Internetfonden final deliverable: CheesePi dashboard

Ian Marsh and Liam McNamara

2016-12-01

CheesePi: Home Internet Dashboard

← Föregående Nästa →

År: 2015

Taggar: bredband, forskning, Ian Marsh, internetuppkoppling, Kista, mätplattform, Raspberry Pi, SICS Swedish ICT

Foto: "Pi Bling" av Drew Morris

För vanliga människor kan problem med internetanslutningen i hemmet vara mystiska och svårbegripliga. Projektet CheesePi handlar om att utveckla ett mätverktyg som på ett användarvänligt sätt möjliggör kontinuerlig och enkel övervakning och diagnostik av folks internetuppkoppling.

Syftet med detta projekt är att utveckla en informativ och användbar "dashboard" för att även internetanvändare utan specialistkunskaper ska kunna få en mer nyanserad uppskattning av hur väl deras internetanslutning fungerar. Ett sådant gränssnitt gör det möjligt för användare att omedelbart se hur deras internetanslutning för närvarande fungerar och hur dess egenskaper förändras över tid. Mätplattformen ska köras på en Raspberry Pi som placeras i hemmet.

Figure 1: IIS's CheesePi project summary page

Project summary

CheesePi (http://cheesepi.sics.se) is a network measurement system that objectively characterises the service users experience from their home broadband connections. Many Internet monitoring systems already exist, some with a focus on home users, however they focus mainly on collecting network-wide information without consideration for providing utility to the person actually hosting the measurement node. We believe that to encourage participation in home broadband measurement systems a node should be as useful as possible to the host, no matter their technical expertise. Through this proposal we will research and develop features, primarily through an informative and easy-to-use visual interface, enabling easy monitoring and diagnostics of a person's home internet connection.

Unfortunately, for members of the public without specialist networking knowledge, problems with their home broadband connection can be mysterious and hard to describe. Indeed, even for network engineers, it can be hard to notice, quantify and record transient problems. To this end, we have built a measurement platform that captures low-level network details, behaviour and service. The platform runs on Raspberry Pis, an extremely popular hobbyist credit-card sized computer. The use of small, cheap, commodity hardware (Raspberry Pis) as monitoring nodes enables easy, flexible and widespread deployment. Over 2 million people already own one, and can simply install our software on it. The system periodically measures and records various attributes of their connectivity, e.g. delay, packet loss, capacity and local wireless signal strength.

Understanding the raw recorded data is not a simple task and, if not carefully presented, requires specialist knowledge. The aim of this project is to create an informative and useful "dashboard" for non-expert users to view the health of their connection. Such an interface would allow users to instantaneously observe how their Internet connection is currently functioning and how its attributes have been changing over time. The project will provide an opportunity for people to move beyond a naive understanding of "connected" and "disconnected" to the Internet towards a more nuanced appreciation.

The results of the measurement system will be high-quality longitudinally measures of Swedish broadband service in a free, neutral and unbiased manner with a focus on gauging Quality of Experience. The resulting network data will be analysed and collated then used to monitor the performance of Swedish infrastructure (e.g. ISP networks and DNS resolution).

The scope of this proposal will deliver the following specific outcomes:

1. Visualisation of existing network measurement tools.
2. Expose possible shortcomings of current measurement tools to capture experienced network behaviour.
3. Provide a detailed evaluation study of user interaction with network measurement data.
4. Investigate how well users are able to interpret such data and whether this educates them about networks.
5. Feedback on potential of visualisation of network tool output to communicate network health to non-expert and expert users alike.
6. Demonstration of the working system, and the data that it produces, to external partners (PTS and .SE)

Background

The preceding 20 years have seen the expansion of Internet based services from niche offerings to fundamental aspects of a majority of people's lives. Applications range from Netflix, Facebook and Skype to more important services such as home security systems and eHealth solutions. All of these are reliant upon a well-functioning Internet connection, some requiring minimum levels of performance for their services to operate with reliability and stability. The performance of a network connection is not static and can change dramatically on a daily and weekly cycle. Network operators may struggle to provide the level of service that they have promised. For consumers this is a challenge, how to know when their connectivity is "good enough" for what they would like to achieve with it, and whether this is due to the network provider (e.g. poor ADSL contention ratio) or local problems with their own network (e.g. WiFi contention or too many connected devices).

We have been developing a platform of measurement tools to understand the wider network, but also to aid individuals understand their own connection better. The initial design for the measurements and code distribution follow the guidelines set out by the IETF's RFC on Large-Scale Measurement of Broadband Performance (LMAP) and was accepted to appear at an IETF/IMC workshop Research and Applications of Internet Measurements (RAIM).

Internet monitoring has been widely studied, though generally with a focus on 'the core' and behaviour between well connected ASs [1,2,3,4]. We are specifically interested in Swedish Internet performance, the main actor being Bredbandskollen (with tptest). However it specialises in measuring from the edge to the core and is only user-initiated, rather than continuous like CheesePi. Some other similar projects for personal network monitoring are NetBeez and SamKnows [5,6]. NetBeez is a distributed network monitoring system, which is mainly targeted at companies owning multiple offices in different locations with a centralized IT department. The agent, a Raspberry Pi, is configured to run ping, DNS resolutions and provides a web-based dashboard.

SamKnows is a project that measures broadband performance for consumers, ISPs, and government (regulators) around the world. Their Whitebox (a modified router) measures throughput and various service application types between the Whitebox test servers. Similar to CheesePi they use a dedicated device and dashboard for displaying results, but users have no freedom to control the Whitebox, nor the dashboard. SamKnows has similar customers and users as this project, which are home internet users, operators, internet regulators, and academia.

[1] BISmark (http://projectbismark.net)
[2] RIPE Atlas (https://atlas.ripe.net)
[3] ARK project (http://www.caida.org/projects/ark)
[4] Leone (http://www.leone-project.eu)
[5] SamKnows (https://www.samknows.com)
[6] NetBeez (http://netbeez.net)

Success factors

The project scope is focussed and bounded and so should be achievable in the project time frame.

SICS has significant expertise in building and using networking tools which can collect the data needed to characterise the service experienced through a home broadband connection. We focus on applied research and practical systems deployments and experimentation, particularly in the area of embedded device networking. Our networking expertise will enable us to build the dashboard in a meaningful fashion. This project has only recently become possible due to the release of powerful visual

data-oriented Javascript libraries.

Uniquely, our platform contains automatic VoIP quality estimators using a suite of tools previously developed at SICS. This project will add more novel network measurement tools, such as Bandwidth Available in Real-Time (BART). We will evolve presentation of network metrics from text-based numerical output into accessible visual forms. For instance the visualisation will interpret the TCP window size into more a meaningful display such as a scrolling graph of "available capacity".

We will provide a more fully-featured mode that gives an abundance of in-depth technical information for more advanced users.

Limitations

There is only provision for purchasing a small deployment of nodes, enough to test the project with users and confirm that it is functional and useful.

- We will not develop a Javascript visualisation library, but use existing libraries, such as D3.
- We will not modify networking tools, simply use standard or existing tools.
- We are not measuring the infrastructure of the core Internet, rather Swedish home connections.
- Our measurement and code distribution is functioning and will not be advanced by this project.

Risks

Failing to find a variety of participants for our user study would limit the ability to demonstrate utility for a range of user technical competency.

Measured metrics do not reflect the experienced behaviour of the network connection. This would be an interesting finding. We should be careful to expose the uncertainty in the reliability of presented measurements, such as through colour coding or opacity of the display.

Progress through the project

The project progressed as planned with the exception of month 7 (July) where one month was lost due to illness. This was conveyed to Erika and hence the final deliverable has been delayed by a month.

Existing Internet dashboards

Browser-based measurement tools offer a simple user interface “click and run” method to assess the Internet “quality” for users. Speedtest around the world, and Bredbandskollen in Sweden, provide a measurement session within 30 seconds. Feedback in the Bredbandskollen tool is given indicating whether the achieved speeds and delays are more or less inline with other users’ measurements. Up and download speeds and delays are given separately.

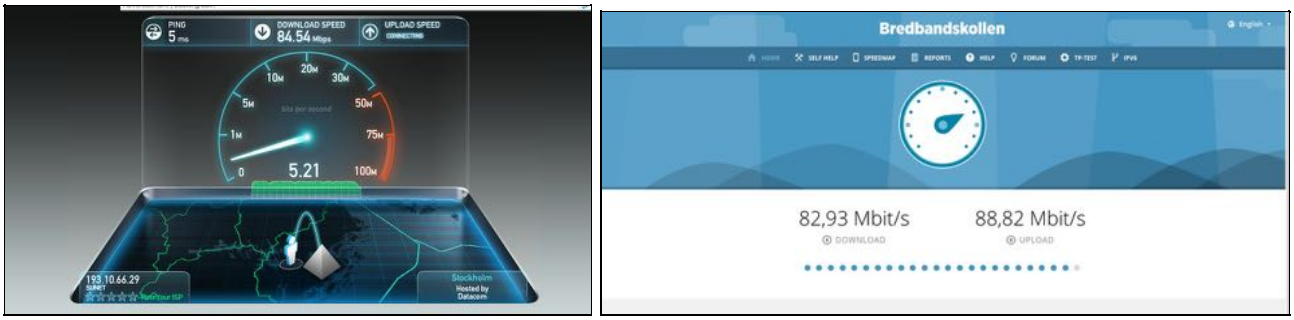


Figure 3: Speedtest (left) and bredbandskollen (right) for testing users connectivity and performance

Although use cases vary, they generally fall into a few broad categories 1) Users verifying whether the new data rates are achievable after an Internet subscription upgrade. 2) Using a tool independent of their ISP to check problems with their connections, 3) curiosity or “just to try” to see my connection 4) gamers who are having problems with multiplayer scenarios.

The technical nature of the output; delay in milliseconds or throughput in Megabits per second is more suitable for advanced users. Bredbandskollen puts such metrics in context by comparing them with others results.

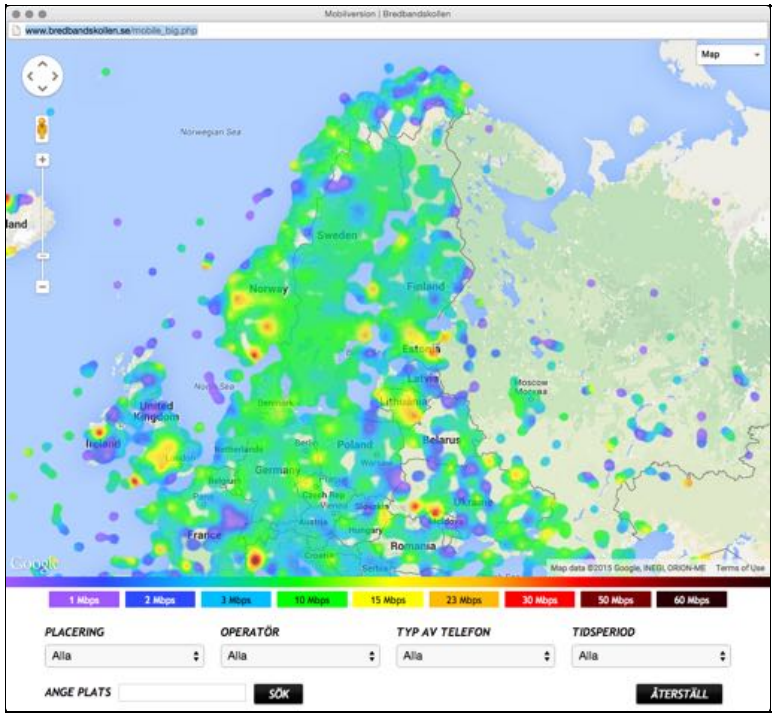


Figure 4: Internet connection speeds in Sweden

From our conversations Speedtest and Bredbandskollen are rarely used by customers on a regular basis to “monitor” their home connection speeds. On the other hand, Ookla (Speedtest) and Internetfonden (Bredbandskollen) do publish reports on best countries or regions, so having users perform such tests for aggregated analysis, for example per region, is very useful. In Sweden there has been partnerships between IIS and the regulator PTS for example.

Technically both Speedtest and Bredbandskollen measure the delay to a close measurement server and the time to download a file. Speedtest has 4000 globally located servers, whereas Bredbandskollen (run and operated by Internetfonden) uses ~5 servers located at Internet exchange points around Sweden.

https://support.speedtest.net/hc/en-us/articles/203845400-How-does-the-test-itself-work-How-is-the-result-calculated-

Although they are widespread, (once people know the web site address) and easy to use, they do not represent an ongoing measurement test that can say how the users quality varies over a longer period. In other words, both Speedtest and Bredbandskollen perform *instantaneous* tests.

Internet Measurements

Measuring the Internet is a complex task and definitely not for the non-technical. Some understanding of the Internet operation is needed, as well as good programming skills as well as statistics to process the inevitable complex and often large amounts of data from the measurement process. Tools like Bredbandskollen perform a good job of running tests, processing data and presenting the data to the users. However, there are some shortcomings of web-based tools. First, they can be limited to what they do, often access to lower level functionality is needed. For example Bredbandskollen sends a request to a server (in Sweden) to see how long it takes and then retrieves and uploads 5 different file sizes. This can be fine, but does not capture someone watching YouTube or browsing the Internet. This is because common tasks can be difficult to automate, also they take time and need to be repeated. In this work, we took a programmatic approach to wrap common tools, but make the results presentable in a simple manner.

CheesePi

The CheesePi project run at SICS aims to address the problem of users only having access to instantaneous measurements. Furthermore, it does not require that a user ‘initiates’ a measurement session, they are scheduled automatically. CheesePi is a software solution that needs to be installed on a device, and rather than tie up a device like a laptop, we chose to mostly base it on a small, quiet, cheap device, like the Raspberry Pi. Furthermore, CheesePi is meant to analyze traffic from devices such as laptops, tablets, mobile phones on the WiFi network. Note CheesePi does not look at the actual contents of the traffic, solely the volumes sent and received.

Previous work within CheesePi focussed on measuring network conditions and presenting detailed data in a dashboard, although the presentation is rather technical, as shown in Figure 5.

Figure 5: Technical dashboard of CheesePi

Project milestone

Visualisation of existing network measurement tools

Visualisation of existing networking tools has been done through our technical dashboard. Each tool is "wrapped" in Python code and can be run and the data put into our dashboard. Currently the dashboard supports 15 measurement tasks : `WiFi Beacon`, `dns lookup`, `Dash`, `httping`, `mtr`, `ping`, `PingB`, `Status`, `TCP handshake`, `throughput (WiFi)`, `traceroute`, `VoIP (Sphone)`, `Wifi`, `iPerf`. An example is shown below:

```
{"taskname": "ping", "period": 900, "offset":0, "landmark": "www.sics.se"}
```

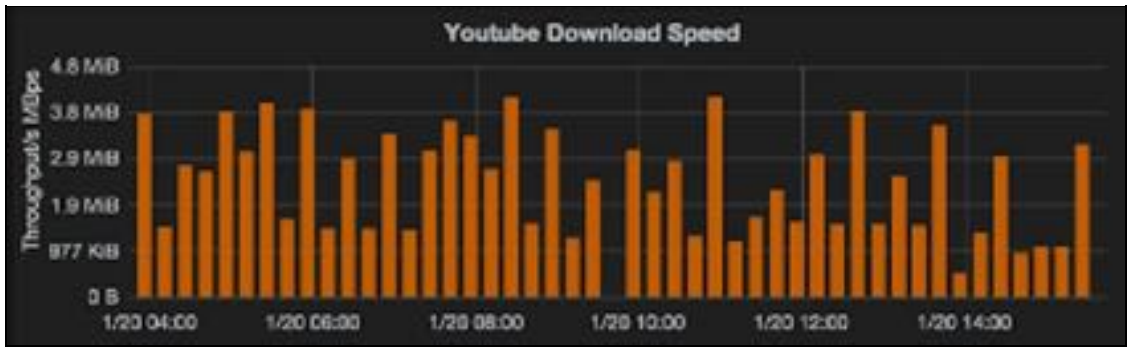
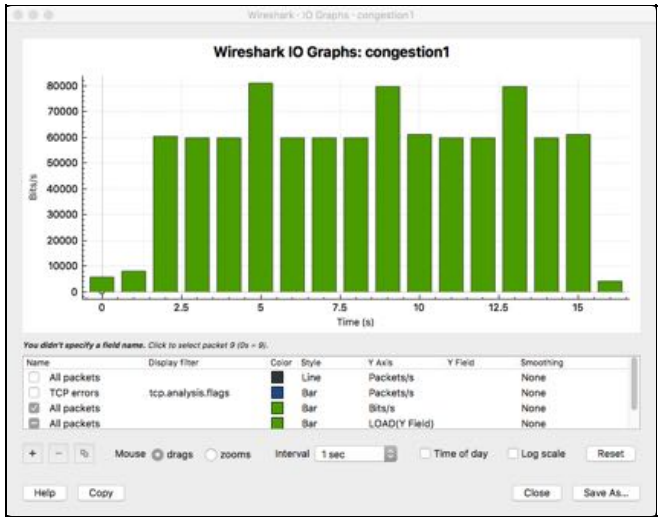


Figure 6: YouTube performance over a 10 hour period

Where we wanted to evaluate data, rather than tools themselves, we used the graphite platform, which is a python-based display of command line data. One can stream data into the `nc` tool and obtain a plot as thus `cat data | nc`. We also used little lower-level packet trace for some of tools in the list above as shown below:

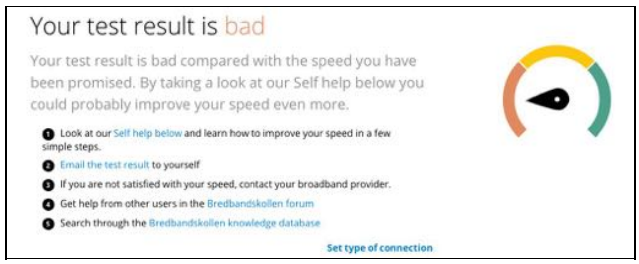
Figure 7: Wireshark trace of traffic flow (an iPerf session)



Expose possible shortcomings of current measurement tools to capture experienced network behaviour.

The current state of the art in network measurements is in window based display with time on the x-axis and some performance measure on the y-axis (e.g. above), they are not entirely natural. For a non-technical user, it is not always clear what the y-axis represents and what should a good value be? Bredbandskollen tries and puts into context if similar tests from close locations produces similar delay and upload/download performance. However the problem remains, what is a good value and which representation this would indicate to a home user *over time*.

Figure 8: Comparative results from Bredbandskollen.



Furthermore, the idea of an accelerometer is not the only representation. Although it represents “speed”, faster is better. For delay however, the lower the better, so mixed messages could be understood. In fact the delay of the connection is normally presented at the same time as throughput as XX ms in the corner. Pernilla Rydmark at IIS also indicated to us, that the accelerometer paradigm could be questioned, which we also thought could be readdressed in a new dashboard.

One further issue is that of flash in Bredbandskollen. Browsers are dropping their support, so HTML5 is the only real option moving into 2017. Some people say that Javascript (for all its pros

and cons) is the “assembly language of the web” which to some degree is true, as one can achieve everything via a browser.

Figure 9: Flash generally accepted as not sufficient for 2017

Provide a detailed evaluation study of user interaction with network measurement data

A running version of the project is available from http://cheesepi.sics.se/webserver/bath. A version of the survey is linked from the page, as well as some information about the project.

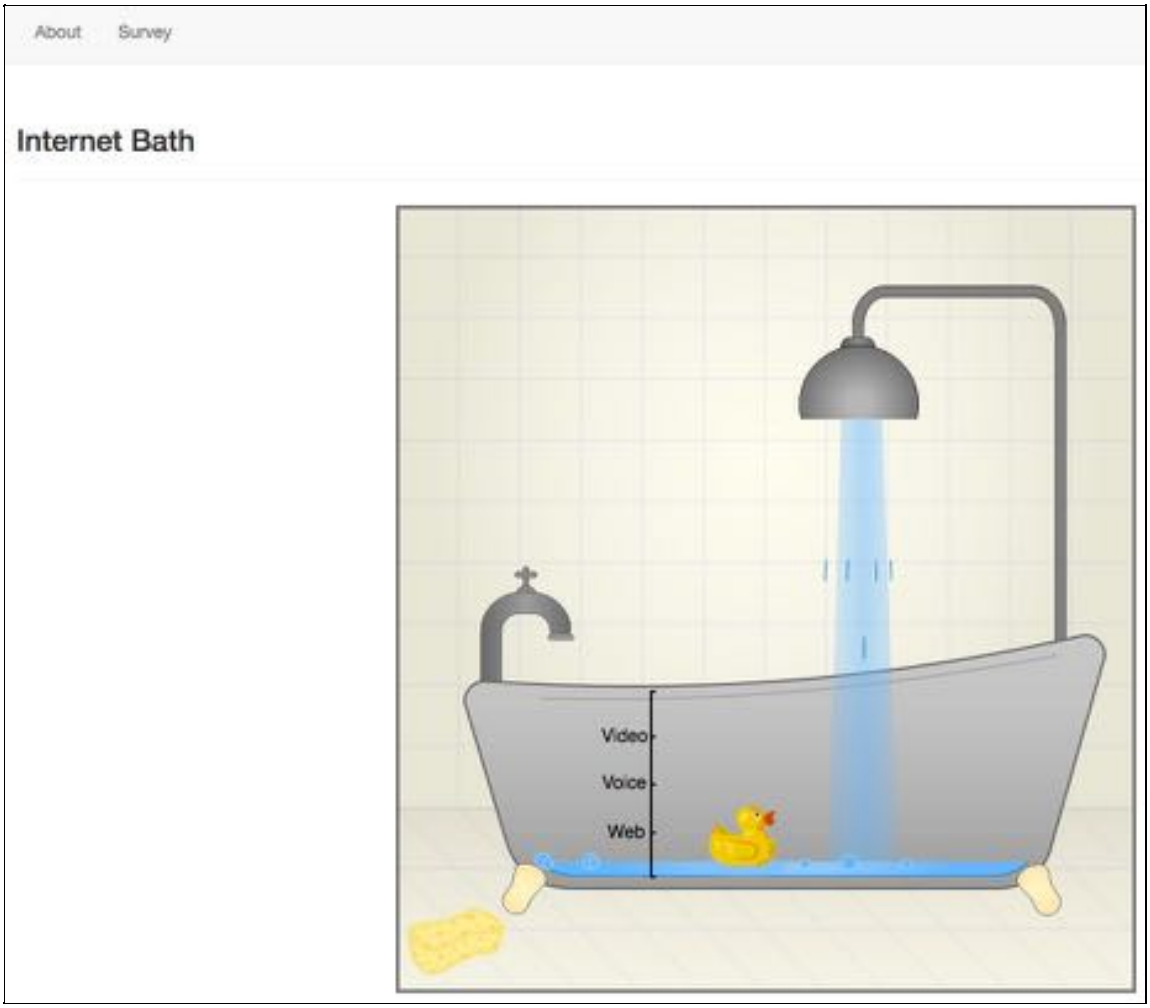


Figure 10: Screenshot of SICS' Internet measurement tool

Investigate how well users are able to interpret such data and whether this educates them about networks

Feedback : user hints (mouseovers)

We expected our bath example to be used for different levels of expertise. Therefore to give users some feedback in our thinking, we added some mouseover events. Hovering the mouse over some objects reveals additional information on the ideas of the visualisation functionality, as shown below.

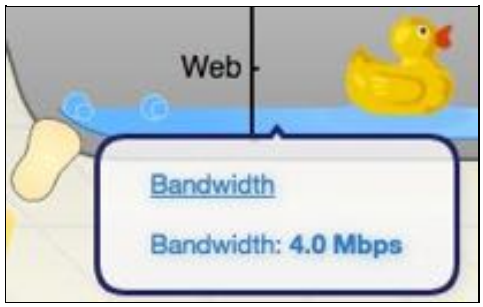


Figure 11: Mouseover hints help users understand the Internet bath

Feedback from our users on these events was not included in the survey, as they were added in the middle of the tests to see if they helped or not as an A/B test. A is without the events and B is with the hints and we can see if they are helpful. From the comments, it is helpful to have feedback to the user.

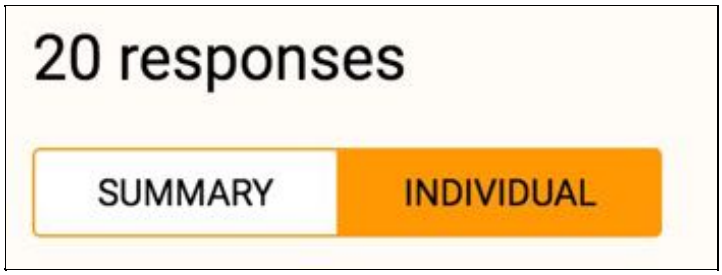


Figure 12: Individual and collective responses can be taken from the questionnaire (note: we chose to make the replies anonymous)

Feedback : Audio instructions, commentary

We have tried to go for a minimum and intuitive type of presentation, but it would be possible to have small hints during the display and even a short instructional sequence before, something as the video demonstration of Internet delay (as shown below).

Figure 13: The EU RITE project has an intuitive video explanation of Internet characteristics (https://riteproject.eu/)

Feedback : In situ tests:

We tested households to see the opinions of groups of people. We did follow the discussions in particular but asked people to enter the general feeling.

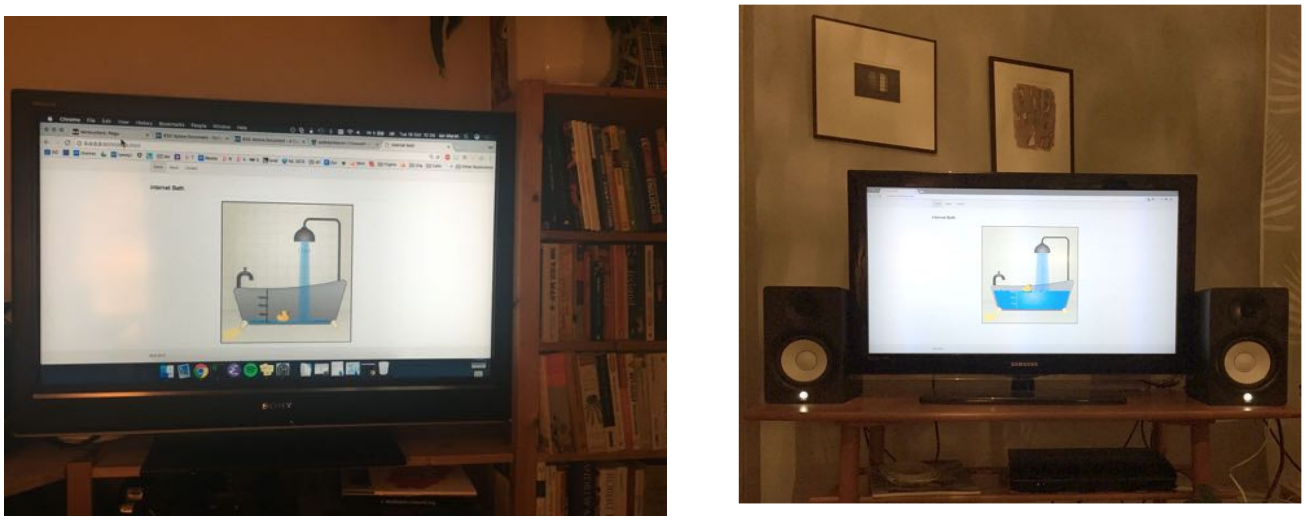


Figure 14: Internet bath under test in users' home

Part of our testing was done in homes, 2 are shown above. CheesePi can measure continuously and present the results of data "flowing" into a users home. We should stress that the bath mockup is a representation of the data flowing into a home network. It is accurate and reliable, but may not exactly match second-by-second of the network.

More precisely, the mean rates is represented by the height and swell represents the variance of the traffic flowing in. Since network conditions tend to vary on small time scales (particularly Wireless) we smooth the data to make the visualisation more pleasing. This is quite common, as can be seen in the WiFi scanner on Android as below:

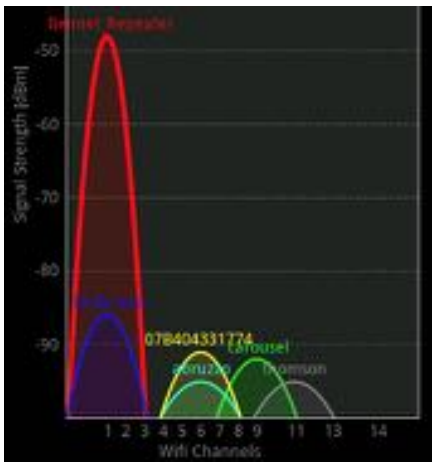


Figure 15: WiFi scanner as example of smoothing network measurements.

The lessons from intuitive but representative animations / visualisations / are that some dynamics should be included to show motion, even if the absence of new measurement data. In Javascript there is support via trigonometric functions (sin, cos). The average measurement value can be a level of something, for example the "available" bandwidth as the water level. In this scenario the variation could be represented as the swell. Typically to have a pleasing animation, we want to show slowish varying quantities. Given some parameters of the system vary quickly either smoothing or removing some values is necessary to produce pleasing results. In CheesePi we smooth some of the values to keep the processing of the browser to a minimum.

Mobile version

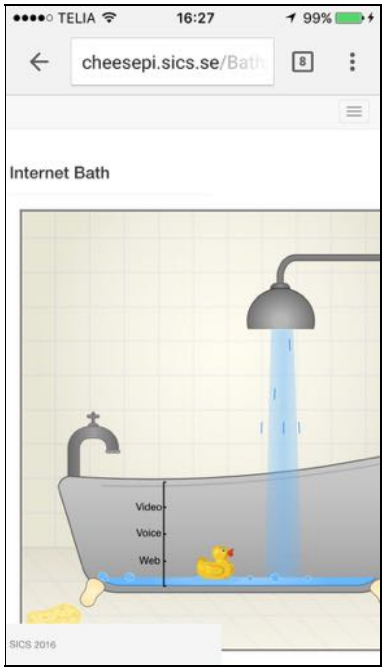


Figure 16: Mobile visualisation of Internet bath

The demo works on mobile devices (Chrome and Safari) as shown above. We expect home users to see the Internet quality either on a HDMI channel on the TV, or more likely via the phone.

Questionnaire design

Questionnaire design is a subtle task, according to () the wording of the survey can severely influence the outcome of the questions being asked.

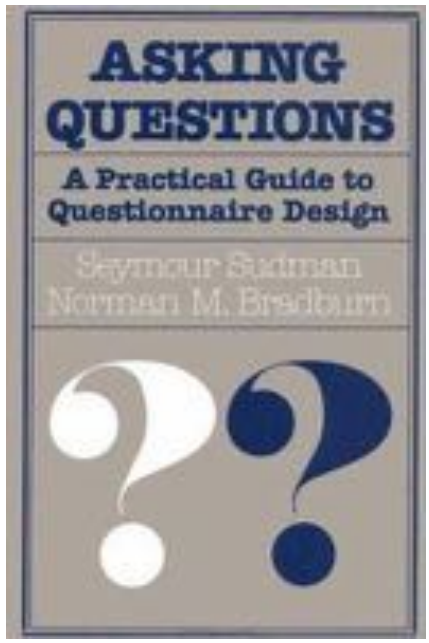


Figure 17: Reference book for the questionnaire design

Therefore we followed the advice given in the book above to construct a questionnaire. http://iss.leeds.ac.uk/downloads/top2.pdf. A Swedish research piece of work relevant to this work is “Questionnaires for Evaluation in Information Visualization by “Forsell and Cooper” from Lingköping, published in 2012.

We also wanted the users to be to complete the questions in a maximum of 15 minutes. We do allow for longer replies, in a free text section, as well as gather feedback verbally and email.

We used Google forms for the actual input and collation of the results, viewable from https://docs.google.com/forms/d/10HP5qcW1dh_eIFSUT4yOJerbsBEn-AtiHn8ABVoc0TY. Forms produces the pie charts as included below, as well as an Excel export.

- Questions pertinent to the bath scenario
    - What does the image suggest in terms of an Internet connection?
    - The water flowing into the bath?
    - The level of the water?

Feedback on potential of visualisation of network tool output to communicate network health to non-expert and expert users alike

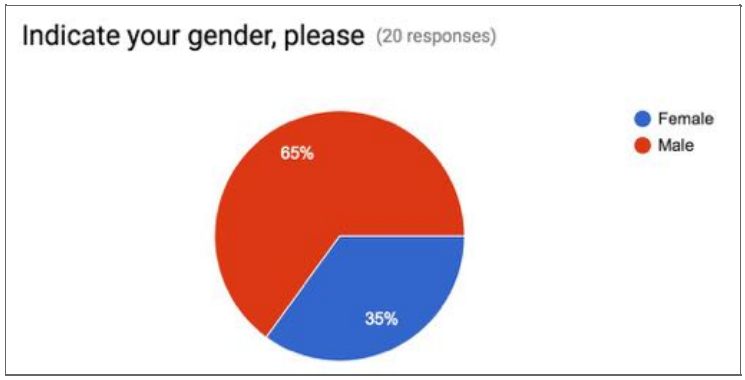
Questionnaire results

1. Date

Good to know, as we have slightly changed the design, for example add mouseovers, so we need to sync responses with those modifications. By having an option such as this it is possible to do some A/B testing of the responses before (A) with those of after (B). These reasons motivated having the date given by the users.

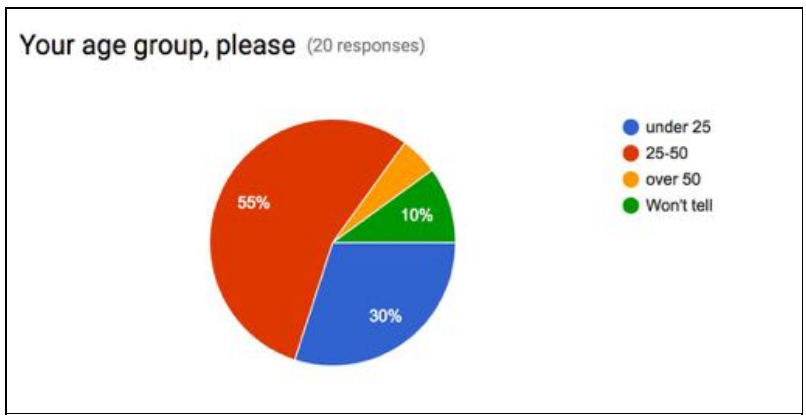
2. Gender

Clearly we would like to have a better gender balance, this will be achieved with a larger test to be conducted on Facebook.



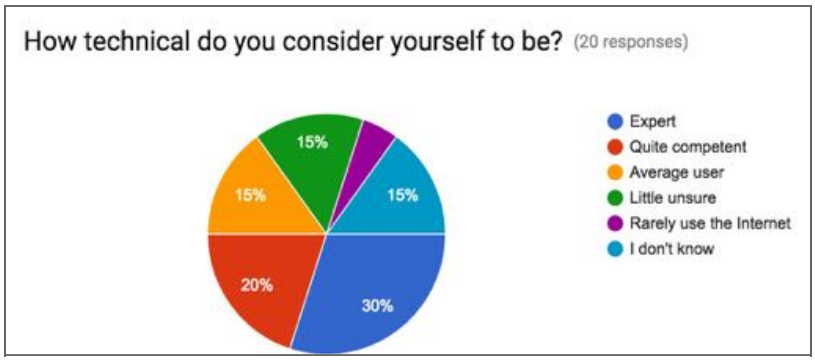
3. Age

Mostly younger users, however we would like to increase the number of older users.



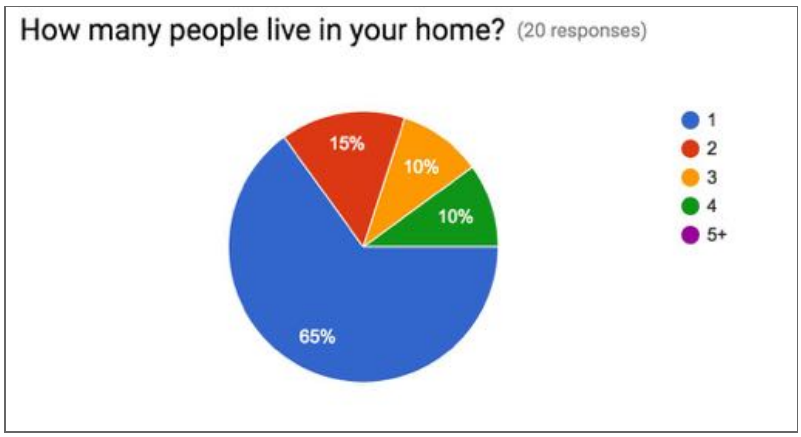
4. (Personal) technical evaluation

It is quite important to estimate the knowledge of the users of the survey. Socio-demography is essential to correlate the answers with the people who took the survey. We would like a reasonably wide coverage of people, will some emphasis on the people who are not experts (ideally). What is of interest is how people *see* themselves in the context of Internet use. People nearly always think that is someone who knows more, and rate themselves in relative terms, rather an absolute.



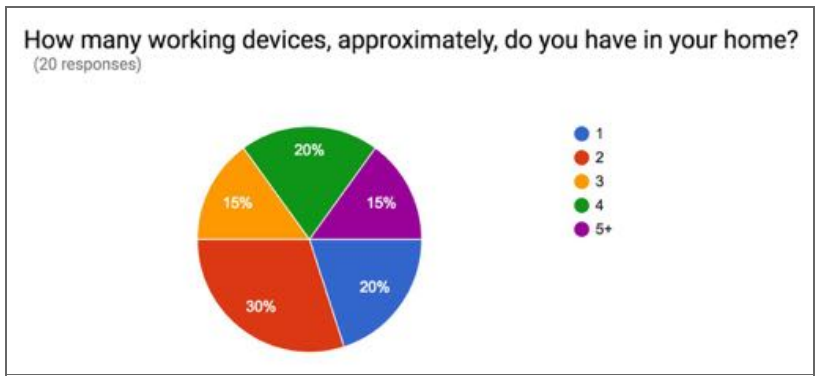
5. **People in the household**

We wonder if many of the homes are single householder, maybe we should focus on families in future studies?



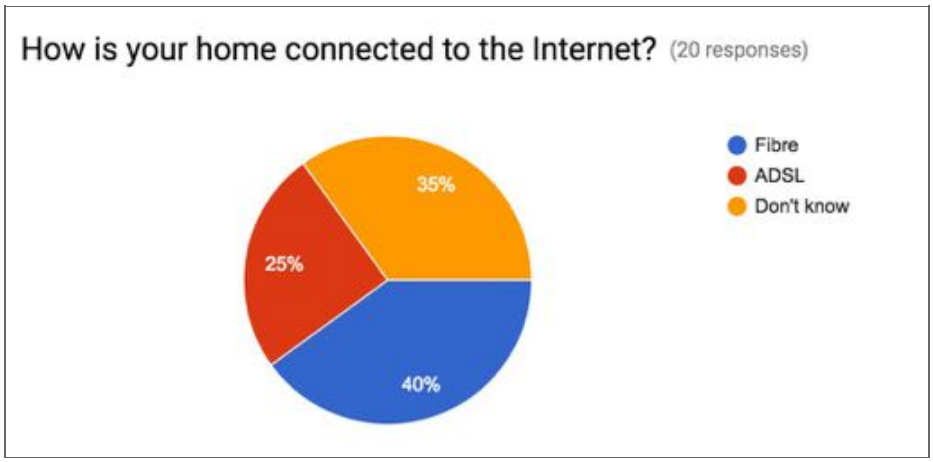
6. **Number of connected devices**

The number of devices with technical competent people might be more, such as things like Raspberry Pis, or families with tablets. Probably more could be done in the questionnaire, but at the risk of adding more time and somewhat laborious questions.



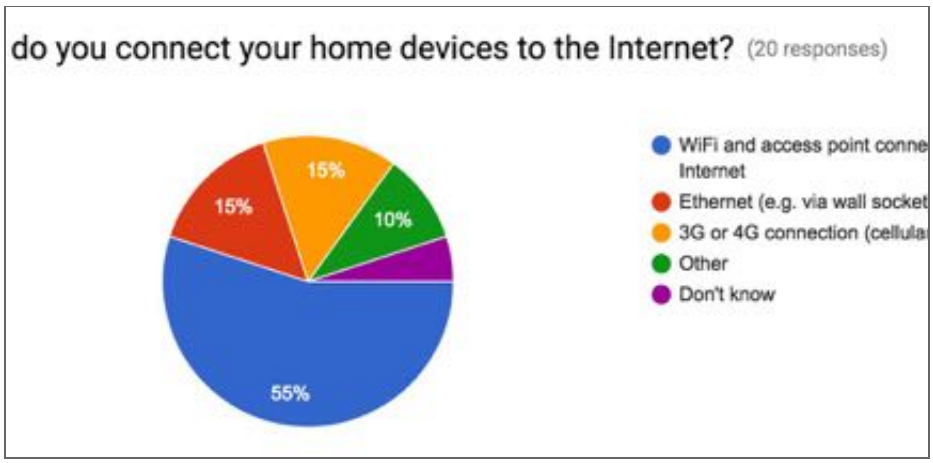
7. **Connection type**

Quite surprising that 7 out of 20 didn't know how they are connected to the Internet, again this should be correlated with some of the other answers, to find out more, for example if ADSL users also have many devices and what quality they obtain (see below).



8. How users connect

Whereas the previous question was about the home to the Internet, this one is about the devices inside the home, i.e. the users' connections. Somewhat unsurprisingly WiFi is predominant, however 1/7 use their cellular connections.

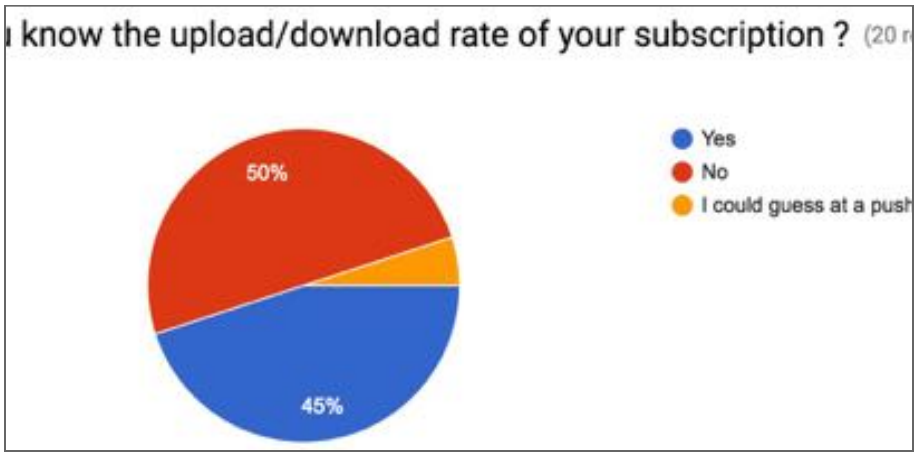


9. Daily usage

Quite an even split from the 20 users who responded to the questionnaire. We suspect the heavy users 8+ hours are the experts, as they use computers at work and home. In fact, almost all those under retirement age use their phones to check messages, but are not always connected AND using the Internet.

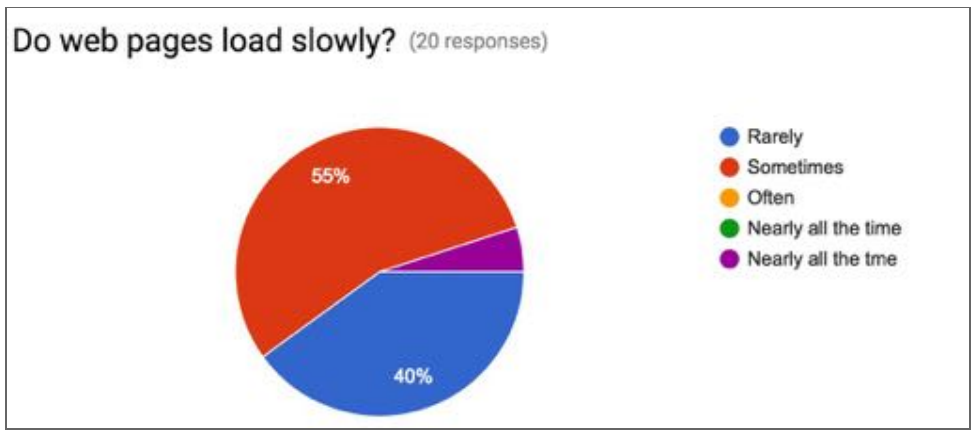
10. Upload and download rates

Simple question on users' Internet rates, somewhat surprising that so many didn't know, however some we questioned were in rented accommodation.



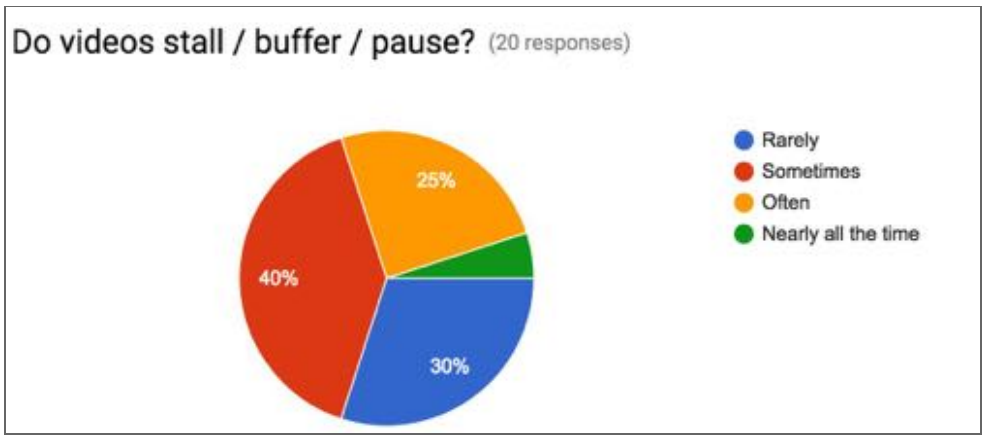
11. Do web pages load slowly?

Yes, essentially, again these results should be correlated with those above to find out, which people, and/or connections could be causing the problems. With only 20 users, cross-correlation will probably only give 2-3 users, but with a larger study, it would reveal a group of users.



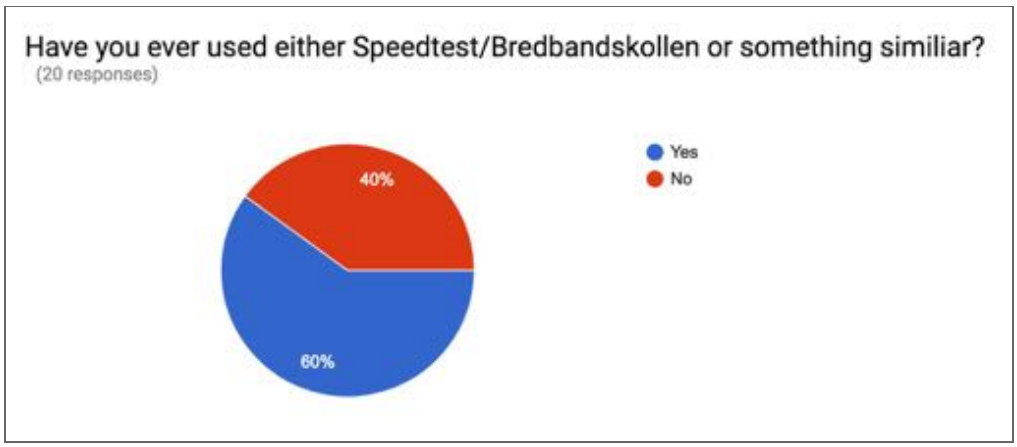
12. Do videos stall / buffer / pause?

¾ of the people experience some video problems. Buffering under heavy loads or poorly dimensioned links is a current research problem being addressed by CheesePi.



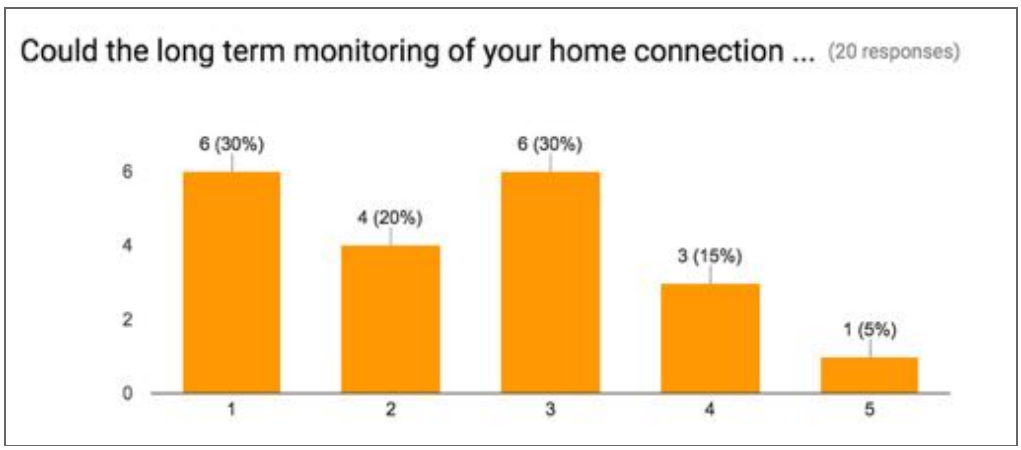
13. Use of other tools

Nearly ⅔ of our respondents have used an alternative to our Dash to measure their Internet connections. This was somewhat surprising and is input to question 14.



14. Use of other tools

Below is a summary of the users thoughts of such a tools. In general it is a good idea, as no such tool exists as of late 2016.



15. 2 sample questions

What does bandwidth mean?

- The rate at which data downloads”
- “Bandwidth is also defined as the amount of data that can be transmitted in a fixed amount of time. For digital devices, the bandwidth is usually expressed in bits per second (bps) or bytes per second. For analog devices, the bandwidth is expressed in cycles per second, or Hertz (Hz).”
- “The speed of my Internet / subscription”
- “Speed”
- 2 NAs

Do you know what Internet delay is?

- “How fast the internet is”
- “The average time for data to be sent from the various sites.”
- “Fördröjning?”
- “The opposite of the bandwidth, one is speed and the other is time, so it is the inverse.”

Conclusions

General

Writing technical applications is relatively straightforward, however working with real users on their daily habits was challenging for computer scientists. The domains of human psychology and behavioural studies requires skills we did not posses at the beginning of the project. However, by reading up and consulting experts we were able to design the appropriate tests and questions.

SICS

It was quite a learning experience for us: user needs, graphic (drawing and illustrator) work, and the questionnaire (questions and feedback) design. It probably needs some refinement, which we will do. Browser support (and phone) very appreciated as well as integration into our CheesePi release(s), the Bath dash included into cheesepi release (1.9.22) via Sudo pip install --upgrade cheesepi.

Non-expert users

1. Non technical users like the idea of data streaming as water flowing into a “home”
2. Acts as some analogy to Internet use
3. Time is shown better in this scenario than Speedtest/Bredbandskollen
    a. It takes time for a site to appear
    b. Video to start
        i. (changing channels on TV is not instantaneous anymore)
4. It would be useful to have the number of people using the connection

5. The idea is useful, ideally on a separate screen
    a. Like a photo frame to correlate Internet usage with the bath
6. It would be useful to see when my connection is full
    a. To blame the kids!

Expert users

1. The ideas are shown are interesting
2. Expert users tend to “add” their own interpretation to what they are seeing
    - For example how is upload represented?
    - Is the shower WiFi or Ethernet?
3. The timings on the flow in and water levels are hard to match to real surfing
4. The markers on the side (web, voice, video) should be adapted to my connection as well

Future and continued work

One of the problems we faced was a good mix of technical and non-technical users. Initially we needed some technical feedback so as ascertain if the scenario is correct as well as iron out small glitches. Given these and the results above we will conduct a larger test via Facebook friends, which represent a larger cross-section of users.

Facebook friends

- Statistical correlation between answers
- Small differences between browsers caused us some small problems
    - Firefox and Edge exposed a SVG/JS/OpenTip bug
- Full screen and reactive mode