Modelling the Tradeoffs in Overlay-ISP Cooperation

Raul Landa, Eleni Mykoniati, Richard G. Clegg, David Griffin, and Miguel Rio

Department of Electrical and Electronic Engineering
University College London
{rlanda, emykonia, rclegg, dgriffin, mrio}@ee.ucl.ac.uk

Abstract. The increasing demand for efficient content distribution using the Internet has fuelled the deployment of varied techniques such as peer-to-peer overlays, content distribution networks and distributed caching systems. These have had considerable impact on ISP infrastructure demand, motivating the development of protocols that enable mutually beneficial cooperative outcomes.
In this paper we propose a parameterised *cooperation utility* that can be used to study the tradeoff between the benefit that an overlay obtains from the ISPs that carry its traffic and the costs that it imposes on them. With this utility, we find a closed-form expression for the optimal resource allocation given a particular cooperation tradeoff, subject to both minimal benefit and maximal cost constraints. Since the model is implementation-independent and has very modest computational demands, it is ideal for large scale simulation. We explore the properties of the proposed model through simulation in both a simple illustrative scenario and a more complete one based on network datasets. The results obtained from the model are shown to be consistent with those of measurement-based studies of overlay-ISP collaboration.

Keywords: Peer-to-peer, virtual and overlay networks; Overlay-ISP cooperation

1 Introduction

User demand for content distributed over the Internet has increased enormously in the last decade. As a result, diverse solutions based on network overlays have been deployed to make content distribution faster and more scalable. These include peer-to-peer systems, content distribution networks and distributed caching infrastructures, and we shall group them under the name of *content distribution overlays* (CDOs). In this paper we present a model that can be used to describe a range of cooperative behaviours between CDOs and their underlying ISPs, taking into account the preferences of both.

If one considers the traffic allocation of greatest benefit to a given CDO, it is clear that it will depend on its preferences regarding cost, QoS, resource availability, replication and data caching. On the other hand, if one considers the traffic allocation of greatest benefit to an ISP, it will depend on its infrastructure and transmission costs, the background traffic that it carries and its traffic engineering policies. Due to these differences, tensions may arise between the preferences of the overlay and those of the the ISP [15]. On the other hand, the existence of mutually beneficial outcomes arising from Overlay-ISP cooperation has been extensively documented [2–5, 7, 8, 10, 16–19]. This has sparked interest not only within the research community, but also within standardisation working groups [14]. Usually, these works investigate particular tradeoffs between overlay optimality and ISP costs in the context of specific protocols or applications. However, a more general cost-benefit model for these tradeoffs, developed from basic assumptions describing the preferences of both ISPs and CDOs, can be a useful tool in the understanding of the common foundations that they share.

The main contributions of this paper are a parameterised *cooperation utility* that can be used to describe the cost-benefit tradeoffs of ISP-aware content distribution overlays, and a closed-form solution for the optimal tradeoff that arises from it. Our model starts from a set of basic assumptions regarding both the benefits that overlays can obtain and the costs that they impose on the ISPs that carry their traffic, as provided by [1, 14, 19], and goes on to provide benefit and cost functions that satisfy them. A utility function is then presented that can be used to describe the tradeoff between CDO benefit and ISP cost. This utility then becomes the objective in a constrained maximisation problem, which is solved to provide a closed form solution for the CDO traffic allocation that embodies the optimal tradeoff. This analytic solution greatly reduces the computational effort involved in performing simulations using our model, allowing it to scale easily to overlays with several million links using modest computational resources. Finally, our evaluation shows that the results produced by this model are consistent with the conclusions of measurement-based studies of overlay-ISP collaboration.

The structure of the paper is as follows. We present the general characteristics of our first-principles model in §2, followed by the model itself in §3. We present a simulation-based evaluation in §4, other contributions in this area in §5, and conclude in §6.

2 A Model for Overlay-ISP Cooperation

We commence by defining the central components of our model, as shown in Fig. 1.

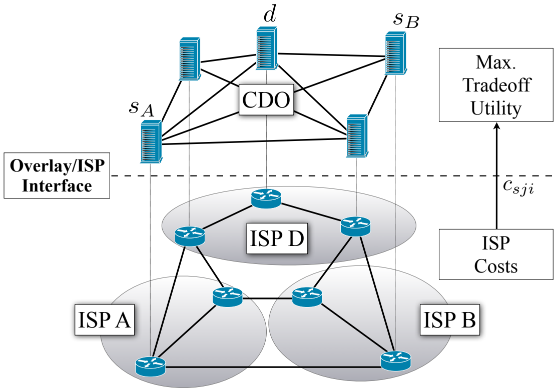


Fig. 1. The Overlay-ISP Boundary between a CDO and its underlying ISPs.

- **Content Distribution Overlays (CDO)** are overlay networks formed by a set of nodes placed across the Internet and providing content retrieval services to end users. Examples would be peer-to-peer networks, managed overlay networks for the distribution of multimedia streams or content delivery networks (CDNs).
- **Internet Service Providers (ISP)** are business organisations that provide connectivity services to end customers and CDOs.

2.1 The Cooperation Utility

We model the tradeoffs in Overlay-ISP cooperation by proposing a utility function that balances the benefit that the overlay derives from the services provided by its underlying ISPs and the costs that it imposes on them. For a given set of CDO and ISP preferences, this utility function can be used to assess the performance of different CDO traffic allocation policies when compared with the optimum. We assume that ISPs reveal their costs only to their local peers, and that they do so using interfaces such as [1, 14, 19]. Since no similar interface exists for overlay-to-overlay communication, we assume that each overlay operates independently and therefore focus on the single-overlay case. We will formulate the Overlay-ISP tradeoff problem so that each peer can solve a local optimisation problem individually, obviating the need for central control.

We denote the set of all ISPs as $\mathcal{I}$, and the set of all CDO nodes as $\mathcal{N}$. For each node in the CDO we consider a utility function U_i that combines the benefits that it can obtain given a particular traffic allocation with the costs that such an allocation will impose on its underlying ISPs. We then maximise this utility, taking as input the relevant flow costs and qualities. This will yield the optimal CDO traffic allocation in terms of a set of bandwidth assignments to traffic flows. We define a *traffic flow* as a 3-tuple (s, i, j) consisting of an ISP s, a *local* node i and a *remote* node j. Conceptually, a flow is a representation for the overlay traffic between a local peer i in ISP s and a remote peer j. Our model admits *multi-homed* nodes; i can be local to a set of ISPs denoted as L_i (of course, $L_i \subset \mathcal{I}\ \forall\ i$). Each flow will be annotated with a *flow volume* b_{sij} which represents the total amount of traffic that the flow carries, a *flow cost* per unit bandwidth c_{sij} provided by s, and a *flow quality* q_{sij} estimated or measured by the overlay. We propose that each one of the nodes of the overlay will have a utility

$$U_i = \alpha_i B_i - \epsilon_i C_i, \tag{1}$$

in which the benefit and cost terms are

$$B_i = \left(\sum_{j \in \mathcal{N}, s \in L_i} b_{sij}^{\beta_i} q_{sij}^{\gamma_i} \right)^{\delta_i} \quad \text{and} \quad C_i = \left(\sum_{j \in \mathcal{N}, s \in L_i} b_{sij}^{\zeta_i} c_{sij}^{\eta_i} \right)^{\theta_i}, \tag{2}$$

and where α_i, β_i, γ_i, δ_i, ϵ_i, ζ_i, η_i and θ_i are cost-benefit node parameters that can be tuned to capture the preferences of both the CDO and its underlying ISPs. The first term in (1) models the benefit that node i obtains from its overlay traffic with all other nodes; the second term, the aggregate preference cost (i.e. in the sense of [1, 14, 19]) that its ISP set L_i is exposed to by carrying this traffic. Rather than modelling intricate protocol specifications or detailed ISP business models, we aim to find a simple, general model based on basic assumptions. For B_i, this led us to select the functional form in (2) because it captures several intuitions about CDO preferences.

- *Increasing benefit with increasing flow volume* ($\alpha_i > 0$, $0 < \beta_i < 1$, $0 < \delta_i < 1$). In a capacity constrained scenario, the best nodes would only be able to provide service to a subset of end users, forcing the rest to rely on less desirable nodes and leading to reduced CDO benefit. Since increased flow volume ameliorates this, it results in an increased benefit for the overlay.

- *Increasing benefit with increasing quality* ($0 < \gamma_i < 1$). We assume that overlay links will be annotated with a *quality* q_{sij}, so that transferring the same amount of traffic between two nodes yields greater benefit if the quality of the overlay link between them increases.
- *Diminishing marginal benefit on increasing flow volume* ($0 < \beta_i < 1$). This models the fact that not all data available in a given node is equally useful. Thus, any given node will experience decreasing marginal benefit from increasing amounts of received traffic from another given node.
- *Non-increasing marginal benefit on increasing flow quality* ($0 \leq \gamma_i \leq 1$). In many cases, such as voice or video streaming, once the quality of the received stream is high enough to decode the stream in time, no further improvement will be achieved by increasing the quality of overlay flows. Thus, benefit increases with quality, but only with diminishing returns.
- *Non-increasing marginal benefit on the number of incoming flows* ($0 \leq \delta_i \leq 1$). We assume that different nodes might have access to different kinds of content of interest to a particular node, so that benefit increases with the number of sender nodes that a given node has. However, it is improbable that all nodes will yield equivalent usefulness to the receiving node. Consequently, the benefit from connecting to an increasing numbers of nodes will increase at a decreasing rate.

Even though the c_{sij} provided by [1, 14, 19] may not represent direct ISP costs, they can be used as proxies for cost-related ISP preferences. We define the preference cost function C_i in (2) so that it captures the following intuitions.

- *Increasing cost with increasing flow volume* ($\epsilon_i > 0$, $0 < \zeta_i < 1$, $0 < \delta_i < 1$). We assume that, for a fixed cost-per-bit, the delivery of increasing amounts of traffic between overlay nodes imposes an increasing cost on ISPs.
- *Increasing cost with increasing infrastructure cost* ($0 < \theta_i < 1$). We assume that transferring the same amount of traffic between two nodes imposes greater costs if the cost of the underlying network infrastructure is higher.
- *Non-increasing marginal flow volume cost* ($0 \leq \zeta_i \leq 1$). This models the fact that Internet connectivity to a particular host imposes fixed costs, usually related to the provision of physical layer infrastructure. Thus, cost increases disproportionally for the first units of provisioned capacity.
- *Non-increasing marginal infrastructure cost* ($0 \leq \eta_i \leq 1$). This models economies of scale in traffic aggregation, which can lead to reduced cost per bit.
- *Non-increasing marginal cost for increasing number of nodes communicating with an overlay node* ($0 \leq \theta \leq 1$). This allows the modelling of economies of scale in colocation and port density. Once a node has been provided with resources, providing resources to other nearby nodes can be done at a reduced cost per node.

3 The Overlay-ISP Cooperation Problem

In our model, the *Overlay-ISP Cooperation Problem* is solved by maximising the CDO cooperation utility taking the preferences of the CDO and ISPs as given by the cost-benefit tradeoff parameters α_i, β_i, γ_i, δ_i, ϵ_i, ζ_i, η_i and θ_i $\forall$ $i \in \mathcal{N}$. We formulate the

Overlay-ISP cooperation problem as

$$\underset{b_{sij} \in \mathbb{R}_{\geq 0}}{\text{Maximise:}} \quad U = \sum_{i \in \mathcal{N}} U_i = \sum_{i \in \mathcal{N}} \alpha_i B_i - \sum_{i \in \mathcal{N}} \epsilon_i C_i. \tag{3}$$

First, we will assume that the CDO has no requirements regarding either the benefits that it obtains or the costs it imposes on its underlying ISPs. Thus, we concentrate our attention on finding the optimal b_{sij} for given q_{sij}, c_{sij}, and additional model parameters. Since only non-restricted optimisation is required, we can apply first order conditions to (3) directly. This leads to the following system of equations

$$\frac{\partial U}{\partial b_{sij}} = \sum_{i \in \mathcal{N}} \alpha_i \frac{\partial B_i}{\partial b_{sij}} - \sum_{i \in \mathcal{N}} \epsilon_i \frac{\partial C_i}{\partial b_{sij}} = 0, \tag{4}$$

where B_i and C_i are given by (2). For clarity reasons, for now we will disregard congestion and economy of scale effects, thus making q_{sij} and c_{sij} constant (we will later re-introduce the notion of costs and qualities as functions of b_{sij}). This makes (4) separable, and the first order conditions become

$$\alpha_i \frac{\partial B_i}{\partial b_{sij}} - \epsilon_i \frac{\partial C_i}{\partial b_{sij}} = 0, \tag{5}$$

where the marginal benefit and cost terms are

$$\frac{\partial B_i}{\partial b_{sij}} = \delta_i \frac{\beta_i b_{sij}^{\beta_i - 1} q_{sij}^{\gamma_i}}{\left(\sum_{j \in \mathcal{N}, s \in L_i} b_{sij}^{\beta_i} q_{sij}^{\gamma_i} \right)^{1-\delta_i}} \quad \text{and} \quad \frac{\partial C_i}{\partial b_{sij}} = \theta_i \frac{\zeta_i b_{sij}^{\zeta_i - 1} c_{sij}^{\eta_i}}{\left(\sum_{j \in \mathcal{N}, s \in L_i} b_{sij}^{\zeta_i} c_{sij}^{\eta_i} \right)^{1-\theta_i}}, \tag{6}$$

for each overlay node i. It is clear that the denominators in (6) do not depend on j, the remote endpoint of the flow, nor on its ingress ISP s. Rather, they are only a function of b_{sij}, the desired traffic allocation, and the other properties of the local node i. By considering $\frac{\partial B_i}{b_{tik}}$ and $\frac{\partial C_i}{b_{tik}}$, the marginal benefit and cost associated with another arbitrary flow (t, i, k) having the same local node i, but with a different ingress ISP t and terminating on a different remote node k, it can be shown that

$$\left(\frac{b_{sij}}{b_{tik}} \right)^{\zeta_i - \beta_i} = \frac{q_{sij}^{\gamma_i}}{c_{sij}^{\eta_i}} \frac{c_{tik}^{\eta_i}}{q_{tik}^{\gamma_i}}. \tag{7}$$

This means that, discounted by a diminishing returns exponent $\zeta_i - \beta_i$, the ratio between the bandwidth allocated to two flows (s, i, j) and (t, i, k) terminating in the same local node i will be equal to the ratio between their **cost-benefits**, defined as the ratios between their qualities and their costs. In particular, if we define the *preference-modified cost-benefit* μ_{sij} as $\mu_{sij} = \left(\frac{q_{sij}^{\gamma_i}}{c_{sij}^{\eta_i}} \right)^{\frac{1}{\zeta_i - \beta_i}}$, we see that (7) can be rewritten so that $\frac{b_{sij}}{b_{tik}} = \frac{\mu_{sij}}{\mu_{tik}}$. Thus, we see that the solution to (3) will provide overlay traffic allocations b_{sij} proportional to the μ_{sij} associated with (s, i, j). Using the previous definitions, the first order conditions can be solved in the standard manner from (5). This solution is cumbersome but straightforward, and is omitted for brevity. For the unconstrained case, we find that b_{sij}, the optimal bandwidth allocation for a flow between nodes j and i using ISP s as an ingress, can be expressed as

$$b_{sij} = \left(\frac{\alpha_i}{\epsilon_i}\psi_i\right)^{\xi_i} \mu_{sij} \tag{8}$$

where

$$\psi_i = \frac{\beta_i\delta_i}{\zeta_i\theta_i} \frac{\left(\sum_{j\in\mathcal{N},s\in L_i} \mu_{sij}^{\zeta_i} c_{sij}^{\eta_i}\right)^{1-\theta_i}}{\left(\sum_{j\in\mathcal{N},s\in L_i} \mu_{sij}^{\beta_i} q_{sij}^{\gamma_i}\right)^{1-\delta_i}}, \quad \xi_i = \frac{1}{\zeta_i\theta_i - \beta_i\delta_i}.$$

The set of flow volumes defined by (8) represent an optimal tradeoff between the CDO qualities q_{sij} and the costs c_{sij} announced by ISPs, given their respective preferences.

We now address the case where the CDO has operational constraints. To this end, we propose an improved model which, as we shall see, is a simple extension of that of the previous section. This new optimisation problem can be stated as

$$\underset{b_{sij}\in\mathbb{R}_{\geq 0}}{\text{Maximise:}}\ U = \sum_{i\in\mathcal{N}} U_i \tag{9}$$

$$\text{Subject to:} \sum_{i\in\mathcal{N}} \left(\sum_{j\in\mathcal{N},s\in L_i} b_{sij}^{\beta_i} q_{sij}^{\gamma_i}\right)^{\delta_i} = \sum_{i\in\mathcal{N}} B_i \geq B_{\min} \tag{10}$$

$$\sum_{i\in\mathcal{N}} \left(\sum_{j\in\mathcal{N},s\in L_i} b_{sij}^{\zeta_i} c_{sij}^{\eta_i}\right)^{\theta_i} = \sum_{i\in\mathcal{N}} C_i \leq C_{\max} \tag{11}$$

where (10) is the minimum benefit tolerable to the overlay, and (11) is the maximum aggregate cost that the overlay is willing to impose on all the ISPs that provide it with connectivity services. The solution to this problem is a simple extension to (3), with a slightly expanded Lagrangean that leads to the first order optimality conditions

$$(\alpha_i + \lambda_B)\frac{\partial B_i}{\partial b_{sij}} - (\epsilon_i + \lambda_C)\frac{\partial C_i}{\partial b_{sij}} = 0, \tag{12}$$

along with the two *complementary slackness* conditions

$$\lambda_B \left(B_{\min} - \sum_{i\in\mathcal{N}} B_i\right) = 0, \quad \lambda_C \left(\sum_{i\in\mathcal{N}} C_i - C_{\max}\right) = 0.$$

In the previous expressions, λ_B corresponds to the Lagrange multiplier associated with overlay benefit and λ_C corresponds to that associated with ISP costs.

We seek an expression for b_{sij}^*, the solution to the budget-constrained problem. The derivation proceeds as in the previous case, and we have that

$$b_{sij}^* = \left(\frac{\alpha_i + \lambda_B}{\epsilon_i + \lambda_C}\psi_i\right)^{\xi_i} \mu_{sij} = \left(\frac{1 + \frac{\lambda_B}{\alpha_i}}{1 + \frac{\lambda_C}{\epsilon_i}}\right)^{\xi_i} b_{sij}, \tag{13}$$

where b_{sij} is given by (8) and represents the flow volume that would have been allocated to a flow from node j to node i entering the network through ISP s, *had no constraints been active*. As expected, if no constraint binds, $\lambda_C = \lambda_B = 0$ and (13) reduces to (8).

The unconstrained problem (3) can be implemented in a decentralised fashion trivially because U is the sum of the individual utilities U_i of each node. The constrained problem (9) is only slightly more difficult to distribute, as it is only coupled by the two constraints (10) and (11). For the simulations in §4, we find λ_B, λ_C and b^*_{sij} using standard dual decomposition techniques [6]. Thanks to (13), and by allocating an independent thread to each node in the optimisation solver, it is possible to take full advantage from multicore architectures and significantly reduce simulation time.

We now include congestion and economies of scale considerations in our model. Having solved (9) for constant q_{sij} and c_{sij}, we now expand our scope to consider $\tilde{q}_{sij}$ and $\tilde{c}_{sij}$, equivalent expressions that are functions of b_{sij}. However, to keep the model as simple as possible and compatible with the solutions that we have already found, we will make two assumptions. The first one is that $\tilde{q}_{sij}$ and $\tilde{c}_{sij}$ are functions with *constant elasticity*, i.e. $E_i^q = \frac{\partial \log \tilde{q}_{sij}}{\partial \log b_{sij}} = 0$, and $E_i^c = \frac{\partial \log \tilde{c}_{sij}}{\partial \log b_{sij}} = 0$. The second one is that their elasticities E_i^q and E_i^c are *functions of i only*, thus keeping the first order conditions separable. Hence, we propose cost and quality functions of the form

$$\tilde{q}_{sij} = q_{sij} b_{sij}^{E_i^q}, \quad \tilde{c}_{sij} = c_{sij} b_{sij}^{E_i^c}, \tag{14}$$

where E_i^q is the *access congestion* elasticity, E_i^c is the *economy of scale* elasticity, and q_{sij} and c_{sij} correspond to the constant quality and cost introduced in (2). The naming of E_i^q and E_i^c is indicative of their function in the model. In particular, $|E_i^q|$ will represent the percent decrease in overlay link quality q_{sij} with a percent increase in b_{sij}, and $|E_i^c|$ will represent the percent decrease of per-unit-bandwidth cost c_{sij} with a percent increase in b_{sij}. Since quality will be reduced with increased traffic flow, $E_i^q < 0$, and since economy of scale effects reduce the cost per bit, $E_i^c < 0$ as well. Therefore, E_i^q determines how the link quality q_{sij} falls off as the access link of the local node i becomes congested; conversely, E_i^c determines how the link cost c_{sij} falls off with economies of scale in the access link of the local node i.

The rationale behind choosing this model of access link congestion and economies of scale is that it allows us to consider the effect of both $\tilde{q}_{sij}$ and $\tilde{c}_{sij}$ as an additive constant. Consider the effect of replacing q_{sij} and c_{sij} in (2) with $\tilde{q}_{sij}$ and $\tilde{c}_{sij}$ from (14): it amounts to using modified $\tilde{\beta}_i$ and $\tilde{\zeta}_i$ so that

$$\tilde{\beta}_i = \beta_i + \gamma_i E_i^q, \quad \tilde{\zeta}_i = \zeta_i + \eta_i E_i^c.$$

To see how substituting $\tilde{\beta}_i$ and $\tilde{\zeta}_i$ into the previous expressions changes our model, we note that, from (8), b_{sij} will increase with increasing $\tilde{\beta}_i$, and it will decrease with increasing $\tilde{\zeta}_i$. As $|E_i^q|$ increases, $\tilde{\beta}_i$ will decrease and b_{sij} will decrease as well; as $|E_i^c|$ increases, $\tilde{\zeta}_i$ will decrease, and b_{sij} will increase. Thus, a propensity for congestion in the access link of i can be modelled with a large $|E_i^q|$, and the overlay will react with a general reduction in traffic with i as flow volume increases. Conversely, if $|E_i^c|$ is large, denoting good cost efficiency, the overlay will increase traffic to and from i.

In order for (8) to remain a valid solution of the optimisation problem, we must impose limitations on some of the model parameters. We now explore these.

- *Positive traffic attraction* ($\tilde{\beta}_i > 0$). When $|E_i^q| = \frac{\beta_i}{\gamma_i}$, $\tilde{\beta}_i = 0$, making b_{sij} equal to zero for all s and j. Conceptually, this means that the overlay has found the access link of node i to be excessively congested, and thus removes it from consideration. Further increases in the magnitude of E_i^q are ignored; traffic will resume when E_i^q decreases.
- *Positive traffic avoidance* ($\tilde{\zeta}_i > 0$). As $|E_i^c| \to \frac{\zeta_i}{\eta_i}$, $\tilde{\zeta}_i \to 0$, and $b_{sij} \to \infty$. This happens when the cost reduction associated with increasing volume on flows towards i is so large that the overlay attempts to continually increase their volume.
- *Concave Utility* ($\tilde{\zeta}_i - \tilde{\beta}_i > 0$ and $\tilde{\zeta}_i\theta_i - \tilde{\beta}_i\delta_i > 0$). These two conditions are related with the existence of a well-defined maximum for U_i over an unrestricted domain. In both cases, violation of these conditions implies that costs grow more slowly than benefits, and it is thus the overlay will continually increase flows towards i.

4 Evaluation

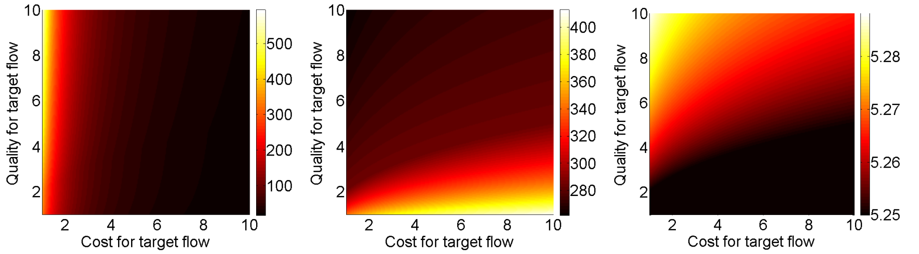
We evaluate our model through simulation. In §4.1, we demonstrate its fundamentals with a simple thought experiment; in §4.2, we perform a dataset-driven simulation.

4.1 Basic Simulation

In this section we explore the basic properties of our model by proposing an illustrative thought experiment. We consider a CDO with presence in at least three different ISPs, which we shall designate as A, B and D (see Fig. 1). We select three overlay nodes, a *client* node d and two *server* nodes s_A and s_B, that connect to the overlay via ISPs D, A, and B respectively. We will consider two overlay flows: (A, s_A, d), the *target flow* through which s_A communicates with d via ISP A, and (B, s_B, d), the *competing flow* through which s_B communicates with d via ISP B. We then modify the cost and quality associated with the target flow, and observe the CDO response.

This simulation makes use of the model with operational constraints presented in §3, with the following parameters: $\alpha_i = 1$, $\beta_i = .1$, $\gamma_i = .2$, $\delta_i = .3$, $\zeta_i = .7$, $\eta_i = .8$ and $\theta_i = .9$ for all nodes i, and we calibrate U by setting ϵ_i to a value such that $U = 0$ if $\lambda_B = \lambda_C = 0$. Regarding the flow-related parameters, we will set $E_i^q = -.2$, $E_i^c = -.1$, $q_{sij} = 1$ and $c_{sij} = 1$ for all flows (s, i, j) except the target flow, for which we will vary both cost and quality uniformly over the range $[1, 10]$. The cost associated with the target flow will be denoted as q_T, and its cost, as c_T. Constraints were set so that $B_{\min} = 5.25$ and $C_{\max} = \infty$; these were chosen to ensure (10) to be active within region of the cost-quality parameter space in which $U_i < 0$. We emphasise that these parameter values *do not correspond to any particular protocol implementation*. Rather, they were selected to be internally consistent, and to aid in the presentation of the basic properties of our model. For the chosen parameter values, the overlay tradeoff preferences favour good-quality connectivity with a large set of nodes, rather than exceptionally high-quality connections with a reduced node set. Therefore, the effect of B_i dominates at lower values of b_{sij}, and that of C_i at higher ones. The results are shown in Fig. 2, in which each pixel in each subfigure is a solution to (9).

Fig. 2(a) shows the optimal tradeoff overlay target flow volume between s_A and d, while Fig. 2(b) shows the competing flow volume allocation between s_B and d. The



(a) Target flow volume (b) Competing flow volume (c) Total overlay benefit

Fig. 2. Simulating quality and cost changes for the CDO tradeoff model (all magnitudes shown are in arbitrary units)

overlay can allocate volume to at least two flows terminating in d, and the costs of these flows will be provided by two different ISPs (A and B). We can see that, for a given target flow cost c_T, as the target flow quality q_T increases, the target flow volume allocated by the optimisation algorithm increases as well. Conversely, for a given q_T, as c_T increases, the allocated target flow volume decreases. Thus, the model provides sensitivity to the requirements of both the overlay and its underlying ISPs.

In Fig. 2(b), we can see that if q_T decreases, the competing flow volume increases. Near the top of the graph, the high quality of the target flow means that the $B_{\min}$ restriction is inactive. Hence, the competing flow increases as ψ_i increases, as required by (8). For the target flow, however, μ_{sij} decreases by a greater amount. This leads to a net reduction in the target flow along with an increase in all other incoming flows to d, including the competing flow. Thus, the model predicts that, when faced with changing ISP-reported costs and varying connection qualities, overlays will respond by substituting volume over expensive, low quality overlay links with volume over links which provide better cost-benefit. However, this effect is small when compared to the response of the model to the minimum benefit restriction becoming active, which is the case near the bottom of Fig. 2(b) and throughout the black region in the lower right of 2(c).

Since we did not introduce any explicit flow volume constraints in the model, we see that the reduction in traffic induced by increasing the target flow is not balanced by an equivalent increase of traffic over the competing flow. The model presented performs traffic substitution proportionally to cost-benefit and with diminishing returns, with traffic *to all other destinations* (many of which are not shown here). Therefore, the availability of overlay links with high-quality and low-cost is always preferable to the CDO, as they significantly increase its utility.

4.2 Dataset-driven Simulation

In this section we compare our model with simplified approximations of tradeoff techniques previously investigated in measurement-based studies. Our objective will not be to accurately describe the tradeoffs made by any single implementation. Instead, we will focus on defining a set of controlled scenarios and exploring the conditions under which previously proposed tradeoffs approach optimality as measured by (1).

The dataset we used was originally obtained for [12], and includes round-trip times and computed AS paths between 1715 users participating in a file sharing application. The round-trip times were obtained by taking samples every two days between all pairs of nodes using the King [9] method. The AS paths were computed from three different sources: RouteViews [20], Looking Glass servers and iPlane [13]. From this dataset, we only kept those pairs of nodes for which latency measurements were consistent across the different samples and a good estimation of the number of AS hops between them was available. For those IP addresses in this subset, we obtained geolocation information by querying the `http://ipinfodb.com` database, and used these geographical coordinates to calculate the great circle (*haversine*) distance between them.

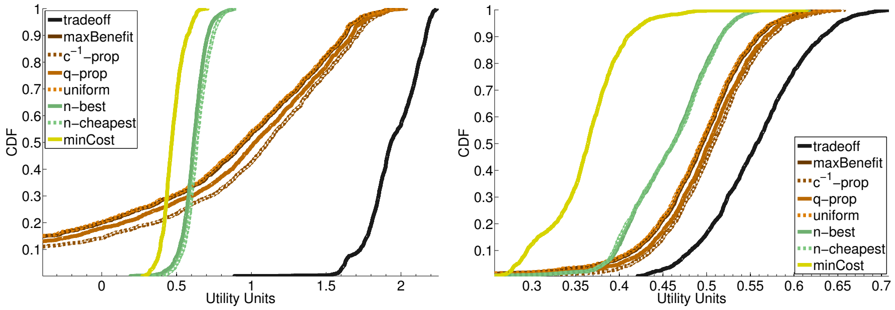
For all cases, we used network-measured round-trip time as a proxy for overlay link quality. This choice is aligned with the performance objectives of real-time content applications, where overlay links with low latency are more desirable. So, for all flows (s, j, i), we have that $q_{sij} \propto t_R^{-1}$, where t_R is the RTT between nodes j and i. Regarding the per-unit-bandwidth cost c_{sij} associated with (s, i, j) and announced by ISP s, we consider two alternatives.

- **HAVERSINE**: In this case, we use the great-circle distance between the geographical coordinates associated with the IP addresses of the nodes as an approximation of the ISP cost to provide connectivity between these endpoints. This choice reflects the intuition that the end-to-end traffic delivery cost is higher the further away two nodes are in physical space. In particular, we have that $c_{sij} \propto d_h$, where d_h is the haversine distance between the nodes.
- **AS HOPS**: In this case, we use the number of AS hops between two node IP addresses as an approximation of the cost incurred by the ISP to provide connectivity between them. In particular, we have that $c_{sij} \propto h + 1$, where h is the number of AS hops between the node IP addresses. This measure is of particular interest because many other Overlay-ISP collaboration works rely in AS-hop distances for node clustering, either in simple *intradomain* vs. *interdomain* terms or using AS-path lengths. Some examples of these include [3–5, 8, 17].

Regarding the cost/benefit tradeoff parameters α_i and ϵ_i, we set $\epsilon_i = 1$ and calibrate α_i so that the reception of 1 unit of bandwidth with a 10 millisecond delay from 4 nodes at a standard level of ISP cost will provide a zero utility U_i. In the case of HAVERSINE, we set this standard cost at 1000 miles; for AS HOPS, we set it at 3 hops. Finally, for the other cost-benefit tradeoff preference parameters, we use the values presented in §4.1. We consider two methods for the construction of our simulated overlay network.

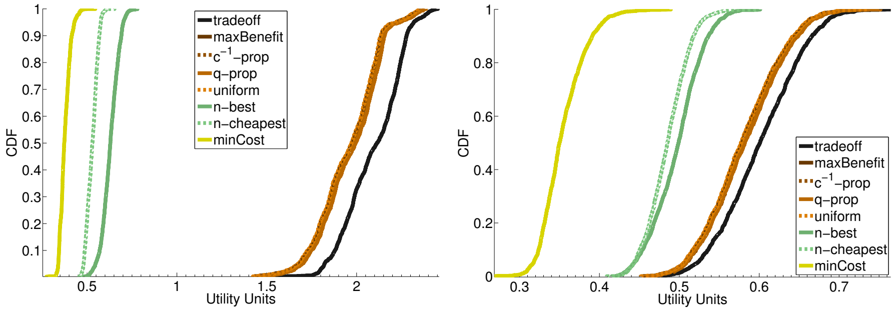
- ***Full Set***: In this case, we assume that the existence of delay measurements between two nodes in the dataset implies the existence of a potential overlay link.When using this method, the overlay network contains $1,545$ nodes and ∼2 million links, whose distribution is sharply bimodal: ∼930 nodes have $\sim 1,530$ overlay links, and ∼620 have ∼920 links.
- ***Synthetic Overlay***: In this case, only a subset of the full delay measurement graph is considered to construct the overlay topology. For each node in turn, m other nodes for which measurements are available were randomly nominated as potential neighbours. For the purpose of this evaluation, we consider m=5, which yields an overlay with $1,545$ nodes and $\sim$15$,450$ links. Each node has between 5 and 20 neighbours, with an average of 10.

In order to explore alternative allocation policies that result from other cost/benefit profiles, we consider the following.



(a) Full Dataset (b) Synthetic Overlay ($m = 5$)

Fig. 3. Cumulative Distribution Functions of the Utility U_i for the HAVERSINE model.



(a) Full Dataset (b) Synthetic Overlay ($m = 5$)

Fig. 4. Cumulative Distribution Functions of the Utility U_i for the AS HOPS model.

- ***Tradeoff***: This is our proposed model, as obtained by solving (9). This solution is used to obtain a per-node traffic volume $\hat{b}_i = \sum_{s,j} b_{sij}$ that is used as a constraint in the other allocation policies.
- ***Uniform***: This is a simple allocation policy in which each node allocates $\hat{b}_i$ equally among all its neighbours (this was typical of early peer-to-peer overlays, *e.g.* Gnutella).
- ***Max. Benefit***: Each node individually maximises its own benefit while adhering to a total traffic flow constraint. In particular, each node solves

$$\text{Maximise: } B_i, \quad \text{Subject to: } \sum_{j \in \mathcal{N}, s \in L_i} b_{sij} = \hat{b}_i.$$

The solution of this problem is standard, and is given by

$$b_{sij} = \frac{\tau_{sij}}{\sum_{j \in \mathcal{N}, s \in L_i} \tau_{sij}} \hat{b}_i, \quad \tau_{sij} = q_{sij}^{\frac{\gamma_i}{1-\beta_i}}.$$

This resource allocation policy approximates the operation of current content delivery networks (CDNs), which operate without explicit consideration of ISP costs.
- ***Min. Cost***: Each node individually minimises the cost it imposes to all ISPs, while adhering to a total traffic flow constraint. In particular, each node solves

$$\text{Minimise: } C_i, \quad \text{Subject to: } \sum_{j \in \mathcal{N}, s \in L_i} b_{sij} = \hat{b}_i.$$

The solution of this problem is standard, and is given by an allocation where $b_{sij} = \hat{b}_i$ for the flow (s,i,j) for which $c_{sij}^{\eta_i\theta_i}$ is minimal, and $b_{sij} = 0$ for all other flows.

- ***n-Best***: Each node chooses the n neighbours for which $q_{sij}^{\beta_i\delta_i}$ is maximal, and allocates $\hat{b}_i$ equally between them. This policy is reminiscent not only of the unchoking behaviour in BitTorrent seeding, but also of [5]. In this case, we chose $n = 4$.
- ***n-Cheapest***: Each node chooses the n neighbours for which $c_{sij}^{\eta_i\theta_i}$ is minimal, and allocates $\hat{b}_i$ equally between them. Again, we set $n = 4$. This policy is suggestive of a situation in which the ISP provides cost-optimised information to the CDO, which then uses it with no regard to its own benefit.
- c^{-1}***-Proportional***: Each node allocates to each neighbour a traffic flow inversely proportional to the ISP cost associated with the overlay link between them. Thus,

$$b_{sij} = \frac{c_{sij}^{-1}}{\sum_{j\in\mathcal{N}, s\in L_i} c_{sij}^{-1}} \hat{b}_i.$$

- q***-Proportional***: In this case, each node allocates to each neighbour a traffic flow proportional to the overlay link quality between them. Thus, we have that

$$b_{sij} = \frac{q_{sij}}{\sum_{j\in\mathcal{N}, s\in L_i} q_{sij}} \hat{b}_i.$$

  This heuristic is reminiscent of the *proportional share* policy presented in [11].

The results of our simulations can be found in Figs. 3 and 4. For the HAVERSINE parameter mapping, we see that all policies tested fall into essentially four groups. If we go from right to left following the top of Fig. 3(a), first we find *tradeoff* by itself, followed by a group formed by *max. benefit*, *uniform*, c^{-1}*-prop* and *q-prop*. We will denote this group as *Group II*. The third group to the left includes *n-best* and *n-cheapest*, and will be denoted as *Group III*; the final group consists of *min. cost* on its own. As shown in Figs. 3(a), 3(b) and Figs. 4(a) and 4(b), these groups appear consistently across the model mapping and overlay construction methods presented. What differentiates these groups is that the policies in *Group II* can allocate traffic between all neighbour nodes of a particular node, those in *Group III* have access to only $n = 4$ neighbours, and *min. cost* uses a single one. Thus, for the parameter values chosen, securing an appropriate number of neighbours (and related flows) is of importance to increase overlay benefit.

Overall, *tradeoff* provides improved utility in all cases. In both AS HOPS and HAVERSINE and for all groups, *synthetic overlay* topologies show more homogeneous utilities that are generally lower than those in *full set*. This is to be expected from the increased node selection opportunities afforded by *full set*. However, since observed tradeoffs are otherwise consistent for these two, we will comment on both of them simultaneously.

Group II exhibits better performance in the AS HOPS scenarios, with HAVERSINE *full dataset* in particular suffering from a long tail of nodes with negative utilities. This makes these policies even worse than those of *Group III* for many nodes in this scenario. Although all policies in this group generate very similar utility distributions in all cases, the differences between them are slightly larger for both HAVERSINE scenarios. As shown in Figs. 3(a) and 3(b), the CDFs generated by policies in *Group III* are much more consistent for the HAVERSINE scenarios, where the lines of *n-best* and *n-cheapest* overlap, than for the AS HOPS scenarios of Figs. 4(a) and 4(b), in which *n-best* exhibits slightly better utility distributions. In all cases, *min. cost* severely impacts the overlay while minimising cost, leading to the worst U_i CDF. The good cost-benefit tradeoff performance of *Group II* provides confirmation that, for some definitions of benefit and cost, there is a convergence of incentives between overlays and ISPs as embodied by the cooperation utility U_i. In particular, as shown by recent wide-scale measurement studies [17], for biased random topologies such as locality-aware BitTorrent, the use of AS-hops yields significant cost savings for ISPs with no reductions in overlay benefit.

5 Related Work

The expression of explicit interactions between overlays and ISPs is receiving increased attention by the research community (see, for instance, [2–5, 7, 8, 10, 16–19]). These studies present particular ways in which CDO construction optimality and ISP cost can be balanced, thus providing specific examples along the spectrum of tradeoffs considered by our model. Some of these works are now described.

The main work in this area is P4P [19], which became the basis for the main standardisation effort in the area [14]. In its original presentation, [19] relied on the ISP aggregating peers into groups (PIDs) and providing a set of end-to-end prices between them. In that case, the overlay was assumed to solve a particular minimisation problem arising from the dual decomposition of the ISP optimisation problem. The present work provides an alternative, overlay-centred view which remains compatible with [14] while being easily re-parameterisable to reflect diverse ISP and CDO preferences. Chronologically, one of the first works to focus in BitTorrent locality was [4], which relies on the ISP tagging peers as either *local* or *external*. This allows the overlay to set a soft limit on *external* peers, biasing peer selection and giving preference to intradomain connections. In [3], the authors present a system in which an *oracle* performs peer ranking according to the preferences of the ISP. Although many possible metrics are presented as candidates (AS path length, IGP metric distance, geographical information, expected bandwidth/delay and link congestion) the authors focus in AS path length. In [2], this system is extended to take into account peer upload bandwidth, and in [16] it is further improved by enriching DNS responses with ISP-provided information.

6 Conclusions

In this paper we proposed a model for the behaviour of ISP-aware overlays that is based only on basic assumptions regarding overlay and ISP preferences. This model is independent from implementation-specific details and can be used to explore the universal aspects of the cooperation between ISPS and overlays. To this end, we presented a *cooperation utility* that captures some basic intuitions regarding the preferences of both overlays and ISPs. This utility is then used to formulate an optimisation problem incorporating minimum benefit and maximum cost constraints, which is solved analytically. It is shown that, if expressed in a specific functional form, congestion and economy of scale effects can be considered in a straightforward way. Finally, simulations are used to show the properties of the proposed model. First, a simple simulation is presented to show the way in which the model responds to a changing overlay link cost.

Then, a more comprehensive simulation based on network measurement datasets is presented. In addition to providing evidence that our model can be used to gain insight into the complex interactions in Overlay-ISP cooperation, this data-driven simulation shows that our model behaves consistently with previous findings regarding the cost-benefit gains provided by some ISP-aware traffic allocation policies.

This research has received funding from the Seventh Framework Programme (FP7/2007-2013) of the European Union, through the ENVISION project (grant agreement 248565).

References

1. ENVISION (2009), http://www.envision-project.org
2. Aggarwal, V., Akonjang, O., Feldmann, A.: Improving User and ISP Experience through ISP-aided P2P Locality. In: Proc. of the Global Internet Symposium (2008)
3. Aggarwal, V., Feldmann, A., Scheideler, C.: Can ISPs and P2P users cooperate for improved performance? SIGCOMM Comput. Commun. Rev. 37(3), 29–40 (July 2007)
4. Bindal, R., Cao, P., Chan, W., Medved, J., Suwala, G., Bates, T., Zhang, A.: Improving traffic locality in BitTorrent via biased neighbor selection. In: Proc. of ICDCS '06 (2006)
5. Blond, S.L., Legout, A., Dabbousa, W.: Pushing bittorrent locality to the limit. Comput. Netw. 55(3) (2010)
6. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press (2009)
7. Choffnes, D.R., Bustamante, F.E.: Taming the torrent: a practical approach to reducing cross-ISP traffic in peer-to-peer systems. In: Proc. of SIGCOMM. pp. 363–374 (2008)
8. Dai, J., Li, B., Liu, F., Li, B., Jin, H.: On the efficiency of collaborative caching in isp-aware p2p networks. In: Proc. of INFOCOM. pp. 1224–1232. IEEE (2011)
9. Gummadi, K.P., Saroiu, S., Gribble, S.D.: King: estimating latency between arbitrary internet end hosts. In: Proc. of SIGCOMM. vol. 32, pp. 11–11 (2002)
10. Jiang, W., Zhang-Shen, R., Rexford, J., Chiang, M.: Cooperative content distribution and traffic engineering. In: Proc. of NetEcon (2008)
11. Levin, D., LaCurts, K., Spring, N., Bhattacharjee, B.: BitTorrent is an auction: analyzing and improving BitTorrent's incentives. In: Proc. of SIGCOMM. pp. 243–254 (2008)
12. Lumezanu, C., Baden, R., Spring, N., Bhattacharjee, B.: Triangle inequality variations in the internet. In: Proc. of IMC '09. pp. 177–183. ACM, New York, NY, USA (2009)
13. Madhyastha, H.V., Isdal, T., Piatek, M., Dixon, C., Anderson, T., Krishnamurthy, A., Venkataramani, A.: iPlane: An Information Plane for Distributed Services. In: Proc. of ACM OSDI. pp. 367–380 (2006)
14. Peterson, J., Marocco, E., Gurbani, V.: Application-Layer Traffic Optimization (ALTO) working group (2009)
15. Piatek, M., Madhyastha, H.V., John, J.P., Krishnamurthy, A., Anderson, T.: Pitfalls for ISP-friendly P2P design. In: Proc. of HotNets '09. ACM, USA (2009)
16. Poese, I., Frank, B., Ager, B., Smaragdakis, G., Feldmann, A.: Improving content delivery using provider-aided distance information. In: Proc. of IMC '10. pp. 22–34 (2010)
17. Rumín, R.C., Laoutaris, N., Yang, X., Siganos, G., Rodriguez, P.: Deep diving into bittorrent locality. In: Proc. of INFOCOM. pp. 963–971. IEEE (2011)
18. Slot, M., Costa, P., Pierre, G., Rai, V.: Zero-day reconciliation of bittorrent users with their isps. In: Proc. of Euro-Par 15. pp. 561–573. Springer-Verlag, Berlin, Heidelberg (2009)
19. Xie, H., Yang, Y.R., Krishnamurthy, A., Liu, Y.G., Silberschatz, A.: P4P: Provider portal for applications. In: Proc. of ACM SIGCOMM (2008)
20. Zhang, B., Liu, R., Massey, D., Zhang, L.: Collecting the internet as-level topology. SIGCOMM Comput. Commun. Rev. 35(1), 53–61 (2005)