On Economic Heavy Hitters: Shapley value analysis of the 95th-percentile pricing

Rade Stanojevic
Telefonica

Nikolaos Laoutaris
Telefonica

Pablo Rodriguez
Telefonica

ABSTRACT

Cost control for the Internet *access providers* (AP) influences not only the nominal speeds offered to the customers, but also other, more controversial, policies related to traffic shaping and discrimination. Given that the cost for the AP is determined by the peak-hour traffic (eg. through the 95th-percentile), the individual contribution to the aggregate cost depends on the behavioral demand pattern of all involved customers. In this paper we propose a metric for evaluating the contribution each individual user has on the peak demand, that is based on Shapley value, a well known game-theoretic concept. Given the computational complexity of calculating the Shapley value, we use a Monte Carlo method for approximating it with reasonable accuracy. We employ our methodology to study a dataset that logs per-subscriber temporal usage patterns over one month period for 10K broadband subscribers of a European AP and report observed results.

1. INTRODUCTION

A large number of Internet Access Provider (AP) adopted flat-rate pricing as a de-facto standard for charging of broadband services as such pricing appears to be preferred by the customers [18]. This creates many difficulties for the APs as it does not allow APs to transparently control the uplinking (transit + infrastructure) costs and forced many APs to create nontransparent rules for traffic shaping and violate net-neutrality as a means for control of their costs [9]. Using the terminology of [21], uplinking costs are the single most expensive component of the costs for broadband connectivity for a majority of currently used technologies, including DSL, cable and WiFi (mash and access point). An important property of the uplinking costs (influenced by the transit costs and the cost of infrastructure) is that they are determined by the peak demand (eg. through the 95th-percentile) rather than average demand, which makes it hard to assess per-customer contribution towards these costs.

Flight tickets typically have different price on the time of travel and hotel rooms in tourist resorts are less expensive during off-season. Similarly, a byte downloaded in peak-hour costs more (for the provider) than a byte of traffic in off-peak hours. In this paper we study per-user contribution in the AP uplinking costs. More precisely, we measure per-customer contribution towards the 95th-percentile of the aggregate demand series[1]. For the purpose of quantifying per-user contribution to the 95th-percentile, we use Shapley value, an intuitive concept from coalitional game theory that characterizes fair cost sharing among involved players (customers). Shapley value framework allows us to: (1) accurately quantify the contribution of each customer towards the peak-hour traffic, (2) analyze the relationship between the aggregate usage (in bytes) and the peak-hour contribution and (3) formally measure how cost of bandwidth is related to the demand pattern. We validate our methodology over a dataset that logs temporal usage of 10K broadband customers of a European AP.

Note that we talk about costs customers generate for the AP rather than the price they pay; retail prices are often strongly impacted by other market, competition and social factors [21]. For various mechanisms for pricing the communication services in the context of revenue (or social welfare) maximization, see [6].

1.1 Toy example

For measuring the peak demand we use the 95th-percentile of the aggregate demand, the most standard measure for billing of the transit traffic and an indicator the network utilization, used for the dimensioning of the infrastructure; see Appendix A for a brief description. To understand the concept of Shapley value and how it applies to the 95th-percentile billing let us consider a synthetic example of an AP ISP providing service to only two users that have demand patterns that are depicted in Figure 1. The user 1 generates a demand of $1Mbps$ during the whole day except for the four-hour period [15-19h]. The user 2 is idle for 22 hours and generates $3Mbps$ traffic during two hours: [16-18h]. The

[1] We stress, however, that the framework is general enough to accommodate any other metric that measures the peak demand.

95-th percentile of the aggregate user demand is the peak-hour traffic $v_{95th} = 3Mbps$ and the price the ISP would need to pay to its transit provider is $v_{95th} \cdot A_0$ (where A_0 is the price in USD per $Mbps$). The following question arises: What is the fair cost sharing among the two involved users? The Shapley value concept gives an answer to this question and the intuition behind it is described bellow.

If there was only user 1 or only user 2 in the system, the 95th percentile would have been:

$$v_{95th}(\{1\}) = 1Mbps \quad \text{and} \quad v_{95th}(\{2\}) = 3Mbps$$

respectively. As we already observed, the 95th percentile of the union of these two users is

$$v_{95th}(\{1, 2\}) = 3Mbps.$$

The Shapley value of user i, ϕ_i is now the average marginal contribution that user i imposes to the coalition cost. In other words:

$$\phi_i = \frac{1}{2} \left(v_{95th}(\{i\}) + (v_{95th}(\{i, i'\}) - v_{95th}(\{i'\})) \right),$$

where $\{i'\} = \{1, 2\} \setminus \{i\}$. In our example the per user Shapley values are:

$$\phi_1 = 0.5Mbps \quad \text{and} \quad \phi_2 = 2.5Mbps.$$

Thus by entering the coalition, the fair cost sharing of the 95th-percentile $v_{95th}(\{1, 2\}) = 3Mbps$ would be the one in which the user 1 is accounted for $\phi_1 = 0.5Mbps$ and the user 2 for $\phi_2 = 2.5Mbps$. The nature of the 95th-percentile pricing is such that even though the user 1 generates in total 3.3 times more traffic than user 2, its contribution to the 95th-percentile is 5 times lower.

COMMENT 1. *We can learn two lessons from the above example: firstly, the user that sends/receives more data does not necessarily have higher impact on the 95th-percentile; and secondly, even if a user does not generate any traffic in the peak hours that does not imply that its impact towards the 95th-percentile is zero. Shapley value balances between these two extremes (aggregate usage and peak-only usage) by evaluating the average marginal contribution of each user (eq. 1).*

1.2 Summary of contributions

Briefly, the main contributions of this paper are the following:

- We develop a new methodology for studying heavy users in an operational ISP. We use the standard concept from cooperational game theory, known as Shapley value, to quantify per-user cost contribution in the context of 95th-percentile pricing.

- Using the Shapley value methodology, we study a month-long dataset that tracks temporal usage patterns from 10K broadband users of a European ISP. We quantify several relevant metrics over this dataset. In particular we find that for approximately 10% of users, the relative cost contribution (Shapley value) is less than half of the relative byte usage (off-peak users), and that for additional 10% of users the relative cost contribution is more than twice of their relative byte usage (peak users). Finally, we use the Shapley value framework to formalize the intuitive wisdom "a byte in the peak-hour has a higher value/cost than an off-peak byte" by quantifying the hourly per-byte bandwidth prices that approximate best the measured Shapley value.

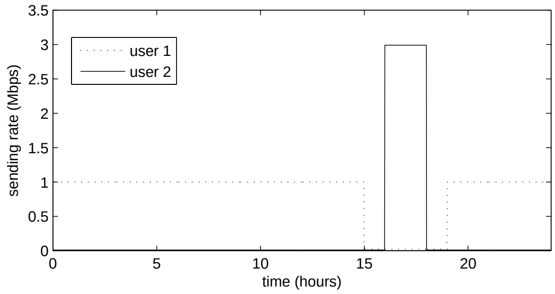

Figure 1: Toy example. Two users with different demand pattern.

2. APPROXIMATING SHAPLEY VALUE

In this section we will briefly introduce the Shapley value (SV) concept for general cooperative games, relate it to our framework in which the cost of a user coalition is determined by the 95th-percentile of the traffic they generate and propose a randomized method for efficiently computing SV for large number of players.

2.1 Shapley value: definition

Consider a set $\mathcal{N}$ of N players[2]. For each subset (coalition) $S \subset \mathcal{N}$ let $v(S)$ be the cost of coalition S. In other words if S is a coalition of players which agree to cooperate, then $v(S)$ determines the total cost from this cooperation.

For given cost function v, the Shapley value is a (uniquely determined) vector $(\phi_1(v), \ldots, \phi_N(v))$ defined bellow that is "fair" in that it satisfies four intuitive properties (see [20, 22]) for sharing the cost $v(\mathcal{N})$ that exhibits the coalition of all players. It can be shown that Shapley value of player i is determined by

$$\phi_i(v) = \frac{1}{N!} \sum_{\pi \in S_N} \left(v(S(\pi, i)) - v(S(\pi, i) \setminus i) \right) \tag{1}$$

[2]We interchangeably use terms player, user, customer and subscriber.

where π is a permutation or arrival order of set $\mathcal{N}$ and $S(\pi, i)$ is the set of players arrived in the system not later than i. In other words, player i is responsible for its marginal contribution $v(S(\pi, i)) - v(S(\pi, i) \setminus i)$ averaged across all $N!$ arrival orders π. Note that the Shapley value defined by (1) satisfies (so called efficiency) property:

$$\sum_{i \in \mathcal{N}} \phi_i(v) = v(\mathcal{N}).$$

2.2 The 95th-percentile cost

The 95th-percentile billing is a method of measuring bandwidth usage based on peak utilization, defined in Appendix A. Informally it measures close-to-peak demand but it also allows usage to exceed a specified threshold for brief periods of time without the financial penalty.

The setup over which we apply the Shapley value framework is the following. We have the set $\mathcal{N}$ of N users that generate traffic over a charging period, say one month. The charging period is split into T sampling intervals, and at time $t \in [1, T]$, user i generates the traffic $Z_i(t)$ (measured in bytes). For a time series $D = (D(1), \ldots, D(T))$, the 95th-percentile $P_{95th}(D)$ is defined as the $\lceil \frac{T}{20} \rceil$-th largest number of the time series. For a coalition S of users the cost they generate is determined by the 95th-percentile of the aggregate demand pattern they generate:

$$v(S) = P_{95th}(\sum_{i \in S} Z_i(1), \ldots, \sum_{i \in S} Z_i(T)).$$

Given the cost function $v(\cdot)$, the contribution of each user to the 95th-percentile of the aggregate traffic $v(\mathcal{N})$ is defined by the Shapley value defined by (1). From the definition, one can notice that the 95th-percentile does not decrease by adding new users to the coalition, therefore implying that the cost function v is monotone:

$$(\forall S \subset \mathcal{N})(\forall i \in \mathcal{N}) \quad v(S \cup i) \geq v(S).$$

The monotonicity of the cost function v implies that the Shapley value of each user is indeed nonnegative.

2.3 Approximating Shapley value

Brute force application of formula (1) is computationally unfeasible once N becomes greater than 100. For APs with thousands (or millions) of subscribers such exact computation is not possible. In this Section we describe a simple randomized method for approximating the Shapley, that can scale with datasets of tens of thousands (if not millions) of subscribers.

The idea of the method is simple. The Shapley value of user i defined by (1) can be seen as the marginal cost increase by user i, averaged over all $N!$ arrival orders. In the example from Section 1.1, $N = 2$ and there are 2 arrival orders: $\pi_1 = (1, 2)$ and $\pi_2 = (2, 1)$ and the user 1 and user 2 Shapley values are

$$\phi_1 = \frac{1}{2}\left((v(\{1\}) - v(\emptyset)) + (v(\{1,2\}) - v(\{2\}))\right) =$$

$$= \frac{1}{2}\left((1 - 0) + (3 - 3)\right) = 0.5.$$

$$\phi_2 = \frac{1}{2}\left((v(\{1,2\}) - v(\{1\})) + (v(\{2\}) - v(\emptyset))\right) =$$

$$= \frac{1}{2}\left((3 - 1) + (3 - 0)\right) = 2.5.$$

While computing the exact Shapley value through the formula (1) is straightforward for small N, it becomes unfeasible for $N > 50$, as the number of different permutation orders grows with $N!$. However, the computational complexity can be significantly reduced by using the Monte Carlo method.

Instead of calculating the exact Shapley value as the average cost contribution across all $N!$ arrival orders, we estimate the Shapley value as the average cost contribution over a set Π_k of k randomly sampled arrival orders (permutations).

$$\hat{\phi}_i(v) = \frac{1}{k} \sum_{\pi \in \Pi_k} \left(v(S(\pi, i)) - v(S(\pi, i) \setminus i)\right) \qquad (2)$$

The parameter k determines the error between the real Shapley value and its estimate: the higher k the lower the error. So basically, one can control the accuracy of the estimators by increasing the number of sample permutation orders.

PROPOSITION 1. *The estimator $\hat{\phi}_i(v)$ is an unbiased estimator of the real Shapley value $\phi_i(v)$.*

PROOF. Indeed, given that each permutation has the same probability of being sampled in Π_k, the expected value of $\hat{\phi}_i(v)$ is:

$$E[\hat{\phi}_i(v)] = E[\frac{1}{k} \sum_{\pi \in \Pi_k} (v(S(\pi, i)) - v(S(\pi, i) \setminus i))] =$$

$$= \frac{1}{k}(k\phi_i(v)) = \phi_i(v).$$

□

Thus the Shapley value estimator (2) is unbiased. However the variance of the estimator is hard to model and in Section 3.2 we present empirical evidence that for reasonably small sample size (say, $k = 1000$) the estimator exhibits small variance, especially for the top users.

PROPOSITION 2. *The estimated Shapley values satisfy the efficiency property:*

$$\sum_{i \in \mathcal{N}} \hat{\phi}_i(v) = v(\mathcal{N}).$$

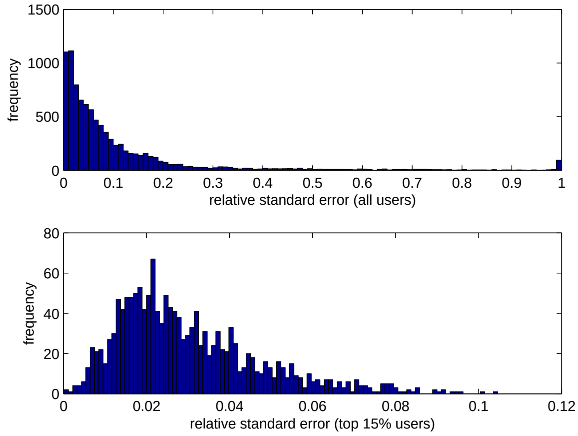

Figure 2: Relative standard errors of the Shapley value estimator (2). Top: all users (top); bottom: users with estimated Shapley value higher than the mean (approx top 15%).

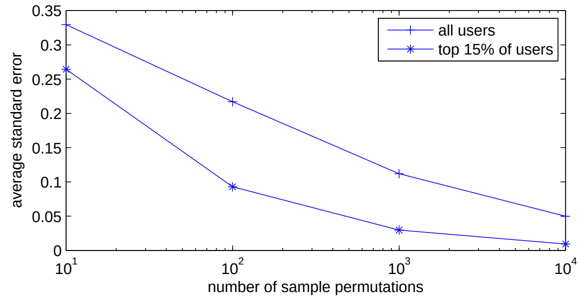

Figure 3: Standard error as a function of number of sample permutations (parameter k).

PROOF. The proof is straightforward:

$$\sum_{i \in \mathcal{N}} \hat{\phi}_i(v) = \sum_{i \in \mathcal{N}} \frac{1}{k} \sum_{\pi \in \Pi_k} (v(S(\pi,i)) - v(S(\pi,i) \setminus i)) =$$

$$= \frac{1}{k} \sum_{\pi \in \Pi_k} \sum_{i \in \mathcal{N}} (v(S(\pi,i)) - v(S(\pi,i) \setminus i)) =$$

$$= \frac{1}{k} \sum_{\pi \in \Pi_k} (v(\mathcal{N}) - v(\emptyset)) = v(\mathcal{N}).$$

□

3. EMPIRICAL RESULTS

In this section we will empirically analyze the dataset of around 10K broadband users of a major European ISP. In Section 3.1 we describe the dataset, then in Section 3.2 we analyze the accuracy of the randomized method for calculating Shapley value. We proceed by analyzing the correlation between per-user aggregate usage and its Shapley value in Section 3.3. The consistency of Shapley value is evaluated in Section 3.4 and then in Section 3.5 we quantify the relative cost of bandwidth in time that would best approximate the Shapley value.

3.1 Dataset description

The dataset consists of around 10K randomly sampled ADSL users of a major access provider in one European country. For each customer, its downstream and upstream consumption (in bytes) is captured during each hour for 30 days (thus spanning 720 hours). These users represent a random sample of ADSL users of the ISP and have diverse uplink/downlink capacities in the ranges of 256Kbps–10Mbps, and 1Mbps–20Mbps, respectively. The downstream traffic dominates the upstream traffic in the ratio 4 : 1, which is consistent with the recent findings from another European access provider ISP [15]. Virtually all ADSL users from the dataset pay flat-fee, without incentives to shift their traffic to the off-peak hours[11]. We stress that the empirical results derived from this dataset are mainly qualitative, used for the purpose of validating the Shapley value methodology and basic properties of Shapley value, and results derived here should not be generalized for other types of environments such as campus, backbone or enterprize networks.

For the computation of 95th-percentile we use 1-hour bins, as this is the granularity of our dataset. Given that we consider large traffic volumes and large number of users, using different bin sizes (eg 5 minutes) would have minor effects on the 95th-percentile [8].

As we said, the downstream traffic dominates the upstream and in the following analysis we will therefore focus on the downstream traffic, as it is the direction that determines the 95th-percentile (see Appendix A). The dataset does not distinguish the per-user share of transit/nontransit traffic, so for the evaluation purposes we assume that all the traffic contributes to the 95th-percentile.

3.2 Accuracy of the Shapley value estimator

Our first step is the evaluation of the accuracy of the Shapley value estimator (2). Given that we do not have the ground truth measurement, to evaluate the error that the estimator exhibits, we use the standard statistical method as follows. Recall that the Shapley value estimator (2) of user i is a mean of k samples of marginal cost contributions $v(S(\pi,i)) - v(S(\pi,i) \setminus i)$. If we denote by $\hat{\sigma}_i$ the estimated standard deviation of the same k marginal cost samples. Then the relative standard error of the estimator (2) is $\frac{\hat{\sigma}_i}{\sqrt{k}\hat{\phi}_i(v)}$. In Figure 2 we plot the histograms of these standard errors when the sample

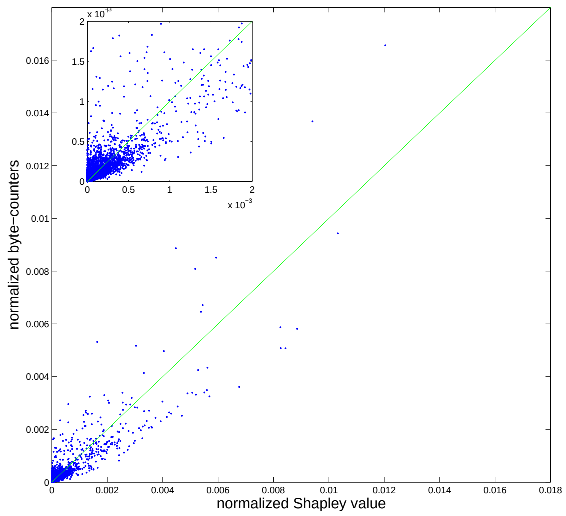

Figure 4: Normalized byte-counters vs. normalized Shapley value.

size is $k = 1000$ permutations[3] for all users as well as the users with Shapley value estimate higher than the mean Shapley value $1/N \cdot v(\mathcal{N})$. One can observe that the relative standard errors are moderately small across all users and are consistently under 10% for top users. In Figure 3 we vary the number of sample permutation orders (parameter k) and evaluate the relative standard errors averaged across all the users and also the top 15% users.

3.3 Aggregate usage vs. Shapley value

Now, that we established the accuracy of Shapley value estimates, we will compare it with the time-oblivious usage measure: bytes downloaded over the whole 30-day period (byte-counters). In Figure 4 we plot the normalized[4] Shapley value (x-axis) against the normalized byte-counters (y-axis) for each user from our dataset. Users with relatively high off-peak usage correspond to datapoints that are far above $x - y = 0$ line. Conversely, users with modest off-peak usage and heavy "peak-hour" usage correspond to datapoints close to x-axis. Finally, the more similar the usage activity of a user is to the aggregate usage pattern, the closer its datapoint is to the $x - y = 0$ line. To measure how different the user's Shapley value and byte-count are we introduce the following metric that basically measures the discrepancy between the user's relative aggregate usage

[3]The computation took under 5 minutes on a PC running Intel Core 2 Duo CPU, 2.33GHz and 2GB of RAM.

[4]Scaled down proportionally to have the sum equal to 1.

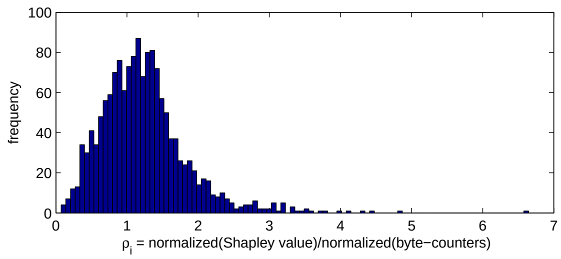

Figure 5: Normalized byte-counters vs. normalized Shapley value.

and its relative contribution to the 95th-percentile:

$$\rho_i = \frac{\text{normalized Shapley value of user } i}{\text{normalized byte-count of user } i},\ i \in \mathcal{N}.$$

As we said above, the users with high off-peak usage (compared to their peak-usage) have low ρ_i and vice versa. In Figure 5 we plot the histogram of ρ_i for all users i with Shapley value greater than the mean (approx 15% of the users) as we have a high confidence in the measured Shapley value for those users (see Figure 2). We see that there are approximately 10% of users whose relative cost contribution is more than twice the relative byte counters ($\rho_i > 2$) and another 10% of users with relative cost contribution less than half of its relative byte count ($\rho_i < 0.5$). The proportion of users with very large or very small ρ_i is even higher for low-Shapley-value users (the remaining 85% of the dataset), but we avoid reporting these numbers because of the accuracy issues for the low-Shapley-value users (see Section 3.2).

Another interesting statistics is that around 30% of top-1% (and around 25% of top-10%) Shapley value users are not in the corresponding top-1% (top-10%) byte-count list.

3.4 Consistency of Shapley value

The next question we ask is: how consistent Shapley value is among individual users from our dataset? In other words, does temporal usage pattern remain similar for the users from our dataset. Our (somewhat surprising) findings suggest an affirmative answer to this question, at least for the top-users (economic heavy-hitters). For that purpose we evaluated Shapley value during 4 weeks for each individual user in our dataset. Figure 6 depicts the observed weekly Shapley values for the top 500 users (ranked using the monthly Shapley value). One can notice that some users have large week-to-week variations, but majority of users' Shapley value remain similar week after week.

3.5 Relative cost of bandwidth

As we already mentioned in the Introduction, the consequence of the 95th-percentile pricing of the transit

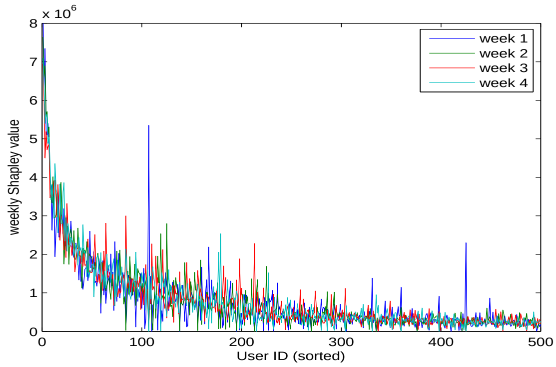

Figure 6: The weekly Shapley value for the top 500 users (based on monthly Shapley value).

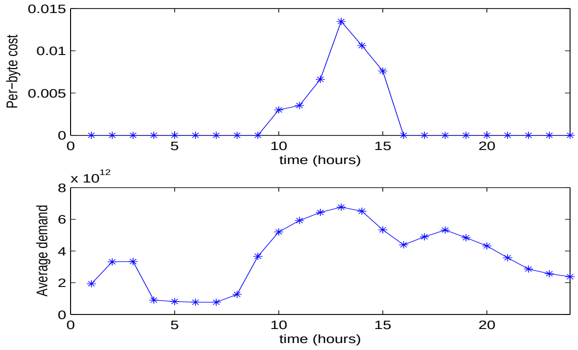

Figure 7: Top: hourly per-byte cost (nonnegative least squares). Bottom: average hourly utilization.

traffic is that the bandwidth is "more expensive" in the peak hours than in the off-peak hours. Here we use the Shapley value framework to quantify how the value of bandwidth changes in time. Namely we seek to find the hourly per-byte prices $c_1, \ldots, c_{24}$, such that if a user i is charged c_h monetary units for each byte downloaded during hour h, then the monthly bill is (approximately) equal to its Shapley value. More formally, if $Z_i(t)$ is the usage of user i at time $t = 1, ..., 720$, we seek for $c_1, \ldots, c_{24} \geq 0$ such that

$$\sum_{h=1}^{24} c_h \sum_{d=1}^{30} Z_i(h + 24(d-1)) \approx \phi_i, \quad \forall i \in \mathcal{N}.$$

Given that the above system of equations is overdetermined (it has 24 variables and 10K equations), we need to seek a fit that matches some optimization criteria. A well know method for approximating solution of the overdetermined linear systems is nonnegative least square (nnls) method that seeks to minimize

$$\sum_{i=1}^{N} \left(\sum_{h=1}^{24} c_h \sum_{d=1}^{30} Z_i(h + 24(d-1)) - \phi_i \right)^2 .$$

Very efficient solutions for nnls problem have been proposed recently and we use [10] to solve our problem. Running nnls over our dataset, we derive time series $c_1, \ldots, c_{24}$ depicted in Figure 7. One can observe that (for the dataset analyzed here) in the 95th-percentile setup, the bandwidth is "free" (has virtually zero impact on the 95th-percentile) for some 18 hours per day and has strictly positive cost during 6 hours per day. Note that even though the monthly 95th-percentile is crossed during only one or two hours per day, the cost of bandwidth is still non-zero for 6 (near-peak) hours.

4. RELATED WORK

Per user analysis of broadband internet traffic was a subject of several recent studies. Cho et al. analyzes broadband traffic from several Japanese ISPs in [4, 5]. They analyze the per-user traffic usage and they show that it is highly skewed[5] (can be modeled with a log-normal distribution). With the duration of the dataset they analyzed, they were able to track upload/download traffic trends over multiple years and quantify the effects different applications have on the traffic aggregates. In [15] Maier et al. perform a measurement study of residential broadband users in one European ISP and analyze several relevant metrics: per-application usage, DSL session duration, and observable round trip-times. Here we take a different look at the broadband traffic by analyzing per-user temporal usage patterns and how they impact the costs for the access provider.

One of the key reasons that influence traffic shaping (also known as traffic discrimination) of Internet traffic is the fact that by throttling some traffic, ISPs control (reduce) their costs [16]. Dhamdhere and Dovrolis [7] and Biczok et al [1] analyze several broadband pricing models that aim to offer a solution that obey the net neutrality rules by discriminating the price of the heavy-hitters (based on the total bytes downloaded/uploaded). In this work we empirically show that the heavy-hitters are not necessarily expensive for the ISP and that temporal usage effects should be taken into account when designing pricing models in the context of net neutrality.

Briscoe [2, 3] argues that fairness mechanisms in computer networks should be judged on "how they share out the 'cost' of each user's action on others" and he offers several heuristics (eg. he suggests that the number of dropped packets over a billing circle is a good indicator of a customer's cost contribution) for evaluating the 'cost'. The Shapley value framework we introduced here can be seen as a formal way to measure user's cost contribution.

[5]The fact that we also observe in our dataset

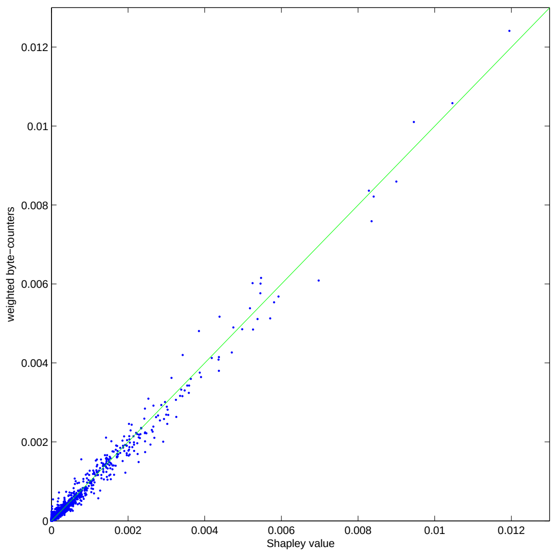

Figure 8: Normalized weighted usage vs. normalized Shapley value

Pricing of communication networks has been extensively studied in the past; see [6] and references therein. These efforts mainly focus on how to use pricing to achieve some form of social welfare (or revenue) maximization. We stress that we do not aim to propose a new pricing scheme in this paper, but rather set to measure the diversity of the broadband usage behavior patterns and their effect on the peak hour consumption.

Recently, several research efforts suggested using Shapley value as a means for providing incentives for optimal resource control. Ma et al. promote use of Shapley value for ISP settlement by proposing revenue sharing among ISPs based on the importance each ISP has on the Internet ecosystem [13, 14]. In the context of peer-to-peer systems, Misra et al. [17] propose using Shapley value to incentivize cooperation in p2p systems.

The 95th-percentile pricing has been analyzed recently by Dimitropoulos et al. [8]. They quantify the dependence between the size of measurement slot and the observed 95th-percentile and show that this dependence becomes weak for large volumes of traffic. Laoutaris et al. [12] use the 95th-percentile pricing to propose ISP-friendly bulk transfers that explicitly avoid to use bandwidth that could increase the 95th-percentile.

5. SUMMARY

Days in which the technological reasons were impacting the performance of the residential Internet users are coming to an end and in the near future, the performance offered to the end users will be predominantly shaped by the economic factors rather than physical bottlenecks. In such environments it is crucial to determine the cost contribution of each individual user to the operation of ISP, as it would be a key metric for evaluating the consumption and accounting in such an ecosystem. Our study is a step towards the fairer usage of the Internet in which economic aspects dominate the per-user performance as it formally quantifies the individual per-user cost contributions in the specific context of burstable (95th-percentile) billing.

APPENDIX

A. THE 95TH-PERCENTILE PRICING

The 95th-percentile pricing is the most prevalent method that transit ISPs use for charging their customers, access providers. A billing cycle, typically one month, is split in constant-size intervals (eg. 5-min or 1-hour) and number of bytes transferred in each interval is recorded, and the 95th-percentile of the distribution of recorded samples is used for billing. Thus, in a billing cycle of 30 days, 36 hours (5% of time) of the heaviest traffic is filtered out, and then the maximal traffic of the remaining 684 hours is used for billing. Usually, the downstream and upstream 95th-percentile are computed independently, and the lower value is neglected.

The 95th-percentile is also a good measure of how utilized the network is, and is often used as an indicator for dimensioning of infrastructure, whose cost is determined by the peak hour demand.

B. REFERENCES

[1] G. Biczok, S. Kardos, T. A. Trinh. "Pricing internet access for disloyal users: a game-theoretic analysis". In Proceedings of NetEcon 2008.

[2] B. Briscoe. "A Fairer, Faster Internet Protocol". IEEE Spectrum, Dec 2008.

[3] B. Briscoe. "Flow Rate Fairness: Dismantling a Religion". ACM Computer Communication Review vol. 37(2), 2007.

[4] K. Cho, K. Fukuda, H. Esaki, A. Kato. "The impact and implications of the growth in residential user-to-user traffic". In Proceedings of SIGCOMM 2006.

[5] K. Cho, K. Fukuda, H. Esaki, A. Kato. "Observing slow crustal movement in residential user traffic". In Proceedings of CoNEXT 2008.

[6] C. Courcoubetis, R. Weber. "Pricing communication networks". John Wiley and Sons, 2003.

[7] A. Dhamdhere, C. Dovrolis. "Can ISPs be profitable without violating "netwotk neutrality"?". In Proceedings of NetEcon 2008.

[8] X. Dimitropoulos, P. Hurley, A. Kind. M. Stoecklin "On the 95-percentile billing method". In Proceedings of PAM 2009.

[9] M. Dischinger et al. "Glasnost: Enabling End Users to Detect Traffic Differentiation". In Proceedings of NSDI 2010.

[10] D. Kim, S. Sra, I. S. Dhillon. "Fast Projection-Based Methods for the Least Squares Nonnegative Matrix Approximation Problem". Statistical Analysis and Data Mining, vol. 1(1), 2008.

[11] N. Laoutaris, P. Rodriguez. "Good Things Come to Those Who (Can) Wait or how to handle Delay Tolerant traffic and make peace on the Internet". In Proceedings of ACM HotNets 2008.

[12] N. Laoutaris, G. Smaragdakis, P. Rodriguez, R. Sundaram. "Delay Tolerant Bulk Data Transfers on the Internet". In Proceedings of ACM SIGMETRICS 2009.

[13] R. T. B. Ma, D.M. Chiu, J. C. S. Lui, V. Misra, D. Rubenstein. "Internet economics: the use of Shapley value for ISP settlement". In Proceeding of CoNEXT 2007.

[14] R. T. B. Ma, D.M. Chiu, J. C. S. Lui, V. Misra, D. Rubenstein. "On cooperative settlement between content, transit and eyeball internet service providers.". In Proceeding of CoNEXT 2008.

[15] G. Maier, A. Feldmann, V. Paxson, M. Allman. "On dominant characteristics of residential broadband internet traffic". In Proceedings of IMC 2009.

[16] M.Marcon, M. Dischinger, K. P. Gummadi, A. Vahdat. "The Local and Global Effects of Traffic Shaping in the Internet". In Proceedings of SIGCOMM 2008.

[17] V. Misra, S. Ioannidis, A. Chaintreau, L. Massouli. "Incentivizing Peer-Assisted Services: A Fluid Shapley Value Approach". In Proceedings of SIGMETRICS 2010.

[18] A. Odlyzko. "Internet pricing and the history of communications". Computer Networks, vol. 36, 2001.

[19] S.Shakkottai, R. Srikant, A. E. Ozdaglar, D. Acemoglu. "The Price of Simplicity". IEEE Journal on Selected Areas in Communications, vol 26(7), 2008.

[20] L. S. Shapley. "A Value for n-person Games". Annals of Mathematical Studies vol. 28, 1953.

[21] R. Tongia. "Why is connectivity in developing regions expensive: policy challenges more than technical limitations?". In Proceedings of WWW 2006.

[22] E. Winter. "The Shapley Value". The Handbook of Game Theory, North-Holland, 2002.