CheesePi: Research, Results, and the Regulator

Ian Marsh
SICS RISE AB
Kista, Sweden
ian.marsh@ri.se

ABSTRACT

In October 2015 we presented CheesePi at the Research and Applications of Internet Measurements (RAIM) workshop in Yokohama, Japan. In this followup paper, 18 months on, we present our findings. These range from simple round trip delay measurements to entire quality campaigns over several weeks. A significant portion of the research has been done in conjunction with the Swedish regulator, which we bring to the fore. CheesePi is a lightweight platform with the goal of performing measurements from within users' homes. Since the goal is to measure Internet quality delivered by ISPs, measurement fairness is of prime concern. This paper covers the regulator's demands, ISPs feedback, our research interests, the users' needs of nearly two years work.

CCS CONCEPTS

- **Networks** → **Network measurement**; *Home networks*;

KEYWORDS

Network measurements, Home networking, Regulation

ACM Reference format:
Ian Marsh. 2017. CheesePi: Research, Results, and the Regulator. In *Proceedings of ANRW '17, Prague, Czech Republic, July 1, 2017,* 6 pages.
https://doi.org/10.1145/3106328.3106337

1 INTRODUCTION

We have presented a distributed measurement system for home Internet quality characterization [15]. The aim of CheesePi was to objectively characterize the quality users obtain from their home Internet connections. It leverages the Raspberry Pi, an always-on, affordable, quiet measurement node acting independently of other devices in users homes. Home networking characterisation requires *continuous* monitoring of connections, an external device satisfies this requirement. Our objective is **not** to aim for a large scale deployment of measurement nodes. Rather, we want to deploy the *minimum* number of nodes that satisfy regulator statistics. This not only keeps the cost down, eases deployment, but also reduces data to be processed, small data can be of value too!

ANRW '17, July 1, 2017, Prague, Czech Republic

ACM ISBN 978-1-4503-5108-9/17/07. . . $15.00
https://doi.org/10.1145/3106328.3106337

2 RELATED WORK

In this paper, we begin with discussing measurement fairness. What exactly fair means is not well-agreed up within the measurement community. Fair sampling has been touched upon in [9], but fair measurements, as far as we know, have neither been extensively researched not clearly mandated by ISPs or regulators. There, has however been some recent work in measurements and policy [6] and measurements and neutrality (from the Finnish regulator) [16].

The Leone EU project, completed in April 2015, assessed 12 new metrics to measure performance factors using the SamKnows infrastructure of 130 "probes" [1]. One metric they identified relevant to CheesePi is the *bitrate reliably streamed* and they trialed it using measurements to Google's YouTube and the BBC's iPlayer VOD platforms. The framework was standardized known as an "Information Model for Large-Scale Measurement Platforms" [12].

RIPE Atlas is a project measuring Internet quality globally as part of their interaction of traffic and network initiative [3]. Around 9400 probes have been deployed by European IP Networks worldwide as of April 2016. ICMP and traceroute measurements are performed using the hardware probes with security support from DNSSEC and SSLcert. The Archipelago Measurement Infrastructure uses 150 Raspberry Pis, globally deployed quantify connectivity and congestion using IPv4/v6 active measurements as an overlay [5]. Quantifying the congestion is done by a Time-Sequence Ping (TSP), a crafted sequence of pings along a selected network path which is similar to our approach.

Internet measurement platforms are certainly not new. A text on Internet measurements by Crovella and Krishnamurthy published in 2006 covers many areas still relevant to measurements today [7]. [4] describes all efforts up to 2015, an updated Dagstuhl seminar paper is [2].

Moving onto work directly inline with this paper, Fathom has been used in home network environments, and has similar goals to ours, but uses a Firefox extension based on Netalyzer [8], whereas we opt for an external hardware approach. It is based on previous work done in [14]. [10] use the TCP three-way-handshake and adaptive `ping` measurements to produce what they claim to be "the most consistent results for base network delays", with a cellular focus, whilst we consider the 3-way handshake as a delay measurement, closer to the application and above the ubiquitous ping. Teacup is a joint platform between Swinburne in Australia and Cisco, using 100's of nodes to investigate TCP, in particular DASH-video.

Table 1: Categorization of measurement strategies.

Type	Description	Comment
No fairness	Measure on random paths to arbitrary destinations	Most common measurement strategy
Data fair	Identical amount of traffic per operator	Akin to a mobile subscription
Scaled data fair	Identical amount of traffic sent per operator	Measure according to the size of the ISP
Time fair	Measure over the same period of time	Traffic could interact at path intersections
Scaled time fair	Operator measured over same time, scaled /size	Akin to data, but in time domain
Rate fair	Time and data fair	Can be harder to measure rates
Period fair	Over period, use same data amount	Fair only over the period
Random fair	Send at varying periods and rates	Uncertain statistically of fairness

3 THE REGULATOR

3.1 Background

This work is in collaboration with the Swedish regulator, the Post and Telecom Authority[1]. Expansion to home Internet measures is our joint goal. The regulator currently sends outs paper surveys to a well-chosen set of consumers, covering their demographic needs. Any CheesePi deployment must satisfy statistical needs, e.g. the appropriate proportions DSL/fiber, rural/urban, young/old etc. PTS also sponsors a Swedish Speedtest[2] which measures from a browser Fig. 1.

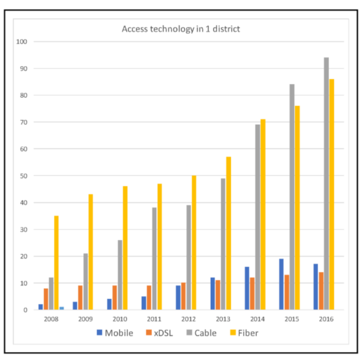

Figure 1: Bredbandskollen's demographic example.

3.2 Fair, unbiased, neutral measurements

Our central tenant is a representative unbiased sampling of home connection quality within Sweden. We want to ensure that each ISP receives a fair share of active measurements. Simple to state, but difficult to achieve in practice, see Table 1. The options in the table are solely to illustrative to show the options one has to decide on *before* one starts to measure. Our work is really at the behest of the regulator for obvious reasons of neutrality and impartiality, and still needs much further research. Orthogonally, SICS/RISE is non-profit organization with no financial nor competition incentives, hence is well-suited to undertake a measurement task with the regulator.

[1]www.pts.se

[2]www.bredbandskollen.se

4 RESEARCH

4.1 Experimental configuration

CheesePi can be used in several configurations, therefore it is described depending on how the nodes are used:

(1) Internet-SinglePi: Video streaming, and dashboard tests are examples of this configuration.
(2) Pi-Pi: TCP protocol performance, capacity measurements, and VoIP experiments using client and server controlled two Pis.
(3) Internet-MultiplePis: In a multi-Pi setting we have used the CheesePi network to monitor several large media events.

4.2 Internet to Single CheesePi node

4.2.1 Investigating buffer stalls. IPTV to the home is commonplace. Buffer playout is a network operation close to the user's perception. Complex dynamics are captured by the buffer: media transmission rate, encoding, session length, X-traffic, buffer adaptation, network delay/loss tradeoffs, and so on. Therefore, our goal is to *capture and identify the major contributors to stalls.*

To quantify correlations in highly dimensional spaces we draw upon Principal Component Analysis. PCA is a variance-based method for reducing and visualizing the dimensionality of data. It uses an orthogonal transformation to convert the HD data into a set of linearly uncorrelated components. The lengths of the components represent their relative contributions, and their relative angle the correlation, see Fig.2. We

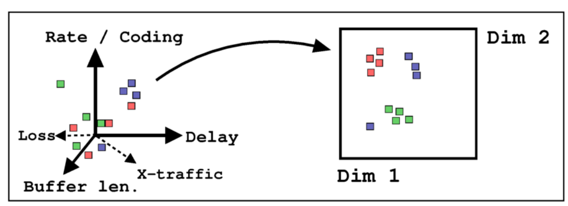

Figure 2: *Multi-2D transformation.* Left: A multidimensional measurement space, the axes represent measurement attributes and colored points represent measurement points. Right: 2D visualization of the same data points.

are using the CheesePi[3] network to identify the magnitude of each feature simultaneously. The combination of each component is learnt using a LSTM neural network [11]. Tests are very much ongoing with VQEQ videos, that already have offline subjective ratings [20], to guide our determination of buffer stalls. This is still ongoing work with further results available by the end of the year.

[3]http://cheesepi.sics.se

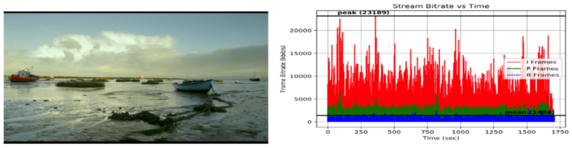

Figure 3: *Rating sequences.* Left: Example frame of 15s 1280x720 24fps AVI. Right: Variable bitrate produced.

4.3 Tasks and Dashboards

We use simple tasks, which actually perform the measurements, after they are scheduled. Once measured data is put into a time-series database on the Pi. Currently we support Influx and Pymongo as databases. Furthermore, 15 class templates for writing measurement tasks are provided (≈100 lines) from WiFi polling to `iperf` suites. We expect developers to contribute tasks as needed, probably starting with modifying an existing Python task.

Where home users are using the Pi to monitor their home connections, we display the results for measurement tasks on a dashboard. Figure 4 shows two representations. The left example is a more technical dash, where row-by-row we show the ICMP delay, the HTTP delay to a Alexa-selected site, a speedtest to the nearest Ookla site, a YouTube[4] video, packet loss, DNS lookup time, the WiFi access points and local Pi information, temperature, load and uptime. The right visualisation is an analogy of data "flowing" into a home as a bathtub filling rate and corresponding water level.

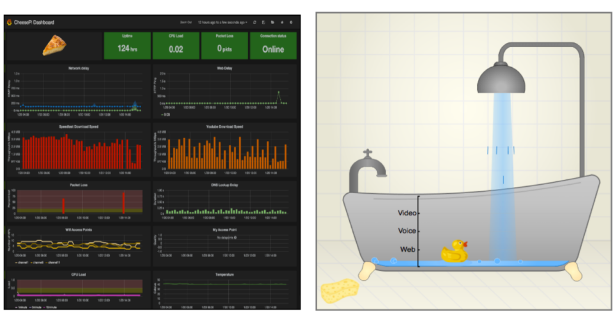

Figure 4: *Dashboard examples.* Left: Technical time-series representation. Right: Throughput as home water flow.

5 CHEESEPI TO CHEESEPI

We describe three scenarios that need control over two endpoints. Examples include: a client-server experiment to determine TCP-level end to end delay, VoIP quality experiments and capacity-throughput tests.

5.1 In-protocol delay measurements

An application to measure the network delay using both ends of a TCP connection is as shown in Figure 5. The idea is to

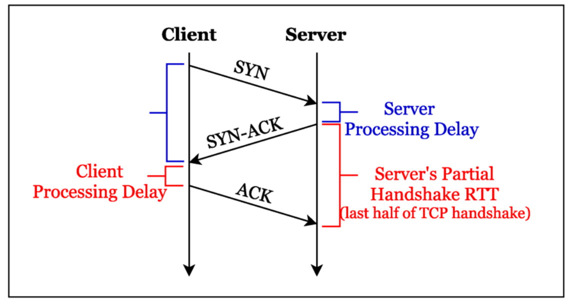

Figure 5: Measurement view of a HTTP request during connection establishment as seen from a client and a server.

use TCP as a "better ping". An insight is to separate the network delay from the object retrieval. During a web page request and retrieval, the latency encountered is caused by a combination of the end-terminal and network processing. With the TCP handshake on can use the protocol to measure the RTT, above ICMP, and closer to, but not including, the application. Figure 6 shows a data set of 3000 ICMP

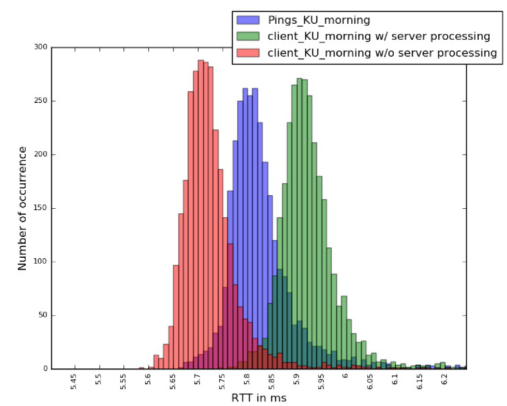

Figure 6: ICMP, TCP and processing delays from Stockholm to Karlstad university (KU) in the CheesePi network.

requests taken at the same time as TCP handshakes. In this experiment, we measured the end-host processing time as seen from the client. Although the processing done by the server between an incoming SYN packet and an outgoing SYN-ACK packet is small, it still contributes to the total round trip time of a connection setup, see [17].

5.2 VoIP quality in the CheesePi network

The E-model [13] is a transmission planning model developed by the International Telecommunication Union. It takes 23 parameters as input and outputs a `Rvalue`, a numerical value ascribing the expected quality of VoIP calls. The model relies

[4] https://www.youtube.com/watch?v=_OBlgSz8sSM

on an additive equation, where each term describes different quality aspects of a call, below.

$$R \quad R_o - I_s - I_d - I_{e-eff} \quad A$$

R_o noise sources such as circuit and room noise

I_s voice impairment to the signal

I_d delay and equipment impairment

I_{e-eff} packet loss impairment

A advantage factor, potential beneficial factors

The plot shows the mean and standard deviation of pre-

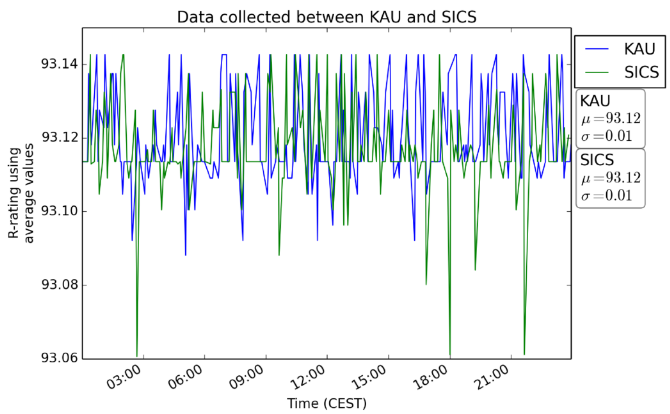

Figure 7: R-rating between two CheesePi nodes over 24 hrs.

recorded calls sent between the nodes using our own tool `Sphone`. It is a 70 second PCM-encoded stored file. The results are comparable in magnitude between two nodes in Stockholm and one node in Karlstad, figure 7. CheesePi predominately uses SUNET, a low-delay, low loss network[5].

5.3 Available bandwidth measurements

Generating traffic to characterize large capacity network links with high accuracy in transport backbones is important for Internet service providers. We used `iperf3` to measure throughput, and indirectly congestion on links within a network. Using short bursts at high utilization can ascertain whether a connection can support capacity sensitive applications, such as video streaming. Repeating the process to capture day-night effects can categorize links for management decisions over busy-quiet periods. As a request from one ISP, where they needed CheesePi nodes to generate higher data rates, we investigated additional single-board computers Table 2. In the throughput case, Figure 8 shows the achievable throughput between two Raspberry Pi nodes, which is at around 80% of the raw speed of the device itself. For more details see [19].

[5]stats.sunet.se

Table 2: Maximum throughput between test machines and measurement nodes (Mbits/sec).

Client	Server	TCP	UDP
Raspberry Pi 1 (B)	Test machine	54.2	69.3
Laptop	Raspberry Pi 1 (B)	84.1	60.0
Raspberry Pi 2 (B)	Test machine	94.1	96.7
Laptop	Raspberry Pi 2 (B)	94.1	95.7
ODROID-XU4	Test machine	837	764
Laptop	ODROID-XU4	785	278

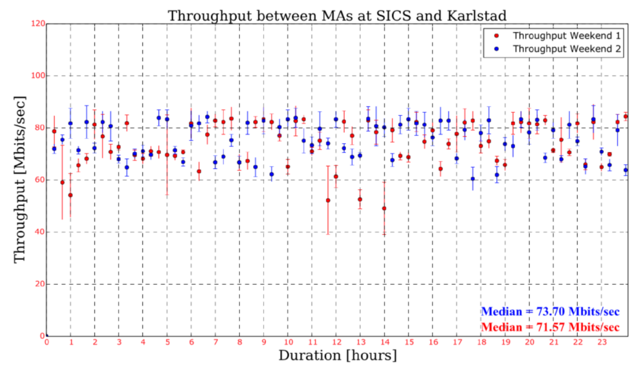

Figure 8: 24 hours (12pm to 12pm) of `iperf` TCP throughput measurements between nodes in the CheesePi network. Conducted on Saturdays of weeks 22 (blue) and 23 (red). Each hour was sampled three times with the median ±1 MAD.

Service provider	Philstar (Phil)	Showtime (Sho)	Sky (Sky)
Domain	philstar.com	sho.com	sky.com
Role	News & entertain. portal	US cable entertain. network	UK-based entertain. network
Server location	Arizona, USA	Ams, NL	Ams, NL
CDN provider	-	Akamai	Akamai
Timezone difference (hrs)	-7	+1	+1
Internet hops	12	8	9
ICMP delay (ms)	196 ±36	29±2	2±3
HTTP delay (ms)	418±128	79±130	24±62

Table 3: Media server characteristics for the most televised Boxing event in history.

6 INTERNET TO CHEESEPI NODES

In the third mode, the Pi acted as a home node to quantify Internet quality during a popular event. One such event was a boxing match between Mayweather-Pacquiao in 2015. We gathered 10 hours of data resulting in 80K measurements from three distributed Pis to three web servers screening the event. The servers are not Raspberry Pis, but rather hosting large severs. IP networking delays using `ping` and the HTTP frontend server responses using `httping` were captured. The latter tool measures a GET request to the remote webserver. `Traceroute` and `mtr` were used as reachability tools.

Figure 9 depicts the HTTP delay to our chosen three sites. Clearly, there is an increase in the delays towards the

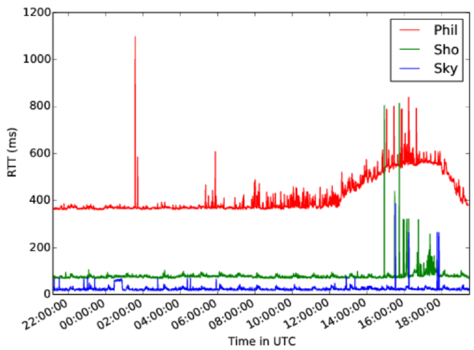

Figure 9: HTTP latencies from CheesePi nodes to 3 global media servers hosting Mayweather/Pacquiao boxing event.

event. The total response times increased up to and including the match itself. One question is can we separate out the difference between the network and server delays? We discuss one approach next.

6.1 Coupled network-server delay

We have seen in the previous results that the major artifact is the delay and its variance. In Fig. 9 we saw the increase in both the mean and variance toward the event in *both* the network and server. There is clear *coupling* between the systems. Clearly the average delay is due to the network,

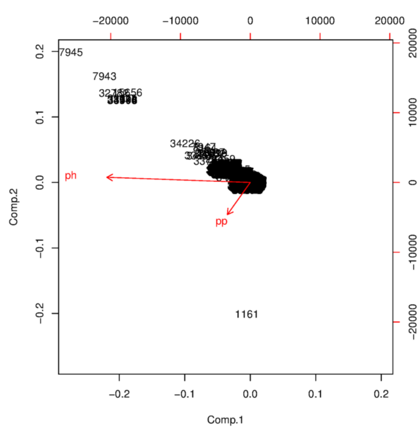

Figure 10: *Example of coupled systems.* The network and server round trip latencies to Philstar as PCA components for the server (PH) and network (PP) latencies.

however, the variance in the delay is due to the server. We show a scatter plot of each delay measurement (as a number) in a visualization space. These are the eigenvalues and eigenvectors of the covariance matrices of the server and network delays. Both vectors are positive and the contribution of the server is around ×4 that of the network (IP) delay. Vectors at 90 degrees indicate independence of the delays, whilst 0 degrees indicates co-dependence between them and in this case is 72 degrees. One advantage of using this approach over time series plots, is it allows additional measurement features/quantities to be included, as was illustrated in Fig. 2. Note, the data is *whitened*, distributions are shifted and scaled to zero mean and unit covariance, no information is lost.

7 THE USER'S PERSPECTIVE

Writing technical applications is straightforward, however working with real users on their daily habits is a different discipline entirely. Domains of psychology and behavioral studies were needed to enquire and interpret the relevant issues, which we briefly show as survey results. Figure **??** shows how our surveyed users connect, motivating video stall research can be done via the answers as well. This will be done again with the rollout of additional Pis. The full report is at https://goo.gl/X5QvTQ.

Figure 11: Example questionnaire responses.

8 DISCUSSION

(1) In the Dagstuhl seminar, January 2016, there was some discussion on the usefulness of measurement platforms [2]. CheesePi has less lofty ambitions. It really is run as the Raspberry Pi project, i.e. as a small group of hobbyists, but be useful.
(2) The challenges of fair, free and unbiased measurements is a significant challenge. Ensuring what one measures can be seen as neutral, and indeed must be fair. We have started out with our regulator onboard, and will bring the ISPs in soon. This is for collaboration and to alleviate any blame in measurement findings.
(3) Perhaps common to all measurement campaigns, planning is never adequate. In our case, it was one ISP wanting to generate higher data rates, and using CheesePI as a free alternative to a commercial offering (NetRounds).
(4) Minimizing the amount of data collected. Includes deploying the least number of nodes, gathering only what is needed, and reducing the dimensionality of the data, through removing features that do not contribute to the statistical correctness.

9 FUTURE WORK

(1) The roll out of Pis, we are looking at around 100 nodes in Sweden to cover the demographics and statistics required by the regulator. The operator sends out 2000

questionnaires to users, although we will not deploy that many nodes.

(2) A better understanding of the root causes of buffering video events. Are "rings of death" due to congestion? buffer adaption, bitrate variability, or actions outside of the connection? e.g. radio interference. Streaming video tests have started in this direction.
(3) Detecting if separate end-systems share a common path to a server and hence explain correlated spatial and temporal congestion. Where `traceroute` cannot reveal topology [5], inference at the end system poses an interesting research challenge [18].
(4) In collaboration with the Université Catholique Louvain, Belgium, and Karlstad university, Sweden we will evaluate multipath TCP. Topics will be path selection and low-latency packet scheduling in the Raspberry Pi kernel, evaluated as part of the CheesePi network.

10 CONCLUSIONS

In this paper, we have described the philosophy, design, implementation, data, analysis, actors, user feedback, future and some conclusions from our home-oriented measurement platform. It is the culmination of around 2 years work. We have conducted around 5 different types of experiments using CheesePi. We have also started to look at experiments exposing the difficulties in measuring coupled systems.

The regulator's knowledge is essential in the deployment of nodes. Interaction with the ISPs is also important, they specified data rates higher than we had tested, motivating using higher performance units such as the Odroid.

Data analysis is an important facet in CheesePi. Not described in this paper, is multi-dimensional measurement visualization of the data and is very much ongoing work. We have compared dimensionality reduction methods for network measurements in a separate paper. Since we are dealing with users, in their homes, data presentation is important.

Worthy of mention is that CheesePi, and Internet measurements has been a good platform in which to educate students. It is relatively simple to learn, extend and experiment with. This is due to the decision to write most of it in Python.

This work was financed by the Swedish Knowledge Foundation project entitled "Research environment for low-delay Internet", under grant number 20130086.

REFERENCES

[1] 2015. EU 7th Framework program, Large-Scale Measurement Platforms. http://www.leone-project.eu. (April 2015). [Online; accessed 01-August-2015].
[2] Vaibhav Bajpai, Arthur W. Berger, Philip Eardley, Jörg Ott, and Jürgen Schönwälder. 2016. Global Measurements: Practice and Experience (Dagstuhl Seminar 16012). *Dagstuhl Reports* 6, 1 (2016), 15–33. https://doi.org/10.4230/DagRep.6.1.15
[3] Vaibhav Bajpai, Steffie Jacob Eravuchira, and Jürgen Schönwälder. 2015. Lessons Learned From Using the RIPE Atlas Platform for Measurement Research. *SIGCOMM Comput. Commun. Rev.* 45, 3 (July 2015), 35–42. https://doi.org/10.1145/2805789.2805796
[4] V. Bajpai and J. Schonwalder. 2015. A Survey on Internet Performance Measurement Platforms and Related Standardization Efforts. *Communications Surveys Tutorials, IEEE* PP, 99 (2015), 1–1. https://doi.org/10.1109/COMST.2015.2418435
[5] K. Claffy, D. Clark, and M. Wittie. 2014. The 6th Workshop on Active Internet Measurements (AIMS6) Report. *ACM SIGCOMM Computer Communication Review (CCR)* 44, 5 (Oct 2014), 39–44.
[6] David Clark. 2017. Policy impact of measurement. In *Workshop on Active Internet Measurements*. http://www.caida.org/workshops/aims/1703/abstracts.xml
[7] Mark Crovella and Balachander Krishnamurthy. 2006. *Internet Measurement: Infrastructure, Traffic and Applications*. John Wiley & Sons, Inc., New York, NY, USA.
[8] Mohan Dhawan, Justin Samuel, Renata Teixeira, Christian Kreibich, Mark Allman, Nicholas Weaver, and Vern Paxson. 2012. Fathom: A Browser-based Network Measurement Platform. In *Proceedings of the 2012 ACM Conference on Internet Measurement Conference (IMC '12)*. ACM, New York, NY, USA, 73–86. https://doi.org/10.1145/2398776.2398786
[9] Nick Duffield. 2012. Fair Sampling Across Network Flow Measurements. *SIGMETRICS Perform. Eval. Rev.* 40, 1 (June 2012), 367–378. https://doi.org/10.1145/2318857.2254800
[10] J. Garcia, S. Alfredsson, and A. Brunstrom. 2015. Delay metrics and delay characteristics: A study of four Swedish HSDPA+ and LTE networks. In *2015 European Conference on Networks and Communications (EuCNC)*. 234–238. https://doi.org/10.1109/EuCNC.2015.7194075
[11] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Comput.* 9, 8 (Nov. 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
[12] IETF. 2015. Large-Scale Measurement Platforms. http://datatracker.ietf.org/wg/lmap/charter/. (2015). [Online; accessed 01-August-2015].
[13] ITU-T. 2014. *The E-model: a computational model for use in transmission planning*. Technical Report. International Telecommunication Union.
[14] Diana Joumblatt, Renata Teixeira, Jaideep Chandrashekar, and Nina Taft. 2011. HostView: Annotating End-host Performance Measurements with User Feedback. *SIGMETRICS Perform. Eval. Rev.* 38, 3 (Jan. 2011), 43–48. https://doi.org/10.1145/1925019.1925028
[15] Liam MacNamara and Ian Marsh. 2015. CheesePi: A Raspberry Pi based measurement platform. In *IRTF and ISOC Workshop on Research and Applications of Internet Measurements (RAIM)*.
[16] K. Nieminen. 2017. Net Neutrality Measurements: Regulatory Use Case and Problem Statement. (Feb. 2017). https://datatracker.ietf.org/doc/html/draft-nieminen-ippm-nn-measurements
[17] Rebecca Portelli. 2016. CheesePi: Delay Characterisation through TCP-based analysis from End-to-End Monitoring. (2016), 107 pages. http://kth.diva-portal.org/smash/record.jsf?pid=diva2:1039071
[18] D. Rubenstein, J. Kurose, and D. Towsley. 2002. Detecting shared congestion of flows via end-to-end measurement. *Networking, IEEE/ACM Transactions on* 10, 3 (Jun 2002), 381–395. https://doi.org/10.1109/TNET.2002.1012369
[19] Sagar Sharma. 2017. Accurate traffic generation within the CheesePi network. (2017), 107 pages.
[20] Samira Tavakoli, Kjell Brunnström, Jesús Gutiérrez, and Narciso N. García. 2015. Quality of Experience of adaptive video streaming. *Sig. Proc.: Image Comm.* 39 (2015), 432–443. https://doi.org/10.1016/j.image.2015.05.001