Broadband Fast-Lanes with Two-Sided Control: Design, Evaluation and Economics

Hassan Habibi Gharakheili†, Vijay Sivaraman†, Arun Vishwanath*, Luke Exton†, John Matthews*, Craig Russell*
†University of New South Wales, *IBM Research-Australia, *CSIRO, Sydney, Australia
Emails: {h.habibi, vijay, luke.exton}@unsw.edu.au, arvishwa@au.ibm.com, {john.matthews, craig.russell}@csiro.au

***Abstract*—Enhancing quality-of-service (QoS) for specific traffic streams by assigning them to "fast-lanes" on the broadband Internet service is a subject of intense ongoing debate. While Internet Service Providers (ISPs) have clear economic imperatives for fast-lanes paid by content service providers (CSPs), proponents of net-neutrality argue that consumer interest will be ignored in the selection of traffic thus prioritized. In this paper we propose a new solution in which ISP fast-lanes have "two-sided" control, i.e. by both consumers and CSPs. Our contributions are two-fold: (1) We develop an architecture in which ISP-operated fast-lanes can be controlled at fine-grain (per-flow) by the CSP and at coarse-grain (per-device) by the consumer, and argue why we think such an architecture can meet the needs of all three parties; and (2) We develop an economic model to guide the ISP in determining fast-lane allocation that balances the needs of the CSP against those of the consumer, and evaluate our model via simulation of trace data comprising over 10 million flows.**

I. INTRODUCTION

The notion of Internet "fast-lanes", whereby certain traffic is given higher priority over others, has been gaining increased traction over the past year [12], [14], spurred by the revelation that Netflix's paid-peering arrangement with Comcast in early 2014 led to significant improvement in Netflix performance for Comcast subscribers [7]. Internet Service Providers (ISPs) argue in favor of fast-lanes as an economic imperative to fund maintenance and upgrade of their access network to cope with growing traffic volumes, without putting undue financial burden on consumers. However, policy-makers and activists are circumspect that such deal-making between ISPs and content service providers (CSPs) can be detrimental to the best interests of the consumer, who is not consulted in the selection of traffic streams that get access to the fast-lanes.

There are surprisingly few proposals that try to bring CSPs and end-users (i.e. consumers) into the fast-lane negotiation. In Oct 2014 AT&T proposed fast-lanes that are controlled by end-users [3]; the proposal unfortunately reveals little technical or business detail, and it remains unclear what interfaces will be exposed to users and how these will be priced. Our proposal in [13], supported by some economic modeling in [8], develops APIs by which the CSP can dynamically request fast-lane creation from the ISP at run-time; this gives per-flow control to the CSP without having to enter into bulk-billed peering arrangements with the ISP. While our prior work does not provide much control (other than an opt-in/out button) to the end-user to control the fast-lanes, in this paper, we seek to fill this important gap by developing and evaluating an architecture that allows both the end-user and the CSP to create, dimension, and use broadband fast-lanes.

The challenges in developing a two-sided fast-lane architecture are manifold: (a) End-users and CSPs will often have different motives for traffic prioritization, leading to conflicts whose resolution needs to be customized per-user based on their desires; (b) Users typically have much lower technical sophistication than CSPs, so the interfaces for control have to be quite different at the two ends; (c) The economic capacity of the two ends is again quite different, with the CSP expected to bear the cost of the fast-lane, but the end-user still being given some means of control over it. Any solution has to take the above sensitivities into account, and yet be attractive to all parties from an economic and performance point-of-view.

In this paper we attempt to develop a new architecture that addresses the above challenges. We begin by devising appropriate APIs that are suitable for the two ends of the fast-lanes, and argue that they are realizable using emerging software defined networking (SDN) technology. We then address the economic aspect of two-sided fast-lanes by devising a model that captures the trade-off between end-user and CSP happiness, and providing the ISP with means to control this trade-off. We evaluate our model using simulation with trace data of over 10 million flows taken from an enterprise network.

The rest of this paper is organized as follows: §II summarizes relevant prior work; §III describes our two-sided fast-lane system architecture and APIs. In §IV we develop a model that captures the economic gains of fast-lanes, and §V evaluates it using real trace data. The paper is concluded in §VI.

II. RELATED WORK

Recent SDN-based approaches have proposed various frameworks to control service quality: APIs have been developed in [2] to allow applications to dynamically interact with the network and set QoS configurations. The work in [13] develops APIs for a content provider to dynamically negotiate QoS with the ISP. However, none of these APIs specifically target home networks and deal with consumer interfaces. In the context of home networks, [15] presents interfaces and apps (similar in spirit to ours) to allow the user to interact with the underlying network to control service quality for different applications, and [9] develops a client hosted application for QoE control. While all the above works are relevant, we distinguish our work in this paper by considering

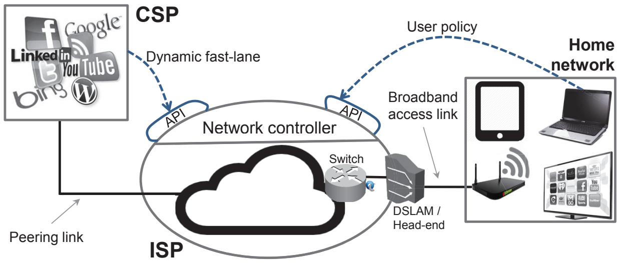


Fig. 1. A typical broadband access network topology.

two-sided control in which both the end-user and the CSP simultaneously exert influence over traffic prioritization, and develop an economic model to support it.

Several different pricing models by ISPs, termed smart data pricing (SDP), have been proposed in the literature, ranging from models for pricing only the end-users [10], [11] to two-sided pricing [4], [5], i.e. pricing both end-users and CSPs. These models consider usage-, time-of-day- and congestion-based pricing to affect user activity (for e.g. deterring usage by charging users more during peak hours than off-peak hours), or propose (semi-)static payment arrangements between ISP and CSP to increase their utility. By contrast, our model does not aim to charge the end-user or affect change in user behaviour, and prices dynamic fast-lanes (at a per-flow level) initiated by the CSP, as opposed to today's (semi-)static payment models between the ISP and CSP.

III. TWO-SIDED FAST-LANE SYSTEM ARCHITECTURE

Consider a representative broadband access network topology shown in Fig. 1. As prevalent today, each household consists of a variety of devices (e.g. laptops, smart phones, tablets, smart TVs, etc.) connecting to the wireless home gateway, which offers broadband Internet connectivity via the DSLAM at the ISP's local exchange. The ISP peers directly with a number of CSPs (such as YouTube, Hulu, and Netflix) or indirectly via CDNs (such as Akamai) and other ISPs. In our proposed architecture, the DSLAM is connected to an SDN Ethernet switch (e.g. OpenFlow switch) which in turn connects to the ISP's backhaul network providing access to the global Internet. The SDN switch is controlled by an SDN controller which is housed within the ISP's network and exposes the APIs to be called – by both the end-user and the CSPs – for the creation of fast-lanes.

A. End-user facing APIs

Consider a family of four living in a household – the father uses his laptop at home for various work-related activities such as video-conferencing and Skyping, the mother uses a smart TV to watch shows or movies (e.g. Internet-TV), the son uses his laptop for gaming and watching videos on YouTube, and the daughter uses her tablet to spend time on Facebook and browse the Internet. To ensure that the users in the household get the required QoS, we permit the subscriber (e.g. the father) to configure a *minimum* bandwidth (on the broadband access link from the ISP to the household) that he deems is necessary for each of the devices in the household. An example of such a configuration could be: 40% of the broadband capacity is assured to the father's laptop, 30% to the smart TV, 15% to the son's laptop, 10% to the daughter's tablet, and 5% for the remaining devices in the house. The key tenets of this approach are as follows:

- *Device-level control:* We have intentionally chosen to configure bandwidth partitions at a device-level, rather than at a service-level (e.g. YouTube, Netflix, Skype, etc.) or flow-level (e.g. specific Skype call or video session). Flow-level control is too onerous for the user, requiring them to interact with the user-interface to configure fast-lane access rights for every session. Service-level control may seem easier to conceive, for example a subscriber could say that Netflix traffic is to be prioritized over BitTorrent. However, we feel that this approach does not capture the fact that the importance of a service often depends on the user within the household accessing it - for example YouTube/Netflix may be more important if the father or mother is accessing it, but less so if the son/daughter is doing so; moreover, it runs the risk that subscribers will strongly favor established content providers (YouTube, Netflix) over smaller lesser-known ones. We therefore believe that device-level bandwidth control is more in line with the subscriber's view on how bandwidth should be shared within the household. Of course device-level control can be combined with service-level control (e.g. give some bandwidth to Skype on the father's laptop), but this requires more configuration on the subscriber's part (cross-product of devices and services), and does not add much value.
- *Single parameter:* We have intentionally chosen the APIs to have only a single control knob (i.e. the minimum bandwidth) because a vast-majority of end-users lack the sophistication to configure a multiplicity of parameters. A single, but intuitive, parameter reduces the barrier for end-users to adopt fast-lanes for improved QoS, and gives them control over it, which has hitherto remained elusive.
- *Proactive approach:* The crux of the QoS problem in a residential setting is bandwidth sharing amongst several household devices. To combat this problem, we advocate a set and forget device centric QoS policy, but leave the door open for end-users to seek additional bandwidth (i.e. create fast-lanes) as and when necessary (e.g. when QoS/QoE is poor) for the duration of a traffic stream.

B. Content Service Provider facing APIs

The APIs exposed by the ISP to a CSP allow the latter to reserve access-link bandwidth at a per-flow level. There are several reasons why we believe such fine-grained control is the most appropriate for CSPs:

- *Economics:* Instead of paying in bulk for all the traffic they are sending via the ISP, the CSPs can exercise discretion in selecting the subset of flows for which they call the bandwidth reservation API into the ISP. For example, they may choose to reserve bandwidth only when there is congestion, or only for certain premium customer traffic. The important point here is that the per-flow API allows the CSP to make dynamic decisions on fast-lane usage, allowing them to align it with their own business models.

- *Control*: Unlike end-users, CSPs have the technical expertise to conduct per-flow negotiations on fast-lane access and the associated pricing, and are indeed expected to have automated their algorithms for doing so. This gives them the flexibility to account for various factors (time-of-day, user net-value, etc.) in making dynamic fast-lane decisions to maximize their returns.
- *Reactive approach:* The CSPs are not obliged to call the API every time a flow request is received from the end-user. Instead, it is left to the discretion of the CSP; the API can be called in a reactive manner (i.e. dynamically) such as when the QoS/QoE of the traffic flow is unsatisfactory.

The API itself for per-flow bandwidth reservation is relatively simple, and specifies the following attributes (much like in [13]): *CSP id*, the identity of the CSP making the request; *Flow tuple*, denotes the IP address and port number of the source and destination, and the transport protocol; *Bandwidth*, the minimum bandwidth that the flow, such as a YouTube video, requires; and *Duration*, the duration for which the bandwidth is requested.

C. Challenges with two-sided control

When an ISP receives a request for creating fast-lanes from the CSPs and/or end-users, the ISP has to decide whether or not to instantiate the fast-lane. On the one hand, satisfying all fast-lane requests from CSPs will generate greater revenue for the ISP, because the CSP pays the ISP for the creation of fast-lanes. On the other hand, creating a dynamic fast-lane for the CSP may violate the minimum bandwidth fast-lanes set by the end-users for their specific devices, causing annoyance to the user and potentially leading to consumer churn. The ISP therefore has to balance the revenue benefits from the CSP against the risk of subscriber dissatisfaction whenever the fast-lane configurations from the two ends conflict.

Consider the following possible scenario: seeing that the video quality of a YouTube stream on the daughter's iPad is not adequate, YouTube calls the API into the ISP network to create a fast-lane for this stream on this subscriber's broadband link. This presents an opportunity to the ISP to charge the CSP for the dynamic fast-lane. However, suppose the bandwidth requested by the YouTube stream is not currently available because the father is doing a Skype session. The ISP then has to decide whether to let YouTube access the fast-lane, in violation of the father's policy that his laptop gets a higher bandwidth share than the daughter's iPad, thereby causing subscriber frustration, or instead to just deny YouTube the requested bandwidth, thereby foregoing the revenue opportunity. Making the appropriate decision requires a cost-benefit analysis by the ISP, for which we develop an economic model in the next section.

We would like to point out that the challenges associated with two-sided control of fast-lanes is not just about resolving the policy conflicts. Indeed, there are existing frameworks (e.g. PANE [2]) that explore various techniques for conflict resolution. Our objective is to evaluate the underlying economic and performance incentives that influence how the conflicts get resolved in this fast-lane architecture with two-sided control.

IV. DYNAMIC NEGOTIATION AND ECONOMIC MODEL

We now present the dynamics of fast-lane creation, and develop an economic model to aid the ISP in making admission decisions that balances the user's needs with the CSP's.

A. Dynamic Negotiation Framework

Broadband fast-lanes are created via two sets of API calls: (a) relatively static policies configured by the end-user that establish per-device fast-lanes, and (b) dynamic API calls coming from the CSP for establishment of per-flow fast-lanes. We assume that the user-facing APIs do not generate revenue, and are given free-of-charge to the end-user. API calls from the CSP are however revenue-generating, with the per-flow fast-lane being associated with a micro-payment dependent on the size and duration of the flow (detailed model to follow). Further, the CSP's request for fast-lane may conflict with the user-set preferences, such as when the bandwidth requested for a video streaming flow exceeds the user-set bandwidth portion for the specific client device. The ISP is still permitted to accept the CSP call, thereby generating revenue; however this leads to violation of the user-set preferences, which can lead to user annoyance – in what follows we will assign a monetary cost to this annoyance by mapping it to a churn probability and consequent loss of revenue for the ISP.

The decision to invoke a dynamic fast-lane via the API call is entirely up to the CSP. The CSP could choose to invoke it for every video stream, or more realistically, when network conditions and/or user importance make bandwidth reservation beneficial. The CSP may even involve the user in this decision, say by embedding a "boost" button in the application that the user can press to trigger fast-lane creation to enhance QoS for this stream (such boosting capability may entail extra payment from the user to the CSP, which could partly or wholly support the cost of the fast-lane API invocation). The ISP charges the CSP each time a call from the latter is admitted. Though the ISP may choose to accept or reject the CSPs fast-lane request, we assume that if accepted, the allocation commitment is maintained over the duration of the flow (indicated in the API call from the CSP) and not modified mid-stream.

The ISP's dilemma on whether or not to accept the CSP's dynamic fast-lane request is illustrated with a simple example: Suppose a dynamic fast-lane of 2 Mbps is requested for a YouTube HD stream to be delivered to daughter's tablet, and further that the father has configured a static fast-lane of only 1 Mbps for that device. If the fast-lane call is accepted, and the daughter's video stream given 2 Mbps, it is likely that other devices that are concurrently online in the house get a lower bandwidth share than configured – this could, for example, cause poor video-conferencing performance on the father's laptop, causing him annoyance even though he had set a higher bandwidth fraction for his device.

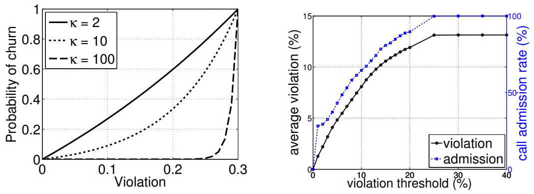
To quantify this user annoyance, we track "violation" metric v, measured as the total shortage of minimum rate across all devices, normalized by the total capacity of the broadband link. For example, in the situation explained above, the shortage of 1 Mbps on father's laptop contributes to 10% of violation, for broadband link capacity of 10 Mbps. We keep track of this violation measure over time via exponential averaging – it rises whenever the ISP accepts CSP API calls for fast-lanes that violate user-set fast-lane preferences, and falls when the ISP rejects such calls from the CSP. Based on this measure, we propose a simple algorithm that the ISP can use to make call admission decisions: for a specific user, the ISP uses a target threshold (v_{th}) to cap the violations, and a call from the CSP is admitted if and only if the current violation measure v is below the threshold v_{th}. It is easy to see that an ISP that never wants to violate the user preference can choose $v_{th} = 0$, whereas an ISP that wants to accept every API call from the CSP irrespective of user preferences chooses $v_{th} = 1$. In general, an ISP could choose an intermediate value, say $v_{th} = 0.2$, that accepts CSP-side fast-lane requests that maintain user-side violations at this acceptable level.

We now attempt to convert the user-preference violation metric above into a measure of damage incurred by the ISP. Prior observations in [1], [6] show that QoE-decay is tightly bound to user-engagement and subscriber churn, though the relationship is not easy to capture mathematically in a succinct way. We resort to a simplified mathematical expression in which the user's probability of churn (i.e. of changing ISP) at the end of the billing period is an exponentially increasing function of the violation measure, given by:

$$P_{churn} = \frac{e^{\kappa v} - 1}{e^{\kappa v_0} - 1} \tag{1}$$

Here P_{churn} denotes the user's churn probability, κ in the exponent corresponds to the user's level of flexibility (discussed below), v_0 denotes the maximum tolerable violation at which the user will undoubtedly leave, and $v \in [0, v_0]$ is the measure of actual violation (computed by the ISP using an exponential moving average). The expression is chosen so that the two end-points $v = 0$ and $v = v_0$ correspond to $P = 0$ and $P = 1$ respectively. Fig. 2(a) depicts the curve for churn probability with three value of $\kappa = 2, 10, 100$ corresponding to increasing levels of user flexibility: at a given violation, churn will less likely occur with a larger κ. The user-flexibility parameter κ can either be explicitly solicited from the user, or learnt by the ISP based on user behavior. Further, the ISP can give users financial incentives to choose a larger κ, since this allows the ISP to make more revenue from CSPs by accepting their fast-lane API calls; however, discussion of such financial incentives is out of the scope of the current paper.

B. Economic Model



(a) Model (b) Simulation result

Fig. 2. (a) Churn probability; (b) Violation of user demands

The fast-lane service offering is free for users, but paid for by the CSP. The pricing structure we employ for dynamic fast-lanes is one in which the cost of the resource changes as a continuous function of its availability. A convenient and commonly used such function is the exponential [8], wherein the unit price of bandwidth is a function of spare capacity available on the broadband access link. The bandwidth cost is therefore set high when the spare capacity (link rate minus load) is low, and we assume it to fall exponentially as the spare capacity increases, expressed by:

$$C = \lambda e^{-\delta x}, \tag{2}$$

where C is the spot cost of bandwidth (i.e. for 1 Mbps over a 1-second interval), x is the variable denoting fraction of available link capacity (computed by the ISP using an exponential moving average), λ is a constant corresponding to the peak spot-price (we use $\lambda = 1, 1.5$ cents-per-Mbps-per-sec in our simulations), and δ is a constant corresponding to the rate at which the spot price of bandwidth falls with available capacity x. Our simulations will employ bandwidth pricing with $\delta = 2$.

Shifting focus to the user-side, the violation of their per-device fast-lane policies by virtue of dynamic fast-lane creation for CSPs will cause annoyance to the subscriber; to capture the economic cost of this, we associate such annoyance with churn, i.e. the user's likelihood of changing ISPs, leading to loss of revenue for the ISP. The ISP's (monthly) change in revenue from fast-lanes will therefore equal the revenue generated from admission of CSP calls, minus the revenue lost from user churn, denoted mathematically as:

$$\sum_k (C.f_k^{rate}.f_k^{duration}) - S.P_{churn}, \tag{3}$$

where f_k^{rate} and $f_k^{duration}$ are the rate (in Mbps) and length (in seconds) respectively for the k-th fast-lane admitted by the ISP. These are multiplied with the spot price C (in dollars-per-Mbps-per-sec) of unit bandwidth (following congestion-based pricing given in Eq. (2), and summed over all calls k admitted over the month; S is the subscription fee (in dollars-per-household per month), and is multiplied by churn probability P_{churn} to derive the loss in revenue from subscribers. Our simulations will use $S = \$60$ for a broadband service of 10 Mbps, consistent with the typical price for a 10 Mbps broadband link in most developed countries.

The objective for the ISP is to operate the fast-lanes in a way that maximizes profit in Eq. (3), by tuning the violation threshold parameter v_{th}: a larger v_{th} allows the ISP to admit more CSP calls (generating revenue), but amplifies user frustration leading to elevated churn probability (with consequent revenue loss): this trade-off, and the various parameters that affect it, are studied via simulation of real trace data next.

V. SIMULATION EVALUATION AND RESULTS

We now evaluate the efficacy of our proposal by applying it to a 12-hour trace comprising over 10 million flows taken from an enterprise campus network. We focus on how two critical parameters – violation threshold v_{th} chosen by the ISP, and user-churn probability exponent κ – influence revenues for the ISP and performance benefits for all the parties involved.

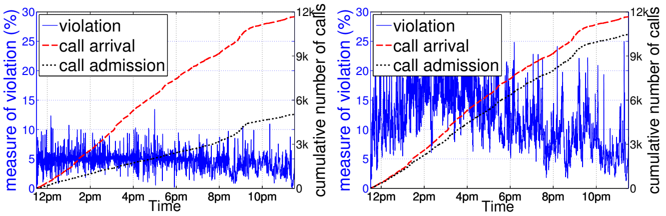
A. Simulation Trace Data

Our flow-level trace data was taken from a campus web cache, spanning a 12 hour period (12pm-12am). Each entry consists of flow attributes such as arrival date and time, duration (in milliseconds), volume of traffic (in bytes) in each direction, the URL, and the content type (video, text, image). The log contains 10.78 million flow records corresponding to 3300 unique end-user clients. Of these flows, $11,674$ were video flows (predominantly from YouTube, identified by the content type field), $9,799$ were elephant flows (defined as transfers of size greater than 1 MB), and the remaining 10.76 million flows were mice (defined as transfers of size 1 MB or less, representative of web pages). Though mice flows dominate by number, three flow types contribute roughly equally by volume (32%, 32% and 36% respectively) to the total traffic downloaded. We found that; 98% of video flows required less than 5 Mbps, and only 0.2% of the flows required more than 10 Mbps, and in terms of duration; 90% of the video flows last under 3 minutes, and only 1% of the flows last for longer than 10 minutes.

B. Simulation Methodology

We developed a native simulation that reads the flows information (arrival time, duration, type, rate/volume) and injects them into the slotted simulation. Flows are serviced slot-by-slot (a slot is of duration 1 second) over a broadband access link of capacity 100 Mbps. For simplicity, we assume this access link emulates a "mega-household" representing a collection of households, each having an average DSL connection of 10 Mbps. The mega-household is assumed to house four premium mega-devices namely family TV, father's laptop, mother's laptop and daughter's tablet and one ordinary mega-device (representing all IoT devices that do not generate high volume of traffic). Each mega-device is serviced at a statically configured minimum rate (assumed to be configured by the user using the user-side API); for our experiments the family TV, father's laptop, mother's laptop, daughter's tablet, and ensemble of IoT devices are respectively set to receive at least 40%, 25%, 25%, 5% and 5% of link capacity. In simulation run, flows are mapped into a randomly chosen mega-device proportionate to the weights mentioned above.

The video flows that are accommodated by the API – assumed to be constant bit rate – are allocated their own reserved queue, while the other flows (mice, elephants, and video flows not accepted by the API) share a best-effort device-specific queue. Within the best-effort queue, the mice flows (that transfer less than 1 MB) are assumed to obtain their required bandwidth first (since they are typically in the TCP slow-start phase), and the remaining bandwidth is divided fairly amongst the video and elephant flows, which are expected to be in the TCP congestion avoidance phase. The scheduling is work-conserving, so any bandwidth unused by any queues are given to the remaining best-effort queues.



(a) $v_{th} = 5\%$ (b) $v_{th} = 20\%$

Fig. 3. Temporal dynamics of violation and call arrival/admission

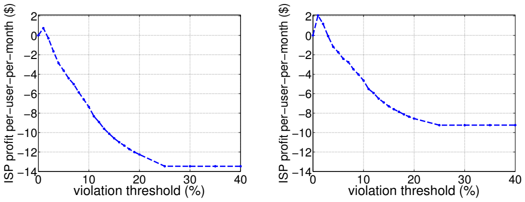
C. Performance Results

1) Impact of Violation Threshold (v_{th}): We first discuss the impact of the ISP-knob v_{th} on the overall experience of both the user and the CSP. In Fig. 2(b) we show by a solid black line the average violation (left-side y-axis) as a function of the chosen violation threshold v_{th}. As expected, when $v_{th} = 0$, no API call from the CSP is accepted, (the ISP therefore makes no money from the CSP); correspondingly, the user's policy is never violated and each mega-device receives its configured minimum rate at all times. As the ISP increases v_{th}, the average violation increases roughly linearly as well, saturating at about 13.14%. That is because the average video load in our trace data is about 13.76 Mbps. Thus, even if all video flows are granted fast-lanes, the bandwidth deficit would be fraction 13.76% of the link capacity, which provides an upper bound for average violation. The call admission rate for dynamic fast-lanes (dash-dotted blue curve, right-side y-axis) increases with threshold v_{th}, meaning that the CSP can exercise more control over fast-lane creation (and pay for it). At $v_{th} = 35\%$, all video flows are reserved. Increasing the violation threshold to $v_{th} = 25\%$ leads to saturation, since at this point 99.85% of CSP requests for fast-lane creation have been admitted; for this reason we truncate the plot at $v_{th} = 40\%$.

Fig. 3 shows the temporal dynamics (i.e. behavior over the 12-hour span of the data) of violation and call admission rate with two sample threshold values: (a) $v_{th} = 5\%$, and (b) $v_{th} = 20\%$. The observed violation rate (solid blue line, left-side y-axis) oscillates around the chosen threshold value, as expected. It is also seen that the gap between the call arrivals (dashed-red line) and call acceptance (dotted-back line) is much narrower when $v_{th} = 20\%$ (Fig. 3(b)) rather than when $v_{th} = 5\%$, since a higher threshold allows the ISP to accept more CSP calls by violating the user-defined policy more frequently.

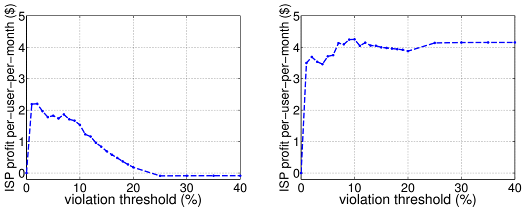
2) Impact of User Flexibility (κ): We now evaluate how the user's flexibility, captured by the parameter κ that translates their policy violation into a churn probability, affects the ISP's economics. For this study the pricing parameter δ is fixed to 2. In Figures 4-6 we show the ISP profit in units of dollars, normalized per-user-per-month. We consider three types of users: (a) inflexible user corresponding to $\kappa = 2$ for whom the probability of churn rises steeply with minor increase in average violations, (b) moderate user corresponding to $\kappa = 10$ who can tolerate violation to some extent, and (c) flexible user corresponding to $\kappa = 100$ who is very permissive in letting the ISP carve dynamic fast-lanes for CSPs.

For the inflexible user corresponding to $\kappa = 2$, Fig. 4(a)



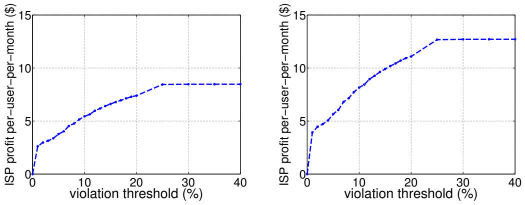
(a) $\lambda = 1$ (b) $\lambda = 1.5$

Fig. 4. ISP profit for inflexible user: $\kappa = 2$



(a) $\lambda = 1$ (b) $\lambda = 1.5$

Fig. 5. ISP profit for moderately-flexible user: $\kappa = 10$



(a) $\lambda = 1$ (b) $\lambda = 1.5$

Fig. 6. ISP profit for flexible user: $\kappa = 100$

shows that the ISP profit largely falls as the violation threshold is increased (bandwidth is priced at a peak rate of $\lambda = 1$ cent-per-Mbps-per-sec for this plot). This is because the risk of losing the customer due to their annoyance at violation of their policy outweighs the revenue obtained from the CSP. Fig. 4(b) shows the situation is roughly the same when the bandwidth peak price is increased to $\lambda = 1.5$, though the numerical profit is less negative. An "inflexible" user therefore poses a high economic risk for the ISP; to retain such users, the ISP has to either reject the majority of CSP API calls pertaining to this subscriber, or offer the customer some incentive (such as a rebate) to increase their flexibility parameter κ.

Increasing the user's flexibility to $\kappa = 10$ (we label such a user as being "moderately-flexible") results in an ISP profit curve shown in Fig. 5(a). In this case the ISP is able to gain an extra maximum profit of \$2.2 per-month per-user by adjusting the violation threshold to $v_{th} = 2\%$, when the bandwidth peak-price is set at $\lambda = 1$ cent-per-Mbps-per-sec. Increasing the violation threshold any higher is however detrimental, since the user annoyance over-rides the gains from the CSP. When the peak-price of bandwidth is increased to $\lambda = 1.5$ cents, Fig. 5(b) shows that the ISP can maximise profit by increasing violations for the user to about 10%, since the dynamic fast-lanes are more lucrative, thereby nearly doubling the profits to \$4.3 per-user per-month, which could even be used to subsidize the user's \$60 monthly bill.

Lastly, we consider an extremely "flexible" user with $\kappa = 100$, for whom the ISP profit is shown in Fig. 6. As expected, we see in this case that the ISP profit rises monotonically with threshold, since the low chance of user churn encourages the ISP to accept all CSP requests for fast-lanes and charge for them. The ISP's substantial profits in this case (\$8.45 and \$12.67 per-subscriber per-month respectively for $\lambda = 1$ and 1.5 cents-per-Mbps-per-sec) can be passed on as a rebate back to the subscriber, though rebate mechanisms are beyond the scope of study of the current work.

VI. CONCLUSIONS

Broadband fast-lanes is being debated vigorously today. Much of the debate has focused on static agreements between ISPs and CSPs while ignoring participation from end-users. In this paper, we advocated broadband fast-lanes with two-sided control, and argued how it benefits all the three entities involved, namely the end-user, CSP and ISP. We developed an architecture, using SDN technology, that permits an ISP to create and operate fast-lanes, provides control of fast-lanes to the end-user on a per-device basis, and allows fast-lanes to be initiated dynamically by a CSP on a per-flow basis. Using simple but representative models for fast-lane economics by ISPs, associated revenue-generation for CSPs, and churn-rates for subscribers, we have shown that our approach can open doors for ISPs to monetize on fast-lanes, assure video quality of flows for CSPs, and adhere to desired end-user quality of service preferences. Using simulations from real traffic traces comprising over 10 million flows, we showed that dynamic-fast-lanes is an attractive revenue stream for ISPs while limiting end-user annoyance to controllable levels. We believe that our solution is a candidate worthy of consideration in the continuing debate surrounding broadband fast-lanes.

REFERENCES

[1] A. Balachandran et al. Developing a predictive model of quality of experience for internet video. In *Proc. of ACM SIGCOMM*, Aug. 2013.

[2] A. Ferguson et al. Participatory networking: An API for application control of SDNs. In *Proc. ACM SIGCOMM*, Hong Kong, Aug 2013.

[3] CNN Money. AT&T wants you to design your own Internet fast lane. http://money.cnn.com/2014/10/13/technology/net-neutrality-compromise/, Oct 2014.

[4] E. Altman et al. Network non-neutrality debate: An economic analysis. In *Proc. IFIP Networking*, Spain, May 2011.

[5] N. Economides and J. Tag. Network neutrality on the Internet: A two-sided market analysis. *Inf. Econ. and Policy*, 24:91–104, 2012.

[6] F. Dobrian et al. Understanding the impact of video quality on user engagement. In *Proc. ACM SIGCOMM*, Aug. 2011.

[7] FinancialTimes. Netflix wants to put Comcast genie back in fast lane bottle. http://www.ft.com/cms/s/0/0bc54d54-639e-11e4-8216-00144feabdc0.html#axzz3KreEKl59, Nov 2014.

[8] H. Habibi Gharakheili et al. Pricing user-sanctioned dynamic fast-lanes driven by content providers. In *Proc. IEEE INFOCOM Workshop on Smart Data Pricing (SDP)*, Apr 2015.

[9] H. Kumar et al. User control of quality of experience in home networks using SDN. In *Proc. IEEE ANTS*, Dec. 2013.

[10] S. Sen et al. A survey of smart data pricing: past proposals, current plans, and future trends. *ACM Comp. Surveys*, 46(2), 2013.

[11] 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 Communications Magazine*, 50(11):91–99, 2012.

[12] The Guardian. The FCC is about to axe-murder net neutrality. Don't get mad get even. http://goo.gl/6kXZqo, Apr 2014.

[13] V. Sivaraman et al. Virtualizing the access network via open APIs. In *Proc. ACM CoNEXT*, Dec 2013.

[14] Wall Street Journal. FCC to propose new 'Net Neutrality' rules. http://goo.gl/ZRYfWc, Apr 2014.

[15] Y. Yiakoumis et al. Putting home users in charge of their network. In *Proc. ACM UbiComp*, New York, NY, Sept. 2012.