On Changing the Culture of Empirical Internet Assessment

Mark Allman
International Computer Science Institute
Berkeley, CA
mallman@icir.org

Categories and Subject Descriptors

C.2.6 [**Computer Communication Networks**]: Internetworking

Keywords

Internet,measurement,science

1. INTRODUCTION

Over the last 20 years empirical assessment of the Internet has become a crucial task with two first-order goals. The first goal is to understand the vast global network as it exists. This informs our mental models on which we can then develop and extend the Internet in myriad ways. The second goal is to understand the problems facing the Internet in a concrete fashion such that the community can then develop solutions. While empirical assessment often appears straightforward at first blush, soundly investigating some facet of the network often results in subtle and nuanced analysis that was unforeseen at the outset of an investigation. While we are getting better as a community at thorough empirical investigation, more work in this direction is needed. In this paper we discuss two (related) problems faced by Internet empiricalists:

1. The global Internet is both massively heterogeneous in many dimensions (as we discuss in detail below) and constantly evolving. Therefore, to understand the scope of particular behavior requires data from myriad vantage points over time. We have noticed an increasing appreciation for this point from program committees, editorial boards and reviewers who yearn for "representative", "typical" or "usual" data and frown upon use of datasets that do not somehow "feel" broad. The general principle behind this yearning is well-placed. However, the need for a large breadth of data to develop sound and general results is often at odds with individual researchers' ability to collect data—which is often quite modest. This creates a situation where the community's quest for "general" data can prevent *any* understanding from being published. Of course, we do not wish to discourage efforts to collect "big data" from broad vantage points. However, we advocate developing coping strategies when reality dictates that gathering broad data is not possible.

2. There is a general understanding that sound science requires reappraisal and confirmation of results. We know from years of our own Internet measurement analysis that mistakes can easily creep into the process. Further, as sketched above, the network varies across vantage point and time and therefore a result developed on one dataset may not hold when analyzing another. Therefore, re-visiting previous results is crucial. While we believe that most of the community sympathizes with the importance of reappraisal in theory, we find only scant examples of the community actually engaging in and fostering such studies. When confronted with reappraisal studies we have witnessed—with few exceptions—many program committees, editors and reviewers taking everything from a dim to an outright hostile view of the efforts.[1] Even more disconcerting, these dim views often seem to hold in places where one might expect the most appreciation for reappraisal of empirical results. The Internet Measurement Conference [3] and the Passive and Active Measurement Conference [6] both welcome reappraisal in their Call For Papers and yet we have found PCs for both conferences—for papers that we have no relationship with—to be disinclined to foster these efforts. Rather, there is a general preference in our community for "novel" work. While there are certainly examples of reappraisal in the literature, it is rare that an empirical study is re-examined by the community. Even the biggest proponents of reappraising previous results rarely if ever do so, likely because it is difficult work with little perceived benefit within the community.

Together these two problems leave our understanding beholden to reviewers' intuition. For a study to be published the dataset has to *feel* "broad enough" to referees. But, after publication a study is oftentimes the final word, given our lack of appetite for reappraisal. This situation leaves us with little idea about whether a particular study is right, wrong, widely applicable or narrowly held.

2. COMMENTS ON DATA SHARING

One mechanism that has been strongly advocated to mitigate the problems sketched above is to broadly share datasets [24, 10, 1, 4, 7]. First, by making a broad array of datasets widely available researchers are no longer constrained by their individual ability to collect data, but rather can draw

[1]We were recently informed by an editor-in-chief of an ACM journal that "novelty is a basic requirement of science" and that reappraisal was in no uncertain terms unwelcome.

from a pool of data collected by the entire community. Second, sharing datasets allows for direct re-examination of results on the actual data used in an original study. Additionally, ready access to a dataset without the attendant costs—both technical and logistical—of gathering the dataset lowers the cost of reappraisal (not to mention empirical work in general).

The community has undertaken a number of efforts in the area of data sharing, such as:

- The community has developed several data catalogs and archives over the years (e.g., [24, 1, 4, 7]). Such structures provide a clearing house such that researchers can readily locate available datasets.
- Even outside the various repositories and catalogs that the community has implemented, researchers sometimes publicly share data (e.g., [5, 2]).
- Additionally, private sharing can be useful in getting data in researchers' hands, but crucially depends on individual relationships.
- The community has developed data anonymization techniques that attempt to balance data privacy issues with research utility (e.g., [29, 23, 22]). While savvy anonymization seeks to preserve research utility the fundamental tradeoff cannot be avoided [22].
- The community has tried to incent data release through awards that are only available to authors who make their data available to the community at conferences (e.g., the Internet Measurement Conference and the Passive and Active Measurement Conference).
- A related suggestion is that data holders could analyze their data on behalf of others in lieu of providing the data itself. While this costs data providers additional resources compared with data sharing, it also provides more control over the data and what is ultimately given out (e.g., [19]).

The community's best intentions are behind all the work described above. Further, these efforts have all largely been beneficial to some degree in terms of getting data into researchers hands. However, ultimately most published studies do not have accompanying data release. We believe there are several reasons for this state of affairs. First, quite often data is considered too sensitive due to privacy or competitive concerns. Second, while anonymization techniques have been developed they are not airtight in terms of not leaking sensitive information (e.g., see [14] and the response [12]). Further, anonymizing to the point required to release some data may remove the necessary utility from the data such that published results cannot be re-created.[2] Third, releasing data is often a large effort. In technical terms, researchers must understand data sensitivities and apply appropriate anonymization techniques for a given threat model. In non-technical terms, researchers must convince their management and legal teams that data release is useful and safe. The fourth reason we believe there is little data release is closely related to the large burden in doing so: there is often little direct benefit to researchers who release data. In short, there are many reasons to not bother to release data and few tangible benefits.

We do not wish to discourage data sharing or diminish its value. However, we believe we must face a reality that suggests wide-scale data sharing is not going to solve the issues sketched in § 1. Therefore, the community needs to have a conversation about how to address these problems in a world without wide-scale data sharing. In the next section we aim to start such a conversation.

3. A PATH FORWARD

Unfortunately, coping with both the heterogeneous Internet being difficult to measure and the lack of reappraisal within our community will take more than simply building a new data catalog, inventing a new anonymization technique, developing a new tool or giving awards to authors who share data. Rather, we must do something decidedly more difficult: *we must change the way we view the problem.* We do not claim the suggestions in the following subsections represent a solution to the problem—or an exhaustive list of strategies—but we hope they are constructive additions to a conversation within the community, editorial boards and program committees about our stance on empirical observation.

3.1 Embrace Insight Over Numbers

While it should be self-evident, we must think about measurement as a process for developing insightful understanding about the system and not just about producing numbers. As an example, [26] seeks to understand whether packet and flow arrival processes in wide-area networks are well modeled by a Poisson process. While the particulars of the exponential processes of various traffic characteristics are given, the conclusions and implications are based on the general nature of the process and not the specifics of the data used in the study. We believe that in general the community appreciates the notion that measurements serve to provide insight. However, we also find that at times we as a community (and as individuals) get caught up on numbers and therefore we state this principle here because much of the discussion below hinges on the community embracing insight over numbers.

3.2 Embrace Reappraisal

As sketched above, re-examination of results is crucial to the scientific process. However, as a community we do not often re-examine previous work. One of the often heard reasons is "we don't have the data". However, this is at least not always the reason[3] and we believe we can use several strategies to cope with the case when data is not available, as follows:

- First, active measurements can be readily re-run and we can expect that any soundly developed invariant or slowly changing behavior will be present in this independent re-execution. If the new measurements do not produce similar insights then that is also useful information, as perhaps the initial analysis was flawed, too narrowly focused or the network is changing rapidly in terms of the particular behavior.

[2]For instance, while we released the dataset used to produce the results in [21], we elide packet payloads and therefore not all results in the paper can be derived from the dataset.

[3]For instance, we released the set of enterprise-level packet traces [5] used for [21] and are not aware of any reappraisals of our analysis.

- While data sharing is in general anemic, there are cases where the data from some study is in fact widely available. For instance, many studies leverage data from RouteViews [9] or skitter [8] to study network topology and this data is available. In these cases we can and should re-examine previous results.
- We sometimes confuse *reproduction* and *reappraisal*. The former focuses on verifying precise answers and would in fact require specific datasets. However, the latter concerns building an experiment anew to revisit some previously identified phenomenon. If we focus on insights rather than numbers—as suggested above—then reappraisal can proceed without specific datasets since we are concerned only with re-examining previously identified behavior and not specific numbers.

While these three strategies can help divorce the task of reappraisal from specific datasets, we need additional effort before such activities will be commonplace, such as:

- Foremost, we must accept that studies reappraising previous results as worthwhile contributions to our field. In other words, we have to place high esteem on re-examinations that both confirm and refute the original studies. We need to stand up for such work in places like program committee meetings, promotion and tenure meetings and with management "up the food chain". If we as a community cannot take this initial step to change our own attitudes and advocate for reappraisal then we cannot even begin to think of empirical Internet measurement as a science.
- Second, we should provide quality venues for publication of reappraisals. In our experience, one reason the reappraisals that are conducted are not enthusiastically welcomed in some of our top tier venues—which happen to be conferences—is sometimes because there is a boredom factor with the thought of using conference time to hear a talk that (say) largely confirms previous findings. However, trading sound science away to avoid boredom seems like a short-sighted strategy.

  It is perhaps natural to use conference time on new and fresh ideas that benefit from real-time interaction. One suggestion is to use our journals to better foster reappraisals. Another idea is for some of our measurement-oriented venues to accept full papers—not abstracts—to the proceedings without accepting them for presentation.[4] Given that we are moving towards electronic proceedings this should not be a significant monetary cost to conferences. A final option is a new venue—likely a journal—explicitly providing a venue for reappraisal of previous results.
- When acting as editors and PC chairs we need to encourage sound science via reappraisal. This is especially true for journals that do not have the strong inclination towards novelty that we often apply to our conferences. Explicitly calling for reappraisals would be a step in the right direction. Further, a venue that explicitly lists reappraisal as desirable should be accepting of these papers and editors and PC chairs have a responsibility to ensure their members and reviewers give such papers a fair shake.

[4]Or, perhaps having a "lightning session" where each such paper is given a few minutes to make the community aware of their results without providing a full treatment during the conference.

3.3 Embrace Heterogeneity

The Internet's behavior is both massively heterogeneous and also constantly evolving. This is argued in detail in "Why We Don't Know How To Simulate The Internet" [27]. The implication of heterogeneity is that getting a "typical" or "representative" view of the network is elusive. That is, every vantage point we measure from has its own peculiarities. As an example, in our own study of residential networks [17] we find peer-to-peer applications constitute roughly 14% of the traffic. However, the paper discusses related work that finds the amount of peer-to-peer traffic from 5–80% in different networks at different times. Of course, heterogeneity does not stop at application mix, but encompasses many facets of networks from routing to protocol behavior to server capabilities and beyond. As empiricalists we must accept this heterogeneous reality and *stop insisting that studies use "representative" datasets*. We offer three coping strategies to deal with this suboptimal reality:

- First, since no one vantage point can capture the breadth of behavior that will be found across the entire network we must embrace the concept that we make progress as a community when we all engage in building a *body of work* to study various Internet phenomena. That is, we accept that no one study is likely going to encompass enough data to convince us that the given behavior is pervasive.
- Second, we should focus on searching for *invariant* network behavior which holds across vantage points. For instance, [25] shows that specific characteristics of traffic vary across network and time, but the shape of the distribution of various characteristics holds constant (e.g., number of bytes in transaction responses).
- A third coping strategy is to ensure we are showing perspective in our work. While we as a community should be open to studies that do not tell the entire story but rather are part of a body of work, we as authors should not over-claim our results by making sweeping conclusions that our data cannot support.

Both reviewers and authors need to take a step back and embrace the heterogeneity of the Internet and not expect or claim too much from each individual study.

As a short case study, consider the community's belief that Internet traffic is well described as heavy tailed. Rather than a single study that conclusively illustrates this invariant, the notion was introduced in [16] and later confirmed by the community through a body of work produced by a number of individuals and datasets (e.g., [25, 15, 13]). This case study is illustrative of the power that lies in a community effort to lend confidence to a particular result. Unfortunately, such efforts are too rare. The community should strongly embrace such efforts as they both (*i*) allow our understanding to capture the heterogeneity fundamental to the Internet and (*ii*) serve to reappraise previous findings and hence play a valuable role in developing the community's confidence in a given result.

Finally, note that while we strongly believe that "representative" datasets are fundamentally elusive, that does not

mean we advocate accepting results from excessively anemic datasets. For instance, we might not be able to ascertain a broad swatch of Internet client behavior in a particular study. However, that is not an excuse for studying only the behavior of a single host in our own basement. While we believe reviewers often wish for overly grandiose datasets, we do not aim to give researchers an excuse to fail to gather as much data as possible in their work.

3.4 Embrace Sharing

As discussed in § 2, the current reality does not support *reliance* on data sharing to help mitigate the issues presented by a heterogeneous Internet nor greatly move us towards a culture that appreciates reappraisal. However, while we advocate for embracing reality, we do not call for abandoning the idea of robust sharing within the community.

First, by sharing both tools and specific analysis processes—i.e., how tools are executed and their results used—we can enable additional researchers to apply the specific analysis methodology to their own data. This both reduces the burden on a researcher reappraising a previous result, but also exposes the analysis procedure to scrutiny for bugs and oversights (much like open source software). Sharing tools eliminates the inherent ambiguity in our descriptions of our analysis processes. For instance, many times we write colloquially about some event like a "retransmission", but exactly how that is defined in our code is not specified.[5] Sharing of analysis tools does not have the same sensitivity burdens present when sharing data. Therefore, the community should make a push for increased tool sharing. In fact, PAM 2013 did just that, see [6].

Second, we note that active measurements often do not have the same sensitivities as passive observation and therefore sharing active measurement data should be encouraged.

Finally, while we have sketched ways to cope without data sharing, getting a broad array of data in the hands of more researchers is often highly useful (as has been discussed previously [10, 4, 1]). We believe the community can make strides towards better science without data sharing, but we do not wish to diminish the value of this activity when it can be made to work.

To aid tool and data sharing we make three suggestions:

- Contributing tools and data to the community is a large undertaking. Collecting a high quality dataset, gathering the meta-data about the measurements, conducting basic calibration of the collection apparatus, anonymizing the data and navigating various organizational procedures takes significant effort. Likewise, writing accurate analysis tools and detailing their use is also a large effort for any non-trivial analysis. We should acknowledge these efforts.
- Further, we should view released tools and datasets as scholarly contributions. As such they should be viewed favorably on job applications or during the promotion process. To the extent we can influence such activities in our various roles we should attempt to give voice to these contributions as scholarly work.
- In service of the above, as reviewers and especially as editors and PC chairs we can ensure tools and datasets are properly cited in the reference list of papers. This provides a way for the use of tools and datasets to be tracked (as we do citations to papers with indexes such as Google Scholar) and hence gives us a way to assess the overall use and contribution of the released resources.

[5] E.g., a packet could be a "retransmission" if its entire payload has been previously sent, or the definition could include the case when only some of the payload is being resent. Alternatively, segments may be considered "retransmissions" if they are sufficiently out-of-order. The specifics of that heuristic are important to reappraisal, but often not well defined in papers.

3.5 Embrace Risk

A final suggestion is for reviewers, editors and PC chairs to take more risks. It is not difficult—and we have observed it many times—for a reviewer to nit-pick empirical studies to death (or at least rejection!). Every study seemingly has some potential blind spot, or some phenomenon that could explain an observation if only a little more data had been collected, or some not-completely-obvious assumption, or a lack of data at some boundary, or the like. We find reviewers too often be adverse to the risk of accepting a paper that does not absolutely "prove" some facet of network behavior. Allowing the perfect to be the enemy of the good then means that the community at-large does not get exposed to the work at all.

One place where we regularly find risk aversion is in terms of validation. Often when assessing network behavior we are making inferences and ideally these techniques are assessed in relation to some ground truth. In reality, we often do not possess the ground truth for such a comparison, but rather rely on a circumstantial case that suggests our techniques are drawing correct inferences. As with assessing network behavior, such circumstantial validations can often be nit-picked to death as they are never perfect.

In some sense, being risk averse may be a rational reaction within a community that does little reappraisal. In other words, if we know that a piece of work will not likely be checked then perhaps we should endeavor to ensure we make the best possible decisions. Therefore, if we follow through on the suggestions in the above subsections we may be able to let our guard down a bit. That is, if we knew there was a reasonable chance of reappraisal of the insights in some submission we may be more inclined to overlook some of the *potential* issues we can conjure in the original analysis. However, we advocate thinking about this in the opposite direction and in fact accepting more risk regardless of whether the community embraces the path we stake out above. Such a stance may in fact drive the community towards reappraisal of insight through developing a body of work. That is, because we welcome papers that do not nail down every last corner, it may coax researchers who are not fully convinced in some result to pursue an additional study—with additional and perhaps ancillary data—on the given phenomenon. This in turn builds the body of work we advocate above.

Note, we are not advocating for the acceptance of sub-par work. Certainly there are sloppy papers that, for instance, do not leverage all the data available to the authors to build the case for some behavior or that make unfounded and outlandish assumptions. We are not encouraging such studies nor suggesting that they should get a free pass. While we believe the community would benefit from being accepting of more risk, the amount of acceptable risk should not be

infinite. Studies that are either clearly broken or highly dubious should not be acceptable.

4. EMPIRICALISM AND SECURITY

The above discussion heavily leans on the notion that different investigators can in fact collect independent data about a given phenomenon. Clearly the Internet evolves and therefore we are never measuring precisely the same system. However, a situation where this is even more acute is in the security arena. In these cases, we are often assessing a malicious adversary who actively does not wish to be understood. Therefore, measurements involving malicious actors tend to capture ephemeral phenomena that may well be one-time events and not amenable to reappraisal by others.

For instance, consider a worm outbreak. While we know that worms tend to fester in the network for long periods of time [11], the behavior manifest at the outset of the outbreak is fleeting [28, 20]. In such cases if there are no measurements of the worm at the time of the outbreak then the opportunity for the community to independently measure the behavior is lost. That said, even continuous generic monitoring (a la NetFlow logs) can often provide a view of these sorts of phenomena—even if limited in terms of fine-grain details. We should leverage these resources as much as possible. Further, data sharing may be more paramount in such situations. Fortunately, it may in some cases also be easier since the traffic is not initiated by users and therefore may be viewed as less sensitive.

Worm outbreaks are hardly the only example of such ephemeral phenomenon. Other investigations that track important but ephemeral aspects of the Internet landscape focus on, e.g., DDoS attacks, BotNet infiltration, phishing web sites employing fast flux names and spam campaigns. We believe the best way to soundly assess these threats is to focus on the invariants across specific attacks. In other words, embrace insights and develop a body of work that focuses less on the specific behavior of each individual attack and more on the general underpinnings (i.e., following the structure of the underground economy, e.g., [18]).

5. SUMMARY

As we discuss in § 3, the coping strategies we offer are largely a change in our attitudes. Crucially, we need to accept that one study cannot be the first and last word on some observed Internet behavior. Further, we must also face the reality that data sharing—while useful where possible—will not be ubiquitous and therefore we must determine how to engage in sound science without relying on possessing others' data. We propose coping with this reality by encouraging researchers to embrace insights that come from a body of work. That is, we first focus on big picture insight instead of nitty-gritty numbers. And, we accept that no one study is going to use data "typical" of the entire Internet, but rather that each individual study helps us understand some corner of the network and is therefore a valuable contribution towards more holistic understanding.

First and foremost this requires a shift in our attitudes towards empirical work as we review papers, and especially as we execute our duties as editors and PC chairs. Further, we may need to re-think our publication venues to better encourage reappraisal and extension of previously discovered behaviors. For instance, we may need new journals for this express purpose or to extend our conference proceedings to include non-presented papers. We encourage our venues that focus on empirical Internet assessment to think about these issues and find a way forward.

Finally, we stress that we do not advocate for a situation where all empirical studies are trivially small or where far-fetched assumptions run wild. While we believe our community can do more to foster sound science by embracing insights and building bodies or work, we do not encourage the acceptance of excessively narrow or sloppy work.

Acknowledgments

The opinions in this piece are the author's alone. However, the thinking benefits from many conversations over many years with too many colleagues to name. In particular, the position in this paper benefits from years of discussions (over many beers!) with Vern Paxson. Additional discussions (and arguments!) with kc claffy, Craig Partridge and Matt Roughan were also particularly helpful. Michael Bailey, Paul Barford, Rob Beverly, Craig Partridge and Vern Paxson all provided insightful feedback on a draft of this note. My thanks to all!

6. REFERENCES

[1] A Community Resource for Archiving Wireless Data At Dartmouth (CRAWDAD). `http://crawdad.cs.dartmouth.edu`.

[2] Click Dataset. `http://cnets.indiana.edu/groups/nan/webtraffic/click-dataset`.

[3] Internet Measurement Conference. `http://www.sigcomm.org/imc/`.

[4] Internet Measurement Data Catalog (DatCat). `http://www.datcat.org/`.

[5] LBNL/ICSI Enterprise Tracing Project. `http://www.icir.org/enterprise-tracing/`.

[6] Passive and Active Measurement Conference 2013. `http://pam2013.comp.polyu.edu.hk/`.

[7] Protected Repository for the Defense of Infrastructure Against Cyber Threats (PREDICT). `http://www.predict.org/`.

[8] The Skitter Project. `http://www.caida.org/tools/measurement/skitter/`.

[9] The University of Oregon Route Views Project. `http://www.routeviews.org/`.

[10] M. Allman, E. Blanton, and W. Eddy. A Scalable System for Sharing Internet Measurements. In *Passive and Active Measurement Workshop*, Mar. 2002.

[11] M. Allman, V. Paxson, and J. Terrell. A Brief History of Scanning. In *ACM SIGCOMM/USENIX Internet Measurement Conference*, Oct. 2007.

[12] M. Allmand and V. Paxson. Issues and Etiquette Concerning Use of Shared Measurement Data. In *ACM/USENIX Internet Measurement Conference*, 2007.

[13] M. Arlitt and C. Williamson. Web Server Workload Characterization: The Search for Invariants (Extended Version). *IEEE/ACM Transactions on Networking*, 5(5), Oct. 1997.

[14] S. Coull, C. Wright, F. Monrose, M. Collins, and M. Reiter. Playing Devil's Advocate: Inferring Sensitive Information from Anonymized Network Traces . In *Proceedings of the Network and Distributed System Security Symposium*, 2007.

[15] M. E. Crovella and A. Bestavros. Self-similarity in World Wide Web traffic: Evidence and possible causes. In *ACM SIGMETRICS*, pages 160–169, May 1996.

[16] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of ethernet traffic. In *ACM SIGCOMM*, 1993.

[17] G. Maier, A. Feldmann, V. Paxson, and M. Allman. On Dominant Characteristics of Residential Broadband Internet Traffic. In *ACM Internet Measurement Conference*, Nov. 2009.

[18] D. McCoy, A. Pitsillidis, G. Jordan, N. Weaver, C. Kreibich, B. Krebs, G. M. Voelker, S. Savage, and K. Levchenko. PharmaLeaks: Understanding the Business of Online Pharmaceutical Affiliate Programs. In *USENIX Security Symposium*, Aug. 2012.

[19] P. Mittal, V. Paxson, R. Sommer, and M. Winterrowd. Securing Mediated Trace Access Using Black-box Permutation Analysis. In *ACM HotNets*, 2009.

[20] D. Moore, C. Shannon, D. J. Brown, G. M. Voelker, and S. Savage. Inferring Internet Denial-of-Service Activity. *ACM Trans. Comput. Syst.*, 24(2):115–139, May 2006.

[21] R. Pang, M. Allman, M. Bennett, J. Lee, V. Paxson, and B. Tierney. A First Look at Modern Enterprise Traffic. In *ACM SIGCOMM/USENIX Internet Measurement Conference*, Oct. 2005.

[22] R. Pang, M. Allman, V. Paxson, and J. Lee. The Devil and Packet Trace Anonymization. *ACM Computer Communication Review*, 36(1), Jan. 2006.

[23] R. Pang and V. Paxson. A High-Level Programming Environment for Packet Trace Anonymization and Transformation. In *ACM SIGCOMM*, Aug. 2003.

[24] V. Paxson. Internet Traffic Archive. `http://ita.ee.lbl.gov/`.

[25] V. Paxson. Empirically-Derived Analytic Models of Wide-Area TCP Connections. *IEEE/ACM Transactions on Networking*, 2(4):316–336, Aug. 1994.

[26] V. Paxson and S. Floyd. Wide-Area Traffic: The Failure of Poisson Modeling. *IEEE/ACM Transactions on Networking*, 3(3), June 1995.

[27] V. Paxson and S. Floyd. Why We Don't Know How to Simulate the Internet. In *Proceedings of the 1997 Winter Simulation Conference*, Dec. 1997.

[28] S. Staniford, V. Paxson, and N. Weaver. How to 0wn the Internet in Your Spare Time. In *USENIX Security Symposium*, 2002.

[29] J. Xu, J. Fan, M. Ammar, and S. Moon. On the Design and Performance of Prefix-Preserving IP Traffic Trace Anonymization. In *Proceedings of the Internet Measurement Workshop*, Nov. 2001.