

Network Neutrality: Avoiding the Extremes

Galina Schwartz, Nikhil Shetty, Jean Walrand
Department of Electrical Engineering and Computer Sciences,
University of California Berkeley, California - 94720
Email: {schwartz, nikhils, wlr}@eecs.berkeley.edu

Abstract—This paper addresses QoS provision in the presence of the regulatory threat of network neutrality. We formulate a technological implementation that enables QoS provision while requiring minimum alteration of the current network. Our proposal permits to avoid the political economic shock that would accompany any major network neutrality regulation. Specifically, we propose an α -network, in which an ISP is explicitly assured in its property rights to a certain fraction of its capacity; the leftover fraction of capacity should continue to function as it does in the existing network, where the risks of public regulations remain substantial. This proposal is based on the fact that current technology permits a reliable estimate of ISPs’ installed (and activated) capacity. The α -network creates better delineated property rights for the ISPs, which improve their incentives to (i) experiment with providing enhanced quality of service, (ii) introduce novel products which require packet differentiation, (iii) invest in capacity expansion. The α -network permits to explore the advantages of a non neutral network, while limiting its disadvantages to a small fraction of the existing network capacity.

I. INTRODUCTION

The network neutrality debate attracts considerable scholarly attention; the existing literature could be subdivided into three strains. The first group of papers focuses on the technological details [1], [2]. The second group of papers approaches network neutrality from a legal (and regulatory) perspective [3]–[7]. A summary of these ideas can be found in the Federal Trade Commission (FTC) report (2007) [8]. The third group of papers provides economic analysis. In turn, these papers could be subdivided into papers with a broad, empirically driven applied economic analysis¹ and papers providing formal economic modeling of network neutrality in a specific context [10]–[12].

On one hand, the imposition of restrictions on ISPs (with an aim to make the network neutral) is supported by major content providers and some public interests groups. On the other hand, many networking researchers and network equipment manufacturers actively argue that the current practice of “best effort service” with no explicit guarantees of packet delivery (latency, jitter, etc.) is an obstacle to network improvement. Clearly, some existing applications benefit from QoS guarantees, and new applications may become viable. But the introduction of QoS necessitates making the network non-neutral.

In short, the current debate about network neutrality embeds many conflicts: conflicting definitions of what constitutes a neutral network, conflicting interests of the involved

parties (content providers, transport providers and users whose needs differ), and conflicting technological avenues for the future Internet. For a collection of arguments in favor and against network neutrality (“pros and cons”), see [8].

By this time, the pros and cons of neutral and non neutral regimes are well known, but the magnitude of gains and losses associated with each regime is nearly impossible to evaluate: the data does not exist. The authors in [4] state that neither neutral nor non-neutral regime is socially optimal. They propose a *third way*: “that will enable the development of enhanced networks while at the same time ensuring a robust, open, best-efforts Internet.” They advocate a moderate approach. They speak of allowing network providers to deliver enhanced services. But, aiming to insure that providers do not abuse their market power, the authors suggest three groups of safeguards: “effective consumer protection measures, sound competition policy oversight, and conditioned tax incentives.” Although these safeguards affect technology, these effects are indirect.

This paper focuses on the technological side of the debate concerning the ISP’s rights and responsibilities with respect to its traffic. We approach the debate from the perspective of ISP property rights. We demonstrate that it is technologically possible to implement the network regime in which ISP property rights are set at any pre-specified (socially desirable) level. Our proposal is based on the fact that current technology permits reliable estimation of ISPs’ installed capacity. We do not attempt to determine how the pre-specified level should be set but we are convinced that in the existing Internet, this level is suboptimal [13].

This paper demonstrates the technological viability of an intermediate regime. Our technology-based proposal gives a regulator the means to achieve his objectives by imposing capacity-based constraints on ISPs. Specifically, we propose to divide the capacity of each ISP into two parts. We let a fraction $1-\alpha$ of an ISP’s capacity continue to operate as currently, and for the leftover fraction α , the ISP is granted ownership rights. Intuitively, $\alpha = 0$ corresponds to the existing network, and $\alpha = 1$ to a private network (i.e., fully secure property rights for the entire network capacity). Any $\alpha \in (0, 1)$ characterizes a mixed (or intermediate) network. The lower the α , the closer is the network to the current one. Thus, lower α reflects more restricted ISP ownership rights.

In the current network, there are no explicit rules imposing common carrier status on the ISPs (i.e., no α is directly imposed on ISPs). While no explicit α is currently imposed in

¹An excellent overview is provided in [9].

the network, the existing network regime tacitly corresponds to some implicit value of α . Perhaps, in today's Internet, α is not exactly 0, because, for example, ISPs interfere by delaying and blocking peer-to-peer traffic. Still, we believe that in today's network α is relatively close to 0, because the aforementioned practices are an exception rather than a commonplace. These practices are subject to discussion in press and public scrutiny. As a result, the ISPs are cautious – due to the threat of regulatory restrictions [1].

One could speculate about the value of α implied in the present network, but obviously, this value is not clearly defined. This reflects the current reality of unclear, poorly delineated property rights of ISPs, and negatively affects ISPs' incentives to invest in capacity expansion and development of new products and services. The regime that explicitly specifies α is advantageous because a declared α clarifies property rights allocation, which improves ISPs' investment incentives. In our view, the network neutrality debate should be more focused on determination of the socially optimal value for α , and less on the extremes. We demonstrate that capacity-based ownership rights are well-defined and easily enforceable. We believe that the proposed α -network, in which an ISP's property rights for α fraction of its capacity are explicitly secured, improves the ISP's investment incentives. The importance of ISP property rights for network functioning is expressed in economic [9] and legal [3] papers.

The rest of the paper is organized as follows. In Section II, we give an economic justification for our proposal, as developed in [13]. In Section II-A, we apply the model in [13] to demonstrate that, in the absence of regulation, a substantial fraction of the existing network users will suffer a loss if an ISP provides two service classes. In Sections III-A and III-B, we formulate our proposal of α -network. In Section IV-A, we look into the technical issues in implementing our proposal and describe the implementation for different access technologies. In Section IV-C, we address the enforcement and discuss the organization of inspection. Discussion and conclusions are presented in Section V.

II. ECONOMIC JUSTIFICATION

Our proposal aims to improve ISP incentives to develop and provide enhanced services. To demonstrate our idea, consider a marginal change of the existing network. Let the major fraction of the capacity operate under the status quo, but let us explicitly assure that each ISP utilizes some pre-specified small fraction (say 5%) of its capacity as it deems most profitable.

Effectively, such a rule endows an ISP with well defined property rights for 5% of its capacity. While no explicit restrictions are imposed on the remaining 95%, a 5% rule will enhance the perception of public good properties for this fraction of capacity. If our proposal is implemented, it will permit to generate data, which will improve our understanding of advantages of non neutral network, while limiting its disadvantages to a small fraction of the network (5%).

We believe that the proposed α -network has the following desirable properties. First, α -network avoids a major regulatory overhaul of imposing network neutrality. Second, it creates incentives for the ISPs to (a) experiment with providing enhanced quality of service, (b) introduce other novel products which require packet prioritization, (c) invest in capacity expansion.

In the next section, we apply the model developed in [13] to demonstrate that, in the absence of regulation, a substantial fraction of the existing network users suffer a loss if an ISP divides its capacity to provide two service classes. We suggest that the existence of such users creates a political economic obstacle to the transition. The proposed α -network permits to circumvent this obstacle.

A. X -model

In [13], we develop a game theoretic model, which permits us to investigate welfare effects and distributional consequences (i.e., how the gains and losses are distributed between the players) of the transition from a single service class to two service classes. Below, we outline the α -model, and apply our results to justify the social desirability of limiting the fraction of capacity that could be used for QoS provision.

Consider a single ISP (a monopolist) who offers connectivity to a user base of fixed size, and let N (which we assume to be large) be the total number of users in the user base. First, the ISP chooses his capacity C , which he builds at a constant unit cost τ . Investment in capacity is irreversible. Second, once the capacity is sunk, the ISP makes a pricing decision after which each user decides whether to adopt the service. The ISP's objective is to maximize his profit Π_{total} which equals his revenue net of his expense on capacity:

$$\Pi_{total} = \max_{C,p} \{pZ - \tau C\}, p, \tau > 0,$$

where Z is the number of users who adopt the service, and p is the ISP access price for users. If the ISP offers multiple service classes, and allocates a capacity C_i for the provision of service i at a price p_i , his objective becomes:

$$\Pi_{total} = \max_{C_i, p_i} \left\{ \sum_i p_i Z_i - \tau C \right\} \text{ and } C = \sum_i C_i,$$

where Z_i is the number of users who adopt service i .

We assume that, on average, each user sends an identical unit amount of traffic. We define the *quality of service* q observed by users as $q = 1 - Z/C$, if Z users are multiplexed in capacity C . This definition of quality reflects the common perception about service quality. As Z decreases, with capacity remaining the same, the quality of service improves, i.e., as the capacity per user increases, the quality increases as well. Further, we assume that each user in the user base is characterized by his type θ , where θ is a random variable uniformly distributed in $[0, 1]$. To characterize user demand, we make the following assumptions. For a user with type θ , the lowest *acceptable* service quality is $q = \theta$; and his highest *affordable* price is $p = \theta$. Thus, this user buys a

service only if this service is acceptable and affordable, i.e., $p < \theta \leq q$. For a user of type θ , if the quality of service $q \geq \theta$, his surplus is the difference between the price and his willingness to pay (which is θ itself). Thus, for this user, the surplus U_θ is given by

$$U_\theta = (\theta - p)I(q - \theta), \text{ where } I(y) = \begin{cases} 1 & \text{if } y \geq 0 \\ 0 & \text{if } y < 0 \end{cases} \quad (1)$$

θ could be intuitively viewed as the quality required by the applications that a user with type θ utilizes. Thus, in our model, user adoption is determined by the availability of the most quality intensive application that his type θ utilizes. Indeed, if a user adopts the network service for e-mail only, he gains no extra surplus from the fact that the actual network quality permits him to use streaming video (which he does not utilize). Since user types are uniformly distributed, from (1), the aggregate surplus U_{total} of all users in the user base is

$$U_{total} = \int_{\theta=0}^1 U_\theta N d\theta.$$

Let c denote capacity per user in the user base ($c = C/N$) and let z denote the fraction of users adopting the service of quality q ($z = Z/N$). Then, $Z/C = z/c$ and from our definition of service quality, we have $q = 1 - z/c$. Also,

$$\Pi = \frac{\Pi_{total}}{N} \text{ and } U = \frac{U_{total}}{N},$$

where Π is the profit and U is the surplus (per user in the base). Thus, the ISP objective and user surplus can be expressed as:

$$\Pi = \max_{c,p} \{pz - \tau c\} \text{ and } U = \int_{\theta=0}^1 U_\theta d\theta. \quad (2)$$

Per user in the base, the social surplus S is the sum of user surplus and provider profit:

$$S = U + \Pi.$$

We analyze the ISP's optimal choices for scenarios where the ISP provides (i) a single service, and (ii) two different services, possibly in the presence of a regulator who constrains the ISP. We assume that the regulator can only cap the fraction of capacity that the ISP allocates to the higher quality service (h). Since the regulator's choice variable is \mathbf{x} , the regulator only affects the ISP by constraining him from dedicating to service h more than a fraction \mathbf{x} of the ISP capacity. Notice that even in the presence of the regulator, the prices and capacities are chosen solely by the ISP. Then, the regulated ISP's profit maximization can be expressed as

$$\Pi = \max_{c_l, p_l, p_h} \{p_l z_l + p_h z_h - \tau c\} \text{ where } c_l \geq (1 - \mathbf{x})c \text{ and } c_h \leq \mathbf{x}c.$$

We consider three regulatory scenarios. Regulator 1 (a social planner) maximizes social surplus (sum of aggregate user surplus and ISP profit), regulator 2 maximizes user surplus and regulator 3 maximizes the surplus of the users served under a single service class (i.e., users with type $\theta \in (\theta_-^\dagger, \theta_+^\dagger)$).

For the regulators 1-3, the respective objectives S_1 , S_2 and S_3 are:

$$S_1 = \max_{\mathbf{x}} \{U + \Pi\}, S_2 = \max_{\mathbf{x}} U, S_3 = \max_{\mathbf{x}} U|_{\theta \in (\theta_-^\dagger, \theta_+^\dagger)}, \quad (3)$$

where U is defined in (2). Let \mathbf{x}_1 , \mathbf{x}_2 , \mathbf{x}_3 be the values of \mathbf{x} that the regulators 1-3 respectively choose to optimize their respective objectives and \mathbf{x}^\ddagger be the value chosen by the unregulated ISP. From (3), it is intuitive that:

$$\mathbf{x}_3 < \mathbf{x}_2 < \mathbf{x}_1 < \mathbf{x}^\ddagger.$$

Fig. 1(a) shows how regulators 1-3 choose \mathbf{x} depending upon the objective functions that they respectively maximize. The results of our simulations match our intuition.

On Fig. 1(a), we plot the provider profit and user surplus, and hence, the social surplus for any \mathbf{x} . Their values are higher with two service classes than with a single service class ($\mathbf{x} = 0$). This confirms that the transition to two service classes is socially desirable. Fig. 1(b) depicts how the ISP optimal prices vary with \mathbf{x} . From Fig. 1(b), we observe that as \mathbf{x} increases, prices of both service classes decrease. The prices affect the fraction of people adopting each service and the qualities of these services as well. From Fig. 1(b), the fraction of users adopting some service increases with \mathbf{x} . This reinforces the inference from Fig. 1(a) that the transition is socially desirable.

In spite of the social desirability, the transition to two service classes could be blocked due to the unfavorable distributional consequences that it inflicts on some fraction of current network users. In Fig. 1(c), we observe that, in the absence of regulation, a considerable fraction (almost 50%) of users in the current Internet will experience a surplus loss when the transition is implemented by the unregulated ISP. This fraction is so substantial, that these losing users may block such a transition. To facilitate this transition, we suggest the imposition of regulation which keeps the fraction of surplus-losing users sufficiently low. This will limit the capability of the losing users to block this transition.

Fig. 1(a) shows ISP profits, social surplus, user surplus and surplus of the existing users as a function of \mathbf{x} . We observe that, in the absence of regulation or with a regulator who maximizes social surplus, the transition to two service classes results in aggregate surplus loss for the existing users. Although, with a regulator who solely maximizes user surplus, the existing users do gain surplus, this surplus gain might be insufficient to make the transition politically feasible. This is especially true if the percentage of losing single service class users is high.

We believe that users who lose surplus as a result of transition are likely to resist the change. Hence, the regulator should assure feasibility by reducing the fraction of such losing users. In particular, we argue that the transition from a single service class to two service classes may be politically feasible only if the existing users gain, at least in aggregate. Hence, we consider a regulator who maximizes the surplus for the existing users only (regulator 3). Though he chooses a smaller value of \mathbf{x} , from Fig. 1(b), we observe that substantial

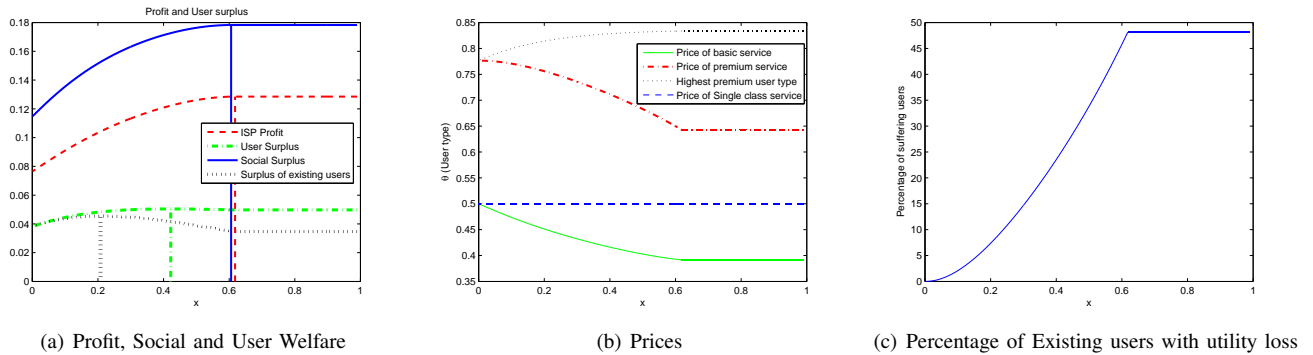


Fig. 1. Results

gains are obtained even if only a small fraction is reserved for the ISP. At the same time, the restriction imposed by regulator 3 permits us to overcome the political economic obstacles to the transition to two service classes.

Below, we sketch an outline for a technical implementation that enables this regulation.

III. TECHNICAL REQUIREMENTS

A. The Proposal

Let us define the following:

- (i) An *ISP* is an entity that provides commercial Internet access. Technically, this entity will be an autonomous system [14] that provides a transit service between networks. (This definition leaves out private networks.)
- (ii) All other entities connected to the Internet are defined as *users* of the network.
- (iii) A *link* is any transmission medium that is characterized by a well-defined information-carrying capacity.
- (iv) A *switch* is a network device that moves data from one link to another.
- (v) *Access links* are links that connect ISPs to *users* of the network. Links between ISPs are excluded from this definition of access links.

We propose the α -network, in which the ISP is explicitly assigned (by the government or the regulator) a fraction α of its installed capacity on each of the links in its network (later we show that *access links* can be excluded). Thus, our proposal explicitly secures an ISP property rights for a fraction α of its capacity. For this α fraction, ISPs acquire the six fundamental rights listed by [9]. We impose no formal requirements on the remaining fraction of capacity. The ownership rights for the remaining fraction α will continue as they are arranged in the existing network.

For the user, the end-to-end performance of her traffic matters. This includes factors like the end-to-end latency, jitter, bandwidth, etc. In our proposal, we suggest to control only the bandwidth seen by the user to the destination. This particular choice encapsulates the fact that ensured bandwidth in most cases also ensures that packets do not suffer delay. ISPs could choose to delay the neutral traffic and hence disrupt latency-sensitive traffic so as to generate demand for premium traffic from these applications. Theoretically,

today, nothing precludes ISPs from delivering differentiated premium services, using such delaying tactics. Practically, we do not widely observe ISPs introducing artificial delays to offer premium services². This can be explained by the uncertain regulatory environment. ISPs do not interfere with traffic due to the fear of a formal regulatory imposition of network neutrality. Our proposed network regime eliminates this uncertainty for α fraction of the capacity, thus further strengthening the communal rights for the remaining $1-\alpha$ fraction.

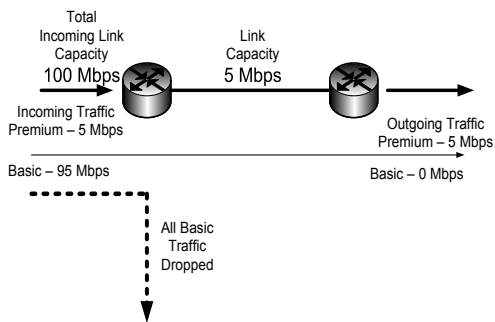
According to our proposal, an ISP could select to maintain the status quo, i.e., it does not alter its operations relative to the current mode. If an ISP chooses to use the newly delineated rights, these premium/enhanced services (i.e., traffic that travels over the ISP's "owned" channel) should amount to at most α fraction of the capacity.

In the following section, we put forward and justify the requirements in our proposal. We will demonstrate that all non-access links in the network must have α fraction of the capacity reserved for the neutral users to ensure that no manipulation is possible. Further, we will show that we can exclude the *access links* from this requirement and argue that doing so could be beneficial.

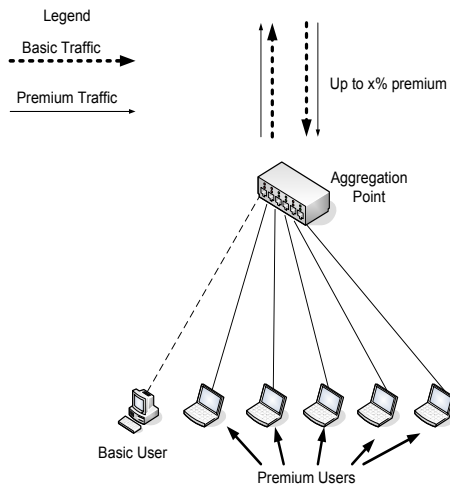
B. Discussion of the Proposed Rules

Consider the example in Fig. 2(a). The link shown in the figure is not an access link of the network and we assume that the regulation does not apply to it. The total capacity of links that bring traffic to this link is 100 Mbps. Assume that $\alpha = 0.95$; this results in a maximum of 5 Mbps of premium traffic. Since the unregulated internal link has a capacity of 5 Mbps, the ISP could choose to transmit the 5 Mbps premium traffic over this link and drop all the packets belonging to neutral traffic. Thus, this extreme example of complete blocking of neutral traffic demonstrates that a rule that is not enforced at even a single internal link can be manipulated by the ISP to support only premium traffic. Hence, we require the regulation to be mandated on all links inside the network.

²Some ISPs delay/block P2P traffic. However, this move appears to be in the interest of other consumers and has been adjudged as a healthy network management practice.



(a) Internal Link Example



(b) Access Link Example

Fig. 2. Examples

Next, we argue that access links can indeed be excluded from regulation without negatively impacting the neutral users. Also, ISPs' incentives remain similar to the situation where access links are covered by the regulation. On the other hand, we will show that excluding access links could indeed benefit some networked parties.

To assure these requirements, we impose that access to the aggregation point be ensured for each of the users. We will come back to this requirement later, when we discuss the various access technologies and show how this may be implemented. In Fig. 2(b), we see a worst-case scenario where all the users, except one, attached to the aggregation point are premium users, i.e., all their generated traffic is premium. In this example, we use a value of $x = 0.95$. In this example, though a large fraction of the traffic on the uplink from and downlink to the aggregation point (not including access links since they are excluded from regulation) is premium, the lone neutral user is protected. Since the amount of premium traffic is restricted to 5% of the capacity in both directions, this neutral user can send his traffic up on the remaining 95% or get his traffic on the remaining 95% of the capacity. Further, it is in the interest of the ISP to not block this user on purpose because he is a paying user and

in his absence, the remaining 95% would be wasted. From this, we conclude that access links can be excluded from the regulation.

If access links are also subject to capacity restriction, our regulation would be easier to formulate. But, regulatory burden would be heavier. Further, we outline two examples demonstrating the advantages of excluding access links from the regulation. First, consider a website/portal. If access links are subject to x -rule, this website cannot offer premium services exclusively. To provide premium services only, the website would have to buy access to a whole lot of bandwidth to maintain the fraction x requirement. Instead, by excluding the access links from the regulation, we provide ISPs the opportunity to aggregate premium and neutral traffic in the mandated fractions and save users the hassles of regulatory enforcement.

If access links were to be included, consumers using DSL would suffer too. Indeed, each DSL connection might be viewed as an access link in itself. But, this definition is restrictive because, then, it means that an end user cannot watch a premium streaming video (1 Mbps say), for example, without having an access bandwidth of 20 Mbps (if $x = 0.95$), which may not be technologically feasible. Hence, we believe that the exclusion of access links from the regulation could be advantageous.

IV. IMPLEMENTATION

Next, we consider enforcement measures to assure the ISPs' compliance with the x -rule. We consider two ways to carry out the enforcement:

- (i) *Hardware-restrained enforcement*: In hardware-restrained enforcement, switch manufacturers are responsible for ensuring that the mandated fraction is maintained. Hence, the hardware (switch) restrains the ISPs' capabilities to send more premium traffic than is mandated. This necessitates that the switches be certified to comply with the x -rule, and the ISPs are mandated to install these certified switches only. In addition, auditing the ISPs' network equipment is required to ensure that each ISP has certified equipment installed only.
- (ii) *Operational enforcement*: In operational enforcement, each switch is mandated to record the capacity usage by premium traffic on links where x -rule is imposed. This information must be available to inspectors whenever requested. Thus, the switches have to maintain genuine information about link usage and allow inspectors to access this information over the Internet (or physically³).

In Section IV-A, we describe further the technical issues for the enforcement options described above while, in Section IV-C, we address costs and benefits.

³We view the option of physical operational enforcement as unrealistic due to its prohibitive costs.

A. Technical Issues

The enforcement requires some restrictions on the switches' hardware/software and some architectural requirements, which we below.

1. Traffic Differentiation: Switches must possess the ability to distinguish between basic and premium traffic. To differentiate between basic and premium traffic, we suggest setting a bit to mark each packet of basic traffic. All ISPs must use this assigned bit in their header to represent the basic traffic. This way, we leave the flexibility to implement further traffic differentiation within the premium traffic to the ISPs. For e.g., within the premium traffic, an ISP may implement multiple service classes and these classes might differ between ISPs. If an ISP does not offer premium services (i.e., stays out of the α -rule), he is required to mark all traffic entering and exiting his network as basic.

2. Tamper-proof Counters: Enforcement of α -rule requires tamper-proof traffic statistics in the switches. Hence, we suggest to mandate that manufacturers provide switches with tamper-proof data counters. At present, network management services already require counters that maintain statistics of data flowing through the switch. In many switches, these statistics are maintained in the line card hardware and can be accessed by the software running on the central processor of the switch. Hence, making these counters tamper-proof requires a software update only (in most cases). Currently, software updates are performed routinely (to get rid of bugs), and tamper-proof counters may be implemented along with these regular updates.

Further, tracking the fraction of link capacity utilized by premium traffic can be accomplished using either exponential averaging or averaging over a sufficiently small time scale.⁴ Counter statistics must be recorded and presented upon the inspector request.

Items 3 and 4 outline the issues of communication between the inspectors and switches. Thus, items 3 and 4 are crucial for operational enforcement only.

3. Switch Accessibility: With SNMP-based monitoring, switches should be accessible to inspectors via the Internet. First, blocking inspector access to switches should be deemed illegal. Second, blocking traffic directed to and from the IP addresses of inspectors should be deemed illegal as well. Further, ISPs have to maintain the mapping between IP addresses and switches and release such information to the inspecting authorities.

4. Authentication and Data Encryption: With SNMP-based monitoring, switch authentication is required to assure that inspectors receive genuine data. Further, inspectors themselves must be authenticated to secure the ISP's switch data from unauthorized access. Hence, authentication is required in both directions. Preventing data intercept and manipulation on the path requires the data to be digitally signed by the switches, necessitating encryption support. The

⁴The time scale may not be too small as it results in lot of overhead. It cannot be too large as that can create fairness issues. A discussion on time scales or parameters for exponential averaging is beyond the scope of this paper.

type and implementation of this encryption is beyond the scope of this paper. As in item 2, a software update may be sufficient to provide data encryption support.

B. Example Technologies

Let us look at major current technologies and show that the α -rule is implementable. We make no attempt to provide a complete implementation. Indeed, we do not even touch upon the complexities of a functional implementation that necessarily includes protocol details and hardware interactions. In this paper, we address the implementation at a conceptual level only. In our related ongoing project, we are looking into these technical issues in depth.

Backbone: At the backbone, irrespective of the technology, the switches have to follow the α -rule. A software update will be sufficient for most switches to achieve compliance with the α -rule, see Section IV-A. Currently, most switches have management information collection implemented in hardware. We require these counters to be made tamper-proof and available to the inspectors.

Access: At the access, the α -rule must be implemented at the aggregation point of data from multiple users. Two issues arise in determining the switch that will be the appropriate aggregation point. The first issue is related to the discussion in Section III-B, where we state that all basic users need to have an assured connection to the aggregation point. An improper definition of the aggregation point opens the possibility for ISPs to circumvent the α -rule. The second issue is related to the switch-inspector communication, see items 3 and 4 in Section IV-A. Due to these communication requirements, the α -rule switches are required to run IP. To sum up, the aggregation points must be defined to have (i) an assured connection to users adopting basic access and (ii) an assured two-way inspector-switch communication via IP.

Below, we consider how to determine the aggregation point for major access-side technologies. We classify access-side technologies into two types: (i) Dedicated access and (ii) Shared access.

Dedicated Access: When the end-users connect through dedicated access links, the analysis in Section III-B holds. Such access links may be excluded from inspection, if the aggregation point is included. The following example will make this clear.

DSL:⁵ The typical DSL subscriber line uses a standard phone line to connect the digital subscriber (256 Kbps to 24,000 Kbps) to the local loop (or central office) [15]. There are about 100-120 DSL subscribers whose traffic is aggregated at the DSLAM (Digital Subscriber Line Access Multiplexer). Here, the incoming traffic lines are independent and do not affect each other. Hence, by ensuring that the internal links of the IP network connected to the DSLAM follow the α -rule, the α -rule can be implemented on a DSL dedicated access link, as shown in Fig. 2(b). From Section IV-A, the α -rule requires the DSLAM to maintain counters to assure that the premium traffic amounts to less than a fraction α of the

⁵Refers only to ADSL. VDSL may be shared access.

link capacity. Note that voice traffic (and increasingly video on the downlink) is also transmitted over the same copper wires. The voice traffic is transferred over to the PSTN at the DSLAM. Since networks are converging, we presume that the voice (and video) data could be combined with the Internet data for the α -rule. We do not argue for inclusion or exclusion of voice (and video) traffic into the α -rule and leave that as a future choice for policy makers. Whatever the adopted policy is, it must be uniform for all ISPs, irrespective of the access technology.

Shared Access: Shared access occurs when multiple end consumers share the bandwidth of the access link. Shared access itself may be of two types: Random and Centrally Scheduled.

Random access occurs when users are allowed to randomly access the channel to transmit their data. Then, if the channel accessibility rules (for e.g., the backoff parameter) are the same for all users, the end user is assured similar contention for a connection to the aggregation point. If premium users are provided preferential access (as in IEEE 802.11e [16]), then the aggregation point must assure access to the basic users by reserving a part of the bandwidth to them. This can be achieved by reserving a fraction of time on the channel for basic users to contend only (for e.g., see the point coordination function (PCF) in IEEE 802.11e [16]).

With centrally scheduled access, the aggregation point must ensure that basic users are assured access. To achieve this, the transmission slots for the premium users must be limited to α -fraction of the available slots. The following example will illustrate this technique.

Cable: In cable, 20 customers are connected to each distribution hub. The distribution hubs are in turn connected to the CMTS (Cable Modem Termination System). The communication between the user's cable modem and the CMTS employs the DOCSIS stack [17]. The CMTS generally handles about 5000 customers, with each linecard handling about 1000-1200 (recommended by Cisco [18]). Technically, the distribution hub is the aggregation point, since the connection to the distribution hub is assured. However, the hub has limited control of transmission over the cable (and does not run IP). Hence, we suggest the CMTS as the aggregation point.

In DOCSIS, the CMTS informs all the Cable Modems (CMs) about the time slots in which they are allowed to transmit. This implies that the CMTS is in complete control of access to both the uplink and the downlink on the cable network. Thus, the CMTS has to be certified/monitored to ensure that all basic users are assured a connection. The premium traffic must be limited to α -fraction of the capacity. The CMTS runs IP and thus, it can communicate with inspectors' machines, possibly enabling operational enforcement. The policy of inclusion of voice and video (cable TV) traffic, as mentioned for DSL, should be consistent.

The case of wireless broadband is even more complicated. In wireless broadband, the utilized spectrum is licensed to network operators via auctions. One could argue that

property rights have already been assigned to these operators through the rules imposed by the FCC. Presumably, the FCC is well aware that wireless broadband competes with existing access technologies. Indeed, the proposed rules create ownership rights for wireless broadband providers on par with the rights for cable and DSL providers. The rules for the new spectrum auctions by FCC in the 700 MHz range seem to corroborate this opinion [19]. Hence, one can contemplate the implementation of the α -rule for wireless broadband.

Cellular Internet: Since mobile stations share the spectral bandwidth and base stations provide mobile stations with transmission opportunities, conceptually, the base station is similar to the CMTS in cable. Base stations must be certified for enforcement purposes. However, operational enforcement may not be possible as base stations do not run IP. In the wireless broadband hierarchy, the Base Station Controller (BSC) aggregates traffic from a number of base stations and traffic from these BSCs is further aggregated at the level of the Mobile Switching Center (MSC) [20]. For example, in GPRS, the BSCs have a connection (independent of the MSC) to the IP backbone. Hence, for operational enforcement, it should be the BSCs, and not the base stations, that restrict premium traffic to the mandated fraction (*to and from* each of the base stations).

To summarize, we have classified the access technologies and given examples of most common technologies (DSL, Cable, Cellular Internet) to demonstrate how the aggregation point for the α -rule can be determined for them. Any emerging technology can be fit into one of these categories. A popular novel technology such as Passive Optical Networks (PONs) can be classified as a dedicated access technology (TDM-PON), while future developments in PONs (such as DOCSIS-PON) could be viewed as shared access technologies [21]. Among future wireless technologies, Worldwide Interoperability for Microwave Access (WiMAX) resembles Cable, and can be classified as a shared access technology [22].

C. Costs and Benefits

Below, we address the pros and cons of the considered enforcement options.

The benefit of hardware-restrained enforcement is that no regular audits of ISP equipment are necessary. But, this benefit comes with an associated disadvantage. Although the switches are certified, during actual operation, they may malfunction and the α -rule may be violated. Such violations may go undetected for days, especially when the ISPs lack incentives to arrange for appropriate repairs.

In operational enforcement, the complication of non-detection does not arise because of regular audits. But, in this case, the inspectors must have online access to all the switches in the ISP's network. The complication of such monitoring is due to the ISPs' reluctance to reveal information about their network structure for competitive reasons.

We do not assess the monetary costs of these enforcement options explicitly. But, the hardware-restrained option appears to be relatively less costly than the operational option. Indeed, both options will require hardware certification of some sort, the costs of which should be similar. But, clearly, operational option entails higher monitoring expenses than the expenses of auditing required for the hardware-restrained option. Hence, we suggest that hardware-restrained option is preferable on cost grounds.

V. DISCUSSION AND CONCLUSION

The focal feature of networking technology that is crucial for viability of our proposal is the ease and reliability with which ISP capacity can be measured. Since capacity is easy to measure, a property regime in which property rights are based on capacity is well defined.

Under our proposal, ISP property rights are secured for a pre-specified fraction of its capacity. The leftover fraction of capacity is assumed to function as in the present Internet. Our numerical analysis demonstrates that, in the absence of regulation, a substantial fraction (see Fig. 1(c)) of the existing network users will suffer a loss if the ISPs divide their capacity to provide two service classes. We suggest that existence of such users creates a political economic obstacle to the transition. The proposed α -network permits to circumvent this obstacle because regulation limits the fraction of losing users.

We believe that our proposal balances the interests of ISPs and other networked parties. Theoretically, now, the ISPs could not only use a mere fraction of their capacity but the entire capacity for enhanced services. Practically, under the threat of regulation, ISPs do not provide any enhanced services. On one hand, under our proposal, for the pre-specified fraction of capacity assigned to the ISPs by the regulator, their rights become more secure than they currently are. On the other hand, our proposal suggests that the regulator guarantees secure rights for this fraction of capacity only. Although not formally restricted, in practice, their rights for the leftover fraction of capacity would be weakened.

Simplicity of implementation, its flexibility and ease of enforcement are clear advantages of our proposal. Though our proposal requires certain enforcement expenses, these expenses are minor relative to the prospect of imposing network neutrality, for which the enforcement, as it is widely held, is prohibitively costly or plain unfeasible. Indeed, to ensure network neutrality, enforcers must be able to prove conclusively that the ISPs are not tampering with data delivery. If ISP tampering is selective, end-to-end measurements would be required for all applications to ensure that no discrimination occurs. Due to the complexity of this task, enforcement of network neutrality would not only require all the measures in our proposal, but also far more on top.

If our proposal is adopted, it would have an effect of creating an experimental (or test) slice for a non-neutral premium network. Such a slice permits to generate actual

data, thus enabling more precise analysis of actual effects of a non-neutral network regime.

REFERENCES

- [1] E. Felten, "Nuts and Bolts of Network Neutrality," Working Paper, <http://itpolicy.princeton.edu/pub/neutralty.pdf>, Center for Information Technology Policy, 2006.
- [2] J. Crowcroft, "Net Neutrality: The Technical Side of the Debate: A White Paper," *SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 1, pp. 49–56, 2007.
- [3] B. M. Owen and G. L. Rosston, "Local Broadband Access: A Property Rights Approach," *SSRN eLibrary*, 2003. [Online]. Available: <http://ssrn.com/paper=431620>
- [4] R. D. Atkinson and P. Weiser, "A 'Third Way' on Network Neutrality," *SSRN eLibrary*, 2006. [Online]. Available: <http://ssrn.com/paper=1004522>
- [5] T. Wu and C. S. Yoo, "Keeping the Internet Neutral?: Tim Wu and Christopher Yoo Debate," *Vanderbilt Public Law Research Paper No. 06-27*. [Online]. Available: <http://ssrn.com/paper=953989>
- [6] R. Frieden, "Network Neutrality or Bias? - Handicapping the Odds for a Tiered and Branded Internet," *SSRN eLibrary*, 2006. [Online]. Available: <http://ssrn.com/paper=893649>
- [7] C. S. Yoo, "Network Neutrality and the Economics of Congestion," *Georgetown Law Journal*, vol. 94, Jun. 2006. [Online]. Available: <http://ssrn.com/paper=825669>
- [8] (Jun.) Broadband Connectivity Competition Policy. Federal Trade Commission. [Online]. Available: <http://www.ftc.gov/reports/broadband/v070000report.pdf>
- [9] N. Economides and J. Tag, "Net Neutrality on the Internet: A Two-sided Market Analysis," Working Papers, 2007. [Online]. Available: <http://ideas.repec.org/p/ste/nystbu/07-27.html>
- [10] B. E. Hermalin and M. L. Katz, "The Economics of Product-Line Restrictions With an Application to the Network Neutrality Debate," Jul. 2006.
- [11] J. Musacchio, G. Schwartz, and J. Walrand, "A Two-Sided Market Analysis of Provider Investment Incentives With an Application to the Net-Neutrality Issue," *Review of Network Economics*, 2008. [Online]. Available: <http://robotics.eecs.berkeley.edu/~wlr/Papers/neutralty.pdf>
- [12] G. Schwartz, N. Shetty, and J. Walrand, "Impact of QoS on Internet User Welfare," in *4th International Workshop On Internet And Network Economics (WINE)*, 2008. [Online]. Available: <http://www.eecs.berkeley.edu/~nikhils/PublishedPapers/SSW-WINE08.pdf>
- [13] Y. Rekhter, T. Li, and S. Hares, "A Border Gateway Protocol 4 (BGP-4)," RFC 4271, Jan. 2006. [Online]. Available: <http://www.ietf.org/rfc/rfc4271.txt>
- [14] (1999, Jun.) Digital Subscriber Line. [Online]. Available: http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/adsl.htm
- [15] "Providing QoS in WLANs: How the IEEE 802.11e Standard QoS Enhancements Will Affect the Performance of WLANs," Intel White Paper, 2004. [Online]. Available: http://www.intel.com/network/connectivity/resources/doc_library/white_papers/30376201.pdf
- [16] *DOCSIS Specifications - DOCSIS 1.1 Interface*, CableLabs Std. [Online]. Available: <http://www.cablelabs.com/specifications/doc11.html>
- [17] What is the Maximum Number of Users per CMTS? [Online]. Available: http://www.cisco.com/warp/public/109/max_number_cmts.html
- [18] FCC Revises 700 MHz rules to advance interoperable public safety communications and promote wireless broadband deployment. [Online]. Available: http://fjallfoss.fcc.gov/edocs_public/attachmatch/DOC-275669A1.pdf
- [19] T. Farely and K. Schmidt. (2006, Jan.) Base Station Subsystem. [Online]. Available: http://www.privatelinet.com/mt_gsmhistory/04_architecture_of_the_gsm_network/ii_base_station_subsystem/
- [20] P. Ossieur, X. Qiu, J. Bauwelinck, D. Verhulst, Y. Martens, J. Vandewege, and B. Stubbe, "An Overview of Passive Optical Networks," *International Symposium on Signals, Circuits and Systems, 2003. SCS 2003.*, vol. 1, pp. 113–116, July 2003.
- [21] Mobile WiMAX - Part 1: A Technical Overview and Performance Evaluation. Wimax Forum. [Online]. Available: http://www.wimaxforum.org/technology/downloads/Mobile_WiMAX_Part1_Overview_and_Performance.pdf