Green Cellular Demand Control with User-in-the-loop Enabled by Smart Data Pricing using an Effective Quantum (eBit) Tariff

Invited Paper

Rainer Schoenen[†], Hamza Umit Sokun[*], and Halim Yanikomeroglu[*]
[†]HAW Hamburg University of Applied Sciences, Hamburg, Germany
[*]Department of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, Canada

***Abstract*—Mobile communications demand for data grows almost two-fold every year and cellular capacity cannot catch up with this in every location and at every point in time. Not only will congestion situations appear more frequently and severely, but also every increase of data transmitted and every new infrastructure point rolled out in order to catch up with that demand is going to consume more total power. One approach towards greener communication is to reduce the actual demand. While technical solutions on Open Systems Interconnection (OSI) layers 1 trough 7 are already being exploited, "layer 8", the user, is the original cause of the exploding demand. It is naive to assume that all data is equally important and valuable. User-in-the-loop (UIL) is a paradigm which can exactly hook in at this point, and it suggests to postpone, relocate, abandon or offload less urgent demand by immediate feedback and strong incentives like discounts and penalties in certain situations (in case of congestion). In particular, this means dynamic prices depending on time, location, and congestion (load) status, similar to smart grid for electricity utilities. Accounting and billing with dynamic prices, value-added services, and smart data pricing can be a challenge, and such accounting and billing are utterly incompatible with the flat-rate (all-you-can-consume) concept. However, cellular customers are involuntarily locked into contracts, but it is the current practice in the market, which is not in accordance with the free market spirit. Flat-rate plans are practically dead, but those still luring with this name, while having caps on the maximum allowance, sell fixed-volume packages under a smoother name instead. Ultimately, we need to re-establish usage-based-pricing as a fairer and more favourable way of charging.**

In this paper we motivate the dynamic pricing of UIL and the upcoming new need for usage-based tariffs. We compare a few status-quo tariff plans with our novel effective quantum (eBit) tariff, which is a usage-based pricing that maps multiple usage dimensions into only one metric, eBits. The eBits will be charged for with one constant unit price ($/eBit) to give the monthly bill. All Quality of Service (QoS)-, application-, and time-dependent dynamic pricing factors are integrated into the effective quantum, so that a user can easily understand his current consumption and the bill upto that point in time. Price differentiation per QoS class is easy with constant scaling factors. This principle can be extended to any other chargeable metric such as kWh of electricity and m^3 of water. This ultimately enables going greener and training awareness of all consumers.

***Index Terms*—Green demand control, smart data pricing, user-in-the-loop, usage-based pricing.**

I. Introduction and Motivation

In mobile communications with demand increasing at a rate of almost 100% per year, going green and reducing power consumption by just hoping for more energy efficient infrastructure is not going to happen. The opposite trends dominate: Higher base station density, more spectrum, more RF amplifiers and antennas, and irresponsible piecewise flat tariffs that heat up the demand increase without sanity. This is of course due to unaware customers and consumers in conjunction with clever operators' marketing of "easy", "no bill shock", "free bundled phone", "2+ year contracts". The new trend of Smart Data Pricing (SDP) is a noble turning point. User-in-the-loop (UIL) [1]–[3] builds upon that fundament. SDP refers mainly to (a) time/location/app/congestion dependent dynamic pricing, (b) usage-based pricing with throttling/booster, (c) two-sided pricing/reverse billing/sponsored content. It can increase users' happiness, alleviate congestion, improve Quality of Experience (QoE), lower capital expenditure (CapEx)/operational expenditure (OpEx), and increase revenue/profit margin [4]. Pricing models for SDP are still not on the market. Instead we currently observe the doom of the flat-rate concept, volume-caps for a fixed monthly price, overage fees for exceeding the caps, and device-plan-bundles as well as family-sharing plans, which all handicap users to make an educated decision in a free market. Additionally, keeping independent billing dimensions for data, voice, and SMS leads to ridiculously high rates such as $30,000/GB for SMS [5]. That is why this paper proposes an one-dimensional mapping of all service volumes into an artificial yet intuitive unit so-called eBit, which will be charged with a fixed and known price in $/eBit, which may differ only per provider. The mapping itself should follow a common rule for all providers, so that a customer can compare all plans on the market by just one scalar number.

The main psychological reason for customers to embrace the flat-rate idea is to stand clear of any bill-shock, which one out of six people have already experienced in the U.S. But they coarsely overpay for volume limits which are usually way beyond their average consumption, including possible times of almost zero consumption. The trend is away from flat-rate models [6]. Related work in SDP is emerging [7], [8]. However, changes on the market only happen gradually. Radical changes are feared by the providers for whatever reason.

In this paper the effective quantum (eBit) pricing model is motivated and proposed, as a one-dimensional metric, which is easy to understand by users, and enables all advanced concepts of dynamic pricing, e.g., the UIL demand-control paradigm [1]; equally important, it enables the use of any monotonic post-processing function, e.g., logarithmic price companding or even (back to) flat-rates with cap models. Sponsored content (reverse billing) is also feasible with this concept. In combination with UIL, eBit pricing allows to give incentives for moving to less congested locations and to less busy time windows [1]. By this way, such combination can reduce not only the actual demand, but also total power consumption, as it is proportional to demand [9].

II. Related Work

The term SDP has recently emerged with a loose reference to the smart grid idea [10]. One key aspect is the ability to control users' traffic demand to adapt to the capacity situation at the current location or time. This has been suggested early in 2010 [9], [11], and applied to SDP in [12]. Empirical studies from surveys provided quantitative metrics how elastic the users' demand reacts to incentives [13]. This has been studied in further papers by the authors, e.g., providing the closed-loop control model for three QoS classes. As incentives must be indicated to the user in a suitable (one-dimensional and scalar) way, eBit was first mentioned in [14]. Basic work on usage-based pricing can be found in [15]. Congestion pricing was first mentioned in [16]. A good survey can be found in [17]. Up-to-date usage statistics can be found in [6].

III. Going Green with User-in-the-Loop

UIL was proposed to influence users' demand behaviour immediately at or before start of consumption [9], [11]. It needs to provide positive or negative incentives (discounts or penalties, payback/airmile points, better service after compliance). As it is closed-loop, the severity of feedback adapts to the conditions of the supply-demand balance at the current radio cell location and time. The outcome is a relocating user (sometimes just a few meters is sufficient), or a demand postponed out of busy hour into evening times. Please imagine for a moment that a dominant use case of young people watching non-urgent streaming videos while unaware parents pay for the flat bill. It is well possible to reduce demand to just below capacity, knowing elasticity of user demand with incentives [13].

A dynamic unit price $p[k] = (1+\chi[k]) \cdot p_{\text{const}}$ [\$/MB] is one way of showing incentives, where $(1+\chi[k])$ is a UIL-enabled dynamic price factor and p_{const} [\$/MB] denotes unit price per eBit. For instance, busy-hour prices can be, e.g., 50%, above regular level [9]. The system dynamics were studied in [18]. For billing purposes at the end of the month, this approach

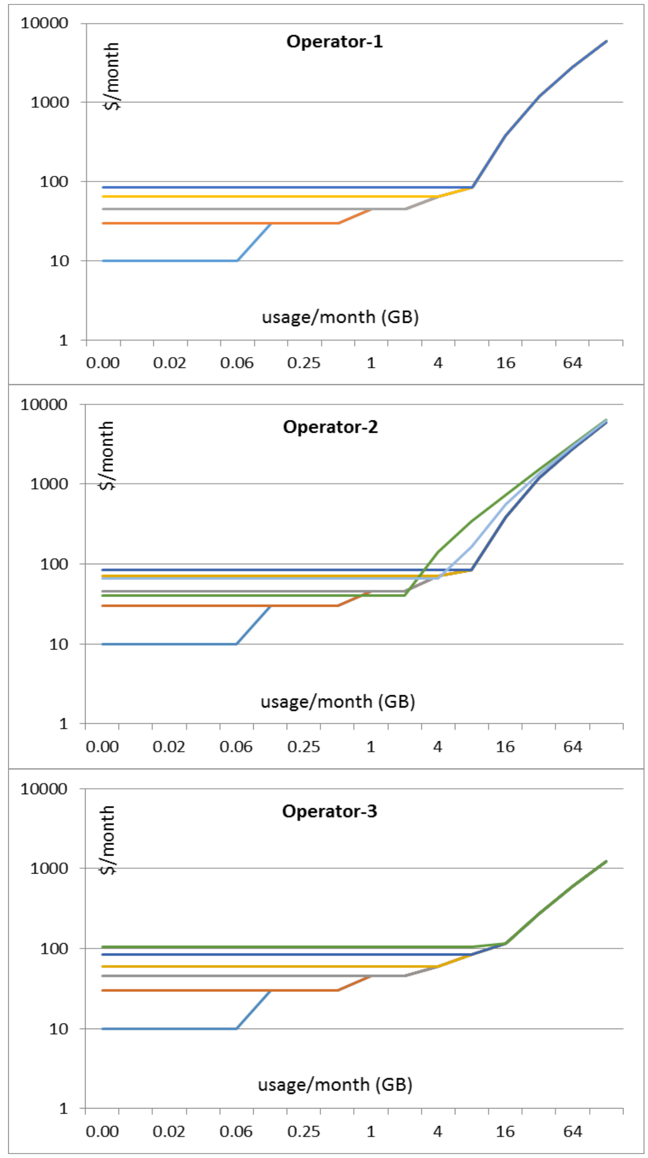


Fig. 1. Typical cellular tiered tariff plans in a developed country with three major operators each with approximately one-third market share. Each line in the figure corresponds to a different plan of the respective operator.

requires accumulating partial volumes $v[k]$ multiplied with their variable unit price $p[k]$. As this can be cumbersome to document for bill review purposes and can undermine the long-term training effect to the user, we need a much simpler system that every user can understand and follow, no matter on which intellectual level. Thus the motivation for better tariff alternatives and our eBit proposal.

IV. Current and Future Tariff Systems

The following evolution of tariff systems can be observed and extrapolated (extended from [6]):

1) Metered Internet dial-up (counting minutes or volume, 1990s),
2) Flat-rate (constant price, free volume, 2000s),
3) Flat-rate with cap (huge overuse fees, 2010s),

4) Tiered pricing (flat rates with auto-jump to next higher package, 2013),
5) Usage-based pricing (proportional to real volume used),
6) Value-based pricing (usage-based plus service differentiation) [6],
7) Dynamic pricing (usage-based plus congestion surcharges) [10],
8) Smart Data or UIL pricing (dynamic pricing with closed-loop demand control) [1],
9) eBit pricing (usage & dynamic value-based with simple one-dimensional "price per eBit").

Currently the tiered pricing is the status-quo for wireless data (e.g., a USB surfstick for a laptop computer). It is not yet the default for smart phone data. In tiered pricing the user selects a plan which includes a fixed volume for a fixed price. If the volume boundary (cap) is exceeded, in that month the contract jumps to the next higher tier bracket, with all associated costs. To demonstrate this system with an example, Fig. 1 shows current pricing packages for wireless data in a developed country with three major operators each with approximately one-third market share. We observe that tiered pricing has a step-wise constant price within the suitable regions (volume per month). Beyond the available tiered plans, an overuse fee of minimum $\$10/\mathrm{GB}$ of data is charged. Seen from a distance, any step function looks like a linear function, so on a coarse scale we can say that this is an usage-based pricing without fine-grain counting. In other words, users are paying more than proportional to their consumption whenever their monthly volume is far from the switching points of their tiered interval. In addition, the limit (excess) value of $\$10/\mathrm{GB}$ is only in effect, if the first 10 GB have already been paid for with approximately $\$100$.

We define the limit price (for very high consumption) $p_{\lim}(v)$ based on the volume v [Bytes] and the charged price p for it as follows:

$$p_{\lim}(v) = \lim_{v\to\infty} \frac{\Delta p}{\Delta v}. \quad (1)$$

In contrast to wireless cellular data with a limit price of $\$10/\mathrm{GB}$, fixed line data comes at a limit price (1) of $\$0.5/\mathrm{GB}$ for any excess of the cap limitation per month, i.e., a factor of 20 cheaper. The same trick here; though, that the first $200/\mathrm{GB}$ are paid for with $\$100$ before it comes into the linear regime. A factor of 20 is manifested here, too, although the aspects are independent. The rule seems to be the following: For the first $\$100$ the consumer pays 10 times the unit price compared to the limit price.

V. One-Dimensional Effective Quantum Metrics

One of the main problems in billing is the transparency for the user. The favourable usage-based pricing naturally comes with monthly consumption in multiple dimensions, e.g., voice, video, data, SMS, and potentially other application-specific volume counts, if they differ in price. This is already hard to track for an average user, and educated choices for plans and options can hardly be made, if contracts come with multi-

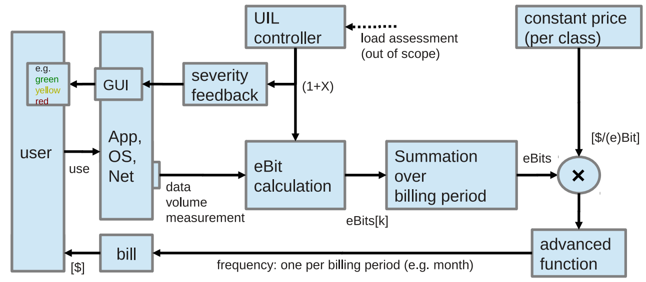


Fig. 2. Block diagram for pricing using effective quanta. This is a closed loop because of the UIL component enabling dynamic prices (specified by $1+\chi$)).

dimensional bundled allowance volumes in a package. It would be even harder to indicate a dynamic price, which we need for congestion alleviation [1]. We must solve this problem and to enable all sorts of dynamic pricing and even non-linear post-processing functions: Therefore, a mapping from all multiple dimensions q into one effective volume (quantum) v_{eff} is suggested, by weighted summation of its components:

$$v_{\mathrm{eff}} = \mathsf{eBits} = \sum_q v_q \cdot f_q \cdot (1 + \chi_q), \quad (2)$$

where v_q, f_q, and $(1+\chi_q)$ denote the measurement of the consumed data volume v per class q, the static price weight factor per QoS class q, and a dynamic tariff factor per QoS class q, respectively. Although it may look trivial, this is a novelty, because the sum is no longer in the base unit of bits and bytes. Instead, it is a dilated version of bits which we call eBits or eBytes (big multiples in KeB, MeB, GeB = Giga-e-Bytes), and in generalized contexts (water, electricity) simply effective quantum of the original measurement unit. (3c) in the next page is the effective volume only for an interval where the dynamic price factor $(1+\chi_q)$ is constant. It must be summed up over all intervals of different dynamic price (index k in (3e)). The constant weight f_q, the different unit price per class q, is agreed on among the providers. For example, consider $f_{\mathrm{data}} = 1$, $f_{\mathrm{video}} = 2$, $f_{\mathrm{voice}} = 10$ is assumed to represent the different QoS requirements and historical emphasis on voice services as a cash cow. With the same method, the operators may set $f_{\mathrm{VoIP}} = 20$ to penalize legacy VoIP third party software (such as Skype). This is better than throttling or blocking it until it cannot be used at all (common practice). The whole differentiated pricing is at conflict with net-neutrality, a concept in which the advocates fight for the same share of the bandwidth for everyone. This discussion is beyond the scope of this paper.

In any moment in time, with eBits the consumer only needs to know one number which is counted up monotonically. Bill shock is avoided by mandatorily setting limits and warning thresholds before any set budget is exceeded. Especially, for reasons of parental control, any user who is not the bill payer must have such limits installed.

Fig. 2 illustrates the principle of the eBit concept. The eBit concept consists of

1) The user block initiates traffic by starting and stopping sessions (use). The user receives immediate UIL feedback (e.g., green, yellow, red indicators) based on the severity of network congestion. Once per billing period (e.g., month) the user receives a detailed bill showing his total consumption in eBits and the corresponding price C (\$). He should be able to see intermediate statements of eBits and \$ at any time (a latency of some minutes may be technically necessary).
2) The "App, OS, Net" box embraces the user's smartphone, its operating system (OS), graphical user interface (GUI) and the network carrying the traffic.
3) The UIL controller [9] assesses the network state and generates a dynamic tariff factor $(1+\chi_q)$ per QoS class q (smart dynamic pricing component).
4) This severity feedback $(1+\chi_q)$ is forwarded to the user's GUI for immediate notification.
5) The severity $(1+\chi_q)$ also goes into the "eBit calculation" box, together with (static) price weight factors f_q per QoS class q and the measurements of the consumed data volume v per class q.
6) All time dependent variables have an index $[k]$ denoting the time-interval slot number, even if not shown everywhere for better visibility.
7) The eBit calculation executes (3a) which just scales each bit with the severity factor. The equation below (3b) is more correctly specified for multiple QoS classes, by considering the (static) price weight factors f_q and (potentially different) $(1+\chi_q)$ per QoS class q.
8) During the billing period (usually starting at the beginning of the month or the beginning of the contract) the eBits are simply summed up (3c).
9) This eBits is a running variable counting the eBits similar to the way regular bits, minutes, or SMSs are counted in the current billing systems. As such, the user should be able to see a "statement of eBits" at any time, for verification purposes, as well as see this number at his end device to assess his consumption behaviour and anticipated costs. It would be good if a parental control option also made this number available to the supervising parent, without showing the concrete type and content of the sessions (privacy).
10) The next step is the bill cost calculation to get C (3d), where the eBits are simply multiplied with a constant "unit price per eBit" p_{const} $[\$/\text{eBit}]$. This price might differ from provider to provider and be subject to change from month to month (or less frequently).
11) For consumer protection it would be very useful if the user could switch providers easily (without changing SIM cards). Ideally the user would be able to switch from month to month or even from session to session for a real competition on the market.
12) The "advanced function" block (3e) allows an extra function on top of the eBits or eBit costs C, e.g., for implementing a non-linear convex function (3f), for example, logarithmic pricing [9]. Parameters a and b are constant scaling parameters.
13) The resulting bill cost C becomes the price on the bill at the end of the billing period.

(Just one QoS class)

$$\mathsf{eBits}[k] = v[k] \cdot (1+\chi[k]) \tag{3a}$$

(Multiple QoS classes)

$$\mathsf{eBits}[k] = \sum_q v_q[k] \cdot f_q \cdot (1+\chi_q[k]) \tag{3b}$$

(Total over time)

$$\mathsf{eBits} = \sum_{k=0}^{K_{\text{now}}} \mathsf{eBits}[k] \tag{3c}$$

(Bill cost linear model)

$$C = \mathsf{eBits} \cdot p_{\text{const}} \tag{3d}$$

(Bill cost generic)

$$C = p_{\text{const}} \cdot f(\mathsf{eBits}) \tag{3e}$$

(Bill cost logarithmic postprocessing)

$$C = p_{\text{const}} \cdot b \cdot \log_{10}(1+\mathsf{eBits}/a) \tag{3f}$$

The behaviour for one QoS class is shown as a graph in Fig. 3. In addition, the eBit concept allows to map a complex multi-dimensional tariff system (different prices and packages for Internet, voice, SMS, MMS, as in value-based pricing) into a one-dimensional "price per eBit", where the service differentiation is shifted into the eBit weight factor f_q. For the following, $8\text{eBit} = 1\text{eByte}$, and one can use the usual multipliers (kilo, mega, giga). This allows

- easy understanding by the consumer (one scalar number, e.g., \$10/GeB = dollars per Giga-eByte),
- harmonization of the eBit mapping values (e.g., $f_q = 1$ [eBit/Bit] for BE data, $f_q = 2$ for NRT video, $f_q = 10$ for RT voice),
- easy comparison of tariff plans between competitors as f_q are standardized among the providers,
- easy to detect flaws and rip-off (e.g., \$30,000/GB for SMS [5]),
- enabling advanced billing functions (3e), e.g., logarithmic tariffs (3f) [9], after metering the monthly volume of eBits.

VI. Application to High-Frequency Closed-Loop Dynamic Pricing

The UIL paradigm or any other dynamic pricing scheme is integrated into the eBit metric by its dynamic factor $(1+\chi[k])$ within any piecewise-constant interval k. Note that for an open session this price factor must not change, as it was negotiated with the user beforehand [1]. For all any new session, the dynamic price is determined by a controller [1] which knows the current load of the particular network location and the elasticity of the users' demand [13]. If the closed-loop UIL is applied, we achieve congestion alleviation [9], spectral efficiency boosts [11], or better load balancing in heterogeneous networks with heterogeneous user distribution

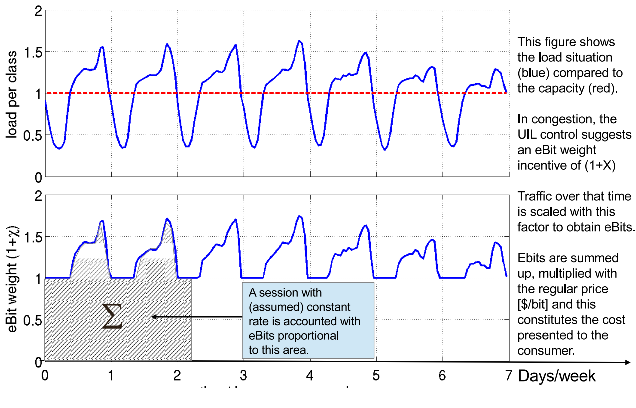


Fig. 3. Graph of eBit dynamic weight parameter $1+\chi(t)$ for dynamic UIL pricing over 7 days.

(HetHetNets) [19]. Fig. 3 shows an example how the dynamic price would change over time if we are in a heavily congested cell. The elasticity of users' demand was determined by a survey [13]. Each time the eBit weight $(1+\chi)$ is greater than a threshold and the application has a significant transaction (e.g., \$0.20, options setting) the user would get an explicit price alert notification at the time the transaction is about to be started. In all less significant cases a permanent indicator in the notification area could show the factor as a color (green to red) or as bars, similar to the RSSI signal strength indicator.

VII. Numeric Examples

It is important to assess any new model with metrics that allow a comparison to the existing ones. If we assume known usage statistics [6] and fit their heavy-tailed statistics into a matching Generalized Pareto distribution, we get the values showing in Fig. 4 for data of 2014 (Fig. 4(a) shows the original data and fitted one; refer to the top of the next page).

For the following analysis we assume that the proportion of video, data, and voice traffic is 53%, 45%, and 2%, respectively for 2014 [6]. Random numbers for each class are generated independently, see Fig. 5. Video and data volumes per month are each Generalized Pareto-distributed with shape parameter 0.4383, scale parameter 1.0175 GB/month, and threshold parameter -0.0164 GB/month and in total described by Fig. 4(a), whereas we model voice volume per month by a normal distribution, with a mean value of 36 MB/month and variance 36 MB/month. Source: [6]+Matlab.

With a sample size of $N = 10^6$ customers we compare the bills of each according to current (Fig. 1) and novel pricing strategies. For a fair comparison we normalized all parameters (unit price p_u) so that the total revenue of the provider is the same in all variants. Fig. 6(a) shows the way the bill depends on the monthly volume for the static strategies.

In order to assess the acceptance of new plans we measure the satisfaction or happiness of each user as $+1$, if the new bill is less then the reference bill, and -1, if it is more. Because of the heavy-tail distribution, those unhappy users at the right end of the consumption constitute a very small minority. Most

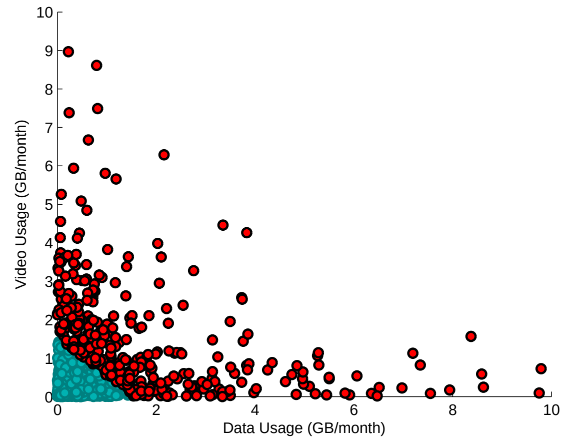


Fig. 5. X-Y plot of random outcome for independent data and video traffic volumes.

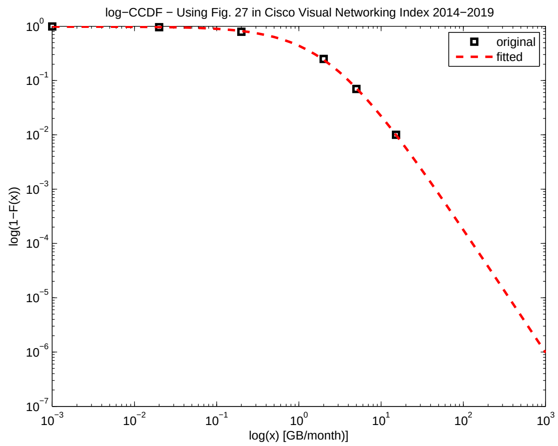
users would benefit from usage-based pricing. Those unhappy would probably reduce their demand to a certain degree (e.g., forced by their paying parents who will set stricter limits). This may be desirable to reduce stress on a congested network. Fig. 6(b) shows all results for comparison. The eBit linear pricing system has 40% more happy users than the existing flat-rates plus cap.

Another metric to assess the suitability is a fairness value according to Jain's fairness index. We can define the bill fairness as $F_b = (\sum b)^2/(N \cdot \sum b^2)$ and see its result in Fig. 7(a). We can as well define a unit price fairness by replacing b by p_u (the unit price in \$/eBit), in which case users feel fairly treated if they pay the same amount per eBit. This can be seen in Fig. 7(b).

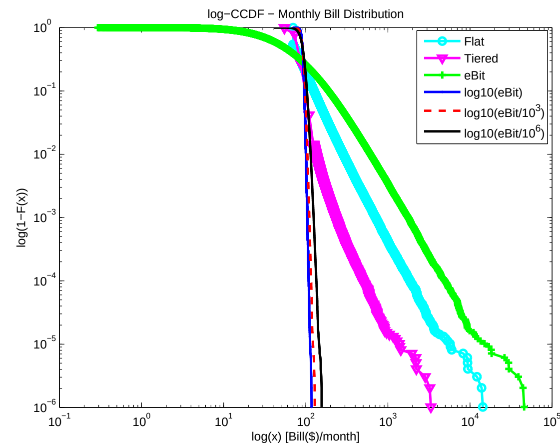
From the results we derive that a usage-based pricing scheme should be viable and acceptable by the majority of users, while having all the benefits of a limiting feedback towards reducing the consumption. Using the dynamic pricing of UIL demand control, this allows extra control of the individual behaviour. Similar to Fig. 3, we observe the individual eBit weights over 30 days and then, we determine an average price increase of only 10% per month, but at the same time we counteract congestion effectively. More fine control of the price function $C = f(\text{eBits})$ can be implemented by a generic monotonic mapping function, e.g., the suggested logarithmic function with scale parameter a (3e). As fairness is concerned, most users would understand well that paying less for consuming less is a logical rule.

VIII. Application to Other Economic Systems

Eventually one day, data traffic transport would be a commodity such as electricity or water supply. To that extent it does not make sense to have such a complex tariff system for the consumer as it is now. Instead, electrical power, water, and bits are counted as they flow in and out of your consuming device. Similar to charging potable water, hot water and cold water supply differently, we count more $\text{eBits}/\mathrm{bit}$ for hot data,

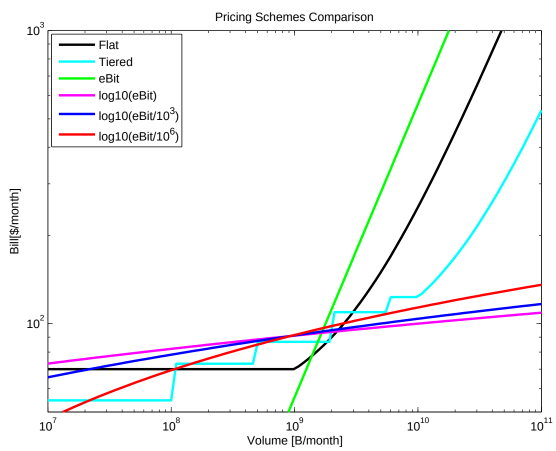


(a) CCDF of the monthly traffic volume per user. Original source [6] and Pareto fit.

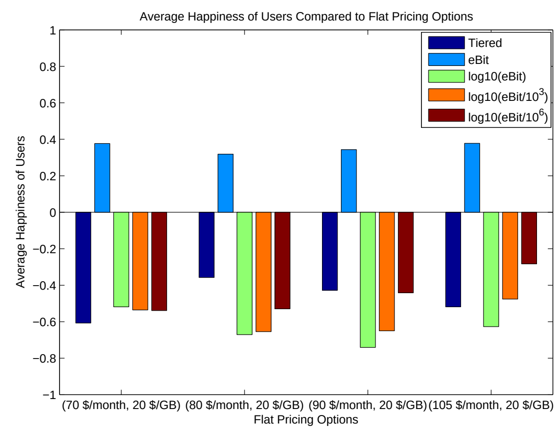


(b) CCDF of the monthly bill per user. The tails for some plans result from overage fees.

Fig. 4. Complementary cumulative distribution function (CCDF) of the monthly traffic volume per user and the monthly bill per user, 2014.



(a) Monthly bill depending on monthly volume for different pricing schemes. Only very vew users are affected beyond 1GB/month.



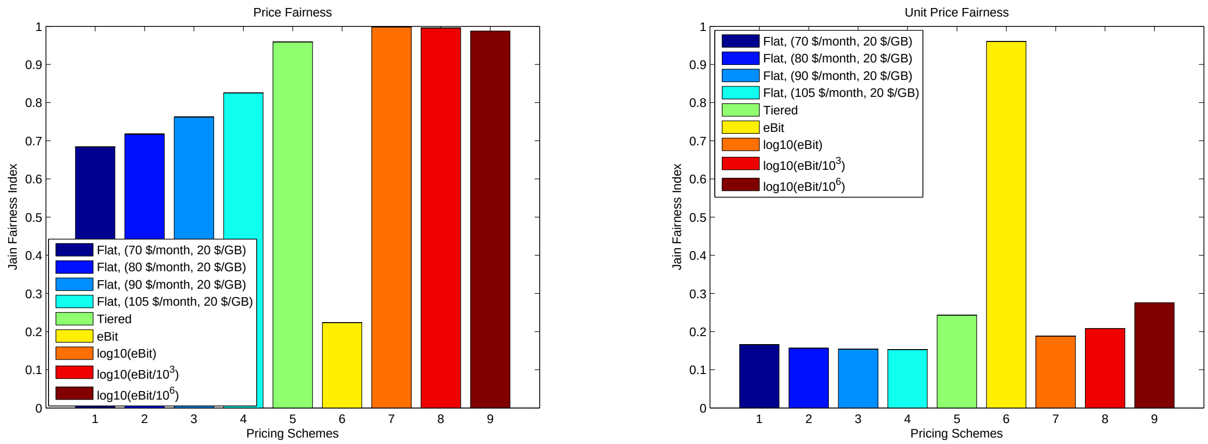
(b) Comparison of the happiness=satisfaction levels over all users for different pricing plans.

Fig. 6. Comparison of monthly bill depending on the monthly volume and the happiness (=satisfaction levels) over all users for the different novel and existing reference pricing plans.

i.e., with stringent QoS requirements (voice, video telephony). Similar to electricity peak hour charging and Smart Grid dynamic tariffs, UIL pricing would motivate consumers to equalize usage patterns over time and space.

IX. Conclusion

This paper proposes an outline for future green communications by introducing demand control, based on incentives that can be mapped to dynamic pricing. A required enabler is a smart data pricing, especially dynamic pricing, which works naturally with usage-based-pricing in general. For better user bill experience we introduce the effective quantum (eBit or eByte) tariff accounting. It combines the distinct costs of all value-based services per QoS class and provides the user with a one-dimensional scalar metric which is (in regular cases) directly proportional to the end price. By mapping multiple dimensions of priceable goods into one, it simplifies the billing interface to the user significantly. Dynamic price elements as needed for UIL demand control [1], are simply incorporated by a dynamic factor between effective volume and real volume (eBits/bits). The indication of dynamic prices reduces to the indication of a dilatation factor, in the easiest case with a traffic light green, yellow, red, depending on severity. The proposed mapping furthermore allows any monotonic function between the effective quantum and the resulting bill. This includes the legacy flat-rates with cap, linear pricing, and a proposed logarithmic pricing. Numeric examples show simulation results of users' satisfaction with different pricing models, based on typical heavy-tailed usage statistics. The complete picture allows controlling for greener resource use and effective long-term user training. At the same time the majority of users favor paying less if they consume reasonably.



(a) Fairness of the bill. This metric favors the same bill for everyone.

(b) Fairness of the unit price. This metric favors the same price per eBit for everyone.

Fig. 7. Fairness of the monthly bill and the unit price.

REFERENCES

[1] R. Schoenen and H. Yanikomeroglu, "User-in-the-loop: Spatial and temporal demand shaping for sustainable wireless networks," *IEEE Commun. Magazine*, vol. 52, no. 2, pp. 196–203, Feb. 2014.

[2] N. Sawyer and D. Smith, "Dynamic stackelberg user-in-the-loop repeated game for device-to-device communications." [Online]. Available: http://arxiv.org/abs/1603.00104

[3] M. Mirahsan, R. Schoenen, H. Yanikomeroglu, G. Senarath, and N.-D. Dao, "User-in-the-loop for hethetnets with backhaul capacity constraints," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 1536–1284, Oct. 2015.

[4] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, "A survey of smart data pricing: Past proposals, current plans, and future trends," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 1–37, Nov. 2013.

[5] "2013 global internet phenomena report," Sandvine Inc., Tech. Rep., 2013. [Online]. Available: http://www.sandvine.com/downloads/documents/

[6] "Cisco visual networking index: Global mobile data traffic forecast update, 2014–2019," Cisco Systems Inc., Tech. Rep., 2015. [Online]. Available: http://www.cisco.com/c/en/us/solutions

[7] M. Andrews, U. Ozen, M. Reiman, and Q. Wang, "Economic models of sponsored content in wireless networks with uncertain demand," in *Proc. IEEE International Conf. Computer Commun. (INFOCOM)*, Apr. 2013, pp. 3213–3218.

[8] Y. Jin and Z. Pang, "Smart data pricing: To share or not to share?" in *Proc. IEEE International Conf. Computer Commun. Wkshps (INFOCOM WKSHPS)*, Apr. 2014, pp. 583–588.

[9] R. Schoenen, G. Bulu, A. Mirtaheri, and H. Yanikomeroglu, "Green communications by demand shaping and User-in-the-Loop tariff-based control," in *Proc. IEEE Online Green Commun. Conf. (GreenCom)*, Sep. 2011, pp. 64 – 69.

[10] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, "Incentivizing time-shifting of data: A survey of time-dependent pricing for internet access," *IEEE Commun. Magazine*, vol. 50, no. 11, pp. 91–99, Nov. 2012.

[11] R. Schoenen, H. Yanikomeroglu, and B. Walke, "User-in-the-loop: Mobility aware users substantially boost spectral efficiency of cellular OFDMA systems," *IEEE Communications Letters*, vol. 15, no. 5, pp. 488–490, May 2011.

[12] S. Ha, S. Sen, C. Joe-Wong, Y. Im, and M. Chiang, "Tube: time-dependent pricing for mobile data," in *Proc. ACM SIGCOMM*, Aug. 2012, pp. 247–258.

[13] R. Schoenen, G. Bulu, A. Mirtaheri, T. Beitelmal, and H. Yanikomeroglu, "Quantified user behavior in user-in-the-loop spatially and demand controlled cellular systems," in *Proc. European Wireless Conf. (EW)*, Apr. 2012, pp. 1–8.

[14] R. Schoenen and H. Yanikomeroglu, "System and method for charging services using effective quanta units," Patent Number: 14/141,143; US2014/0 195 394 A1, WO2014/106 476 A1.

[15] C. Courcoubetis, F. Kelly, V. Siris, and R. Weber, "A study of simple usage-based charging schemes for broadband networks," *Telecommunication Systems*, vol. 15, no. 3, pp. 323–343, Dec. 2000.

[16] J. MacKie-Mason and H. Varian, "Pricing congestible network resources," *IEEE J. Select. Areas Commun.*, vol. 13, no. 7, pp. 1141–1149, Sep. 1995.

[17] L. A. DaSilva, "Pricing for QoS-enabled networks: A survey," *IEEE Commun. Surv. Tutorial*, vol. 3, no. 2, pp. 2–8, Nov. 2000.

[18] R. Schoenen and H. Yanikomeroglu, "Dynamic demand control with differentiated QoS in user-in-the-loop controlled cellular networks," in *Proc. IEEE Veh. Tech. Conf. (VTC-Spring)*, Jun. 2013, pp. 1–6.

[19] Z. Wang, R. Schoenen, H. Yanikomeroglu, and M. St-Hilaire, "Load balancing in cellular networks with user-in-the-loop: A spatial traffic shaping approach," in *Proc. IEEE International Conf. on Comm. (ICC)*, Jun. 2015, pp. 2638–2643.