Design Trade-offs of Crowdsourced Web Access in Community Networks

Emmanouil Dimogerontakis*, Roc Meseguer*, Leandro Navarro*, Sergio Ochoa† and Luís Veiga‡

* Universitat Politècnica de Catalunya, Barcelona, Spain
{edimoger,jlneto,meseguer,leandro}@ac.upc.edu
† Universidad de Chile, Santiago, Chile
sochoa@dcc.uchile.cl
‡Tecnico Lisboa/INESC-ID Lisboa, Lisboa, Portugal
luis.veiga@inesc-id.pt

Abstract—Internet access has become a requirement to participate in society; however, the majority of the world's population is not yet online. Citizens can self-organize cooperatively to crowdsource community network infrastructures and achieve Internet access. In order to help address that challenge, this paper provides an analysis of a crowdsourced Internet access mechanism: the distributed Web proxy service in one of the largest community networks in the world. Several perspectives were considered in this analysis, e.g., data traffic, networking issues, and proxies responsiveness. The evaluation results show how the current manual proxy choice, based on social clues, becomes a popular service plagued with hot spots and inefficiencies, which opens several opportunities for improving these infrastructures. By taking advantage of it, our research shows that the trade-offs between informed proxy selection and admission control in proxies, could alleviate imbalances and uncertainty, and also improve the service with little additional burden. This represents an explicit and direct mechanism for improving the service provided by these community networks, and a clear benefit for its members.

I. INTRODUCTION

Internet access has become a requirement to participate in society; for instance, to access public services, education material, social media and also to support everyday work of millions of organizations. However, the majority of the world's population is not online [1] yet, far from the vision of "universal service". This situation aggravates the digital divide between several communities/regions/countries, and the rest of the world. The cost and availability of this service seem to be the main limitation for becoming Internet access a generally available service in our society. Network infrastructures that provide these services are, in most cases, under control of former monopolies, now telecom incumbents. These entities control the offer and have the strength to influence regulation and discourage competitors. Except in developed urban areas, the typical situation is a lack of competitive offers, defined as "market failure" [2]. This negatively affects people by reducing their capability of improving their digital literacy and increasing the digital divide with them. Rural and poor communities, and also age brackets (like the older adults) are particularly vulnerable to this situation.

As a way to mitigate this challenge, in many regions worldwide the citizens self-organize to explore alternative models for getting Internet access under reasonable conditions. An example of it are the community networks (CN), that are crowdsourced data network infrastructures built by citizens and organisations, who pool their resources and coordinate their efforts [3] to provide an Internet community access service to their members. These communities are *open*, *free*, and *neutral*. They are open since everyone has the right to know how they are built. They are free because the network access is driven by the non-discriminatory principle; thus, they are universal. Moreover, they are neutral in terms of technology solutions to extend the network, and neutral for supporting data transfers.

The community networks are quite new, and they represent an alternative paradigm for developing network infrastructures and services in a broad sense. Communities can propose locally adapted self-organized cooperative schemes for developing self-provided data networking solutions, sharing wireless links and spectrum, optical fibre, and Internet gateways; and even sharing Internet connectivity with other members of the community. When these fundamental principles are applied to an infrastructure, they often result in networks that are *collective goods*, *socially produced*, and governed as *common-pool resources* (CPR). Natural CPR, also called commons (such as, communal pastures, fisheries, forests), were studied in depth by E. Ostrom [4]. According to that we use the term *network infrastructure commons* [2]. These infrastructures developed cooperatively become regional IP networks that enable inexpensive interaction and access to local digital content and services. In addition, there exists the issue of access to the global Internet, that can be reached through Internet Service Providers (ISP) in these regional network infrastructures.

There are many examples of community networks that can fit in this scheme. In [5] we outline 18 cases, with 9 described in detail, and 267 potential cases in 41 countries. There are also several studies that consider structural [6], [7], [8], technological [9], [10], [11] and organisational [3], [12], [5] points of view of these networks.

In this paper we focus our study on the guifi.net community network, one of the largest worldwide. Particularly we analyze crowdsourced provision of Internet access using a pool of shared Web/Internet proxies, distributed over nodes in a regional network. This is an inclusive and cost effective model to provide limited Internet (Web) access, complementary to commercial offerings. However, crowdsourcing requires motivation and incentives for the participants, regulation of contributions and consumption to achieve operational and sustainable outcomes, such as in [13].

The contribution of this paper is the analysis of the real-world distributed Internet access service, implemented as a set of Web proxy servers in the guifi.net community network. This analysis allows us to identify potential weaknesses that limit the service provision, its quality and the way in which the community evolves. This study considers several data inputs; e.g., the patterns of usage from service logs, the design choices and implications (considering client and proxy choices) according to patterns of usage, and the relative location of users and proxies in the network topology. The results show the key metrics, the design space for cooperative choices, the involved trade-offs, and the effects on the service cost and performance.

Next, we first introduce the proxy service in the guifi.net CN. Section III looks at the behavior and clustering of users according to content and network locality, and it also analyzes the impact on the criteria for proxy selection. We present an analysis of the current scenario, limitations and potential for improvement from the perspective of the access network in Section IV, proxies in Section V and users in Section VI. Section VII presents the conclusions and the future work.

II. The Web through a proxy service

Global access to Internet for everybody requires not only to increase the service availability, but also a dramatic reduction of its cost, especially in geographies and populations with low penetration [14]. This cost reduction can be achieved by sharing; e.g., a large population of C clients can browse the web taking advantage of the aggregated capacity of a pool of P web proxies, with $C \gg P$, over a regional network infrastructure, at a fraction of the cost of C Internet connections.

The providers of these Internet connections can be quite diverse, such as commercial ISP, cooperatives or associations of users sharing costs [15], content or service providers promoting their offer [16], citizens sharing their unused Internet access capacity with neighbors and friends [17], public or private organizations sponsoring Internet access for complimentary interests. However, telecom regulation authorities in many countries limit publicly subsidized Internet access to preserve market competition. Internet access through web proxies is clearly a limited service compared to an IP tunnel, as the service is usually restricted to a set of protocols/ports; however, it can help enhance privacy as the origin IP addresses may be hidden. The most popular application in community networks is Web access. Many citizens, private and public organizations involved in community networks, such as freifunk.net or guifi.net, have chosen to provide that service within their community network. Using Web proxies through local networking infrastructures (e.g., community networks) that provide local or regional connectivity, the citizens can reach Internet content and services at no additional cost.

To understand the impact of these proxies in the behavior of these networks, we focus our study on the distributed web proxy service of the guifi.net community network; a free, neutral and open access CN with more than 32,000 nodes mostly in Spain [9]. The service has more than 300 Web proxy servers, however this study is based on measurements of 30 days on a network sub-zone that consists of 4 proxies shared among more than 500 users.

Without access to one of these proxies or a guifi.net connected ISP, community members can still share contents and access applications within the same community network, but not to external resources. In order to get Web access, the clients manually specify a list of proxies, the main proxy and the secondary ones. Access control to the proxy is performed through federated authentication credentials. In case a proxy does not respond (timeout) or rejects the connection, the client automatically switches to the next proxy in the list. The choice of proxies is manual and the list usually comes from acquaintances in the community or personal experience.

By considering the available information (anonymized proxy log files, and topology and link datasets of the guifi.net community[1]), the next section presents the methodology used to identify patterns of usage of the Web proxy service, which could be leveraged to improve such a service.

III. Clustering of Users

We started the study exploring data concerning the service usage, in order to group users according to their behavior. Then, we identified the graph communities that exist in the network to analyze the factor of network locality. Thus, we tried to understand the trade-offs of grouping users according to similarities in their behavior and/or according to their location in the network.

A. Clustering according to usage

For the analysis of service usage according to patterns of data traffic, and based on [18], we considered three different types of clustering algorithms: K-means, suitable for generic applications, DBSCAN and Ward's hierarchical clustering (HC) that can trace complex patterns. The input used by the algorithms was the total data transferred per user in bytes, as well as the corresponding amount of traffic for contents that constitute a large amount of the total service traffic, like video (20%), image (6%) and HTML (2%). We experimented with various cluster sizes for K-means and Ward's HC, including well-known empirical estimation methods like the 'elbow method', as well as many parameters for DBSCAN. Table I presents the optimal results for each method in terms of cluster validation. For the validation we used the coefficient Shilouette score that has values in the [-1,1] range. As described in Table I, for all the cases there is a big cluster of 450-480 users with a Shilouette score of 0.9, indicating a very strong cluster density. Nonetheless, the rest of the users belong to overlapping clusters, with scores close to 0. After manually reviewing other results of the algorithms for getting a better insight, since it is the standard process in these cases, we chose Ward's HC method with 3 clusters, that partitions the users in one large consistent cluster and two small overlapping clusters, minimizing thus overlapping elements.

Table II presents the characteristics of the clusters, as formed using Ward's HC for 3 clusters. We find two consistent clusters of users with distinct properties. The Figure 1 depicts the comparison of the clusters in terms of traffic and size (number of users). Cluster 1 of *light* users, includes the majority of users and their profile consists of generating very low traffic, as low as 1% of the maximum noticed per user

---

[1]Datasets available at: http://dsg.ac.upc.edu/proxy-guifi

Method	Clusters #	Clusters Size	Clusters Shilouettes
DBSCAN	2	7, 499	0.03, 0.90
Ward's	2	33, 473	-0.04, 0.89
	3	4, 29, 473	0.30, 0.01, 0.87
K-Means	2	21, 485	0.09, 0.89
	3	10, 44, 452	0.07, -0.04, 0.88

ID	Size	Shilouette	Characteristics	Alias
1	473	0.87	Low total traffic	Light
2	29	0.01	Medium total and video/images traffic	Medium
3	4	0.30	High total and video/images traffic	Heavy

value, mostly HTML browsing. Cluster 3, *heavy* users, consists of only 4 users and it is characterized by high total traffic, where most of it is spent on downloading video and images. Cluster 2, *medium* users, presents an intermediate behavior; nevertheless, following the patterns of the *heavy* users. *Medium* users create a significant portion of the total traffic, around 20% of the maximum value, which they consume mostly on videos and images. This cluster has low consistency, with users presenting a behavior similar to cluster 3, but with traffic level close to cluster 1.
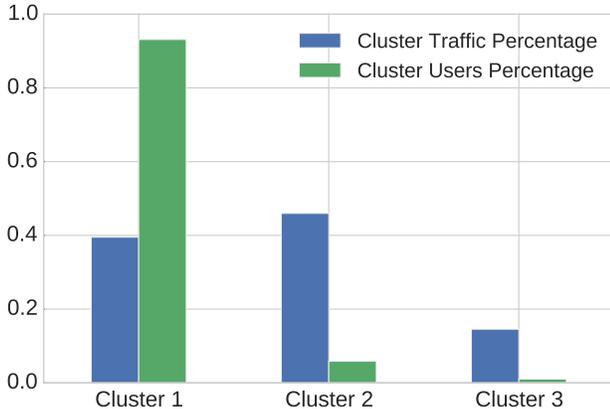

Figure 1.   Traffic and Users Percentage per Cluster

B. Clustering according to Network Locality

For the analysis of user groups according to network locality, we use graph community detection techniques. Based on [19], we choose three of the most prominent detection algorithms: Spinglass, Multilevel and Infomap. The data input for the algorithms is the backbone graph, consisting of 48 nodes. Moreover, since the studied guifi.net zone has a small well-connected backbone, with many clients connected to the routers of the backbone, we used the number of clients using those routers to establish the graph weight for the InfoMap algorithm. The weight for each link is defined as the average time to transfer a single byte according to our topology dataset. The results of the different algorithms can be seen in Table III. We compare the algorithms using the modularity score, which lies in the range [-1/2, 1), where the higher the value, the more consistent the community. Experimenting with the algorithms we noticed that the node size argument of the Infomap does not

affect significantly the output, thus Infomap does not offer any additional information. Therefore, we choose the Multi-level Algorithm that has the highest modularity score and smaller number of clusters, considering the small backbone.

	Infomap	Multilevel	Spinglass
Modularity	0.699	0.712	0.702
Clusters	12	9	15

Figure 2 shows the resulting graph for the Multi-level algorithm. The squares represent the routers that operate also as proxies. As depicted, the proxies are not well positioned relatively to the network clusters, considering that most of the clusters have no proxies, while one of the clusters has two proxies. Additionally, we observe that there are clusters poorly connected to their neighbouring clusters, resulting in an infrastructure far from ideal. For the rest of this work we assume that all the clients of a router belong to the cluster of that router.
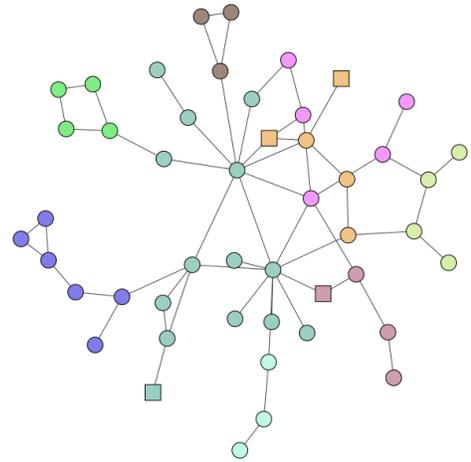

Figure 2.   Multi-level Community Detection for the backbone network (colors)

C. Influence of the criteria for proxy selection

According to our clustering analysis, we present simulations that exploit the two clustering techniques in algorithms for proxy selection, in order to provide alternatives to the current manual proxy selection. The objective is to demonstrate the impact of network locality and user traffic behavior on the performance of the proxy service and user experience. Thus, we show how they can be used to inform the design of an improved service.

Next, we present an initial evaluation of the mentioned techniques under the perspectives of the network, the proxies and the users. It is important to clarify that our algorithms implement one of several ways to use the information from user behavior clustering and community detection. The first algorithm we implemented, referred as $data\_cluster$, uses the clustering of user behavior to assign equivalent user load to each proxy by equally distributing the users of each cluster. In the cases where a new user has to be assigned to a proxy and all existing assignments from the clusters are equally balanced, the algorithm selects a proxy randomly. The second algorithm,
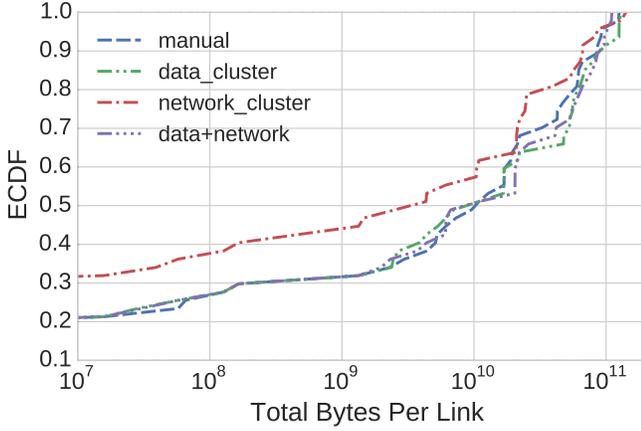
Figure 3. Comparison of Total Links Bytes Per Strategy ECDF

referred as $network\_cluster$, uses graph community detection to assign users to proxies according to the proximity of their community. For instance, a user with an available proxy in his community will be assigned to this proxy, while in the opposite case, it will be assigned to the proxy that is located in the closest community. In case of equal proximity, the proxy selection is random. Finally, we implemented an algorithm that combines both solutions in one of the possible ways. The algorithm $data + network$ is mainly based on the $data\_cluster$ algorithm, but in case it encounters equal assignments, it uses the $network\_cluster$ algorithm to decide. All these algorithms are compared with the $manual$ manual service selection.

IV. Network perspective

The impact on the network is studied according to the total bytes transferred through each link during the simulation. We do not take into account possible retransmissions, and we assume that the links cannot be saturated and have always the same performance, even across different links.

As shown in Figure 3, the $network\_cluster$ algorithm outperforms significantly the other algorithms in distributing the load in the links. It maintains the total traffic of 50% of the links, one order of magnitude lower than the other algorithms without compensating that by overloading a few links, as we would expect for the links that connect the clusters. The other algorithms present a similar, but shifted, distribution. Moreover, considering that each algorithm is using different number of links to send the traffic, it is worth mentioning that $network\_cluster$ transfers the lowest total amount of bytes, 1 Terabyte, while $data\_cluster$ is the most expensive transferring 1.7 Terabytes. We also find that $data + network$ lies between $network\_cluster$ and $data\_cluster$, with 1.4 Terabytes, while $manual$ transfers 1.3 Terabytes.

Overall, we observe that network locality plays a significant role in distributing the load on the network. Even in the case of existing communities without proxies, like the studied case, a locality-aware service can reduce its network impact.

V. Proxy perspective

From the perspective of the proxies, it is important for both the service performance and the user's experience to distribute the load according to the capacity and performance of each proxy.
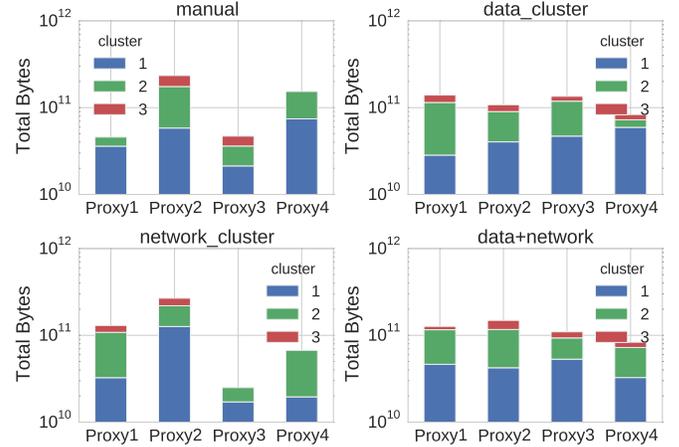

Figure 4. Comparison of Strategies: Traffic per Proxy and Clustering
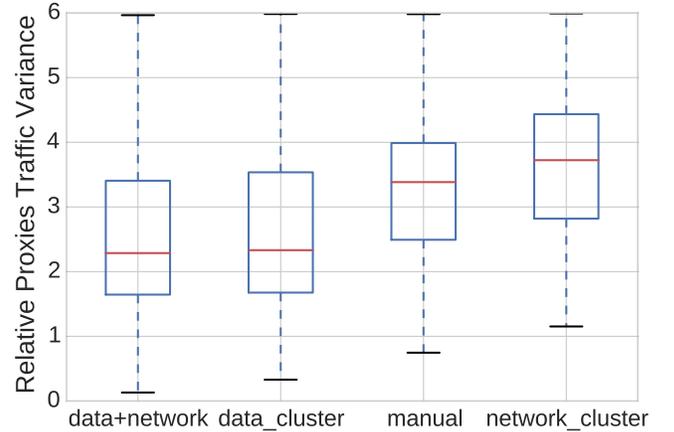

Figure 5. Proxies per second Relative Traffic Variance

In our simulations we start by assuming that all proxies have infinite capacity and the same processing performance (i.e., unlimited throughput). We evaluate the different algorithms by the total amount of bytes sent to each proxy per strategy, with information of the corresponding clusters, as seen in Figure 4. We initially observed that the *heavy* users occupy an important percentage of the traffic, even though they are nearly the 1% of the total users. Nonetheless, *light* users generate the majority of traffic despite the fact that each of them use the service comparatively much less. Therefore, as a result of manual selection the proxy load is very unbalanced, but the $data\_cluster$ and $data + network$ algorithms succeed in balancing the traffic. The $network\_cluster$ approach can result to an imbalance in the load among proxies, due to sub-networks with an uneven number of clients and the proxies inconveniently placed with respect to the clients. It is worth noting that, from the proxy perspective, the $data + network$ algorithm achieves its goal very successfully since it is mainly

based on the $data\_cluster$ algorithm; however, it also achieves a better performance than $data\_cluster$ from the network perspective. Moreover, as shown in Figure 5, the sum of distances of traffic values for each proxy to the mean at each instant is clearly smaller for the $data\_cluster$ or $data+network$. This small variability implies that these algorithms work well over both short and long term periods. We can therefore deduce that an algorithm that combines both, user clustering and network graph community detection, can be used for tuning the trade-off of impact of uneven proxy load and excessive network impact, due to long network paths. This lesson is applicable to server selection in a decentralized service.

If we take into consideration the limited capacity and throughput in proxies, then balancing the traffic across them according to the capacity of each proxy becomes a key issue. For example, in the case of a large number of users the clustering information could be used to perform admission control and therefore congestion control in the proxy.

In the current scenario proxies have a rough admission control based exhaustion of limits, and they do not on congestion control according to load or performance. Proxies take new requests based on a maximum number of concurrent clients, even when the proxy service is already under-performing for ongoing responses. This results in poor performance during peaks of large requests that cause congestion or a service timeout. In our decentralized scheme, clients have a list of several proxy choices. Clients make an initial choice, proxies can reject connections, and clients can just make a new local choice, transparently retry and continue from there, with no major visible effect to the user. The combination of clients using a list of proxy choices, proxy admission control, and network routing choices results in a simple, decentralized and cooperative regulation scheme that requires little coordination.

Admission control is important in large user populations, e.g., wide-area networks with many proxies, since proxies have a limited Internet access capacity. Any hotspot or imbalance in a massive system can easily lead to congestion, either in the access network, any proxy or the Internet access, resulting in a dramatic reduction of service throughput for many users of that proxy.

In addition to the local choices at each client and proxy, there is potential for global optimization in balancing global choices, across all proxies, by combining the user traffic behaviour, user proxy choices, and proxy capacities. Thus, we can help avoid globally imbalanced scenarios, where a proxy is saturated or providing low throughput, while at the same time another proxy is underutilized.

VI. USERS PERSPECTIVE

The evaluation of impact on service performance from the user perspective is the most complex, as users have different metrics to assess their service according to their diverse usage habits. While exploring these metrics is future work, here we present a first simple cost model to estimate how users perceive the impact of the presented algorithms. We assume that users try to minimize the transfer time in the local network, combined with the processing time in the proxy server.

As far as the network is concerned, we define as c_l the cost of the link l, in terms of time, to transfer one byte, assuming

that the links have infinite capacity, although we plan to study more sophisticated models in the future.

For each user we calculate the total cost of the network transfer as $\sum_{l=0}^{n} c_l * b_u, l \epsilon L_u$, where L_u is the set of links and b_u the total number of bytes attributed to user u.

The users' perception of the proxy performance is modeled similarly to the network performance. We define c_p as the cost of proxy p to process one byte, from the time it receives the request from the user, until it sends the last byte. We calculate the cost c_p of each proxy p separately for every strategy as $t / \sum b_u, u \epsilon U_p$, where t is the total measurement time, b_u the total number of bytes sent by user u and U_p the set of users of proxy p. Based on that, the proxy perceived cost for each user is: $c_p * b_u$.
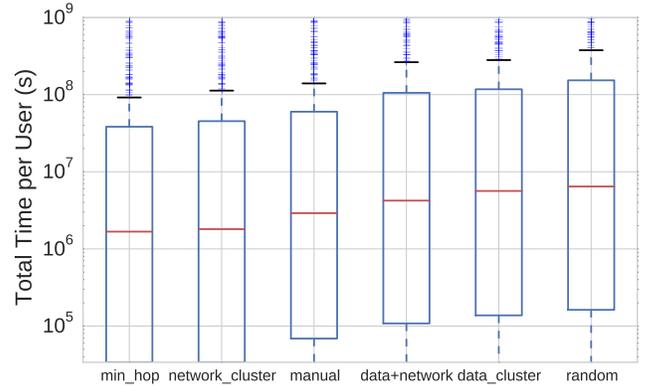

Figure 6. Cost per user ECDF

Considering that the costs are linear and independent, we can assume that the overall cost perceived by a user u is: $C_u = \sum_{l=0}^{n} c_l * b_u + c_p * t_u, l \epsilon L_u$. Hence, the objective of user u would be to minimize C_u. Figure 6 presents the distribution of the users' costs for each of the presented strategies.

While the distributions have very similar behavior, we can observe that for 80% of the users, the network community detection strategy performs slightly better than the current situation, and the rest of the strategies follow. The community strategy achieves equivalent results to a $min-hop$ strategy, only differing when proxies are not in the center of its zone. The $random$ strategy achieves equivalent results to cluster, as the latter only cares about contents and none about infrastructural aspects.

The network efficiency of community-based proxy selection, and therefore the impact of network locality, appears as an important factor. Studying the individual costs we observe that the network transfer time cost is, in average, significantly higher than the proxy processing cost. This fact explains why the community solution performs better overall, even though it is an inefficient option for load distribution in the proxies. The (clustering according to) user behavior appears to have an influence on the user perceived performance (cost), since it presents a differentiated behavior from the current situation (manual proxy selection). However, the simplicity of the model does not allow us to draw more conclusions.

In contrast, the current situation is that clients (Web browsers) have a list of proxy servers manually defined or

adjusted. The initial configuration is based on hints from other nearby users, or by downloading the list from a local guifi.net forum. The adjustments come from similar sources, personal usage experience, hints from other users or news about new proxies being offered. Web browsers switch to another proxy server just when a proxy fails to respond and do not provide load balancing, or more effective choices considering to degradation, congestion signals or relative performance. These models enables us to design a service selection algorithm that takes into account the characteristics of the users and the local network, confronting thus the inefficiencies caused in the service and the user experience by the manual static proxy selection.

VII. Conclusions and Future Work

The paper presents an analysis of how crowdsourced network infrastructures can provide very cost effective ways to access the Internet. We look at a distributed proxy service in guifi.net, one of the largest community networks in the world. The analysis of service logs shows patterns of usage and network topology grouping users and proxies, that can influence the criteria for proxy selection. The currently manual and not well-informed choice of proxies by clients work rather well for its users, but it result in inefficiencies that affect the service cost and shows episodes of degraded performance. Considering that situation, this paper explores alternatives for cost reduction and service improvement when going from a simple but rigid mapping between users and proxies, towards coordinated informed choices based on several metrics. Design trade-offs lie in considering infrastructural aspects (e.g., reduce network cost, avoid network and proxy congestion) and service aspects (e.g., good response time or QoE).

The combination of server alternatives in clients, finer grain proxy admission control, and the underlying network routing decisions result in a decentralized cooperative regulation scheme that can provide a crowdsourced proxy service, with good performance and requiring little coordination. Moreover, that scheme allows scaling up the network to larger sizes. As part of the future work, we plan to evaluate these strategies in detail with experiments under several real conditions, where clients perform more informed choices and proxies perform more fine grained admission control.

Acknowledgments

References

[1] Internet Society, "Global internet report 2015," Internet Society, Tech. Rep., October 2015. [Online]. Available: http://www.internetsociety.org/globalinternetreport/

[2] L. Navarro, F. Freitag, R. Baig, and R. Roca, *Community Connectivity: Building the Internet from Scratch.* Internet Governance Forum, 2016, ch. A commons-oriented framework for Community Networks, pp. 25–71.

[3] R. Baig, R. Roca, F. Freitag, and L. Navarro, "guifi. net, a crowdsourced network infrastructure held in common," *Computer Networks*, vol. 90, pp. 150–165, 2015.

[4] E. Ostrom, *Governing the commons: the evolution of institutions for collective action.* Cambridge University Press, Nov. 1990.

[5] netCommons, "Deliverable D1.2, Report on the Existing CNs and their Organization (v2)," Sep 2016, tech. Rep. D1.2. [Online]. Available: http://netcommons.eu/?q=content/report-existing-cns-and-their-organization-v2

[6] L. Cerda-Alabern, "On the topology characterization of guifi.net," in *Proc. Int. Conf. Wireless and Mobile Computing, Networking and Communications (WiMob)*, Barcelona, Spain, 2012, pp. 389–396.

[7] D. Vega, L. Cerda-Alabern, L. Navarro, and R. Meseguer, "Topology patterns of a community network: Guifi.net," in *Proc. Int. Conf. Wireless and Mobile Computing, Networking and Communications (WiMob)*, Barcelona, Spain, 2012, pp. 612–619.

[8] L. Maccari and R. Lo Cigno, "A week in the life of three large wireless community networks," *Ad Hoc Networks*, vol. 24, Part B, pp. 175–190, 2015.

[9] D. Vega, R. Baig, L. Cerdà-Alabern, E. Medina, R. Meseguer, and L. Navarro, "A technological overview of the guifi. net community network," *Computer Networks*, vol. 93, Part 2, pp. 260–278, 2015.

[10] L. Maccari, L. Baldesi, R. Lo Cigno, J. Forconi, and A. Caiazza, "Live Video Streaming for Community Networks, Experimenting with PeerStreamer on the Ninux Community," in *Proc. Workshop on Do-it-yourself Networking: An Interdisciplinary Approach (DIYNetworking)*, Florence, Italy, 2015, pp. 1–6.

[11] L. Baldesi, L. Maccari, and R. Lo Cigno, "Improving P2P streaming in Wireless Community Networks," *Computer Networks*, vol. 93, Part 2, pp. 389–403, 2015.

[12] R. Lo Cigno and L. Maccari, "Urban Wireless Community Networks: Challenges and Solutions for Smart City Communications," in *ACM Int. Workshop Wireless and Mobile Technologies for Smart Cities (WiMobCity), part of MobiHoc*, Philadelphia, PA, US, 2014, pp. 49–54.

[13] D. Vega, R. Meseguer, F. Freitag, and S. F. Ochoa, "Motivating the non-technical participation in technical communities," in *Proc. Computer Supported Cooperative Work in Design (CSCWD)*, 2015, pp. 259–264.

[14] G. WG, "Global access to the internet for all research group," 2016, [Online; accessed 14-September-2016]. [Online]. Available: https://irtf.org/gaia

[15] C. Rey-Moreno, Z. Roro, W. D. Tucker, M. J. Siya, N. J. Bidwell, and J. Simo-Reigadas, "Experiences, challenges and lessons from rolling out a rural wifi mesh network," in *ACM Computing for Development (ACM-DEV)*, Bangalore, India, 2013, p. 11.

[16] R. Sen, H. A. Pirzada, A. Phokeer, Z. A. Farooq, S. Sengupta, D. Choffnes, and K. P. Gummadi, "On the free bridge across the digital divide: Assessing the quality of facebook's free basics service," in *Proc. ACM on Internet Measurement Conference (IMC)*, Santa Monica, CA, USA, 2016, pp. 127–133.

[17] E. Dimogerontakis, J. Neto, R. Meseguer, L. Navarro, and L. Veiga, "Client-side routing-agnostic gateway selection for heterogeneous wireless mesh networks," in *IFIP/IEEE Int. Symp. Integrated Network Management, IM*, Lisboa, Portugal, 2017, pp. 1–8.

[18] P. Berkhin, *Grouping Multidimensional Data: Recent Advances in Clustering.* Springer, 2006, ch. A Survey of Clustering Data Mining Techniques, pp. 25–71.

[19] A. Lancichinetti and S. Fortunato, "Community detection algorithms: A comparative analysis," *Phys. Rev. E*, vol. 80, p. 056117, Nov 2009.