Experience in using MTurk for Network Measurement

Gokay Huz
Naval Postgraduate School
ghuz@cmand.org

Steven Bauer
MIT CSAIL
bauer@mit.edu

kc claffy
CAIDA
kc@caida.org

Robert Beverly
Naval Postgraduate School
rbeverly@nps.edu

ABSTRACT

Conducting sound measurement studies of the global Internet is inherently difficult. The collected data significantly depends on vantage point(s), sampling strategies, security policies, or measurement populations – and conclusions drawn from the data can be sensitive to these biases. Crowdsourcing is a promising approach to address these challenges, although the epistemological implications have not yet received substantial attention by the research community. We share our findings from leveraging Amazon's Mechanical Turk (MTurk) system for three distinct network measurement tasks. We describe our failure to outsource to MTurk an execution of a security measurement tool, our subsequent successful integration of a simple yet meaningful measurement *within* a HIT, and finally the successful use of MTurk to quickly provide focused small sample sets that could not be obtained easily via alternate means. Finally, we discuss the implications of our experiences for other crowdsourced measurement research.

CCS Concepts

•**Networks** → **Network measurement;**

Keywords

Crowdsourcing, Network measurement, Mechanical Turk

1. INTRODUCTION

Difficulties in gathering representative and otherwise scientifically sound Internet measurement data sets include: the sheer technical complexity and scope of the infrastructure; scaling measurement technology, algorithms, and analysis; security and privacy concerns; and basic Internet architectural limitations [15]. A key concern is the selection of vantage points when conducting networking research, as the network or attachment point from which data is collected can strongly influence the resulting data. Unfortunately, researchers typically have access to only a small number of vantage points relative to the size of the Internet.

When conducting networking research there are at least three instances where access to additional vantage points would be beneficial: 1) during the exploratory phase, to gather a small set of samples from interesting or varied locations; 2) when attempting to gather larger amounts of data; and 3) when attempting to confirm, validate, or correlate results obtained via other methods (e.g. other tools). Crowdsourcing measurement is an approach to providing these additional vantage points.

Researchers have a growing interest in how to effectively use, and incentivize, crowdsourced measurements. Crowdsourcing can be done directly or as a by-product of other actions by a crowd. An example of the later is the successfully crowdsourced topology [12] and outage [9] data collection via a BitTorrent plugin that improves download performance; from the user's perspective, the data collection is incidental.

We hypothesized that directly crowdsourcing network measurements would facilitate network research, increase the size and representativeness of the resulting data set and enable one to quickly gather target data from some otherwise inaccessible network locations.

We report our findings and lessons learned from crowdsourcing measurements of the Internet itself using Amazon's Mechanical Turk (MTurk) platform [5]. While MTurk has traditionally been used for psychological, behavioral, and annotation experiments and surveys [19, 20, 10], we seek to understand the challenges and potential in using it to conduct networking research. In particular, we design and run crowdsourced network measurement experiments for each of our three envisioned scenarios: data exploration, collection, and validation. Our primary contributions include:

ACM acknowledges that this contribution was authored or co-authored by an employee, or contractor of the national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only. Permission to make digital or hard copies for personal or classroom use is granted. Copies must bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. To copy otherwise, distribute, republish, or post, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

C2B(1)D'15, August 17, 2015, London, United Kingdom
© 2015 ACM. ISBN 978-1-4503-3539-3/15/08…$15.00
DOI: http://dx.doi.org/10.1145/2787394.2787399

1. Description of our inability to crowdsource a network security measurement: BCP38 compliance.
2. Development of a successful simple network measurement task integrated within a HIT.
3. Characterization of the network measurement diversity that MTurk affords, e.g., biases in the distribution of worker's vantage points.

While our initial results are mostly anecdotal, crowdsourced measurements could, in the future, allow macroscopic inference of network properties not otherwise observable, yet vital to our understanding of the Internet.

2. BACKGROUND

Amazon's Mechanical Turk ("MTurk") is a marketplace and a crowdsourcing platform where *workers* from around the world connect with *requesters* who publish micro-tasks known as *Human Intelligence Tasks (HITs)* [5]. Workers receive monetary compensation for completing HITs. In January 2015, there were over 245,000 available HITs and Amazon claims more than 500,000 registered workers across 190 countries [2]. The life cycle of a HIT includes:

1. Assignable: During HIT creation, the requester assigns the compensation amount (minimum $0.01 USD), expiration date, and minimum qualifications. Qualifications can limit HITs to workers from particular countries or with a specified approval rating. Requesters can create multiple assignments of a HIT, i.e., offering the HIT to multiple distinct workers. Once uploaded, the HIT becomes assignable and qualified workers can accept it to begin work.
2. Unassignable: Accepted HITs become unassignable; other workers cannot accept or work on the HIT.
3. Reviewable: After a worker completes and submits a HIT, it becomes reviewable, whereby the requester can review the worker's results and approve or reject the HIT.
4. Reviewing (optional): The requester can review completed HITs manually, via an API, or automatically accept all results. Only accepted results receive compensation.

While Amazon does not publish information about their workers, several studies have analyzed its user demographics. Across 20 months and more than 3800 users surveyed, Ross et al. [24] found in November 2009 that the MTurk worker population was skewed to the U.S. and India (56% and 36% of the total respectively). Possible contributing factors to this geographic skew are HIT language (almost exclusively English) and the form of compensation. While U.S. and Indian workers can receive direct financial compensation, due to tax reporting requirements, workers from other countries can only earn credit toward purchases from Amazon [4].

Researchers have used MTurk for user studies [19, 13], behavioral research [20], and other experiments [23, 10] that require rapid and affordable collection of many tasks. Mason and Suri [20] discuss conducting behavioral research on MTurk, praising the platform for offering workers with diverse ages, genders, and incomes.

MTurk has also been used to conduct computer security studies. To demonstrate the potential to identify and infect the computers of the large worker population, Kanich successfully embedded javascript within a HIT to infer the type and version of various web browser plugins [18]. More than 85% of worker clients were using versions of plugins with known vulnerabilities.

Most closely related to and inspiring our effort, Christin et al. [13], in an eye-opening security experiment, successfully hired > 950 MTurk workers (out of 2,854 HIT views) to download and run an executable on their local computer with administrative privilege for one hour – ignoring traditional security advice and policy against doing so – in exchange for payments as low as $0.01.

3. METHODOLOGY

We conducted three new experiments on MTurk corresponding to three distinct measurement goals: exploratory research of broadband speed testing from select vantage points (§3.1), large data collection of a network security property (§3.2), and validation of edge IPv6 adoption (§3.3).

3.1 Broadband Speed Tests

Over the last decade residential broadband speed testing has received considerable attention. At times the results from these consumer testing tools have been relied upon by government regulators and academics as accurate reflections of broadband access link performance. However, our prior work [7] found many of the available tests systematically underestimated access link performance. A primary cause was that the tests were conducted by downloading a large file over a single HTTP (and thus TCP connection) that often ended up being bottlenecked by the end user's TCP receive window. In 2010, we validated our analysis of a large amount of data from a specific tool (NDT [11]) by enlisting a small and targeted set of MTurk workers (approximately 20 users running three different testing tools).

Although our ad-hoc 2010 MTurk experiment results were never published, they inspired us to perform follow-on work that similarly leverages MTurk for targeted exploration of active measurements from *vantage points otherwise inaccessible to our research team.* In particular, the fidelity of results from speed testing tools when run on very high speed broadband access links (100 Mbps to 1 Gbps) has not been extensively explored. However, we currently lack access to these networks and none of the academic measurement projects that we collaborate with have access either. We therefore designed a HIT to investigate network performance measurement tools run on these emerging high speed access links. We targeted US-based workers on select

broadband networks, in order to sequentially run and compare the results of a series of web-based speed tests.

We created two HITs, one targeting workers in the US and one targeting users on Google Fiber broadband networks. We used written instructions to restrict a HIT to workers on Google Fiber networks; we did not construct an MTurk qualification test to enforce this restriction, since we believed most workers would not take the time to run the qualification test given the limited number of assignments available. To receive compensation, the worker had to open a specific web-based speed test (e.g., Ookla-based sites, NDT, or DSLreports), run the test by clicking the start button and report the results by copying the upload and download speeds and any "share results" links that were available.

3.2 (Not) Executing a measurement tool

Given Christin et al.'s [13] successful use of MTurk to have users run their own executables, we explored whether we could use this capability to measure a network-layer infrastructure vulnerability: compliance with BCP38 [16], a best practice for network operators to filter traffic with invalid source addresses. Unfortunately, BCP38 compliance (also called *anti-spoofing*) is incentive-misaligned, in that networks only help other networks by implementing the practice; they do not directly help themselves. Measurement of BCP38 compliance is also incentive-incompatible; operators are unlikely to volunteer measurements that reveal their networks to be violating best security practices given the visibility of cybersecurity vulnerabilities today. The most viable method of explicitly measuring BCP38 compliance at the macroscopic level is by crowdsourcing. Previous voluntary measurements of spoofing [8] have acknowledged the need to increase coverage with improved incentives. MTurk provides just that opportunity.

We designed a HIT that required users to download, install, and run the IP Spoofer network measurement tool [8], which tests BCP38 compliance of a hosting network. We posted our spoofability HIT in November, 2013 with a compensation of $0.26.

Within hours, our HIT was reported and subsequently removed from the MTurk site. We received an email from Amazon citing the Mechanical Turk Terms of Service (ToS) [3]. Although the ToS does not explicitly address the permitted nature or content of HITs on MTurk, at the time of our experiment, Amazon explicitly prohibited any "HITs that require workers to download software" [22]. Through private correspondence with an author of [13], we learned that their experiment was similarly removed, but then unblocked after an appeal indicating that the HIT was part of a university research project. Unfortunately, our attempt at appealing to Amazon was not successful.

Interestingly, Amazon recently relaxed their policies to forbid only "HITs that require Workers to download software that contains any malware, spyware, viruses, or other harmful code" [6], suggesting that network measurements conducted by MTurk users running an executable may be viable in the future.

Figure 1: Ball-counting HIT: red balls are fetched via IPv4 while the (missing) blue balls are only available via IPv6, thereby enabling inference of the worker's IPv6 capability.

3.3 IPv6 Adoption HIT

We sought to explore what kinds of network properties we could investigate by working *within* the (then existing) confines of the MTurk platform. Another globally important question is whether a client (or its hosting network) supports IPv6 at the network layer. We constructed a HIT that would indirectly measure this property, as well as matching the clients' IPv4 and IPv6 addresses. From the user's perspective, the HIT is a simple survey to report the number of red and blue balls that appear on their browser screen. Each individual HIT retrieves a random number (from 1 to 4) of red and blue balls, from IPv4 and IPv6 web servers that we maintain. The blue balls are only served by an IPv6 web server, which creates an opportunity to measure a host's IPv6 capability and corresponding IPv4 address. This high-level approach has been previously used to mitigate measurement bias in inferring IPv6 penetration [25], but never within the MTurk platform.

Figure 1 shows our HIT as it appeared to a worker lacking IPv6 connectivity. In this example, the system provided the user's browser with links to three red balls and one blue ball, but the IPv4-only client could not fetch the IPv6-only blue ball image. When the blue ball image cannot be retrieved, web clients display either no image or a broken image icon. However, if the blue ball images are retrieved, we infer that the worker has IPv6 connectivity. The user completes the HIT by entering the number of red and blue balls displayed.

We use two host names for the experiment. We host the red ball images and web form on an external server (e.g., http://ipv4.example), and host the blue ball images on a separate domain (http://ipv6.example). Our DNS server provides only A (IPv4) resource records for the IPv4 host and only AAAA (IPv6) resource records for the IPv6 host. Thus, a host attempting to resolve the A record for http://ipv6.example receives a negative response (NXDOMAIN).

The images for the red balls are small bitmap files hosted on the IPv4 domain's web server. The survey page uses a simple PHP script where the IPv4 address of the client that fetches the main page is embedded in the URLs for the blue balls. Each blue ball URL is also

Table 1: Distribution of worker's geolocated IPv4 addresses; USA and India dominate.

Country	IPv4 Req.	Country	IPv4 Req.
United States	322 (60.8%)	Canada	7 (1.3%)
India	148 (27.9%)	Ireland	3 (0.6%)
Great Britain	13 (2.5%)	Others	28 (5.6%)
Japan	7 (1.3%)		
		Total	530 (100%)

a PHP script that returns a blue ball with the correct HTTP header. The blue ball URLs are of the form:

```
http://ipv6.example/img.php?1.2.3.4
```

where the query string of the URI is the worker's IPv4 address represented in ASCII dotted-quad notation (here, IPv4 address `1.2.3.4`).

Every time a worker displays the survey, their browser attempts to fetch the images hosted on our server as part of the HIT completion process. If the worker's host and network has IPv6 support, their browser also fetches the blue ball image(s). By recording each HTTP request the web server handles, we can match a client's IPv4 and IPv6 addresses. For example, if a worker with IPv6 connectivity and address `2001:dead::beef:cafe` resolves the previous blue ball URL to ipv6.example and successfully fetches the blue ball image, we log:

```
2001:dead::beef:cafe - - [11/Mar/2014:01:17:36]
"GET /img.php?1.2.3.4 HTTP/1.1" 200 37977
"http://ipv4.example/?assignmentId=XXXXXX
&hitId=YYYYYY&workerId=ZZZZZZ
```

and infer that IPv4 address `1.2.3.4` corresponds to the client with IPv6 address `2001:dead::beef:cafe`.

4. RESULTS

4.1 Broadband Speed Tests

We designed our broadband speed testing HIT to gather a small set of exploratory measurements using existing web-based tools on emerging very high speed residential access networks. Our goal was not to rigorously investigate MTurk, but rather to explore its potential for obtaining *particular, targeted* measurements.

For the first HIT we paid $0.20 for running a single web based performance test. We hoped each user would complete the five assignments that they were eligible to complete by running tests on five different speed testing web sites. Of the 45 assignments completed during ≈8.5 hours by 14 unique workers, five workers completed all tests on five different websites, while five completed only a single assignment.

On the second HIT targeting workers on Google Fiber, we anticipated the need to pay more per assignment. We offered $0.50 per individual test, $2.00 if four tests were run to services we selected for the geographic diversity of their test servers (Ookla-based tests and NDT). None of these HITs were completed within a twenty hour period, perhaps due to a low number of MTurk workers with Google Fiber or a lack of demographic match between workers and high speed broadband networks.

While this later HIT did not succeed, the results were worthwhile overall in time (≈2 hours) and cost (≈$12.00) given the exploratory state of this research. The gathered data helped confirm our hypothesis that there is wide variance in speed test results run from the same vantage point using different sites and tools.

Table 2: Published IPv6 adoption rates vs. MTurk experiment inference; our results are roughly consistent with these independent data.

	Google [17]	Akamai [1]	Cisco [14]	MTurk
Overall	2.72%	1.50%	2.72%	3.21%
USA	5.27%	3.2%	5.25%	5.28%
India	0.16%	0.05%	0.14%	0.00%

4.2 Measuring IPv6 Adoption HIT

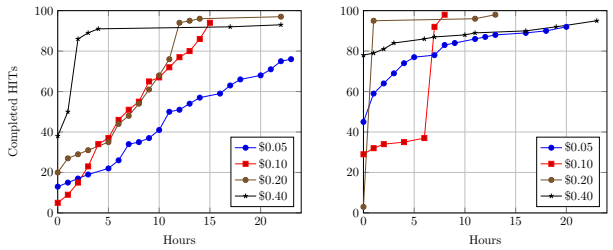
We first sought to assess the feasibility of measuring native IPv6 adoption using MTurk. We began by publishing a batch of 200 HITs that awarded workers $0.26. Workers completed this initial batch in under 10 hours. In the second batch, to maximize HIT completion for our budget, we lowered the compensation but raised the number of available HITs. We published 300 of the same HIT, but reduced the award to $0.11. After three days, 142 of the 300 available HITs were completed and we prematurely expired the remaining 158. We allowed users to participate in only one batch of HITs.

From the HTTP requests corresponding to this experiment, we extracted a total of 530 unique IPv4 and 38 unique IPv6 addresses. We geolocated IPv4 and IPv6 addresses using Maxmind [21] and mapped IPv6 addresses to their providers using whois. Table 1 shows the country distribution of the IPv4 addresses of workers completing the HIT. Workers from the United States and India dominate our results, constituting 89% of the total completed HITs. The next most frequent country in our data was Great Britain at $\simeq 2.5\%$.

We observed 37 unique IPv6 source addresses fetching the blue ball images as part of this experiment. Of these 37 IPv6 addresses, nine used Teredo and 11 used 6to4; we ignored these non-native (auto-tunneling) forms of IPv6 in our analysis. All remaining 17 IPv6 addresses belonged to U.S. ISPs (9 in Comcast, 4 in AT%T, 2 in Verizon, and 1 in Time Warner). Our MTurk-enabled measurement results were consistent with Google and Cisco's public measurements for the USA and India (Table 2), the two countries that generated a statistically significant number of sampled measurements.

4.3 HIT Price Sensitivity

Given our findings in §4.2, we next sought to better understand the relationship between compensation and workers willing to complete our HIT. We used the same HIT as before, and published batches of HITs incrementing in compensation from $0.05 to $0.10 to $0.20 to $0.40. Because of the concentration of workers in two



(a) India (b) United States

Figure 2: Relationship between compensation and HIT completion time

countries, each batch consisted of 100 HITs that could be completed only by U.S. workers, and 100 HITs available only to Indian workers. Each user was allowed to complete only one assignment in each batch of 100 assignments, but could participate in multiple batches. In total, we obtained approximately 800 samples.

The tests ran for 24 hours on MTurk, after which they expired and were not available for submission. For some of the HITs, not all of the 100 tests were completed due to workers failing to completed accepted HITs before our experiment's expiration.

Figures 2(a) and 2(b) show the effect of compensation on HIT completion time for India and the U.S., respectively. The x-axis shows the number of hours from the time the HITs were assignable, while the y-axis shows the count of completed HITs. In general, as we increased compensation, HITs were completed more quickly. The $0.40 HIT was completed more quickly than other HITs. However, there were irregularities in the price sensitivity plots. For the HIT that targeted only workers in the U.S., the $0.20 HIT was completed faster than the $0.40 HIT. The $0.05 HIT reached 80% completion within 10 hours – faster than the $0.10 HIT. We postulate that these differences are attributable to day-of-week and time-of-day differences. Future work should more carefully control for these variables.

Table 3 shows the number of completed HITs and the number of HITs that were approved, i.e., the worker identified the correct number of red balls. HIT workers in the USA generally produced more correct answers than those in India, for a variety of possible reasons: language barriers, prioritizing speed over correctness, ignorance, or automation.

4.4 Previewing and Over-Constrained HITs

Recall that the elements of our IPv6 HIT were hosted externally as we must control IPv4 vs. IPv6 access, and perform logging and correlation. Another measurable aspect of Mechanical Turk HIT processing is that a worker can *preview* a HIT before choosing to perform it. We accidentally discovered that our HIT design enabled our intended measurement even when a worker only previewed the HIT, i.e., we still captured the user's IP addresses and whether her computer was IPv6 enabled. Of the 3339 HTTP requests to our web server, 1485 of them were due to workers previewing the HIT.

Table 3: Compensation amount vs. number of completed and approved HITs

Country	$0.05		$0.10		$0.20		$0.40	
	Total	Correct	Total	Correct	Total	Correct	Total	Correct
USA	94	92	100	100	100	100	100	100
India	99	95	100	93	100	97	93	88

Further, we discovered that requesters can design *over-constrained* HITs. For instance, a HIT can require the user to be located in both the USA and India. With such constraints, no worker can accept and perform the HIT. However, over-constrained HITs allow requesters to gather measurements such as ours as part of the HIT preview process, without providing compensation.

Leveraging HIT previews and over-constrained HITs can be more effective by using large compensation amounts. For example $25.00 is a high price for a simple HIT, when most HITs are priced for pennies. Many MTurk users sort HITs by compensation so that their visibility is proportional to the HIT award. We did not investigate over-constrained HITs as means of obtaining samples, and future work should examine the ethical considerations of doing so.

5. CONCLUSIONS

Crowdsourcing platforms like MTurk offer a novel approach to network measurements. Our research demonstrated the utility of MTurk for a number of tasks associated with conducting networking research: 1) exploratory data collection 2) targeted collection of small to medium size data sets, 3) small scale validation of results obtained by other means. Some limitations and considerations specific to MTurk bear attention by the research community.

5.1 Geographical (Non)-Diversity

Despite Amazon's claim of workers from 190 countries, we found it difficult to get results from workers in specific countries. We took advantage of HIT location enforcement in MTurk and targeted a separate experiment to workers in Japan, Turkey, and USA only with an award of $0.26. We uploaded three separate HITs, each targeted to a different country. After two days we received only 56 results, all from the USA; no HITs targeted for workers in Japan and Turkey were completed.

Our findings suggest that it is difficult to obtain network measurements from specific countries at this point in time. Consistent with prior studies [24], we obtained workers from around the world – but the majority of the active workers were in India and the USA. While our English language HIT may partially explain the lack of results from Japan and Turkey, workers can (and do) use automated language translation software. The more fundamental problem appears to be the skewed geographic distribution of workers. In future work, we plan to investigate publishing country-targeted measurement HITs in the country's native language.

5.2 Ethical Considerations

While Amazon's conditions of use [3] explicitly prohibit tasks that collect personally identifiable information (PII), protecting workers is the responsibility of the requesters. We received an official determination from our institution's Institutional Review Board (IRB) that our experiment was not human subjects research.

Discussion with our IRB centered around three factors: i) the quantity of monetary reward; ii) the information collected; and iii) whether the experiment collects data "about whom" or "about what." Because the total per-worker compensation was negligible, our IRB had no concerns that workers would be unduly coerced. Although our experiment collected IP addresses, we have no reasonable way to map such addresses to individuals. Indeed, requesters cannot obtain any PII about workers other than what the workers disclose, and in our case, we did not ask workers for any information other than counting the number of balls. Last, workers as humans were incidental to our network measurement goal, therefore the experiment was "about what."

Further, in consideration of beneficence and respect for persons, counting balls is a relatively innocuous task with no expectation of harm, while the concomitant opportunity for Internet measurement created by this task provides a substantial societal benefit, by increasing our understanding of the evolution of the Internet network layer's capability to serve all users equally.

5.3 Future Work

We demonstrated how crowdsourced network measurements can reveal simple macroscopic properties of parts of the Internet, which inspired us to consider other measurements that MTurk could facilitate in the future. Amazon's recent policy change to allow workers to download software executables means that we could use our method (§3.2) to re-attempt to measure the deployment of source address validation, or other edge-visible security technologies, e.g., DNSSEC. Finally, we plan to use MTurk to validate IP geolocation techniques.

Acknowledgments

We thank Erin Kenneally and Larry Shattuck for ethical guidance and review. This work supported by National Science Foundation grants CNS-1213155 and CNS-1111445. Views and conclusions are those of the authors and should not be interpreted as representing the official policies, expressed or implied, of the U.S. government.

6. REFERENCES

[1] https://blogs.akamai.com/2013/06/world-ipv6-launch-anniversary-measuring-adoption-one-year-later.html.
[2] AWS Developer Forums: MTurk CENSUS. https://forums.aws.amazon.com/thread.jspa?threadID=58891.
[3] Mechanical turk participation agreement. https://www.mturk.com/mturk/conditionsofuse, 2015.
[4] Amazon. Getting paid. https://www.mturk.com/mturk/help?helpPage=worker.
[5] Amazon. Mechanical Turk. https://www.mturk.com.
[6] Amazon. Mturk policies. https://www.mturk.com/mturk/help?helpPage=policies, June 2015.
[7] S. Bauer, D. D. Clark, and W. Lehr. Understanding Broadband Speed Measurements. TPRC, Aug. 2010.
[8] R. Beverly, A. Berger, Y. Hyun, and k claffy. Understanding the efficacy of deployed internet source address validation filtering. In *Proceedings of Internet Measurement Conference (IMC)*, Nov. 2009.
[9] Z. S. Bischof, J. S. Otto, M. A. Sánchez, J. P. Rula, D. R. Choffnes, and F. E. Bustamante. Crowdsourcing ISP characterization to the network edge. In *Proceedings of the ACM SIGCOMM W-MUST*, 2011.
[10] M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1):3–5, 2011.
[11] R. A. Carlson. Developing the web100 based network diagnostic tool (ndt). In *Proc. PAM*, 2003.
[12] D. R. Choffnes, F. E. Bustamante, and Z. Ge. Crowdsourcing service-level network event monitoring. In *Proceedings of ACM SIGCOMM*, 2010.
[13] N. Christin, S. Egelman, T. Vidas, and J. Grossklags. It's all about the benjamins: An empirical study on incentivizing users to ignore security advice. In *Financial Cryptography and Data Security*. 2012.
[14] http://6lab.cisco.com/stats/, December 2013.
[15] D. D. Clark. The design philosophy of the DARPA internet protocols. In *Proceedings of ACM SIGCOMM*, 1988.
[16] P. Ferguson and D. Senie. Network ingress filtering: Defeating denial of service attacks which employ ip source address spoofing. BCP 38, May 2000.
[17] Google. IPv6 Statistics. http://www.google.com/intl/en/ipv6/statistics.html, December 2013.
[18] C. Kanich, S. Checkoway, and K. Mowery. Putting Out a HIT: Crowdsourcing Malware Installs. In *WOOT*, 2011.
[19] A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing user studies with Mechanical Turk. In *Proceedings of SIGCHI*, pages 453–456, 2008.
[20] W. Mason and S. Suri. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, 44(1):1–23, 2012.
[21] Maxmind - ip geolocation and online fraud prevention. http://www.maxmind.com/en/home, March 2014.
[22] Mturk policies (archived). http://web.archive.org/web/20150210034052/ https://www.mturk.com/mturk/help?helpPage=policies.
[23] J. Oh and G. Wang. Evaluating Crowdsourcing through Amazon Mechanical Turk as a Technique for Conducting Music Perception Experiments. *Proceedings of ICMPC*, 2012.
[24] J. Ross, L. Irani, M. Silberman, A. Zaldivar, and B. Tomlinson. Who are the crowdworkers?: shifting demographics in mechanical turk. In *Proceedings of SIGCHI*, pages 2863–2872, 2010.
[25] S. Zander, L. L. Andrew, G. Armitage, G. Huston, and G. Michaelson. Mitigating sampling error when measuring internet client IPv6 capabilities. In *Proceedings of IMC*, 2012.