Internet Access for All: Assessing a Crowdsourced Web Proxy Service in a Community Network

Emmanouil Dimogerontakis, Roc Meseguer, Leandro Navarro

Universitat Politècnica de Catalunya
{edimoger,meseguer,leandro}@ac.upc.edu

Abstract. Global access to the Internet for all requires a dramatic reduction in Internet access costs particularly in developing areas. This access is often achieved through several shared web proxy gateways in commercially and community driven local or regional access networks. We look at individuals or organisations sharing the full or spare capacity of its Internet connection with other users. We present a measurement study of a crowdsourced Internet proxy service in the guifi.net community network that provides free Internet access to a large community with a very high ratio of users to proxies. The dataset consists of Squid proxy logs for one month, combined with data about the network topology and traffic. Our study focus on a representative subset of the whole network with about 900 nodes and roughly 470 users of the web proxy service. We analyze the service from three viewpoints: web content traffic from users, performance of proxies and influence of the access network. We find clear daily patters of service usage, excess capacity and little reuse of content which makes caching almost unnecessary. We also find uneven and inefficient distribution of traffic load across the access network and proxies. We argument how manual proxy selection by users brings to some issues. Finally, users experience an overall usable Internet access with good throughput despite being a free crowdsourced service.

Keywords: Community network, guifi.net, User experience, Proxy service

1 Introduction

The majority of the world's population does not have none or any adequate Internet access [12]. This implies that the Internet cannot provide service to the general public, to reach anyone without discrimination. Global access to the Internet for all requires a dramatic reduction in Internet access costs especially in geographies and populations with low penetration [21]. Community Networks (WMNs) [20] allow local communities to build their own network infrastructures and provide affordable inter-networking with the Internet including the deepest rural communities worldwide [17]. Internet companies have also tried to address the digital divide with initiatives such as FreeBasics [19] from Facebook or the Google Global Cache. Sharing resources, such as local access infrastructure or global Internet transport, is encouraged at all levels [11,7] to lower the cost of network infrastructures and Internet services.

Among many other community networks, guifi.net exemplifies how communities can develop their own network infrastructures as a commons [2], using several wired and wireless links to create a regional IP network, and sharing several Internet gateways among all their participants. A costly option is buying transit from global carriers, and the cheapest is using Internet gateways. These gateways are usually web proxies for web access, the most popular traffic, but can accommodate other traffic through HTTP CONNECT, SOCKS or tunneling. Proxies, not exempt from the drawbacks of middleboxes, have also additional advantages: some content and DNS resolution can be shared in caches, and most important, proxies can protect the privacy of end users if they trust the proxy provider. For instance guifi.net, probably the largest community network in the world has more than 32,000 network nodes. 12,500 registered users can use any of the 356 web proxies (May 2016) offered inside that network. The links of the regional network connecting clients and proxies are crowdsourced: contributed and managed by the participants. Therefore, paths may be less predictable than networks from commercial ISPs and more susceptible to high load and congestion. Access to the Internet through web proxy gateways relies on individuals or organisations sharing the full or spare capacity of its Internet connection with other guifi.net users. However, these crowdsourced gateway nodes have limited processing and Internet transfer capacity and may be overloaded by the demand.

In this paper we provide an analysis of a large and crowdsourced proxy service in a regional community network. A large population of C clients can browse the web taking advantage of the aggregated capacity of a pool of P contributed web proxies, with $C \gg P$, over a regional network infrastructure, at a fraction of the cost of C Internet connections.

The contributions are twofold. First, we describe and provide detailed datasets collected from the guifi.net network and its proxy service, Sections 2 and 3. Second, we provide an analysis of the service from three viewpoints: 1) the users, Section 4, in terms of patterns of usage of the service, type of content and cache performance, 2) the proxy, Section 5, in terms of performance and cyclic daily variance, and 3) the local network, Section 6, in terms of topology and usage. Our measurements describe the effectiveness of a simple setup of a regional network sharing a set of Web proxies in delivering free basic Web access to a large population despite some hotspots.

2 The guifi.net Proxy Service

Guifi.net is an open, free, and neutral network built by its members: citizens and organisations pooling their resources to build and operate a local network infrastructure, governed as a common pool resource [2]. The network infrastructure is mostly wireless [20] with a fiber backbone. Participants accept a community license, can extend the network to reach new locations and use the network to reach intranet services. The web proxy service is one of them.

The most popular application in community networks is web access and guifi.net is no exception. Web proxy nodes connected both to guifi.net and an ISP act as free gateways to the Internet to the community network users. Proxies run on simple servers and take advantage of individuals or organizations (like libraries or municipalities) offering their Internet access to other guifi.net users. Using web proxies and controlling their maximum network throughput, public entities can provide free Internet access without infringing telecom market competence regulations.While some of the web proxies in the network are kept as a private service, 356 out of the 477 registered web proxy servers in the network (May 2016) are shared with all the network registered participants (12,500). A registered member is allowed to choose any proxy of their convenience and use it. Users can select or change its choice based on quality of experience. Therefore, while some proxies may become popular and degrade, others may remain underused.

3 Data Collection

For our analysis we choose to study the guifi.net zone Llucanes, a region in the Osona county of Catalunya, Spain. As explained in [6], this zone is representative of other rural guifi.net networks.

Furthermore, Llucanes is the only guifi.net zone with published anonymized logs for all involved operational proxies. Even-day proxy log entries hide the client IP address and show information about the requested URLs, while odd-day proxy logs show the opposite. We assisted in the preparation and publication of these logs. The logs combined with topology and network traffic information, provide a consistent and complete view of the regional network.

Link network description: expressed in CNML, an XML schema to describe the registered nodes and links in the network with snapshots published by guifi.net every 30 minutes. It has been used to build the topology graphs.

Daily Proxy logs: we used anonymized logs of one month of duration for the four proxies operating in the Llucanes zone provided by guifi.net.With their consent we have made these logs public[1]. The logs follow the native Squid log format registering information per request concerning the amount of bytes delivered to the client, time elapsed, users, URL, HTTP status codes and cache status codes, but not the upload traffic.

Usage and availability of paths: guifi.net has a distributed hierarchical monitoring system, called SNPServices, to collect metrics about the in/out traffic, delay and availability on all registered nodes and link interfaces. In each zone there exist one or more nodes –called graph servers– that continuously collect information though SNMP requests, pings and traceroutes to nodes in that zone. In our case, the proxy nodes were co-located with these graph servers, therefore we were able to collect the number of hops and average delays between the proxies and every other router in the zone under analysis.

Proxy Internet bandwidth and other characteristics: We had access to the proxies to measure Internet bandwidth and collect hardware details such as CPU or RAM. The bandwidth estimation was performed using speedtest-cli [16].

4 Service Usage Viewpoint

The behaviour of the users and the service can be described at macro-level as a set of different time series concerning metrics that can be extracted from the monthly logs, namely bytes per request, number of requests and number of users. Figure 1 presents the traffic time series for the aggregate of the set of proxies showing a daily repetitive pattern, but also strong aperiodic negative spikes.

[1] `http://dsg.ac.upc.edu/anon_guifi_proxy_logs`

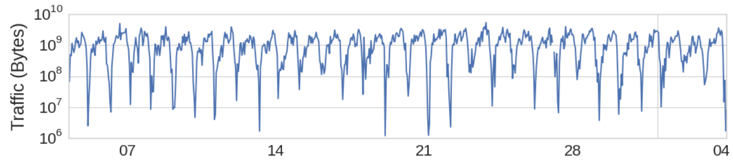

Fig. 1: Web proxy time series (days in April 2016)

We first investigate the stationarity and seasonality of time series. We present results of this analysis only for the bytes time series but other metrics behave almost identically. First we calculated the most important frequencies of the bytes time series using a periodogram. The Lomb-Scargle Periodogram of the bytes time series shows the highest peak at frequency 0.0417 which corresponds to a dominant period of 1 day. The second largest peak is at frequency 0.0833, which corresponds to 12 hours. The results are almost identical for users and requests time series. The dominant period is used to calculate the standard seasonal decomposition by moving averages based on Loess. The seasonal component presents a clear 24 hour period validating our initial observation of a daily pattern. As far as the residuals are concerned, we verified there is not any other hidden information.

Furthermore, we tested the stationarity of the time series using the Augmented Dickey-Fuller unit root test and we find with great confidence that the bytes time series are stationary which practically means that we can analyze and compare periods of the series using representative values such as mean and variance. It is worth mentioning here that the time elapsed per request time series, not mentioned earlier since it depends on the bytes and would not add new information, shows similar behaviour and can also be compared.

Service Usage: Now we look at service usage from distributions of the metrics per request and per user.

Figure 2 shows the distribution of bytes transferred per request. The majority of the traffic is due to a relatively small number of large requests. For example, 80% of the requests have size smaller than 10KB, but represent only the 3% of the traffic. Between 20% and 80% of the requests size varies steadily from 1-10KB implying an average close to 5KB. Indeed, the calculated mean is 6.8KB.

While the distributions per request show small variations around an average number, the distributions per user are far more disperse. For instance, the number of requests per user varies importantly between 1 and hundreds of thousands with a heavy weight at the top: 25% of the users requested 90% of bytes. This behaviour translates into the fact that few users generate the majority of requests. A similar behaviour is noticeable in the bytes per user distribution in figure 3, where 15% of users create 90% of traffic in bytes.

The time elapsed per request is a metric that depends on the bytes transferred. The log values measure the time from having received a request until the proxy finishes sending the last byte of the response. Thus, the time elapsed can be a metric for the proxy throughput in terms of processing, but it is unclear if it can be an indicator for the proxy network throughput, since this would imply the assumption that each request needs zero processing time and the proxy never saturates. The times vary from hundreds of nanoseconds to tens of milliseconds

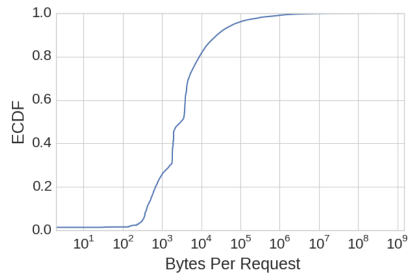

Fig. 2: Distribution of the request size

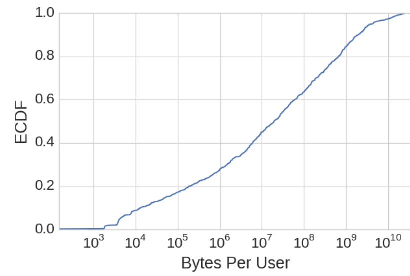

Fig. 3: Distribution of the user traffic

but for approximately 95% percent of the requests the time spend to process one byte is between 0.1ms to 1ms with a calculated average of 0.7ms. The most probable reason for the values higher than the mean is that they occur when the proxy is saturated, either in terms of computing resources or of network resources. The reverse relation, bytes per time elapsed per request, represent the **process throughput** of the server.

Additionally, we present an indicator of the network throughput achieved by the proxies. We calculated that 98% of the requests have elapsed time less than 100 seconds. Therefore, if we use aggregation time spans with granularity much larger than 100 seconds, we can safely assume that the vast majority of the requests will be completed inside this time span. As a result, using the aggregated data per hour we can assume that all the bytes present in the included requests have been requested by the remote servers inside this hour. This assumption allows us to estimate the **average network throughput** during this hour, which would be the total bytes of the hour divided by 1 hour.

Figure 4 shows the different throughput metrics presented above. We notice that the distributions of both throughput metrics are not very different in the overall behaviour, though the average network throughput rises more slowly showing more variance. Moreover, the network throughput is almost orders of magnitude greater than the process throughput which is expected, since downloading the net object is just a part of processing a request, and the logged elapsed time includes sending the entire requested object to the client.

In summary, we have observed that few users generate the majority of traffic and requests. Similarly, few requests create the majority of traffic. Additionally, we have introduced the metrics of request process throughput and network throughput as they apply, which can be used for any load processing service that has similar logging datasets.

Content analysis and cache effectiveness. Even-day proxy logs show the target URL of users requests. The analysis considers the requested domains, HTTP content types, HTTP methods and cache statuses. Based on these metrics we present what type of content users are accessing, we discuss the limitations of the analysis resulting from the SSL traffic and we evaluate the cache service provided by the Squid-proxy.

The type of content of interest to the users can impact the behaviour of the service, since different types are sensitive to different parameters. Despite the fact

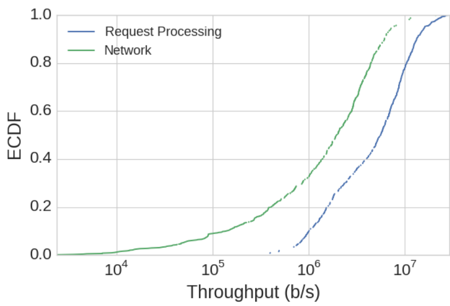

Fig. 4: Request processing and network rates

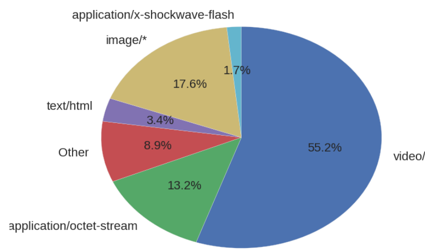

Fig. 5: Distribution of HTTP traffic types

that the most common request type is 'text/html' (26%), the majority of the traffic (61%) is spent on the HTTP CONNECT method. This is used to tunnel TCP based protocols through web proxies. Tunneling has various applications, but mainly as described in RFC2817 [14], it supports SSL-encrypted communication (HTTPS) through an unencrypted web proxy. We could verify that the domain names involved had HTTPS service (port 443). As a result, we cannot go in further detail for this 60% of the traffic. For the remaining 40%, most of the traffic, merely 20%, is video (MP4,FLV,MPEG,Flash), which is followed by images (PNG,JPEG) which take around 6% of the traffic. Regarding the percentage of number of requests, HTML pages are the most requested, which is expected since users start requesting an HTML page, which in turn generates HTTP requests for the additional contents. HTML pages are followed in the ranking by images, that constitute more than 24% of the number of requests. The distribution of web traffic per URL in Figure 6 can roughly approximate a Zipf distribution, equivalent to results [15] with domestic Internet connections.

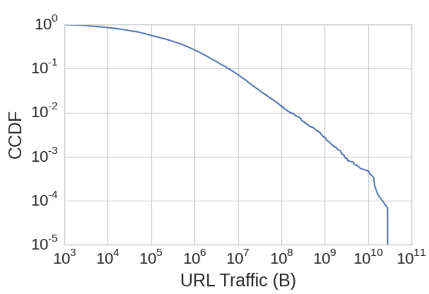

Fig. 6: Distribution of traffic per URL

Domain	Traffic Fraction
Video Portal	36.25%
CDN	6.59%
Video Portal	3.96%
Video Portal	3.39%
Video Portal	3.20%
Software Company	3.17%
Software Updates	2.56%
Search Engine	2.39%
CDN	1.57%
Web Portal	1.38%

Table 1: Top Domains by traffic

In Table 1 we observe that the top video portal traffic occupies 36% of the traffic, which is an unexpectedly large amount. For completeness, we mention that this is not reflected in the number of requests, therefore it is attributed on the size of the objects requested. Since video is by far the HTTP type with most traffic, it is not surprising to find that 4 out of 10 top domains are video portals.

Concerning cache performance the logs reflect that the vast majority of requests result to a cache MISS (approximately 97% of traffic and 66% of requests). Only 1.7% of the requests - corresponding to 1% of the traffic - results to a cache HIT. The bad cache performance is expected as the majority of the traffic is dynamic media content, a content type with poor caching support from providers. The default allocated cache size in guifi.net proxy settings, if available, is 10GB (leading to less than 5% HIT rate according to [22]), while in some guifi.net proxies caching is not enabled.

5 The Proxy Viewpoint

In this section we investigate the capabilities and influence of the proxies involved in the studied service. We demonstrate the behaviour of the different representative metrics presented and investigate daily characteristics of the service.

Performance Comparison: Our dataset concerns the only 4 proxies operating in the Llucanes zone. Table 2 shows the CPU and RAM characteristics of the proxy servers, as well as the nominal maximum throughput of the Internet connection they offer. The characteristics of the hosts are very diverse, with greater deviation in terms of the Internet throughput ranging from 4Mbps to 80Mbps. We also observe that proxy 11252 has the slowest combined characteristics. It is also worth noting that the servers with proxies provide other services, like the SNMP service mentioned earlier. The interference caused by other services is expected to be negligible and beyond the scope of this work.

Id	CPU	RAM	Max Throughput
3982	Intel amd64 2-core 2.6GHz	2GB	80Mbps
10473	Intel x86 2-core 2.6GHz	0.5GB	6Mbps
11252	AMD Athlon(tm) XP 1700+	0.5GB	4Mbps
18202	Intel amd64 2-core 2.7	2GB	8MBps

Table 2: Description of Proxies

Before evaluating the performance of the different proxies, it is necessary to explain how users select their proxies. Web proxy services are configured in client web browsers based on information provided by other trusted community network users, obtained from the public list of proxies, or based on previous user experience. As all federated proxies use the same authentication service, users are free to choose whatever proxy they prefer. The choice of proxy is rather fixed and prioritized, only switching to lower priority proxies when the top choice fails. Nevertheless, this manual choice is mostly based on social factors. Social bonds and service reputation are key factors in the selection.

Figure 7 presents the distribution of the average number of users per hour. The different proxies show similar distributions, though we observe that proxy 10473 has a differentiated demand, with 40% of time without any user and a maximum of 10 users per hour. For the rest of proxies, the majority of time (60%) have an almost linear distribution between 5 and 25 users, with near equally distributed values, and an average of around 17 users per hour for proxies 11252 and 18202, and an average of 12 users for proxy 3982. The difference in distribution among proxies comes as a result of manual selection.

The user's distribution among proxies has a clear impact in the distribution of the number of requests in figure 8. The ordering of proxies with respect to the number of users remains visible in the distribution of requests. Also, there is close-linear behaviour between 20% and 60% for all proxies except 10473. For proxies 11272 and 18202 the number of requests per hour is typically between 1K and 10K requests, with a mean of 8187 and 6716 respectively. In proxy 3982 typical values are between 500 and 1K requests per hour.

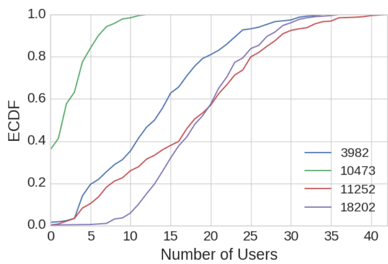

Fig. 7: Hourly average number of users per Proxy

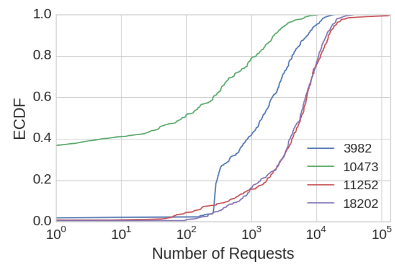

Fig. 8: Hourly average number of requests per Proxy

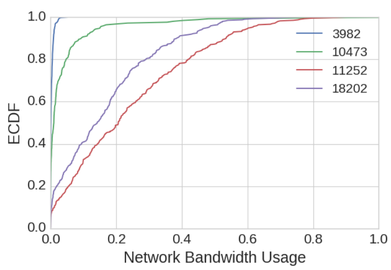

Fig. 9: Network usage per Proxy

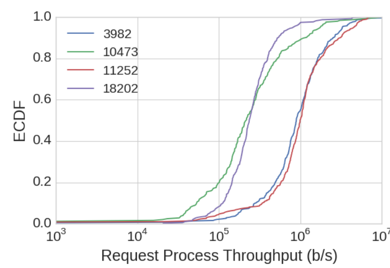

Fig. 10: Hourly average processing throughput per Proxy

The network performance of each proxy can be more accurately evaluated in terms of its network capacity in order to estimate the network usage. Figure 9 normalizes the network throughput of each with the maximum bandwidth available provided in Table 2. All proxies show low utilization of their network resources, being approximately less than 0.3 (30%) for all the proxies for 80% of the time. Nevertheless, proxies 11252 and 18202 have higher utilization.

Figure 10 shows the distribution of the request processing throughput, as defined in Table 2. We observe that all proxies have almost identical distribution but around different mean values, depending on the individual characteristics of the proxy. Moreover, we can see that a significant percentage (>60%) of the time proxies serve at a very narrow range of throughput, meaning they can offer a stable service. Even in the worst cases, the service does not suffer from extreme degradation, while remaining higher than 100Kbps 80% of the time. We also observe that for proxies 3982 and 11252, the processing throughput distribution resembles the number of requests distribution possibly indicating, as before, that the proxies are not saturated.

To gain a more complete perspective we also studied the daily aggregates of the traffic, users and requests clearly observing not only expected human daily pattern but also a clear view of the different way each proxy receives and serves request as a result of the users' manual proxy selection. Moreover, studying the mean daily patterns, we noticed that, as seen in figure 11, that the process throughput presents very small variations implying a stable service behaviour. On the other hand, the traffic size varies more than 1.5 order of magnitude. The fact that the process throughput is not affected by the traffic size confirms our observation that the servers are not saturated.

From all the above we can conclude that the proxies are able to offer a stable service, with respect to the traffic load, allowing them to be used as an alternative domestic Internet connection. Moreover, in our concrete scenario, the network resources of the proxies resources are underutilized assuming that no external clients are using the service (as stated in 3) and that no other services co-located in the host of the proxy are heavily using the network capacity.

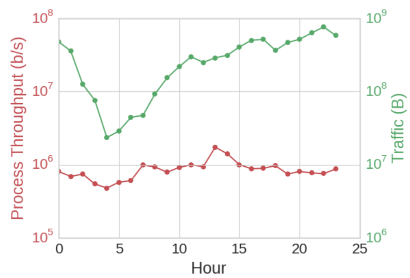

Fig. 11: Daily average request processing throughput compared to traffic

graph	nodes	edges	degree max/mean /min	diameter
base-graph	902	914	98/2.04/1	11
proxy-clients-graph	463	472	60/2.04/1	10
backbone graph	47	56	10/2.38/1	9

Table 3: Summary of base-graph

Even in that small scale, we observe the daily cycle of human activity which shows preference for evenings and really reduced traffic during the first hours of the day. The pattern is visible in all the described metrics but in different degrees. The network throughput metric is heavily affected by the number of users, while the request processing throughput and number of requests are affect more mildly. This is an expected result since the processing throughput represents more the ability of the proxy to process the load than the load itself.

6 The Local Network Viewpoint

Here we present an overview of the local network infrastructure, as well as an estimation on how it affects the final user experience. For the analysis we used information extracted from odd day logs, as explained in Section 3.

Network overview: For the local network we have considered all operational nodes and links of the Llucanes guifi.net zone. Similarly to the rest of Osona county zones, and in general to many rural community network deployments, the network consists of a small set of interconnected routers, the backbone graph, where each router is connected with a large number of end nodes, most of all wireless links. The end nodes are the points from which users access in the entire guifi.net network. Some of the routers act also as hosts for various guifi.net services, including the proxy service. We refer by Proxy-Clients graph to the part of the Llucanes network including only the nodes (clients, routers, proxies) that participate in the proxy service. Table 3 describes the main characteristics of the aforementioned graphs. We notice that the mean degree of the base-graph and of the proxy-clients-graph is very low since the end-nodes with degree 1 dominate the distribution of degrees. The low mean degree value in the backbone-graph is more interesting though, since it implies that the majority of the routers have only two neighbours.

Network usage: Since selection among proxies is static (manual), the analysis of network usage can show how the selection reflects on network usage and the perceived user experience. Towards that end, we first analyze metrics of distance between the users and the servers. Figure 12 shows the distribution of the number of hops between the users and the selected proxies. The distribution is almost uniform for 95% of the users with values between 1 and 6 hops. The remaining 5% is split between 7 and 8 hops. Nevertheless, we observe that the manual choices result to increased number of hops, therefore possibly introducing unnecessary overheads. The latency involved, depicted in figure 13, shows a different behaviour. Almost 80% of the users experience an average latency smaller than 15ms to reach their proxy. The remaining 20% lies between 20ms to 35ms. Despite the almost uniform distribution of hops, latency values are vary much less, implying that during normal network conditions, the distance between the users and proxies is not significantly deteriorating the user experience for web services not requiring very low latency.

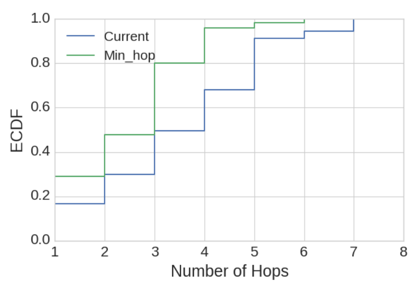

Fig. 12: Number of network hops between users and their selected proxies

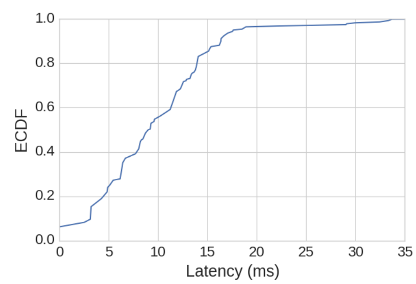

Fig. 13: Average latency between users and their selected proxies

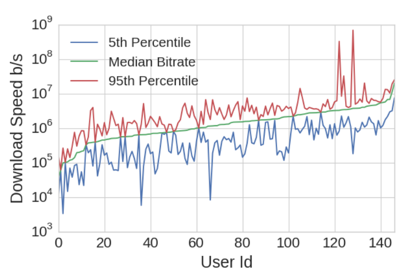

Fig. 14: Estimation of user experienced throughput with objects larger than 1MB

As we described earlier, the request processing throughput is calculated based on the request time elapsed time, which includes the time the proxy requires until he sends the last byte of the web object to the client. Therefore, any significant local network deterioration affects the throughput behaviour. Based on this observation we can utilize the request processing throughput metric for objects larger than 1MB, in order to estimate significant deterioration on the user experience. Including smaller objects would give unreliable throughput results due to the major influence of proxy buffering, DNS caching and network latency in short connections. Figure 14 illustrates the individual user experience in throughput. We estimated from proxy logs the download speed for objects larger than 1 MB. A bold assumption is that users focus on few or a single large object at a time. If so our measures could be taken as a lower bound for individual throughput experience. Median values of download appear quite stable with speeds ranging from 0.1Mb/s to 10Mb/s for different users. Quite good result for many users of a free crowdsourced service.

7 Related Work

Most work on wireless networks focuses on usage traffic patterns, link level characteristics and topologies, but not user experience, e.g. MadMesh [4], Google WiFi [1] and Meraki [3] networks. In all these studies, Internet access is direct instead of using proxies, and these wireless networks are homogeneous. Thus, measurements and results here cannot easily be compared with those.

The web proxy business has significantly changed over the recent years. The percentage of cacheable content has been decreasing. At the same time, web caching remains relevant, as the data volumes in the networks continue to grow and network operators are seeking ways to cope with this demand. The performance of web proxies is not only about high-level metrics such as hit rates. Low-level details such as HTTP cookies, aborted connections, and persistent connections between clients and proxies as well as between proxies and servers have a strong impact on performance, particularly in heterogeneous bandwidth environments [8]. Recent studies, e.g. [18,5,19], show that the use of web proxies in wireless networks and mobile networks is still an open research question.

Wireless network user experience has been characterized in previous studies. The first [13] focuses on web traffic and the use of proxies to access Internet content in rural areas. We have also analyzed traffic using proxies inside the guifi.net network, but the lessons learned are more complementary than comparable as scenarios are too different. The second study [10] looks at web traffic patterns and content caching, relevant in an scenario with a single and very limited Internet link, but guifi.net has more than 350 proxies, with excess overall capacity and quite different usage profiles. The last study [9] focuses on network performance of gateways on wireless-based home networks.

8 Conclusions

We analyze the overall performance of the crowdsourced proxy service in guifi.net, a mostly wireless-based community network. We separately study the different local components: service usage, the proxies and the edge network.

Our analysis indicates that the user behaviour expressed as traffic time series has a clear daily pattern. The content, in contrast, does not follow any such repetitive pattern, making the cache service irrelevant. We also find important evidence that the traffic load is not distributed efficiently across the links of the network, nor across the participating proxies. Finally, we argument that the static manual selection of proxies by the users plays an important role is the main cause creating the above issues. As a result, this crowdsourced service is comparable to a more expensive traditional basic domestic Internet access, and it fulfills the goal of an usable and satisfactory free basic Internet service for all.

References

1. M. Afanasyev, T. Chen, G. Voelker, and A. Snoeren. Usage patterns in an urban wifi network. *IEEE/ACM Transactions on Networking*, 18(5):1359–1372, 2010.
2. R. Baig, R. Roca, F. Freitag, and L. Navarro. guifi. net, a crowdsourced network infrastructure held in common. *Computer Networks*, 90:150–165, 2015.
3. S. Biswas et al. Large-scale measurements of wireless network behavior. *ACM SIGCOMM Computer Communication Review*, 45(4):153–165, 2015.
4. V. Brik et al. A measurement study of a commercial-grade urban wifi mesh. In *Internet measurement conference*, (IMC), pages 111–124, 2008.
5. D. Catrein et al. An analysis of web caching in current mobile broadband scenarios. In *New Technologies, Mobility and Security*, (NTMS), pages 1–5, 2011.
6. L. Cerdà. On the topology characterization of guifi. net. In *Wireless and Mobile Computing, Networking and Communications*, (WiMob), pages 389–396, 2012.
7. European Parliament and Council. Directive 2014/61/EU on measures to reduce the cost of deploying high-speed electronic communications networks., May 2014.
8. A. Feldmann et al. Performance of web proxy caching in heterogeneous bandwidth environments. In *INFOCOM*, pages 107–116, 1999.
9. S. Hätönen et al. An experimental study of home gateway characteristics. In *Internet measurement conference*, (IMC), pages 260–266, 2010.
10. S. Ihm and V. S. Pai. Towards understanding modern web traffic. In *Internet measurement conference*, (IMC), pages 295–312, 2011.
11. International Telecommunication Union. Trends in telecommunication reform 2008: Six degrees of sharing (d-pref-ttr.10), July 2009.
12. Internet Society. Global internet report 2015, October 2015.
13. D. L. Johnson, V. Pejovic, E. M. Belding, and G. van Stam. Traffic characterization and internet usage in rural africa. In *World Wide web*, pages 493–502, 2011.
14. R. Khare and S. Lawrence. Upgrading to TLS Within HTTP. RFC 2817, 2013.
15. G. Maier et al. On dominant characteristics of residential broadband internet traffic. In *Internet Measurement Conference*, (IMC), pages 90–102, 2009.
16. M. Martz. speedtest-cli. https://github.com/sivel/speedtest-cli, 2015.
17. C. Rey-Moreno et al. Experiences, challenges and lessons from rolling out a rural wifi mesh network. In *ACM Computing for Dev. (ACM-DEV)*, page 11, 2013.
18. T. Sangwongthong and P. Siripongwutikorn. Proxy caching in wireless mesh networks. In *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, (ECTI-CON), pages 1–4, 2012.
19. R. Sen et al. On the free bridge across the digital divide: Assessing the quality of facebook's free basics service. In *Internet measurement conference*, (IMC), pages 1–7, 2016.
20. D. Vega, R. Baig, L. Cerdà, E. Medina, R. Meseguer, and L. Navarro. A technological overview of the guifi.net comm. n. *Computer Networks*, 93:260–278, 2015.
21. G. WG. Global access to the internet for all research group, 2016. [Online; accessed 14-September-2016].
22. M. Zink et al. Characteristics of youtube network traffic at a campus network. *Computer Networks*, 53(4):501 – 514, 2009.