects lead to circumstances that threaten the reproducibility, reliability, and ethics of–especially–active network measurement research, while also leading to instances of measurements affecting network operations at large, negatively affecting the connection between academic research and operations.

Contributions: Here, we make the following contributions:

- We self-critically analyze the practical conditions of academic network measurement research, considering how practicalities inflict on ethics, reliability, and reproducibility of active network measurements in academia.
- From our analysis, we derive challenges we, as a community, have to overcome.
- Finally, we propose a solution in the form of an open measurement platform and organization.

Disclaimer: In this paper, we make bold and critical statements on the way academia and research works. We do this self-critically, and ensure the reader that these points do not come from an intention of pointing fingers at others, but have been learned painfully through experience while running network measurements. For the very same reason, i.e., to avoid unjustly pointing fingers for things we all might have done, we will refrain from providing specific citations for examples we use; While some illustrative examples may be recognizable despite a lack of citation, we deeply encourage the reader to reflect on their own experiences when considering the challenges we describe. We are confident that–when one either is an academic themselves or has experience working with academics–these challenges are relatable. Or, to put it into an idiom:

"We're all just cooking with water."

Structure: The remainder of this paper is structured as follows: We first introduce general challenges of academic network measurement in Section 2. Next, in Section 3, we discuss technical challenges in the operation of an active measurement test-bed, and discuss challenges for operators facing network measurements. We then introduce our proposal for creating infrastructure that addresses the outlined challenges in Section 4, and conclude in Section 6.

ANRW '23, July 22–28, 2023, San Francisco, CA, USA

ACM ISBN 979-8-4007-0274-7/23/07.
https://doi.org/10.1145/3606464.3606483

2 ACADEMIC CHALLENGES

Despite empirical assessments of networks being a part of building the Internet since its inception [5], network measurement, as a distinct sub-discipline, emerged in the late 1990s, and the establishment of the ACM SIGCOMM Internet Measurement Workshop (now conference) with its first event in 2001 can be seen as a landmark of when the discipline 'arrived'. Naturally, due to the applied nature of running networks, measurement research at universities is often interwoven with industry collaborations, may it be to access existing (passive) datasets [7, 23], or to connect to practical challenges, test-beds, and systems to evaluate solutions in [19]. Since then, network measurements–active and passive–found their way as a expressed methodology into many sub-fields, ranging from security [8] to the social sciences [6, 12]. Especially for the systems and network security field, the introduction of zMap [9]–enabling researchers to scan the whole Internet within feasible time–was an accelerating factor.

Here, however, we will take a step back, and discuss challenges more inherent to the way academia grew to work, how the Internet grew, and how those two interact.

2.1 Reliable Measurements

Before we get into academia, we have to ask ourselves what it takes to run truly reliable and reproducible active network measurements, where one has a sufficiently high certainty that they will not cause unintended harm.

Conducting *reliable* network measurements is a *complex* task. Given one of the fundamental rules of the Internet– *'if it can be built, somebody did'*–this complexity is enriched with an abundance of corner-cases. To accomplish the goal of creating a measurement toolchain that is robust, reliable, and does not cause harm requires intricate knowledge of the (part of) the protocol stack one intends to measure. Nevertheless, even experienced researchers may cause unintended harm with a standard compliant implementation.

To give an example from the email domain, OpenSMTPd based mailservers using a MySQL backend setup with default settings may experience a denial of service situation if presented with UTF-8 characters in authentication statements, as a default setup does not use a UTF-8 character-set for newly created tables [10]. While this is an easily fixed oversight on the side of the mail-system's operator, a measurement–for example, of UTF-8 support in the mail Ecosystem–here may still cause an unintended disruption when sending UTF-8 characters, even though these `SHOULD` not be causing issues. Even though this issue can not be eliminated, the likelyhood of such side effects obviously gets reduced with researchers using off-the-shelf software that has been tested and proven in production as much as possible. Nevertheless, measurements will always require implementations. To increase these components' reliability, they should obviously be developed using best practices like test driven development [3], including regular tests against common implementations of the protocol stack under test.

Another major issue of network measurements is that one has to have an end-to-end understanding of the whole measurement platform, while also being able to *explain* each individual piece of data collected, and how it has been collected. For example, when measuring email related protocols, one should *also* understand how email protocols interact with the DNS, public key infrastructure/TLS, and–when we go down the MTA-STS rabbit hole–HTTP; And mail is certainly *not* as simple as it used to be [13]. Explainability gets easier by simplifying systems and removing abstraction layers where possible–a common best practice for measurement systems–while tracing all parts of the system involved is also necessary.

To follow this up with another example: If we were measuring something about the resolution behavior of open DNS resolvers, we would want to make sure that all DNS requests we observe at our vantage points have been caused by remote systems, and are not a fragment of our measurement system; This may easily get difficult if we are using a public DNS resolver itself, so we want to make sure that our measurement setup uses a dedicated resolver of which we do have a query log.

Naturally, this list of–potential–side effects goes on. In another example, when we are dealing with RPKI data and look at the evaluation speed of said data, we might be constraint by the throughput or latency of our storage device, especially when dealing with a case including many small files. Hence, we need to have historic monitoring in place for our measurement platform to be able to correlate our results with potential bottlenecks.

Monitoring, in general, is a good point, because we should also monitor the liveness of each component of our measurement setup, because otherwise we might find that a critical component silently failed, introducing an artifact into our system's behavior that is not necessarily apparent from the results. For example, in the context of mail measurements, think about forward and reverse DNS diverging. Some mail servers terminate connections from remotes lacking forward confirmed reverse DNS (fcrDNS, i.e., the reverse DNS entry listing a name that resolves to the IP address of the reverse DNS entry); Hence, we may now exclude a set of hosts from our measurements, even though in our final data, this may not be apparent, apart from a fraction of hosts having–for some reason–terminated connections early.

Hence, to design robust and effective measurement artifacts while reducing the potential for harm, researchers must be able to:

- Thoroughly understand the protocol stack they are measuring, including operational lore and lived experience since the inception of these protocols.
- Be versed in the domain of available implementations to identify components they can use to construct their measurement setup.
- Be experienced programmers and versed in software development in general to follow development best practices and produce tested and reliable code.
- Be experienced system administrators–or have such institutional support–to setup the measurement system, including all basic services the system depends on, including historic and real-time monitoring of all components.

2.2 Ethical Measurements

The ethics of network measurement are tangential, yet related, to their reliability and potential to cause harm. In general, the matter of ethics has already been broadly discussed and best practices documented have been documented, see, e.g., the Menlo report [2] and the probe attribution draft [26]. The gist can be summarized in three simple rules:

- Ensure measurements can be attributed, are documented, and people can opt-out
- Ensure to not cause (unnecessary) harm
- Weigh the benefits of your work against the impact it creates on the Internet

Of these three points, the interaction between the latter two is actually not algorithmically 'simple', but a difficult reflection task. In disciplines traditionally involving human subjects like psychology or medicine–a clear indicator of ethical considerations having to be taken–researchers can usually leverage a process involving an ethics committee. This support through frameworks and external perspectives is crucial, as harm in itself is a wide term, and there are usually no *'good vs. bad'* simple answers. This necessitates consideration and weighting different points, as, e.g., law, usually, is also not a good first principle, as what is ethical may not even be legal, and what is legal may not be ethical. Consider investigating whether a corporation engages in–in the applicable jurisdiction legal, yet clearly unethical–discrimination, and the measurements violate the terms of service of said corporation. The law would be rather clear on this not being feasible.

2.3 Academic Realities

To contrast the idealized requirements above with academic realities, we will now take a look at how lived practice often tends to deviate. We acknowledge that this may sound–in parts–cynical, but we claim that we have to openly address these issues in order to advance our field. We would also like to stress–again–that we take a no-blame approach here [14, 15], i.e., we see what we describe not as the fault of individuals but a result of a system being as it is.

Doing a PhD: In most regions, a PhD is done by students who have obtained a masters' degree, and is intended to take between three and six years. Usually, students start their PhD directly after completing a masters' degree, and do not collect industry experience prior to their PhD [20]. Industry regularly claims that university programs often lack an in-depth practical component. Furthermore, depending on the local academic system, there may be (regular) evaluations to determine if the student can continue their PhD.

Graduation-requirements for a PhD differ between universities and advisors, but usually summarize to 'demonstrating one is able to conduct independent scientific research'. In metrified academia [16], this usually translates to first authoring 3-6 papers in reputable venues, depending on the esteem of these venues and the extend of these papers, woven into a consistent story.

Doing the math between the average length of a PhD and the number of required papers, it becomes clear that a PhD student should usually have their first submitted paper after roughly one year. This means that–if working with active measurements–after one year they have built a new measurement toolchain, ran the measurements, analyzed the results, and–usually with support from their advisor–wrote a 10-15 page text on their findings, while also embedding them in the (academic) related work.

Connecting this to the requirements for reliable measurements means that for a student working on 'just' DNS, they have but a few months to fully read through and understand the DNS camel [4] (DNS grew far beyond RFC1034 [17] & RFC1035 [18] in the past few decades), be an expert in software development, become a 'full-stack' system administrator, and have the decades of operational experience working with DNS to anticipate a multitude of corner-cases. This, obviously, is not feasible.

Instead, students will–and we acknowledge that this even gets sometimes (unjustly) romanticized–struggle. And, being struggling people, they will do what struggling people do: Optimize for getting the task–somehow–done; This includes (uncritically) following stackoverflow and howtos on the web when building systems [1, 27], essentially copying examples and iterating until 'it' works, or using automated setup toolchains and frameworks that make 'things work' without reflecting on how those components actually work.

The result will be a setup where a major component may be built on code never intended for production use, or a system where core-components utilize a docker-stack (including far more components than necessary) whose behavior may deviate from the expected due to some automation, leading to researchers missing components failing in subtle ways.

And, to reiterate, we do not see this as a failure on the students' part, but simply as the result of people doing their best within a system under a set of expectations.

Being an Advisor: A common reply to students' challenges might be that their advisor should supervise, audit, provide the–preexisting–infrastructure for measurements, and generally enable their students' work. However, in practice, advisors are under a different, yet similar, set of requirements influencing the outcomes that emerge. More junior people are usually on a tenure clock, juggle as many students as they can find funding for ('more students' are good for tenure), constantly struggle to find funding ('money' is good for tenure), handle service ('reviewership and departmental committees' are good for tenure), while facing the issue that 'scientific programmer' and 'group system administrator' are not common in start-up packages. Tasks for tenured researchers may (slightly) differ, yet we still have to see evidence that time becomes an abundant resource after tenure.

In that situation, people will again do people things. What one learns to deal with over-load is prioritizing. And we argue that the time budget of advisors would be already filled by thoroughly auditing measurement systems and results created by their students; And then there *also* has to be space for students to fall and grow. Post-docs–the layer in the academic pyramid who might be able to invest sufficient time in these tasks–are usually also in a comparably time starved situation, in so far that their priority is qualifying for and finding their (ideally permanent) next position. Similarly, infrastructure built by prior PhD students–recall, documentation is a part of development best practices usually falling of the ledge–may no longer be sufficiently accessible, functional, or present all together. Ultimately, we find ourselves in a situation without the room to apply necessary care [11], even if one wants to.

Getting Ethics Approval: With ethics becoming a relevant aspect of measurement research–see some conferences now requiring dedicated ethics appendixes–more and more researchers started leveraging their institutions ethics committees. However, with these boards originally often more concerned with human-subject research, there tends to be a lack of technical understanding, sometimes underestimating the impact of measurements in terms of

ethics, as demonstrated by some papers in the past that actually received ethics approval while later being critically acclaimed in the community. Similarly, the estimation of impact may differ, even between parts of the operations community. For example, DNS operators will most likely just shrug at 30 packets one-time hitting their daemon as long as they only contain valid DNS requests. Still, mail operators may easily jump on the fence if those 15 packets constitute one unsolicited email, depending on the source of the destination address.

3 TECHNICAL CHALLENGES

As we saw in Section 2, academia already holds a multitude of challenges for conducting reliable network measurements. However, besides these practical challenges of academia being academia, there are also more technical aspects of doing network measurements, which can prove difficult in an academic context. In this section, we will look at these more technical aspects.

3.1 Infrastructure

Running network measurements needs a certain infrastructure. Besides dedicated machines, ideally with IPv4 and IPv6 addresses, one needs the ability to configure reverse DNS and forward DNS and be able to run a webserver on the systems used in the scans that is reachable from the outside to follow probe attribution best practices [26]. Even more ideally, one should even be able to create INET(6)NUM objects (objects in the database of the responsible Regional Internet Registry (RIR) accessible via `whois`) also providing attribution information and an abuse contact.

Furthermore, and this is likely a classic in terms of Internet measurement researchers' experiences with IT departments, the measurement network should not interact with any middle boxes altering packet flow. On the one hand, this is essential to not influence measurement results, but on the other hand 'overloading a state table' is also a frequent reason for researchers to get more acquainted with their IT department than they wanted to. Of course, in general, a network should be able to handle the packet-per-second rates researchers may originate; But in practice 'researcher using zMap via WiFi' is usually not in operators' load plans.

While some established groups have dedicated setups for these tasks, this is not the case everywhere. Especially more junior PIs or PIs who just recently moved to an institution so far not involved in network measurements may find it difficult to communicate their needs to the IT department. Furthermore, with the progressing cloudification of universities [12], researchers may receive infrastructure, e.g., in the Amazon Cloud, adding another variable to the measurements. Finally, from the researchers' side, providing such infrastructure is–to a degree–costly, and hence may limit groups and institutions with less extensive financial backing from participating in active network measurement research.

On the side of operators running universities' networks, hosting a dedicated scanning segment may also introduce additional issues. For example, anecdotally, operators reported that having a single IPv4 /24 prefix dedicated to measurements led to their whole announced prefix (up to a /16) being blocklisted by some external operators. As such, enthusiasm for provide scanning infrastructure may be limited.

3.2 Abuse Handling and Opt-Out

While we discussed reliable network measurements that do not cause harm in Section 2, the abundance of corner-cases on the Internet ensures that one can never say with certainty that specific measurements may not cause harm. For this reason, it is important to have an–ideally 24/7 reachable–abuse handling while measurements are running, which is also able to stop measurements in case of an incident. Similarly, best practices require researchers to refrain from measurements for networks where the operators opted out, and a measurement system needs to provide a low-effort opt-out possibility for operators.

For consistency over time, groups often maintain opt-out lists. However, these are a) usually not shared *across* groups, and b) need maintenance, as opt-out desire may change over time, or the networks that should be excluded change. For example, we are familiar with an opt-out containing prefixes for a major European hoster due to an undocumented, possibly automated, complaint several years ago. Since then, said operator acquired several /16 of additional IPv4 space, which is not listed in the opt-out list. Furthermore, it might be that–when approached–the operator could be convinced to permit measurements all-together. Hence, maintenance of the list should entail considering such circumstances, e.g., as reasonable, contacting the operator and either updating the list, or–if they can be convinced of permitting measurements–removing them from the list. Hence, across different groups, measurements will differ depending on their opt-out lists.

3.3 Receiving Network Measurements

Turning to the receiving side, i.e., networks being measured, we find that operators also face challenges with the state of active network measurement. With the diversity of groups conducting measurements, operators can not easily block measurements of their network if they so desire. Especially with researchers using public cloud infrastructure, this also induces a burden of maintaining ones' block-lists, as a system used by researchers one week may host a component of the university's learning management system a few weeks later.

Furthermore, operators may find the information on measurements and contact details to be incomplete, unsuitable in their situation, or unintentionally broken and unavailable.

4 `MEASUREMENT.NETWORK`

To address the challenges from Section 2 and 3, i.e., to reduce the load on researchers, enable them to focus on the core of their work, while reducing the impact of active measurements on the Internet, we propose to build a neutral organization facilitating active measurements. In this section, we outline a concept for such an entity by detailing applicable requirements in terms of governance and infrastructure.

4.1 Overview

The basic idea for this entity is creating an organization that provides research infrastructure to academics, ensuring measurement best practices are followed, and that researchers receive the support they need to execute active measurements while limiting harm. For that the entity should:

- Provide the necessary measurement infrastructure to conduct active measurements, especially for measurements not feasible with platforms like RIPE Atlas [22].
- Provide ethical and feasibility review, via a mixed panel of industry domain experts and academics.
- Aid researchers in the open-data publication of their results and measurement artifacts.
- Assist artifact review processes by providing replication infrastructure anonymously.

In addition, depending on resources, such an organization could provide further services to researchers, like distributed and anycasted vantage points.

4.2 Governance

The organization should be non-profit, open, and governed by rough consensus. Additionally the following should apply:

Independence: At the moment, several established groups have their own comparable infrastructure. Some of these systems as, for example, OpenINTEL, are also open to researchers not affiliated with the host institution. However, we argue that the nature of academia as a system of people makes it beneficial to (ultimately) co-locate the governance and oversight of the platform with an independent entity that is *not* directly affiliated with a publication driven research organization or university.

The idea here is that for the platform to be trustworthy among academics this independence is a necessity; No matter the relevance, the fear of getting scooped is a common issue for academics. Hence, submitting one's plans to a 'competitors' organization is unlikely.

Review: The review process should be comparable to the common peer review process. However, in the pursuit of being useful, the goal of the process should not be acceptance or rejection. Instead, it should be a process aimed at moving the planned research into an executable state, covering ethics and reliability aspects alike.

Hence, reviews should be conducted by a diverse group of experts, i.e., covering academics, but also domain experts from industry. Especially for the latter it is instrumental to find experts in the specific sub-aspect of measurements planned; A DNS expert might overlook an oversight in a DMARC (email) related measurement, while an expert in mail may be oblivious to something missed in the context of DNS.

Finally, the past years have sometimes shown researchers finding oddities and issues they would consider vulnerabilities. However, operators may consider said issue to be rather well known. For example, what is a bug shutting down networks *may* be a very well known thing for which there just happens to be an unwritten consensus that *'ah, we all know about it, that's why we don't talk about it; Ultimately, the Internet is held together by bubble gum and duct-tape.'* Then, early practice input might be really good to let researchers in on that observation, before they excitedly run a 'measurement' on uncommon extended attributes in BGP.

Opt-Out & Abuse Handling: The organization should maintain an opt-out list shared across all measurements conducted via the platform. Furthermore, it should provide abuse handling for running measurements, i.e., provide a consistently monitored contact via phone and email, which can take action in case measurements cause harm. Additionally, in handling abuse, the organization should be mindful of caused issues which may, for example, indicate the measurements (accidentally) triggering a so-far undocumented vulnerability, and coordinate improvements.

Specifically, *if* ultimately a vulnerability is found (accidentally), the whole issue of 'how to notify whom' comes into play. I guess just in the last year alone we can all think of several cases where that process did not 'necessarily go ideal' in the context of network measurements. For example, researchers may draft summaries of vulnerabilities perceived to be over-hyping; This is especially an issue when it pertains to one of those 'well known' things, and becomes critical if–what the researchers want to communicate–actually is *"Ups; We may have broken the 'don't talk about it' rule, because we didn't know it was there."*

Open Research & Open Data: In general, the platform should find a feasible way to promote open access to research data. This includes promoting researchers to share their artifacts and collected data (as feasible), but also encouraging authors to utilize open-access opportunities for publications. The question of how to ideally balance various trade-offs here is, however, an open one.

4.3 Infrastructure

In terms of infrastructure, requirements are driven by containedness (blockability) and control, i.e., the organization should be able to operate a network as independently as possible. Specifically, we suggest the following components to be (at least) in place:

Dedicated AS Number & Prefixes: The organization should be able to use a dedicated AS number and dedicated IPv4 and IPv6 prefixes, and run all services and measurements below one domain. Per address family, there should be at least two publicly routed prefixes with different routing policies so core services like DNS can be operated contained within the AS. Furthermore, the prefixes should come from continuous netblocks, enabling operators to block all measurements by blocking one AS path and one prefix per address family.

This does, of course, make it easier to block the prefixes and ASN used by the project. However, this is part of the appeal, as it allows operators to easily and permanently opt-out of all measurements. Dropping all AS-paths containing the measurement ASN will make their networks unable to be measured, while also ensuring that the resources are not used for other (production) purposes in the future.

Basic Services: Besides the functions outlined before, the platform should support the base services necessary for the organization. This includes authoritative DNS for the chosen domain, a review system, an inventory of running and concluded measurements, and an inventory of collected data and used artifacts if available. Additionally, basic services should include a historic and real-time monitoring platform.

Measurement Services: For each measurement to be run, the platform should provide researchers with:

- A dedicated prefix per address family, for which appropriate reverse DNS and INET(6)NUM objects are in place.
- Network uplink to handle the configured line-rate of the measurement machines provided.

- A sub-domain dedicated to the specific measurements.
- A listing of the ongoing experiment, including a descriptions of the experiments and stated abuse contacts.
- Already setup base-services (recursive/forward DNS etc.) within the measurement prefix.
- Support in setting up/configuring measurement machines, so researchers retain control of all parts of their measurements.
- Historic and real-time monitoring for the measurement platform, including end-to-end tests.

Additionally, it may be feasible to also provide dedicated solutions if resources permit, e.g., for anycast measurements.

5 DISCUSSION

Here, we briefly touch upon comments and suggestions received from reviewers and colleagues in the preparation of this paper.

Leveraging an Existing Organization: One might wonder why we propose to start from scratch, instead of leveraging existing structures. This, in general, might be a viable idea. However, in practice, this is likely to face several issues:

- Convincing an organization to adopt an idea is significantly harder than convincing an organization or group of people to adopt something that already works.
- Starting a project potentially involving 'toys' (an AS, IPv4 and IPv6 prefixes, servers, systems, routers) with associated 'big plans' has a certain risk of causing a major instance of bike-shedding among all too interested engineers and potential collaborators. We are convinced that first building the park while worrying about the bike-shed later might be more productive.
- Starting with resources associated *with* an organization also carries the risk that it might–ultimately–be harder to move them to an independent organization.

Hence, for now, we decided to pursue the creation of this idea independently.

Funding: Connected to the question of cooperating with an existing organization is the matter of funding. Running systems costs, ultimately, money. As outlined in Section 4, for now, the project is self-funded. Additionally, we receive support from operators in terms of (indirect) upstream and networking resources (see Acknowledgements). Again, as with the question of an organization to govern the project, funding is ultimately a question more easily resolved when something is less of an idea and more of an implementation.

Passive Data: Another interesting item brought up in response to our work was the matter of passive traces. It was suggested that active network measurements might not always be necessary, if there was more widely available access to (industry) data.

However, getting access to (industry) passive datasets can be a matter of (social) connections and networks. For PhD candidates, it is most likely always dependent on their advisors prior access to such data sources, again leading to an equity issue in terms of who can do what science. Connected to this is the issue that a group that is strongly rooted in a specific passive/industry dataset may create a 'walled garden', in which it becomes increasingly challenging to work outside these boundaries. Furthermore, if such a walled garden exists, it becomes more challenging for junior researchers to build an independent research agenda[1].

The `measurement.network` we propose, might in fact be a good 'proxy' entity for facilitating access to measurement data. Similar to reviewing requests for access to active measurement infrastructure, we could (and should) strife to motivate organizations holding interesting data to collaborate in facilitating access. Of course, this is a vision, and how to actually implement this–how to approach industry partners to convince them of this idea–a whole different challenge; But it is an option we argue is worth considering.

6 CONCLUSION

In this paper, we reflect on the interaction between practical circumstances in academia and the growing complexity of network protocols. Based on our reflections, we argue that this development creates a situation catering towards mistakes, potentially limiting the reliability of network measurement campaigns, while also increasing the likelyhood of unintended harm occurring. Following learnings from the safety sciences, we consider this not as the result personal failures or mistakes, but see it as the conclusion of systemic composition.

We argue that the most feasible path forward to at least address some symptoms of metrified academia is creating infrastructure that improves (junior) researchers' access to communal operational knowledge (lore) and practical domain experience. At the same time, infrastructure and a framework which reduces the organizational overhead of running network measurements, may improve the accessibility of active measurements overall, while giving researchers more time to focus on their research.

We conclude by outlining our ideas on how an entity providing such infrastructure could be created, and would like to engage into a discussion on realizing this vision. While we will give realizing these plans a try, and happily invite anyone interested to join, we can of course not promise success; A future reader may test whether we failed by checking whether https://measurement.network/ still exists and if AS211286 can be found in the global routing table.

```
  ASN:  AS211286
 IPv4:  141.39.220.0/22
 IPv6:  2a0d:8d04::/32
Domain: measurement.network
```

Acknowledgements

I would like to thank the communities of operators of various protocols with which I interacted over the years, shaping my perspective on operating systems (not as in OS). Especially LWLcom (AS50629) and OpenFactory (AS58299) generously provided support on the path towards this project and for public services like https://www.email-security-scans.org/ [13], https://bttf-whois.as59645.net/ [24], and https://v6only-resolver.measurement.network/ [25]. Furthermore, I want to acknowledge my past co-authors for putting up with me having the learning experiences I discussed in this paper. Finally, I would like to thank Simone Ferlin-Reiter for her initial feedback on an early draft of this paper, as well as the reviewers for their insightful comments in the reviews.

[1]We would like to credit Simone Ferlin-Reiter for noting this in her feedback.

REFERENCES

[1] Yasemin Acar, Michael Backes, Sascha Fahl, Doowon Kim, Michelle L Mazurek, and Christian Stransky. 2016. You get where you're looking for: The impact of information sources on code security. In *IEEE Symposium on Security and Privacy (SP)*. 289–305.

[2] Michael Bailey, David Dittrich, Erin Kenneally, and Doug Maughan. 2012. The menlo report. *IEEE Security & Privacy* 10, 2 (2012), 71–75.

[3] Kent Beck. 2003. *Test-driven development: by example*. Addison-Wesley Professional.

[4] Bert Hubert. 2018. *Herding the DNS Camel*. https://www.ietf.org/blog/herding-dns-camel/

[5] Ramón Cáceres, Peter B Danzig, Sugih Jamin, and Danny J Mitzel. 1991. Characteristics of wide-area TCP/IP conversations. *ACM SIGCOMM Computer Communication Review* 21, 4 (1991), 101–112.

[6] Ray M Chang, Robert J Kauffman, and YoungOk Kwon. 2014. Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems* 63 (2014), 67–80.

[7] Nikolaos Chatzis, Georgios Smaragdakis, Jan Böttger, Thomas Krenc, and Anja Feldmann. 2013. On the benefits of using a large IXP as an Internet vantage point. In *Proceedings of the 2013 conference on Internet measurement conference*. 333–346.

[8] Jakub Czyz, Michael Kallitsis, Manaf Gharaibeh, Christos Papadopoulos, Michael Bailey, and Manish Karir. 2014. Taming the 800 pound gorilla: The rise and decline of NTP DDoS attacks. In *Proceedings of the 2014 Conference on Internet Measurement Conference*. 435–448.

[9] Zakir Durumeric, Eric Wustrow, and J Alex Halderman. 2013. ZMap: Fast Internet-wide Scanning and Its Security Applications. In *USENIX Security Symposium*, Vol. 8. 47–53.

[10] T. Fiebig. 2022. *Send it (or rather not): Stupid things with OpenSMTPd and mysql*. https://doing-stupid-things.as59645.net/mail/opensmtpd/mysql/2022/08/30/receiving-an-email.html

[11] Tobias Fiebig and Doris Aschenbrenner. 2022. 13 Propositions on an Internet for a "Burning World". In *Proceedings of the ACM SIGCOMM Joint Workshops on Technologies, Applications, and Uses of a Responsible Internet and Building Greener Internet*.

[12] Tobias Fiebig, Seda Gürses, Carlos H Gañán, Erna Kotkamp, Fernando Kuipers, Martina Lindorfer, Menghua Prisse, and Taritha Sari. 2023. Heads in the Clouds? Measuring Universities' Migration to Public Clouds: Implications for Privacy & Academic Freedom. In *Proceedings on Privacy Enhancing Technologies Symposium*, Vol. 2023.

[13] Florian Holzbauer, Johanna Ullrich, Martina Lindorfer, and Tobias Fiebig. 2022. Not that Simple: Email Delivery in the 21st Century. In *2022 USENIX Annual Technical Conference (USENIX ATC 22)*. 295–308.

[14] Mannat Kaur, Harshini Sri Ramulu, Yasemin Acar, and Tobias Fiebig. 2023. "Oh yes! over-preparing for meetings is my jam:)": The Gendered Experiences of System Administrators. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–38.

[15] Mannat Kaur, Michel van Eeten, Marijn Janssen, Kevin Borgolte, and Tobias Fiebig. 2021. Human Factors in Security Research: Lessons Learned from 2008-2018. *arXiv preprint arXiv:2103.13287* (2021).

[16] Chris Lorenz. 2015. The metrification of 'quality' and the fall of the academic profession. *Oxford Magazine* 355 (2015), 7–11.

[17] P.V. Mockapetris. 1987. Domain names - concepts and facilities. RFC 1034 (Internet Standard). https://doi.org/10.17487/RFC1034 Updated by RFCs 1101, 1183, 1348, 1876, 1982, 2065, 2181, 2308, 2535, 4033, 4034, 4035, 4343, 4035, 4592, 5936, 8020, 8482, 8767.

[18] P.V. Mockapetris. 1987. Domain names - implementation and specification. RFC 1035 (Internet Standard). https://doi.org/10.17487/RFC1035 Updated by RFCs 1101, 1183, 1348, 1876, 1982, 1995, 1996, 2065, 2136, 2181, 2137, 2308, 2535, 2673, 2845, 3425, 3658, 4033, 4034, 4035, 4343, 5936, 5966, 6604, 7766, 8482, 8490, 8767.

[19] Jonathan Perry, Amy Ousterhout, Hari Balakrishnan, Devavrat Shah, and Hans Fugal. 2014. Fastpass: A centralized "zero-queue" datacenter network. In *Proceedings of ACM SIGCOMM*. 307–318.

[20] Estelle Phillips and Colin Johnson. 2022. *Ebook: How to Get a PhD: A Handbook for Students and Their Supervisors, 7^{th} Edition*. McGraw-Hill Education (UK).

[21] Philipp Richter and Arthur Berger. 2019. Scanning the scanners: Sensing the internet from a massively distributed network telescope. In *Proceedings of the Internet Measurement Conference*. 144–157.

[22] RIPE Atlas. 2023. *RIPE Atlas*. https://atlas.ripe.net/

[23] Arjun Roy, Hongyi Zeng, Jasmeet Bagga, George Porter, and Alex C Snoeren. 2015. Inside the social network's (datacenter) network. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*. 123–137.

[24] Florian Streibelt, Martina Lindorfer, Seda Gürses, Carlos H Gañán, and Tobias Fiebig. 2023. Back-to-the-Future Whois: An IP Address Attribution Service for Working with Historic Datasets. In *International Conference on Passive and Active Network Measurement*. Springer, 209–226.

[25] Florian Streibelt, Patrick Sattler, Franziska Lichtblau, Carlos H Ganán, Anja Feldmann, Oliver Gasser, and Tobias Fiebig. 2023. How Ready is DNS for an IPv6-Only World?. In *International Conference on Passive and Active Network Measurement*. Springer, 525–549.

[26] E. Vyncke, B. Donnet, and J. Iurman. 2023. *Attribution of Internet Probes*. https://datatracker.ietf.org/doc/draft-ietf-opsec-probe-attribution/03/

[27] Tianyi Zhang, Ganesha Upadhyaya, Anastasia Reinhardt, Hridesh Rajan, and Miryung Kim. 2018. Are code examples on an online Q&A forum reliable? a study of API misuse on stack overflow. In *Proceedings of the 40th International Conference on Software Engineering*. 886–896.