NISTIR 7590

A Model of Joint Congestion Control and Routing Through Random Assignment of Paths

Fern Y. HUNT
Information Technology Laboratory
National Institute of Standards and Technology
Gaithersburg, Maryland 20899, USA

Vladimir MARBUKH
Information Technology Laboratory
National Institute of Standards and Technology
Gaithersburg, Maryland 20899, USA

February 16, 2010

Abstract

We investigate the trade-off between utility and path diversity in a model of congestion control where there can be multiple routes between two locations in a network. The model contains a random route allocation scheme for each source s where the degree of randomness and therefore path diversity is controlled by h_s, the entropy of the route distribution. After deriving the model from a network utility maximization problem we analyze it in detail for two sample topologies. We conclude that, starting from an allocation with maximum robustness and path diversity, one can always increase the utility by decreasing h_s. However it can only decrease until a critical value of the entropy is reached. The value depends on the topology and link capacities of the network and can be explicity computed for our examples.

Keywords:TCP/IP protocols, multipath routing,discrete dynamics, Morse theory

1 INTRODUCTION

In recent years, network protocols have been interpreted as algorithms that solve a convex optimization problem. In this set-up overall congestion control is formulated as a network utility maximization problem (NUM) that is solved in a distributed fashion by the various network layers. The network topology and the capacity of its links introduces constraints on the optimal solution. Then algorithms solve the problem by computing and modifying the primal and dual variables based on efficient communication between users, link and router layers of the network. Since the work of Kelly et al [7] congestion control protocols have been seen as regulating user network transmission rates so that the objective function i.e., the aggregate utility, is maximized subject to capacity constraints ([8], [3]). The utility function incorporates efficient utilization and fair allocation of resources among users. Significantly, such functions have been identified for existing protocols such as TCP and BGP (Border Gateway Protocols) through a process of "reverse engineering", thus opening up opportunities for analysis and improvement of existing protocols as well as the development of new ones.

The paradigm just described has been extended to the problem of characterizing protocols that jointly control congestion and routing ([12],[3], [4], [5],[6]). Aside from the obvious improvement in the utilization of network resources that could be gained by such an approach, there are obvious benefits in this time of cybersecurity concerns about adequate network robustness against route disruptions. However implicit in such a protocol is a tradeoff between robustness through path diversity on the one hand and network performance or utility on the other. Single path routing based on e.g. the OSPF (Open Shortest Path First) protocol can lead to route flapping instability and reduced utility. However splitting traffic equally across all paths regardless of cost would also imply reduced utility. The best tradeoff if it exists would have to navigate between these extremes. Thus the design of stable, implementable algorithms that achieve maximum aggregate utility remains a challenge.

Our purpose is to propose a generalized NUM optimization problem for joint congestion control and routing where we can examine the behavior of the time averaged utility as a function of entropy, a model parameter that can be interpreted as a measure of path diversity. This paper is not presenting an algorithm to be implemented on a real network but is rather, a theoretical study of issues that must be addressed by any algorithm that is based on the NUM problem we discuss. Our model contains a routing scheme based on random allocation where the degree of randomness is controlled. The randomness of any route or path allocation for a class of network users s, can be measured in terms of the entropy h_s of the probability distribution defined by the allocation.

Treating h_s as a parameter, we find that as we increase h_s from 0, the model equations fail to converge to an equilibrium until the entropy reaches a critical value (see Section 5.1 for TwoLinks and Section 6.1 for Diamond). Stable equilibria exist for values of h_s greater or equal to the critical value and we show that the corresponding equilibrium pair (x_s^*, β_s^*) are solutions of the optimization problem.

We also calculate the time averaged aggregate utility for a range of values h_s for each topology (see Figure 8, Figure 9). These graphs show how the utility decreases as the entropy increases and they illustrate the fact that one can always increase the utility for a route allocation that has entropy greater than the critical value by changing the allocation so that the entropy decreases. This automatically decreases the mean route cost. A protocol based on the model we presented does not allow one to decrease the entropy below the critical value. We conjecture that the behavior of the model for entropy values less than the critical value corresponds to the unstable behavior seen in the single path TCP/IP protocols as discussed in reference [12].

We now turn to a description of the organization of the paper. The NUM optimization problem is presented in the next section and is followed by a description of the model and the notation used in Section 3. The model is an algorithm for solving the NUM optimization problem and is derived following the work of [8], [12] in Section 4. Then the model is specialized to the TwoLinks topology (see (29)) in Section 5, and there follows a detailed description of the model behavior in 5.1-5.3. Readers who want to skip the details and see the results and conclusions can go to the beginning of sections 5 and 6. The model equations for the Diamond topology (equations (38)-(42)) can be found in Section 6. A summary of the model behavior for this topology can be found at the beginning while details can be found in the subsections 6.1-6.3. A discussion of how the aggregate utility varies as a function of the parameter h_s for these two topologies can be found in Section 7. The conclusion can be found in Section 8. Finally the last section, Section 9 is an appendix, containing proofs of some results needed in the detailed subsections of 5 and 6.

2 OPTIMIZATION PROBLEM

Consider a planar network to be a graph with nodes representing physical nodes in the network and edges representing links. The link capacities are defined by the vector $\mathbf{c} = (c_1, c_2, \cdots, c_L)$. A user who requires bandwidth to transmit from one node to another in the network (or a single TCP session between those two nodes) is indexed by s, the index of the source-destination pair. For each s , the utility function is a twice differentiable strictly concave function $U_s : [m_s, M_s] \to \mathbb{R}$. Here m_s and M_s are lower and upper bounds respectively on the bandwidth rate x_s. $U_s(x_s)$ measures the degree of user satisfaction, network fairness and efficiency for the part of the network defined by s. Users in a source-destination class s are assigned a path by edge routers using a probability distribution with a minimum degree of randomization as specified by h_s. The distribution is constrained so that the traffic on any link does not exceed the link capacity.

The optimization problem that our protocol seeks to solve is:

$$\max_{\beta \geq 0, \mathbf{x} \in X \, \mathbf{x} \geq 0} \sum_s U_s(x_s) \tag{1}$$

$$\sum_s \sum_{r \in R_s(l)} \beta_{sr} x_s \leq c_l \tag{2}$$

$$\forall s \sum_{r \in R_s} \beta_{sr} = 1, \quad \beta_{sr} \geq 0 \tag{3}$$

$$- \sum_{r \in R_s} \beta_{sr} \log \beta_{sr} \geq h_s \tag{4}$$

$\mathbf{x} = \{x_s : s = 1, 2, \cdots S\}$ is the vector of source rates with each $x_s \in [m_s, M_s]$ where $m_s \geq 0$. In this paper we will take $m_s = 0$ for all s and $U_s(x_s) = w_s(1 - \alpha_s) x_s^{1-\alpha_s}$ with $\alpha_s = 2$. The form for U is commonly used in the literature [3]. The elements $\beta = \{\beta_{sr}\}$ comprise a matrix whose rows are the route probability distributions for source s. The constraints (2) require that all routes r in $R_s(l)$, the set of routes of source s that use link l, be assigned bandwidth rates $\beta_{sr}x_s$ so that the total link load does not exceed the capacity c_l. Finally if R_s is the set of all paths joining the source-destination pair s, then (3) is the usual requirement for probability distributions and (4) places an lower bound on the degree of randomness for the distribution $\{\beta_{sr}\}$ for source s. Small values of h_s indicate that the problem is closer to the single path case and as h_s increases we are closer to the equiprobable case- a version of the multipath problem. Following the approach of [8] we developed an algorithm based on a gradient projection iteration method for the dual optimization problem. We will discuss the convergence of this algorithm and its stability as a function of h_s for two sample topologies. As in [8] and [12], the dual variables are link costs changing in response to congestion. The route allocation distributions follow dynamics that minimize the average route cost.

In [9] where the idea of random route allocation was first proposed, the authors introduced an adaptive algorithm involving non-constant values of h_s. Through simulations they demonstrated the trade-off between stability and utility. They found that the maximum utility occurs near the boundary between stability and instability.

3 DESCRIPTION OF MODEL

To describe the results of this paper, the model equations are presented next. They are based on a gradient projection algorithm for solving the dual form of a NUM problem and they are derived in Section 4. The utility funcion $U_s : [m_s, M_s] \to \mathbb{R}$ where $0 < m_s < M_s$, is a strictly concave, twice continuously differentiable function.

We will as usual represent a computer network as a planar graph with nodes representing locations in the network. In the literature, the total amount of bandwidth to send information from one part of the network to another is controlled by an agent called a source. There are several routes a source can take to its destination and we assume each route is defined by a set of uni-directional links indexed by $l = 1, \cdots L$. These links are represented in the graph as weighted edges, where an edge weight is the link capacity. Sources are indexed by s, and traffic is assigned to route $r \in R(s)$ with probability β_{rs} where $R(s)$ is the set of all routes used by source s. If $p_l^{(k)}$ is the link cost at time k and c_l is the capacity of the lth link then the equations of the model are

$$p_l^{(k+1)} = \left[p_l^{(k)} - h \left\{ c_l - \sum_s x_s(k) \sum_{r \in R_s(l)} \beta_{sr}^{(k)} \right\} \right]^+ \quad l = 1, \cdots L \tag{5}$$

$$\beta_{sr}^{(k)} = \exp(-\gamma_s(k) d_r(k))/Z_s(k) \tag{6}$$

where $d_r(k) = \sum_{l \in r} p_l^{(k)}$ is the cost of route r at time k, h is a step size , $[a]^+ = a$ if $a > 0$ and is 0 otherwise. $Z_s(k) = \sum_{r \in R(s)} exp(-\gamma_s^{(k)} d_r(k))$ is the normalization factor for the route distribution and, the variable $\gamma_s^{(k)}$ is the solution of the implicit equation,

$$\gamma_s^{(k)} D_s(k) + \log(Z_s(k)) = h_s, \quad D_s(k) = \sum_{r \in R(s)} \beta_{sr}^{(k)} d_r(k) \tag{7}$$

The model equations are completed by a relation between the bandwidth rate $x_s(k)$ and $D_s(k)$, the mean route cost at time k for positive constants w and M.

$$x_s(k) = \min \left(\left(\frac{w}{D_s(k)} \right)^{1/2}, \ M \right) \tag{8}$$

Equations (6) and (7) force the route distribution $\beta_{sr}^{(k)}$ to be the unique distribution of entropy h_s with the smallest mean route cost at each time step k. Recall that for any probability distribution $\beta_{\mathbf{s}} = \{\beta_{sr}\}_{r \in R(s)}$ the entropy of the distribution is

$$\mathsf{H}(\beta_{\mathbf{s}}) = - \left(\sum_{r \in R(s)} \beta_{sr} \log \beta_{sr} \right). \tag{9}$$

Thus the condition on the route distributions is $\mathsf{H}(\beta_{\mathbf{s}}) = h_s$. This constant entropy requirement will constrain the set of values $\{p_l \mid l = 1 \cdots L\}$ for which a bounded $\gamma_s^{(k)}$ exists. The precise set depends on the network topology and capacity of the links. In this paper we present a detailed discussion of two very simple sample networks where these regions can be determined.

4 DERIVATION OF THE MODEL

The problem (1)-(4) is not convex in (x, β) but through an invertible change a variables $(x, \beta) \mapsto (x, y)$ it can be transformed to one. The new problem is:

$$\max_{x,y} \sum_s U_s(x_s) \tag{10}$$

$$\sum_s \sum_{r \in R_s(l)} y_{sr} \leq c_l \tag{11}$$

$$x_s = \sum_{r \in R_s} y_{sr}, \quad y_{sr} \geq 0 \tag{12}$$

$$-\sum_{r \in R_s} \frac{y_{sr}}{x_s} \log \frac{y_{sr}}{x_s} \geq h_s \tag{13}$$

This is a convex program because the functions in (10)-(13) are clearly concave. If we assume the conditions

$$\sum_{s'} m_{s'} < c_l$$

hold for each l, where the summation is over all source-destination pairs s' that use link l, and note that a $\beta = \{\beta_{sr} : r \in R_s, s = 1 \cdots S\}$ can always be found so that $y_{sr} = x_s * \beta_{sr}$ and

$$-\sum_{r \in R_s} \beta_{sr} \log(\beta_{sr}) \geq h_s$$

then there is a Slater point([11]) and the Slater constraint qualification is satisfied. We note here that $x_s \in [m_s, M_s]$ can be chosen independently of β. Thus (10)-(13) is superconsistent and Lagrange multipliers exist. The Lagrangian for the problem is:

$$\begin{aligned} \mathfrak{L}(x, y\lambda, p) = \sum_s [U_s(x_s) - \sum_{r \in R_s} \left(\sum_{l \in r} p_l \right) y_{sr} - \lambda_s^1 \left(\sum_{r \in R_s} y_{sr} \log(\frac{y_{sr}}{x_s}) + h_s x_s \right) \\ -\lambda_s^2 \left(x_s - \sum_{r \in R_s} y_{sr} \right) + \sum_{l=1}^{NL} p_l c_l \end{aligned} \tag{14}$$

where $\{p_l, \ l = 1 \cdots NL\}$ are the Lagrange multipliers for (11) and $\lambda = [\lambda_s^1, \lambda_s^2 : s = 1, \cdots S]$ are Lagrange multipliers for constraints (13) and (12). To obtain solutions x, y we consider the problem;

$$\max_{x,y} \mathfrak{L}(x, y, \lambda, p) = \max_x \max_y \mathfrak{L}(x, y, \lambda, p)$$

where maximization is first performed in y with x and p held fixed. The problem $\max_y \mathfrak{L}(x, y, \lambda, p)$ is equivalent to the constrained problem

$$\max_y \sum_s U_s(x_s) - \sum_{r \in R_s} (\sum_{l \in r} p_l) y_{sr} \tag{15}$$

$$x_s = \sum_{r \in R_s} y_{sr}, \quad y_{sr} \geq 0$$

$$-\sum_{r \in R_s} \frac{y_{sr}}{x_s} \log(\frac{y_{sr}}{x_s}) \geq h_s$$

Problem (15) has a Slater point so it too is superconsistent. A solution y^* to (15) exists if and only if there is a λ^* that satisfies the Karush-Kuhn-Tucker conditions ([11] p.182,183). Reintroducing $\beta_{sr} = \frac{y_{sr}}{x_s}$ where β_{sr} is independent of x_s, we obtain the following form of the optimal solution of (15) for each fixed x_s and p.

$$\beta_{sr} = \frac{\exp(-\gamma * d_r)}{Z_s} \tag{16}$$

where

$$Z_s = \sum_{r \in R_s} \exp(-\gamma * d_r), \quad d_r = \sum_{l \in r} p_l$$

and $\gamma = \frac{1}{\lambda_s^2}$ is the solution of the implicit equation,

$$\gamma * D_s + \log Z_s(\gamma) = h_s \ , \quad D_s = \sum_{r \in R_s} \beta_{sr} d_r \tag{17}$$

Recognizing that (15) is equivalent to the problem of minimizing $\sum_{r \in R_s} y_{sr} d_r = x_s D_s$ subject to the constraints, we see that the optimal solution can be obtained by setting γ equal to the unique positive root of (17). Indeed there are in general two roots of this equation when $h_s < \log |R_s|$ and the positive root provides the smallest value of D_s. Given $\gamma > 0$, call any β_{sr} subject to (16) and (17), β^*_{sr}. Then we have $y^*_{sr} = x_s \beta^*_{sr}$. The $\{y^*_{sr} : r = 1, \cdots R_s\}$ are unique since for each s, (17) has a unique positive solution. Set

$$L(x,p) = \max_y \mathfrak{L}(x,y,\lambda,p) = \mathfrak{L}(x, y^*(x,p), \lambda^*(x,p), p)$$

where $y^*(x,p)$ and $\lambda^*(x,p)$ are the optimal solution and Lagrange multipliers of (15) for fixed x and p. Using the KKT conditions L can be written as,

$$L(x,p) = \sum_s [U_s(x_s) - x_s \sum_{r \in R_s} (\sum_{l \in r} p_l) \beta^*_{sr}] + \sum_{l=1}^{NL} p_l c_l \tag{18}$$

L is strictly concave on the product interval $X = \prod_s [m_s, M_s]$ and therefore $\max_x L(x,p)$ exists for a unique x. Moreover L is convex in p. To see this note that

$$-\Lambda_s(p) = -\sum_{r \in R_s} (\sum_{l \in r} p_l) \beta^*_{sr} \geq -\sum_{r \in R_s} (\sum_{l \in r} p_l) \beta_r \tag{19}$$

and is the supremum over all vectors β; $\beta : \sum_{r \in R_s} \beta_r = 1, -\sum_{r \in R_s} \beta_r \log \beta_r \geq h_s$. Thus $L(x,p)$ is the supremum of set of convex functions over an infinite (convex) set (p. 81 Boyd and Vandenberghe). Applying the same reasoning to L we can conclude that $\mathcal{Q}(p) = \max_x L(x,p) = \max_x \max_y \mathfrak{L}(x,y,\lambda,p)$ is convex in p. The remainder of this derivation follows the work of [8] and [12]. First the dual problem for p is formulated as:

$$\min_{p : p_l \geq 0} \mathcal{Q}(p) \tag{20}$$

If $\{\beta^*_{sr}\}_{s,r}$ is differentiable in p and λ, then Danskin's theorem [1] applied to $\mathfrak{L}$ implies the differentiability of $\mathcal{Q}$. Indeed the unique maximizing $\bar{x}$ and $y = x\hat{\beta^*}$ are differentiable in p and λ. At $\bar{x}$, either $\bar{x_s}(p) = M_s$ or $U_s^{'}(\bar{x_s})(p)) = \Lambda_s(p)$. If $\bar{x_s}(p) < M_s$, then a condition introduced in [8], $U^{''}(x_s) \geq \delta_s > 0$ for all $x_s \in (0, M_s]$, together with the differentiability of Λ_s with respect to p and λ, imply that $\bar{x}_s$ is differentiable(see Appendix) and thus $\mathfrak{L}(\bar{x}, y^*, \lambda, p)$ is differentiable. We conjecture that Λ_s is differentiable at p if p lies in a convex subset of $\mathcal{R}^L$. The exact nature of the subset depends on the existence of a finite positive root of (17). We have determined the region for sample problems discussed in this paper. In general however we do not expect $\mathcal{Q}$ to be everywhere differentiable. These considerations lead to the following subgradient algorithm for solving (20).

$$p^{k+1} = [p^k - hg^k]^+ \tag{21}$$

If $x_s(p) < M_s$, for all s, where g is a subgradient of $\mathcal{Q}$. When it is differentiable $g^k = \bigtriangledown \mathcal{Q}(p^k)$. Our interest lies in understanding the dynamics of the latter equation when this condition is satisfied at least initially and the boundary of X can only be approached from the interior thus for the rest of our discussion we assume that for each k:

ASSUMPTION For all s , $\bar{x_s}(p^k) < M_s$.

Under this assumption, (21) can be written as

$$p_l^{k+1} = \left[p_l^k - h \left\{ c_l - \sum_s x_s^k \sum_{r \in R_s(l)} \beta^k_{sr} (\sum_{l' \in r} p^k_{l'}) \right\} \right]^+ \quad l = 1 \cdots NL \tag{22}$$

Equations (22) is supplemented by the relations coming from the optimality conditions.

$$\beta^k_{sr} = \frac{\exp\left(-\gamma^k_s (\sum_{l \in r} p_l)\right)}{Z(\gamma^k_s)} \tag{23}$$

$$x^k_s = \min(\left(\frac{w}{\mathbb{E}[d]}\right)^{\frac{1}{\alpha}} M) \ , \quad \mathbb{E}[d] = \sum_{r \in R_s} \beta^k_{sr} (\sum_{l' \in r} p^k_{l'}) \tag{24}$$

where γ^k_s is the solution of (17) using β^k_{sr} , and $Z(\gamma^k_s)$ is defined analogously to $Z(s)$. Equation (17) effectively insures that at each time step k the route distribution $\left\{ \beta^{(k)}_{sr} \ : \ r \in R_s \right\}$ has entropy h_s for every s.

5 TWO LINKS TOPOLOGY

A network consisting of a single source-destination pair of nodes connected by two links is depicted in Figure 1. In this section we describe the dynamics of equations (22) particular to this case as seen in equation (29). We find the values of the entropy h_s where the iterations converge and these can be used to express the solution of the original optimization problem. As h_s is varied it will be convenient from a conceptual as well as mathematical point of view to express the variation in entropy in terms of the route allocation distribution.

In Section 5.1, we discuss the existence of stable equilibria for (29) at an intrinsic (critical) entropy value with route probabilities,

$$\beta_1^* = \frac{c_1}{c_1 + c_2} \;,\; \beta_2^* = \frac{c_2}{c_1 + c_2} \tag{25}$$

where c_1 and c_2 are the link capacities and without loss of generality we assume $c_1 > c_2$. The corresponding critical value of the entropy is therefore

$$h_T(c) = -\left[\frac{c_1}{c_1 + c_2}\log\frac{c_1}{c_1 + c_2} + \frac{c_2}{c_1 + c_2}\log\frac{c_2}{c_1 + c_2}\right] \tag{26}$$

$$\tag{27}$$

Each initial point in a region $\mathcal{H}$ in the p_1, p_2 plane, converges along a straight line orbit to a point on the line of constant mean cost $\bar{x^*}$ (see the line L_1 Figure 2). Although there are infinitely many such points in the plane there is a unique equilibrium bandwidth rate $xs^* = \left(\frac{w}{x^*}\right)^{1/2}$. These equilibria are neutrally stable.

There is a corresponding equilibrium route allocation as well. In the latter part of our discussion here and Theorem 9.1 in the Appendix we show that there are just two probability distributions with a given entropy $h_s < \log 2$. It is shown that within the quadrant sector $\{\mathsf{p} = (p_1, p_2,) \;:\; p_1 < p_2\ \}$, the route probabilities $\beta_1^{(k)} > \beta_2^{(k)}$ are constant and equal to the values in (25) for $k \geq 1$. Thus the optimal route probabilities are reached in a single step although of course the solution of the implicit equation varies and depends on $\mathbf{p}$. The reverse inequality holds in the other quadrant sector. The particular choice of solution depends on the sector the intiial values of $\mathbf{p}$ are in. The proof that the equilibrium values (xs^*, β^*) are solutions of the optimization problem for h small enough is omitted here but follows along the same lines as the proof of Theorem 1 in reference [8] and page 214 of [1], because orbits of (29) remain in a single sector so the route probabilities are constants. Therefore the equations take the same form as those discussed in [8]. We note that it is also at this point that the lower bound on U'' is needed.

If $h_s \neq h_T(c)$, then the route probabilities in analogy with equation (25) are:

$$\beta_1^{(hs)} = \frac{c_1 + \mu c_2}{(1+\mu)c_2 + c_1} \quad \beta_2^{(hs)} = \frac{c_2}{(1+\mu)c_2 + c_1} \tag{28}$$

where $\mu \neq 0$. If $\mu = 0$ then $h_s = h_T(c)$, while $\mu < 0$ implies that $h_s > h_T(c)$. The dynamics of the iterations for these values are discussed in Section 5.2. There it is proved, that if initial points start in $\mathcal{H}$, iterates converge to a asymptotically stable equilibrium point on the p_2 axis. Figure 3 shows some of these orbits. This equilibrium corresponds to a solution of the optimization problem but here the utility, i.e. $U(x_s^*)$ is smaller than the value obtained at $h_T(c)$. The discussion of TwoLinks ends in Section 5.3 where we treat the case $\mu > 0$. Here $h_s < h_T(c)$. The algorithm does not converge but instead orbits approach the line $p_1 = p_2$ and the solution of the implicit equation becomes unbounded.

We now introduce the equations for the TwoLink topology and establish some preliminary facts needed for their analysis in the later subsections. We are interested in identifying sufficient conditions for the existence of equilibria for our algorithm in the region where the dual function $\mathcal{Q}$ is differentiable. In this topology note that a finite solution of (17) fails to exist if $h_s < \log 2$ but $p_1 = p_2$. Therefore we will restrict our investigation of (22) to the set $\{p = \langle p_1\ ,\ p_2\rangle \;:\; p_1 < p_2\}$. We will denote by xs the bandwidth rate with the single source in this network. In this setting equations (22) become:

$$\begin{aligned} p_1^{k+1} &= \left[p_1^k - h\left\{c_1 - \beta_1^k xs(k)\right\}\right]^+ \\ p_2^{k+1} &= \left[p_2^k - h\left\{c_2 - \beta_2^k xs(k)\right\}\right]^+ \end{aligned} \tag{29}$$

Here $\beta_1^{(k)}$ is the probability of being assigned to route 1 at the kth step and $\beta_2^{(k)} = 1 - \beta_1^{(k)}$. Now the entropy of a two point distribution is defined as

$$\mathsf{H}(\beta) = -\left(\beta\log(\beta) + (1-\beta)\log(1-\beta)\right)$$

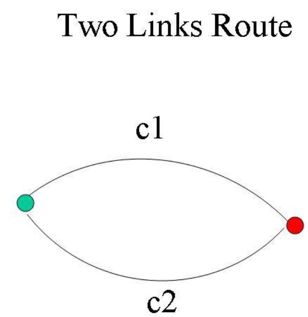

Figure 1:

The equation $\mathsf{H}(\beta) = h_s$, for some fixed $h_s < \log 2$, is satisfied by exactly two distributions $\hat{\beta} = \{\beta^{(hs)}, 1-\beta^{(hs)}\}$ or $\{1-\beta^{(hs)}, \beta^{(hs)}\}$. Since the left hand side of (17) can be interpreted as $\mathsf{H}(\hat{\beta}^k)$, where $\hat{\beta}^{(k)} = \{\beta_1^{(k)}, \beta_2^{(k)}\}$, this means there are only two possible values of $\beta_1^{(k)}$ ($\beta_2^{(k)}$ respectively).The optimality conditions require that the positive root be selected, so we conclude that $\beta_1^{(k)} \equiv \beta^{(hs)}$ where $\beta^{(hs)} > (1-\beta^{(hs)})$ if $p_1^{(0)} < p_2^{(0)}$ and the reverse inequality when $p_1^{(0)} > p_2^{(0)}$ (see Theorem 9.1 in the Appendix).

In what follows the sets $W_i = \{p_i | p_i - h\{c_i - \beta_i xs\} > 0\}$ will be mentioned frequently.

5.1 Route Distribution defined by Link Capacities

Given capacities $c_1 > c_2$ where c_1 and c_2 are the capacities of link 1 and link 2 respectively suppose,

$$\frac{c_1}{c_2} = \frac{\beta_1^{(h_s)}}{\beta_2^{(h_s)}} = \lambda > 1. \tag{30}$$

Here $\beta_1^{(h_s)}$ and $\beta_2^{(h_s)}$, the proportion of traffic allocated to links 1 and 2 respectively define a probability distribution with entropy $h_s = h_T(c)$ given by (27). The entropy and thus the possible path allocation probabilities are determined by the capacities of the links and are therefore intrinsic to the network. The ratio in (30) is greater than 1 so we make the choice of assigning the smaller initial price to link 1, i.e. $p_1^{(0)} < p_2^{(0)}$. We also make the realistic assumption that $M > \frac{c_1}{\beta_1^{(hs)}}$. There exists a family of steady states (equilibrium values) $(xs^*, p_1^*, p_2^*, \beta_1^{(h_s)}, \beta_2^{(h_s)})$. Here $xs^* = \frac{c_2}{\beta_2}$ is the unique equilibrium source rate, where p_i^* are the link costs for $i = 1, 2$ such that $\beta_1^{(hs)} p_1^* + \beta_2^{(hs)} p_2^* = \bar{x^*}$, where $xs^* = \left(\frac{w}{\bar{x^*}}\right)^{1/2}$. The equilibrium values of p_i are not unique as we will see. To simplify notation in what follows we write $\beta_i^{(hs)} = \beta_i \ \ i = 1, 2$.

The dynamics of (29) are most conveniently described by dividing the (p_1, p_2) plane into 3 regions. Let $\mathcal{L}_0$ be the line perpendicular to the line $\mathcal{L}_1 : \bar{x} = \bar{x^*}$ at the point $(0, \tilde{p_2})$. It is clear that $\tilde{p_2} = \frac{\bar{x^*}}{\beta_2}$. Similarly let $\mathcal{L}_2$ be the line perpendicular to $\mathcal{L}_1$ at the point of intersection of the lines $\mathcal{L}_1$ and $\{p : p_1 = p_2\}$. We will begin with the region bounded above by $\mathcal{L}_0$, and below by $\mathcal{L}_2$ and on the left by the p_2 axis. Denoting this region by $\mathcal{R}$ we assume that initial value $p^0 = (p_1^{(0)}, p_2^{(0)})$ is in $\mathcal{H} = \mathcal{R} \cap W_1 \cap W_2 \cap \{p : p_2 > p_1\}$. As long as $p_2^{(k)} > p_1^{(k)}$, $\beta_i^{(k)} = \beta_i \ i = 1, 2$ (see the discussion at the end of the introductory part Section 5). Equations (29) and (30) imply that $p_2^{(k)} - p_1^{(k)}$ increases so this inequality holds while $p_i^{(k)} \in W_i$. The orbit of p proceeds in a direction parallel to $\mathcal{L}_0$, i.e. perpendicular to $\mathcal{L}_1$. To see this let, Δp equal the vector $< \ h(\beta_1 xs(k) - c_1), \ h(\beta_2 xs(k) - c_2) >$. Then the dot product with $< \ -\beta_2, \ \beta_1 \ >$ is $\Delta p \cdot < \ -\beta_2, \ \beta_1 \ >= 0$. If p_1 leaves W_1, (p_2 cannot leave W_2 because $p_2 > \beta_2 \bar{x}^*$), then the orbit will move to the p_2 axis and $p_2^{(k)} \to \tilde{p_2}$ as $k \to \infty$. If p_1 remains in W_1 we will show that $p^{(k)}$ will exponentially converge to a point on the line $\mathcal{L}_1$. From (29) we derive the following equation for the evolution of $\bar{x}$ for $p_i^{(k)} \in W_i \ \ i = 1, 2 \ \ k \geq 1$.

$$\bar{x}^{k+1} = T(\bar{x}^{(k)}) = \bar{x}^k + h\beta_2^{(hs)}(1+\lambda^2)\left(xs(k)\beta_2^{(hs)} - c_2\right) \tag{31}$$

First suppose that $xs(0) < xs^*$, ($\bar{x}^0 > \bar{x}^*$). Equation (31) shows that $\bar{x}$ decreases as long as $xs^{(k)} \leq xs^*$ and is fixed when $xs^{(k)} = xs^*$ ($\bar{x}^{(k)} = \bar{x}^*$). Moreover, an application of the the Mean Value theorem shows the the map T is a uniform contraction in within the interval, $[\bar{x}^*, \bar{x}^{(k)}]$ and the approach to $\bar{x}^*$ is exponential. In fact we have,

$$\bar{x}^{(k+1)} - \bar{x}^* = (\bar{x}^{(k)} - \bar{x}^*) - \frac{h\beta_2^2}{2w}(\frac{w}{\xi})^{3/2}(\bar{x}^{(k)} - \bar{x}^*) \tag{32}$$

where $\xi \in [\bar{x}^*,\ \bar{x}^{(k)}]$. Thus $p^{(k)}$ converges to a point $\hat{p}$ on $\mathcal{L}_1$ that is the intersection of $\mathcal{L}_1$, with the line perpendicular to it and passing through through $p^{(0)}$. It is clear then, that the equilibrium values are not unique. However the equilibrium value of the bandwidth xs^* and therefore the equilibrium mean cost $\bar{x}^*$ is unique. If $xs(0) > xs^*$, then $\bar{x}$ increases and eventually reaches an interval where T is a contraction i.e. (32) holds with $\xi \in [\bar{x}^{(k)},\ \bar{x}^*]$. Since $p^{(0)} \in W_1 \cap W_2$ the orbit remains in this set because $p_1^{(k)}$ and $p_2^{(k)}$ are both increasing while $c_2 - \beta_2 * xs^{(k)}$ decreases with k. As in the previous case, $p^{(k)}$ exponentially approaches a point on $\mathcal{L}_1$. It is not hard to show by local stability analysis that the limiting point is neutrally stable along the $\mathcal{L}_1$ direction (with eigenvalue 1) and is stable along the $\mathcal{L}_0$ direction, with eigenvalue less than 1 in absolute value.

Let us now consider the cases where $p^{(0)} \in W_1 \cap W_2$ is located outside of $\mathcal{H}$. If the initial point is above (and to the left of) $\mathcal{L}_0$, then the orbit of p lies along a straight line perpendicular to $\mathcal{L}_1$ and approaches a point on the p_2 axis. At some time k, $p_1^{(k)} \notin W_1$, then $p_1^{(k+1)} = 0$. Since $xs^{(m)} \leq xs^*$, for $m \geq k$, then $p_1^{(m)} = 0$ for $m \geq k+1$ and the orbit is confined to the p_2 axis. $p_2^{(m)} \to \tilde{p}_2$ as $m \to \infty$. If $p^{(0)}$ is located to the right of the line $\mathcal{L}_2$ but still satisfies, $p_2^{(0)} > p_1^{(0)}$, then the orbit moves in a direction perpendicular to $\mathcal{L}_1$. As in previous cases however, it will reach the $p_1 = p_2$ before it reaches $\mathcal{L}_1$ so that the orbit fails to converge to an equilibrium.

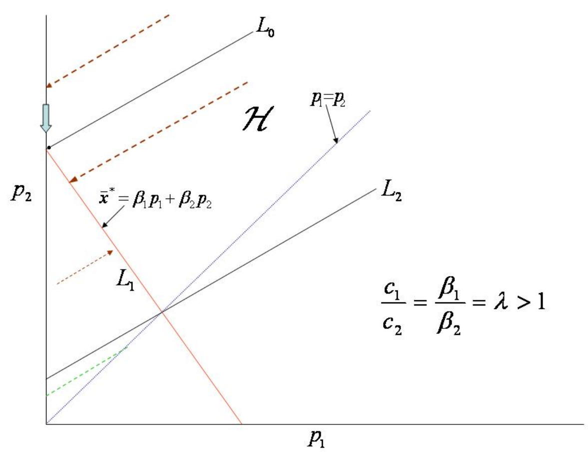

Figure 2: (p_1, p_2) plane when $h_s = h_T(c)$

5.2 Route Distribution with Less Weight on the Larger Link (increased h_s)

We now consider the choice of parameters

$$\frac{c_1}{c_2} = \lambda \ \ , \frac{\beta_1^{hs}}{\beta_2^{hs}} = \lambda + \mu.$$

The parameter μ expresses the degree to which the route allocation distribution's entropy differs from what we can call an intrinsic entropy defined in (27). In this instance λ is fixed by the intrinsic capacities of the links but the controller can vary traffic allocation by varying μ. The first case we consider, $\mu < 0$, decreases the portion of traffic assigned to link 1. As β_1 decreases while maintaining the inequality $\beta_1 > \beta_2$, we have $h_s > h_T(c)$ and the entropy increases. It approaches the distribution that assigns 1/2 to each i.e. $h_s = \log 2$. We assume that $p_i^{(0)} \in W_i,\ i = 1, 2$. If $p_i^{(k)} \in W_i\ 1 \leq k \leq K\ \ i = 1, 2$,where K, is an arbitrary but fixed positive integer, the right hand sides of of the two equations in (29) can be used to show the following:

Lemma 5.1 *If* $p_1^{(0)} < p_2^{(0)}$ *and* $xs^{(0)} \in \left[\frac{c_2}{\beta_2^{hs}}, \frac{c_1}{\beta_1^{hs}}\right]$, *then* $p_1^{(k)} < p_2^{(k)}$ *and* $xs^{(k)} \in \left[\frac{c_2}{\beta_2^{hs}}, \frac{c_1}{\beta_1^{hs}}\right]$ *for* $1 \leq k \leq K$.

Proof: We use mathematical induction on $0 \le l \le K-1$. The induction hypothesis is that $p_1^{(l)} < p_2^{(l)}$ and $xs^{(l)} \in \left[\frac{c_2}{\beta_2^{hs}}, \frac{c_1}{\beta_1^{hs}}\right]$ with $\beta_i^{(l)} = \beta_i^{(hs)}$, $i = 1,2$. Equation (29) then implies that $p_1^{(l+1)} < p_2^{(l+1)}$. Using the argument in the appendix, we then see that $\beta_i^{(l+1)} = \beta_i^{(hs)}$, $i = 1,2$. To show that $xs^{(l+1)} \in \left[\frac{c_2}{\beta_2^{hs}}, \frac{c_1}{\beta_1^{hs}}\right]$ we rewrite (29), simplifying the notation for β_i by dropping the hs superscript. Note that the right hand sides of (29) can be rewritten as $[p_1^{(l)} - h\{c_1 - \beta_1^{(l)} xs^* + \beta_1^{(l)}(xs^* - xs^{(l)})\}]^+$ and $[p_2^{(l)} - h\{c_2 - \beta_2^{(l)} xs^* + \beta_2^{(l)}(xs^* - xs^{(l)})\}]^+$ respectively. Here $xs^* = \theta \frac{c_2}{\beta_2}$ where $\theta = \frac{1+\lambda(\lambda+\mu)}{1+(\lambda+\mu)^2}$. Therefore after some algebra the equations for $p^{(l+1)}$ are:

$$p_1^{(l+1)} = p_1^{(l)} + \frac{h\mu c_2}{1+(\lambda+\mu)^2)} + h\beta_1^{(l)}(xs^{(l)} - xs^*) \tag{33}$$
$$p_2^{(l+1)} = p_2^{(l)} - \frac{h\mu(\lambda+\mu)c_2}{1+(\lambda+\mu)^2} + h\beta_2^{(l)}(xs^{(l)} - xs^*)$$

The fact that $\beta_i = \beta_i^{(l)} = \beta_i^{(l+1)}$ $i = 1,2$ is then used to derive an equation for the mean value $\bar{x}^{(l+1)}$,

$$\bar{x}^{(l+1)} = \left[\bar{x}^{(l)} + h(\beta_1^2 + \beta_2^2)(xs^{(l)} - xs^*)\right] \tag{34}$$

If $xs^{(l)} \ge xs^*$, then (34) shows that $\bar{x}^{(l+1)} \ge \bar{x}^{(l)}$ and thus $xs^{(l+1)} \le xs^{(l)}$. It is clear then that $xs^{(l+1)} \le \frac{c_1}{\beta_1}$. To see that $xs^{(l+1)} \ge \frac{c_2}{\beta_2}$ for h small enough, suppose the contrary, $xs^{(l+1)} < \frac{c_2}{\beta_2}$. We would then have

$$xs^{(l)} - xs^{(l+1)} > xs^{(l)} - \frac{c_2}{\beta_2} \ge xs^* - \frac{c_2}{\beta_2}.$$

We compute the last difference as $xs^* - \frac{c_2}{\beta_2} = \frac{-\mu(\lambda+\mu)}{1+(\lambda+\mu)^2}\frac{c_2}{\beta_2}$. However the differentiability and in particular the continuity of U' shows that $xs^{(l+1)} - xs^{(l)}$ can be made much smaller than this (it is $O(h)$) for h small enough). Indeed we have

$$|xs^{(l+1)} - xs^{(l)}| \le \mathsf{C} \max_{\bar{x_1^*} \le x \le \bar{x_2^*}} \left(\frac{w}{4x^3}\right)^{1/2} h$$

where $\bar{x_i^*} = \left(\frac{\beta_i w}{c_i}\right)^{1/2}$ for $i = 1,2$ and $\mathsf{C} = \max_{i=1,2} |\frac{c_i}{\beta_i} - xs^*|$. Thus the contradiction. A similar argument can be made when $xs^l < xs^*$. □

If $p_1^{(k)}$ remains in W_1 for a sufficiently long interval then $xs^{(k)}$ will be close enough to xs^* so the $(xs^{(k)} - xs^*)$ terms in the equations of (33) can be neglected. Note also that in this situation $p_2^{(k)}$ is always in W_2. We then have $p_1^{(k)}$ decreasing until it leaves W_1 and thereafter $p_1^{(k)} = 0$. Meanwhile $p_2^{(k)}$ increases as the orbit moves along the p_2 axis until $p_2 \to p_2^*$, where $\beta_2 p_2^* = \bar{x_2^*}$. This conclusion still holds true if $p_1^{(k)}$ leaves W_1 at some time we can assume to be $k = K+1$. Using the fact that $p_i^{(l)} \in W_i$ $0 \le l \le K$ we can show that $xs^{(k)} \in \left[\frac{c_2}{\beta_2}, \frac{c_1}{\beta_1}\right]$. This means that the orbit lands on the p_2 axis on the next step $k = K+2$ and $p_2^{(k)}$ continues to increase and approach p_2^*.

5.3 Distribution with More Weight on the Larger Link (decreased h_s)

We now discuss the case $\mu > 0$, where $h_s < h_T(c)$. First note that for any μ the equation for $p_1^{(k)}$ can be rewritten as,

$$p_1^{(k+1)} = \left[p_1^{(k)} - h((\lambda+\mu)\beta_2\{x_1^* - xs(k)\}\right]^+ \;,\;\; x_1^* = \frac{\lambda}{\lambda+\mu} c_2/\beta_2 \tag{35}$$

We have $x_1^* < xs^* < c_2/\beta_2$ since $xs^* = \theta\frac{c_2}{\beta_2}$ and $\theta < 1$. As before, we first assume that $p_i^{(0)} \in W_i$. If $xs(0) \in [x_1^*, \frac{c_2}{\beta_2}]$ then $p_1^{(k)}$ increases and since $p_2^{(k)} > p_1^{(k)}$, it follows that $p_i^{(k)} \in W_i$ $i = 1,2$ for all $k \ge 0$ and xs will approach xs^*. Rewriting (29) to reveal the behaviour of $p^{(k)}$ near xs^*, we have,

$$p_1^{(k+1)} = p_1^{(k)} + h\left\{\frac{\mu}{(\lambda+\mu)^2+1} + \beta_1(xs(k) - xs^*)\right\} \tag{36}$$

$$p_2^{(k+1)} = p_2^{(k)} - h\left\{\frac{\mu(\lambda+\mu)c_2}{(\lambda+\mu)^2+1} + \beta_2(xs(k) - xs^*)\right\}. \tag{37}$$

Therefore it follows that $p_1^{(k)}$ increases and $p_2^{(k)}$ decreases in this situation so that $p_2^{(k)} - p_1^{(k)} \to 0$ as $k \to \infty$. This means that $\gamma(k) \to \infty$, where $\gamma(k)$ is the solution of (17) at the kth step and the iterations fail to converge.

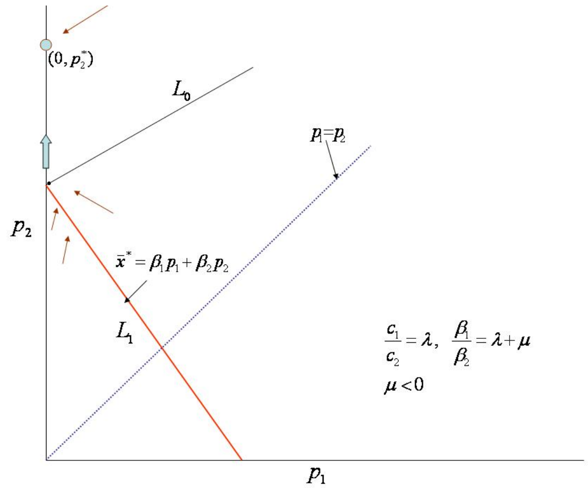

Figure 3: (p_1, p_2) plane when $h_s > h_T(c)$

If $xs(0) < x_1^*$, then $p_1^{(k)}$ and $p_2^{(k)}$ decrease. The dot product $\Delta p \cdot \langle -\beta_2, \beta_1 \rangle = -h\mu\beta_2 c_2$ shows that when $p_i^{(0)} \in W_i$ then $p_i^{(k)} \in W_i$ because in this region, the increment Δp is pointing away from the p_2 ($p_1 = 0$) axis. At $xs(k) = x_1^*$, $p_1^{(k)}$ remains fixed but $p_2^{(k)}$ decreases, i.e. Δp is pointing downward. $xs(k)$ increases until it enters the interval $[x_1^*, \frac{c_2}{\beta_2}]$, and approaches xs^*. Therefore p_2 decreases and p_1 increases. Consequently $p_2^{(k)} - p_1^{(k)} \to 0$ as in the previous case. If $xs^{(0)} > \frac{c_2}{\beta_2}$ then p_2 and p_1 both increase. Thus, $\bar{x}^{(k)}$ increases and $xs(k)$ decreases and eventually enters the interval $[\bar{x}_1^*, \frac{c_2}{\beta_2}]$. Suppose therefore that this is the case. If for some k, $p_1^{(k)} = 0$ then , since $p_2^{(k)}$ decreases and $p_1^{(k)}$ increases, either $p_2^{(k)} \to p_2^*$ while $p_1^{(m)}$ remains 0 for $m \geq k$ or $p_1^{(m)}$ becomes positive and is therefore in W_1. In the latter case we can use the previous arguments for the case $p_i^{(k)} \in W_i$ that show that $p_2^{(k)} - p_1^{(k)} \to 0$.

Our final case for the parameter choice $\mu > 0$ is to suppose that $xs(0) > \frac{c_2}{\beta_2}$. Clearly xs will decrease until it enters the interval $[x_1^*, \frac{c_2}{\beta_2}]$. Previous arguments can then be used to determine the fate of $p_i^{(k)}$, $i = 1, 2$. We in fact see the divergence of $\gamma^{(k)}$ in our numerical computations. The cases discussed in this section correspond to the selection of a distribution that allocates a larger fraction of traffic to link 1 than the allocation defined by (30)and the corresponding critical entropy in equation(27). Thus $h_T(c)$ defines the outer limit of how far one can decrease entropy in order to increase utility.

6 DIAMOND TOPOLOGY

We turn our discussion now to another simple topology (see Figure 5) with 5 links attached to a single source destination pair. As seen in the figure there are 3 routes, using the links indicated in the figure.

Inspection of the figure leads to the system:

$$p_1^{(k+1)} = \left[p_1^{(k)} - h\{c_1 - (\beta_1^{(k)} + \beta_3^{(k)})xs(k)\}\right]^+ \tag{38}$$

$$p_2^{(k+1)} = \left[p_2^{(k)} - h\{c_2 - \beta_2^{(k)} xs(k)\}\right]^+ \tag{39}$$

$$p_3^{(k+1)} = \left[p_3^{(k)} - h\{c_3 - \beta_1^{(k)} xs(k)\}\right]^+ \tag{40}$$

$$p_4^{(k+1)} = \left[p_4^{(k)} - h\{c_4 - (\beta_2^{(k)} + \beta_3^{(k)})xs(k)\}\right]^+ \tag{41}$$

$$p_5^{(k+1)} = \left[p_5^{(k)} - h\{c_5 - \beta_3^{(k)} xs(k)\}\right]^+ \tag{42}$$

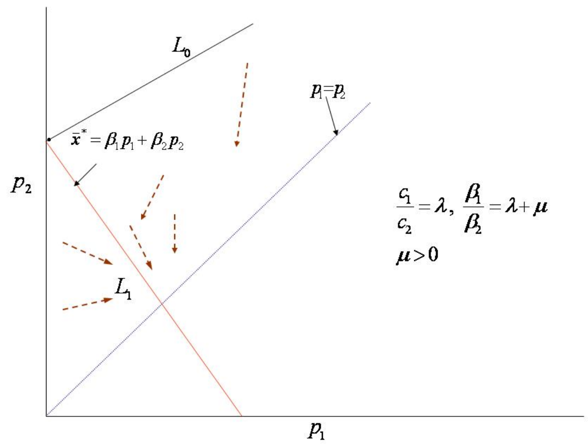

Figure 4: (p_1, p_2) plane when $h_s < h_T(c)$

where, as in Section 5, the bandwidth rate $xs(k) = \min(\left(\frac{w}{\mathbb{E}[d^{(k)}]}\right)^{\frac{1}{2}}, M)$, with $\mathbb{E}[d^{(k)}] = \sum_{r=1,2,3} \beta_r^k d_r^{(k)} = \sum_{r=1,2,3} \beta_r^{(k)} \left(\sum_{l' \in r} p_{l'}^k\right)$. The route probabilities are updated according to the following equations:

$$\beta_i^{(k+1)} = \frac{\exp\left(-\gamma^{(k)} d_i^{(k)}\right)}{Z(k)} \quad i = 1, 2, 3 \tag{43}$$

where $Z(k) = \sum_{i=1}^{3} \exp\left(-\gamma^{(k)} d_i^{(k)}\right)$ and

$$d_1^{(k)} = p_1^{(k)} + p_3^{(k)} \ , \quad d_2^{(k)} = p_2^{(k)} + p_4^{(k)} \ , \quad d_3^{(k)} = p_1^{(k)} + p_5^{(k)} + p_4^{(k)}. \tag{44}$$

As was the case with the twolinks topology, we will examine the dynamics for a special choice of capacities. Here, led by the search for strictly interior equilibria of (38)-(42), we take

$$c_2 = c_3, \ c_1 = c_3 + c_5, \ c_4 = c_2 + c_5. \tag{45}$$

The Subsections 6.1-6.3 contain discussions of the dynamics (38)-(42) for values of the route distribution entropy h_s in relation to the critical entropy for this topology,

$$h_D(c) = -\left[\frac{c_2}{c_2 + c_3 + c_5} \log \frac{c_2}{c_2 + c_3 + c_5} + \frac{c_3}{c_2 + c_3 + c_5} \log \frac{c_3}{c_2 + c_3 + c_5} + \frac{c_5}{c_2 + c_3 + c_5} \log \frac{c_5}{c_2 + c_3 + c_5}\right]. \tag{46}$$

Equation (47) expresses the corresponding route distribution. In Section 6.1, (see Figure 6), we prove the existence of a set of equilibria in (d_1, d_2, d_3) space that correspond to solutions of the optimization problem. Propositions 6.2 and 6.3 show that these values lie on the line $\bar{x^*} = \sum_{i=1}^{3} \beta_i d_i^*$, with β_i, the optimal route distribution given by (47). The equilibrium bandwidth rate is unique and can be expressed in terms of the equilibrium route costs by $xs^* = \frac{w}{x^*}^{1/2}$. These solutions are neutrally stable and do not persist when h_s is perturbed. New unique and stable equilibria for (38)-(42) arise when h_s is increased. Note however that the value of the aggregate utility will be less than the value achieved at $h_D(c)$. In sections 6.2 we study the dynamics of the iterations when $h_s > h_D(c)$. Changes in h_s are reflected in changes in the route allocation distribution as seen in equation (53), where $\nu < 1$. In Section 6.3, $h_s < h_D(c)$ and $\nu > 1$. It is shown there that the iterations fail to converge and in fact the line $d_1 = d_3$ is approached.

The proofs that the equilibrium values of subsections 6.1 and 6.2 are solutions of the original optimization problem again rely on the proof of Theorem 1 in reference [8] . The arguments below show that the route allocation distributions $(\beta_1^{(k)}, \beta_2^{(k)}, \beta_3^{(k)})$, are constant for $k \geq 1$ when $h_s \geq h_D(c)$. Thus the proofs of our statements closely follow those in [8] and [1] therefore they are omitted.

We now choose initial conditions for $p_i^{(0)}$, $i = 1 \cdots, 5$, that greatly simplify the dynamics as seen in the following result.

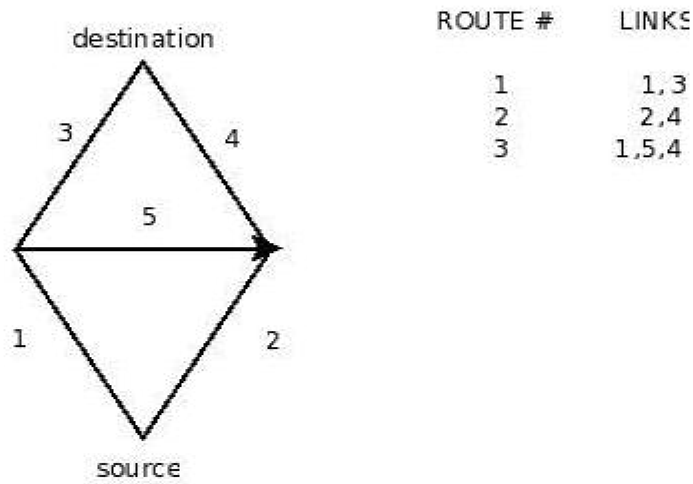

Figure 5: DIAMOND NETWORK

Lemma 6.1 *Suppose the capacities in (38)-(42) satisfy (45). Further suppose the initial conditions*

$$p_2^{(0)} = p_3^{(0)}, \quad p_1^{(0)} = p_3^{(0)} + p_5^{(0)}, \quad p_4^{(0)} = p_2^{(0)} + p_5^{(0)}$$

hold, with $p_i^{(0)} \in W_i, \ i = 1, ...5$. *Then*

$$p_2^{(k)} = p_3^{(k)}, \quad p_1^{(k)} = p_4^{(k)}, \, k \geq 1$$

and $\beta_1^{(k)} = \beta_2^{(k)}$ *for* $k \geq 1$.

Proof of Lemma: The proof is by induction. Suppose the conclusions of the lemma hold for some k. We then have $d_1^{(k)} = d_2^{(k)}$ by (44). Now at stage k ,$\gamma^{(k)}$ must satisfy the implicit equation(17) with $d_r = d_r^{(k)}$ $r = 1, 2, 3$. Since β_1 and β_2 satisfy (43), we must have $\beta_1^{(k+1)} = \beta_2^{(k+1)}$. The other conclusions follow from the right hand sides of (38)-(42). □

Remark: If $p_i^{(k)} \in W_i$ for $i = 2, 3, 5$, then $p_1^{(k+1)} = p_3^{(k+1)} + p_5^{(k+1)}$ and $p_4^{(k+1)} = p_2^{(k+1)} + p_5^{(k+1)}$. When $p_i^{(k)} \notin W_i$ for some $i = 2, 3, 5$ we cannot claim these equations are valid. Nevertheless it is always true that $p_1^{(k+1)} = p_4^{(k+1)}$.

As a corollary of Lemma 6.1 we can show that $\beta_i^{(k)}$ $i = 1, 2, 3$ is constant in k and therefore the dynamics of (38)-(42) are greatly simplified. To see this let us call the simplex of probability distributions on the three routes $\mathsf{S} = \left\{ \hat{\beta} = (\beta_1, \beta_2, \beta_3) \mid \quad \beta_i \geq 0 \ i = 1, 2, 3 \ , \ \beta_1 + \beta_2 + \beta_3 = 1 \right\}$. For a fixed $h_s \geq \log(2)$, the subset defined by $C = \left\{ \hat{\beta} = (\beta_1, \beta_2, \beta_3) \in \mathsf{S} \mid \quad \mathsf{H}(\hat{\beta}) \geq h_s \right\}$ is a closed convex region enclosed by a smooth simple closed curve as boundary. It is the constraint region of the optimization problem (3)-(4). Its boundary is the level curve $\ell = \left\{ \hat{\beta} \in \mathsf{S} \mid \ \mathsf{H}(\hat{\beta}) = h_s \right\}$. When $h_s < \log(2)$, the level curve breaks up into 3 pieces at the points where some $\beta_i = 0$ (see Figure 10). When $0 < h_s < \log 3$ there are precisely two points in C with distribution h_s that satisfy $\beta_1 = \beta_2$. Each point corresponds to one of the conditions $\beta_1 > \beta_3$ or $\beta_1 < \beta_3$. It is not hard to show that in the region where $d_3^{(k)} \geq d_1^{(k)} = d_2^{(k)}$, we have $\beta_1^{(k)} = \beta_2^{(k)} \geq \beta_3^{(k)}$ so that over any time interval where this inequality is satisfied, $\beta_j^{(k)}$ $j = 1, 2, 3$ is constant. When $h_s < \log 2$, $\beta_1 = \beta_2 > \beta_3$ occurs when $\beta_3 = 0$. Therefore the corresponding value of h_s must therefore be $\log 2$ which is infeasible.

Corollary 1 *Assume the hypotheses of Lemma 6.1 hold and let* h_s *be given with* $h_s > \log(2)$. *Further suppose that* $d_3^{(l)} > d^{(1)}$ *for* $1 \leq l \leq k$. *Then* $\hat{\beta}^{(l)} = \hat{\beta}^{(h_s)}$ *where* $\hat{\beta}^{(h_s)}$ *is the distribution with entropy* h_s *satisfying* $\beta_1^{(h_s)} = \beta_2^{(h_s)} \geq \beta_3^{(h_s)}$. *The latter inequality is strict when* $h_s < \log(3)$.

Remark: The dual function $\mathcal{Q}$ in (20) is differentiable in the region $\{\mathbf{p} \mid d_3 > d_1 \ \}$.

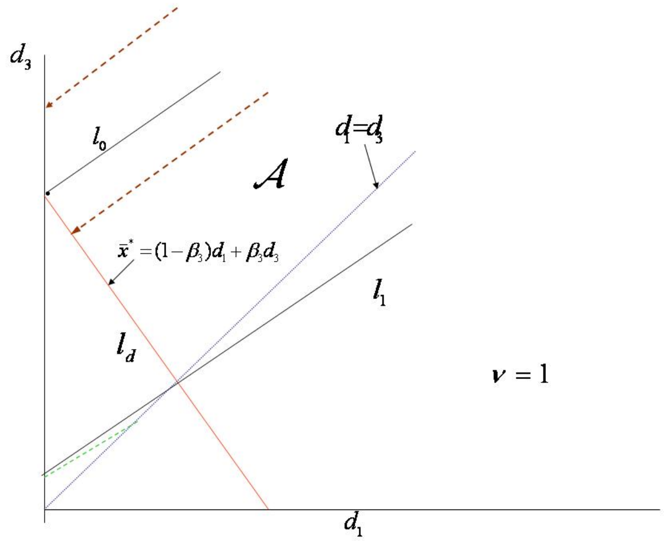

Figure 6: (d_1, d_3) plane for $h_s = h_D(c)$

6.1 Diamond Route Distribution Using Link Capacities

We assume a value of h_s for which

$$\hat{\beta}^{(h_s)} = \left\langle \frac{c_3}{c_2 + c_3 + c_5}, \frac{c_2}{c_2 + c_3 + c_5}, \frac{c_5}{c_2 + c_3 + c_5} \right\rangle \tag{47}$$

As was true with the twolinks topology, the values of xs for which $\Delta p_i^{(k)} = 0$ when $p_i^{(k)} \in W_i \;\; i = 1 \cdots 5$, will play an important role in defining the dynamical behaviour of the link and route costs. Here $W_i = \{ \; p_i \geq 0 \mid p_i - h\Delta p_i > 0 \; \}$ where Δ_i is the part of the expression on the right hand side of (38)-(42) inside the curly brackets. Since the entropy in (47) is expressed in terms of link capacities that satisfy (45) it is not hard to show that $xs_i^{(*)} = xs^{(*)} = c_2 + c_3 + c_5$ where $xs^{(*)}$ as in Section 5.1, is the equilibrium bandwidth rate. Equations for the route costs can be obtained from (38)-(42) when $p_i^{(k)} \in W_i$.

$$d_1^{(k+1)} = d_1^{(k)} - h\left[C_1 - xs(k)b_1\right] \tag{48}$$

$$d_3^{(k+1)} = d_3^{(k)} - h\left[C_3 - xs(k)b_3\right] \tag{49}$$

where $C_1 = c_1 + c_3$, $b_1 = \beta_1 + \beta_2 + \beta_3 = 1$, while $C_3 = c_1 + c_4 + c_5$, and $b_3 = \beta_1 + \beta_2 + 3\beta_3$ with $\beta_i = \beta_i^{(h_s)}$. Recall that Lemma 6.1 implies that $d_1^{(k)} = d_2^{(k)}$ so the equation for d_2 is unecessary. We assume that initial values of p_i begin in W_i. When (48)- (49) are valid it can be seen that for $\Delta d_i^{(k)} = d_i^{(k+1)} - d_i^{(k)} \;\; i = 1, 2, 3$, the dot product $< b_3, -b_1 > \cdot < \Delta d_1^{(k)}, \Delta d_3^{(k)} >$ is,

$$b_3 \Delta d_1^{(k)} - b_1 \Delta d_3^{(k)} = 2(1-\nu)(c_2 + c_3) \tag{50}$$

where in the present case $\nu = 1$ (see equation (47) and (53)). Let $\bar{x}^*$ be the mean route cost corresponding to $xs^* = \left(\frac{w}{\bar{x}^{(*)}}\right)^{1/2}$ and let l_d be the line in the d_1, d_3 plane given by $l_d: \;\; \beta_1 d_1 + \beta_2 d_2 + \beta_3 d_3 = \bar{x}^*$ (see Figure 6) It intersects the d_3 axis at the point $(0, \frac{\bar{x}^*}{\beta_3})$. Let l_0 be the line through this point, parallel to the vector $< -b_3, b_1 >$ and given by the equation,

$$l_0: \;\; b_1 d_3 - b_3 d_1 = b_1 \frac{\bar{x}^*}{\beta_3} \tag{51}$$

Finally, if $\mathcal{P}$ is the intersection of l_d with the line $d_1 = d_3$, let l_1 be the line passing through $\mathcal{P}$ and parallel to l_0. We call $\mathcal{A}$ the region in the d_1, d_3 plane bounded above by l_0 and below by l_1. Propositions 6.2 describes the dynamics of (48)-(49) when $p_i^{(k)} \in W_i \; i = 1 \cdots 5$ for all $k \geq 1$ and Proposition 6.3, discusses equations (38)- (42) when at some time k and for some i, $p_i^{(k)} \notin W_i$.

Proposition 6.2 *Assume $\nu = 1$ and suppose $\mathbf{p}^{(0)}$ is chosen so that $(d_1^{(0)}, d_3^{(0)})$ lies in $\mathcal{A} \cap \{(d_1, d_3) | d_3 \geq d_1\}$. Further suppose that $p_i^{(k)} \in W_i$ for all $k \geq 0$ and the hypotheses of Lemma 6.1 are satisfied. Then the orbit $\mathbf{d}^{(k)} = (d_1^{(k)}, d_3^{(k)})$ moves along a straight line through $\mathbf{d}^{(0)}$ parallel to l_0 , and converges to a point $\mathbf{d}^* = (d_1^*, d_3^*)$ on l_d. We have $\beta_1 d_1^* + \beta_2 d_2^* + \beta_3 d_3^* = \bar{x}^*$ with $d_i^* > 0$.*

Proof: Since $p_i^{(k)} \in W_i,\ k \geq 0$, we can compute the dot product $< -b_3, b_1 > \cdot < \Delta d_1^{(k)}, \Delta d_3^{(k)} >= 0$ for all k, by equation (50) where $\Delta d_i^{(k)} = d_i^{(k+1)} - d_i^{(k+1)}$ $i = 1, 2, 3$. Therefore iterates proceed in a direction parallel to $< b_1, b_3 >$, i.e. parallel to the line l_0. Since $p_i^{(k)} \in W_i$, we can write down the following equation for $\bar{x}^{(k+1)}$, the mean route cost at time $k + 1$,

$$\bar{x}^{(k+1)} = T\bar{x}^{(k)} = \bar{x}^{(k)} - h(\beta_1 b_1 + \beta_2 b_2 + \beta_3 b_3)(xs^* - xs(k)) \tag{52}$$

where $xs^* = c_2 + c_3 + c_5$ and $xs(k) = (\frac{w}{\bar{x}^{(k)}})^{\frac{1}{2}}$. As in Section 5.1 we can see that T is a uniform contraction when $\mathbf{d}^{(k)} \in \mathcal{A}$ so that $\bar{x}^{(k)} \to \bar{x}^*$, where $xs^* = (\frac{w}{\bar{x}^*})^{\frac{1}{2}}$. When $\mathbf{d}^{(k)}$ is above the line l_d, $d_1^{(k)}$ and $d_3^{(k)}$ decrease and approach the line from above. Conversely when $\mathbf{d}^{(k)}$ is below the line, the $d_1^{(k)}$, and $d_2^{(k)}$ both increase so the approach is from below. □

Remarks: Note that $d_1^{(k)} < d_3^{(k)}$ for all $k \geq 1$ in this situation so we can legitimately treat the $\beta_j^{(k)}$ $j = 1, 2, 3$ as constant.
If $(d_1^{(0)}, d_3^{(0)})$ is below the line l_1, the orbit will reach the line $d_1 = d_3$ before it reaches l_d and hence $\gamma(k)$ will diverge to ∞.

The fate of orbits beginning at points in the region $\mathcal{A} \cap \{(d_1, d_3) | d_3 \geq d_1\}$ where for some k $p_i^{(k)} \notin W_i$, is the same as that of orbits that begin above the line l_0 because in both cases, the orbits approach the d_3 axis. The next proposition discusses the latter situation. Here some of the link prices will converge to zero, an indication the traffic on those links have reached their capacity.

Proposition 6.3 *If $\mathbf{p}^{(0)}$ is chosen as in the previous proposition but $(d_1^{(0)}, d_3^{(0)})$ is above the line l_0, then $\mathbf{d}^{(k)}$ approaches the d_3 axis. Moreover:*

1. *if for some k, $p_1^{(k)} \notin W_1$, and $p_3^{(l)} \in W_3$ $l \geq 0$, then $\mathbf{p}^{(k)}$ converges to $\mathbf{p}^* = (0, p_2^*, p_2^*, 0, p_5^*)$, with $\beta_1 p_2^* + \beta_2 p_2^* + \beta_3 p_5^* = \bar{x}^*$.*
2. *if for some k $p_3^{(k)} \notin W_3$, and $p_1^{(l)} \in W_1$,$l \geq 0$, then $\mathbf{p}^{(k)}$ converges to $\mathbf{p}^* = (p_1^*, 0, 0, p_1^*, p_5^*)$ where $\beta_1 p_1^* + \beta_2 p_1^* + \beta_3(2p_1^* + p_5^*) = \bar{x}^*$.*
3. *if for some k $p_i^{(k)} \notin W_i$ $i = 1, 3$, then $\mathbf{p}^{(k)}$ converges to $\mathbf{p}^* = (0, 0, 0, 0, p_5^*)$ where $\beta_3 p_5^* = \bar{x}^*$.*

Proof: The proof that the orbit $\mathbf{d}^{(k)},\ \ k \geq 0$ proceeds along a straight line parallel to l_0 is also valid for this case. Therefore the orbit must approach the d_3 axis. We can also assume that h is small enough so that any point above the line l_0 has a d_3 value large enough so that either $p_1 \in W_1$ or $p_3 \in W_3$ or $p_5^{(k)} \in W_5$. Since $d_1^{(k)} = p_1^{(k)} + p_3^{(k)}$, there is some k, where either $p_1^{(k-1)} \notin W_1$ and/or $p_3^{(k-1)} \notin W_3$.

(i) If $p_1^{(k-1)} \notin W_1$ we must have $p_1^{(k)} = p_4^{(k)} = 0$. From (38)-(42), one can therefore see that $p_1^{(l)} = 0$ $l \geq k$ as long as $xs(l) \leq xs^*,\ \ l \geq k$. During this time we can restrict our attention to the equations for p_2 $(= p_3)$ and p_5. Here we assume that $p_2^{(k)} \in W_2$. The mean route cost at time k, $\bar{x}^{(k)} = \beta_1 p_2^{(k)} + \beta_2 p_2^{(k)} + \beta_3 p_5^{(k)}$, continues to decrease and converges to $\bar{x}^*$. We have $\bar{x}^{(l)} \geq \bar{x}^*$ so $xs(l) \leq xs^*$ for all $l \geq k$. Meanwhile, $p_2^{(k)}, p_5^{(k)}$ converge to values p_2^* , p_5^* for which $\bar{x} = \bar{x}^*$ while $p_1^* = p_4^* = 0$. Note that $p_2^{(l+1)} - p_5^{(l+1)} = p_2^{(l)} - p_5^{(l)} + h(\beta_3^{(h_s)} - \beta_2^{(h_s)})(xs^* - xs(k))$. Therefore if $p_5^{(k)} > p_2^{(k)}$, the inequality remains true for $l \geq k$ and $d_3^{(l)} \geq d_1^{(l)}$.

(ii) If $p_3^{(k-1)} \notin W_3$, we will also assume in analogy with (i) that $p_1^{(k)} \in W_1$. Since $p_2^{(k)} = p_3^{(k)} = 0$, and $p_1^{(k)} = p_4^{(k)}$, a system for p_1 and p_5 remains. Arguing as before it can be seen that since $\bar{x}^{(l)} \geq \bar{x}^*$, for $l \geq k$, then $p_1^{(l)} = 0$ $\ l \geq k$. The mean route cost at time k, $\bar{x}^{(k)} = (1 + \beta_3) p_1^{(k)} + (\beta_2 + \beta_3) p_5^{(k)}$, decreases and $p_1^{(k)}, p_5^{(k)}$ converge to p_1^*, p_5^* where $\bar{x}^* = (1 + \beta_3) p_1^* + (\beta_2 + \beta_3) p_5^*$. The remaining limiting p_i values are $p_2^* = p_3^* = 0$. In this case we also have $d_3^{(l)} \geq d_1^{(l)}$, for $l \geq k$.

(iii) If both $p_3^{(k-1)} \notin W_3$ and $p_1^{(k-1)} \notin W_1$ hold, then we can repeat the reasoning in (i) and (ii) and conclude that the system reduces to the evolution of just p_5 . In this case $p_5^{(k)}$ converges to p_5^* where $\beta_3 p_5^* = \bar{x}^*$. The remaining limiting values are $p_i^* = 0$ $i = 1 \cdots 4$. As in the previous cases, $d_3^{(l)} > d_1^{(l)}$ for $l \geq k$. □

Proposition 6.2 shows that there are infintely many equilibrium points with $\mathbf{d}^* > 0$ for the system (38)-(42) with each point depending on the initial condition $\mathbf{d}^0$. However the equilibrium value xs^* is unique. In the $\mathbf{d}$-plane, linearization about any such point shows that each has a neutrally stable direction (eigenvalue 1) along the direction of l_d, and an asymptotically stable direction parallel to l_0 and so these points are neutrally stable. Thus it is not surprising that they do not persist when a parameter,in particular the entropy h_s , is changed. The equilibrium points in the $\mathbf{p}$ plane discussed in Proposition 5.3 are more difficult to analyze because the right hand side of the system fails to be differentiable when any $p_i \notin W_i$. Nevertheless linearization of the reduced system discussed in (i) and (ii) in the proof of Proposition 6.3 shows the equilibrium points there are neutrally stable and the equilibrium point in (iii) is asymptotically stable. A complete proof would have to show that the stability demonstrated in the reduced system is also valid in the entire $\mathbf{p}$ plane. The fact that, in each of these cases, the plane for the reduced system is invariant under (38)-(42) leads us to conjecture that this extension can be made.

In the next section we will discuss the effect of perturbing the route allocation away from the distribution defined by the critical entropy. Under such a perturbation the neutrally stable points will disappear but the asymptotically stable point in Proposition 6.3 (iii) with a single positive price for link 5 will persist.

6.2 Diamond Route Distribution with increased h_s

In our discussion of the behavior of (38)-(42) for perturbed entropy values, it is useful to introduce a parameter ν that decreases (increases) the weight on links 2 and 3 as the perturbed entropy increases (decreases). Since we will continue the assumptions on the capacity values displayed in equation (45), $\beta_1^{(hs)} = \beta_2^{(hs)}$ will still hold. Therefore, it is assumed that the entropy is the value h_s for which the route allocation distribution is,

$$\hat{\beta}^{(hs)} = \left\langle \frac{\nu c_3}{c_2 + c_3 + c_5}, \frac{\nu c_2}{c_2 + c_3 + c_5}, \frac{c_5 + (1-\nu)(c_2 + c_3)}{c_2 + c_3 + c_5} \right\rangle, \tag{53}$$

where $\nu < 1$. The entropy of the distribution with $\beta_1^{(hs)} = \beta_2^{(hs)} = \beta$, is $H = 2\beta\log(2) + h_\beta$ where $h_\beta = -2\beta\log(2\beta) - (1-2\beta)\log(1-2\beta)$. It is not hard to show that if $\beta > 1/4$, decreasing β increases H and vice versa. Since we confine our analysis to the region $d_1 \geq d_3$ (see the discussion at the beginning of Section 6), we must consider a distribution where $\beta_1 = \beta_2 \geq \beta_3$. This implies that $\beta \geq 1/3$.

The case $\nu < 1$ corresponds to an entropy value h_s larger than the critical entropy where $\nu = 1$. Here in contrast to Section 6.1, the values of xs for which $\Delta p_i^{(k)} = 0$ in (38)-(42) are not all equal and so any orbit for which $p_i^{(k)} \in W_i,\ k \geq 0$ as in Proposition 6.2 will not become a stationary point once $\mathbf{d}^{(k)}$ reaches the line l_d . The equilibrium value $\bar{x}^*$ satisfies:

$$\begin{gathered} xs^* = (w/\bar{x}^*)^{1/2}, \text{ where } xs^* = (c_2 + c_3 + c_5)\left(1 - \tfrac{(1-\theta)\beta_3 b_3}{\beta_1 b_1 + \beta_2 b_2 + \beta_3 b_3}\right) \text{ and} \\ \theta = (c_2 + c_3 + 3c_5)\,/\,(c_2 + c_3 + 3c_5 + 2(1-\nu)(c_2 + c_3)) \end{gathered}$$

The neutrally stable points discussed in that proposition no longer exist. The same situation also holds for the neutrally stable points discussed in parts 1 and 2 of Proposition 6.3. However, as the next proposition shows, for each $\nu < 1$, there is a unique equilibrium bandwidth rate and a corresponding unique equilibrium to which orbits of (38)-(42) converge.

In what follows we let $\mathcal{B}'$ be the region in the d_1, d_3 plane bounded above by the line l_0 , below by the line l_1 and on the left by d_3 axis. The lines l_0 and l_1 are described just as they are in Section 6.1. We then define $\mathcal{B} = \mathcal{B}' \cap \{(d_1, d_3) \mid d_1 \geq d_3\}$.

Proposition 6.4 *Let h_s be the entropy defined by the distribution in (53) and suppose the initial values of $\{p_i^{(0)}\ i = 1 \cdots 5\}$ satisfy the hypotheses in Lemma 6.1. Then* $\mathbf{p}^{(k)} = (p_1^{(k)}, p_2^{(k)}, p_3^{(k)}, p_4^{(k)}, p_5^{(k)}) \rightarrow (0, 0, 0, 0, p_5^*)$

Proof: Since $p_i^{(0)} \in W_i\ i = 1, \cdots 5$, the initial dynamics of the route cost vector $\mathbf{d}^{(k)}$ is governed by the equations in (48)-(49). First suppose, $\mathbf{d}^{(0)} \in \mathcal{B}$. The mean route cost, $\bar{x}^{(k)} = \beta_1 d_1^{(k)} + \beta_2 d_2^{(k)} + \beta_3 d_3^{(k)}$ becomes stationary (i.e. $\Delta\bar{x}^{(k)} = 0$) when $xs^{(k)} = xs^*$. We can repeat the arguments in the derivation of (32) from (31) to this situation to show that the mapping T, defined by

$$\bar{x}^{(k+1)} = T\bar{x}^{(k)} = \bar{x}^{(k)} - h\bar{b}\{xs^* - xs(k)\} \tag{54}$$

where $\bar{b} = \beta_1 b_1 + \beta_2 b_2 + \beta_3 b_3$, is a global contraction in $\mathcal{B}$ with a fixed point at $\bar{x}^*$. Consequently orbits in the $\mathbf{d}$ plane starting in $\mathcal{B}$ will converge to the line l_d. Once the orbit approaches the line, it remains there and moves toward the d_3 axis i.e. d_1 decreases while d_3 increases. This is because,

$$xs_3^* < xs^* < xs_1^* \tag{55}$$

where $xs_1^* = C_1/b_1 = c_2 + c_3 + c_5$ and $xs_3^* = (c_2 + c_3 + c_5)\theta$ are values for which $\Delta d_i^{(k)} = 0$, for $i = 1, 3$. If $\mathbf{d}^{(0)}$ lies above the line l_0, then $\bar{x}^{(k)}$ decreases since $xs(k) < xs^*$. It is clear therefore that regardless of where the orbit starts, d_1 will eventually become small enough so that $p_1^{(k-1)} \notin W_1$ or $p_3^{(k-1)} \notin W_3$ for some k for the first time.

We examine the case when $p_3^{(k-1)} \notin W_3$ first. It follows that $p_3^{(k)} = 0$. We will show that $p_3^{(k+1)} = 0$ and that the subset $\{\mathbf{p}|p_3 = 0\}$ is invariant under the iteration of (38)-(42). By Lemma 6.1, $p_2^{(k)} = 0$, and $p_1^{(k)} = p_4^{(k)}$, so we will focus on p_1 and p_5 . If $p_i^{(k)} \in W_i$ $i = 1, 5$, then an equation for $\bar{x}^{(k+1)}$ can be obtained. The value let us call it $xsp_{1,5}^*$,for which $\Delta\bar{x}^{(k+1)} = 0$ is $xsp_{1,5}^* = u * xsp_1^* + (1-u) * xsp_5^*$ where $u = \frac{(1+\beta_3)(\beta_1+\beta_3)}{(1+\beta_3)(\beta_2+\beta_3)+\beta_3^2}$. Here, $xsp_1^* = (c_3 + c_5)/(\beta_1 + \beta_3)$ and $xsp_5^* = c_5/\beta_3$ are xs values for which $\Delta p_i^{(k)} = 0$, $i = 1, \cdots, 5$. The inequality $xsp_5^* < xsp_{1,5}^* < xsp_1^* < xsp_3^*$ holds. It can also be shown that the mapping $\bar{x}^{(k+1)} = T\bar{x}^{(k)}$ is a contraction in the p_1, p_5 plane with a fixed point $\bar{x}_{1,5}^*$ corresponding to $xsp_{1,5}^*$. Now $p_3^{(k-1)} \notin W_3$, implies that $xs(k-1) \le xsp_3^* = (c_2 + c_3 + c_5)/\nu$ and if $xs(k-1) > xsp_{1,5}^*$, then xs will decrease and $xs(k) < xs(k-1)$. If $xs(k-1) < xsp_{1,5}^*$ then $xs(k) > xs(k-1)$ but since h is small $xs(k) < xsp_3^*$. Moreover, the distance between $xs(k)$ and $xsp_{1,5}^*$ decreases because of the contraction property in the iteration of $\bar{x}^{(k)}$. Thus $p_3^{(k+1)} = 0$. Even though, subsequent iteration of p_1 and p_5 leads to increasing xs when $xs(k) < xsp_{1,5}^*$ the corresponding value of xs will approach but cannot exceed $xsp_{1,5}^*$. Thus p_3 remains 0. Eventually $(p_1^{(k)}, p_5^{(k)})$ converges to the line $xsp_{1,5}^* = \beta_1 p_1 + \beta_3 p_5$ and remains on the line as $p_1^{(k)} \to 0$, and $p_5^{(k)} \to p_5^*$, where $xs^* = xsp_5^*$.

If $p_1^{(k-1)} \notin W_1$ and $p_3^{(k-1)} \notin W_3$, then since $xsp_1^* > xsp_5^*$, $xs(k) > xsp_1^*$ implies that $\bar{x}^{(k)} = \beta_3 p_5^{(k)}$ will increase and thus xs decreases and eventually, $xs(l) < xsp_1^*$ for some $l \ge k$. If $p_1^{l)} = 0$, it remains 0 and the orbit remains on the p_5 axis and converges to p_5^*. If $p_1^{(l)} \in W_1$ for some l then we are back in case discussed in the previous paragraph. In all these cases the orbits eventually converge to the point $(0, 0, 0, 0, p_5^*)$ where $\bar{x}^* = \beta_3 p_5^*$ with corresponding $xs^* = c_5/\beta_3$.

Finally we turn to the case where the equations for $\mathbf{d}$ lose validity because for some k, $p_1^{(k-1)} \notin W_1$ but $p_i^{(k-1)} \in W_i$ $i = 3, 5$. We are supposing that up to this point, $p_i^{(l)} \in W_i$ $l = 0, 1 \cdots k-2$, $i = 1, 3, 5$. From the Remark following Lemma 6.1, we have $p_1^{(k-1)} = p_3^{(k-1)} + p_5^{(k-1)}$. But $p_3^{(k-1)} > h\{c_3 - \beta_1 xs(k-1)\}$ and $p_5^{(k-1)} > h\{c_5 - \beta_3 xs(k-1)\}$ so it must be that $p_1^{(k-1)} > h\{c_1 - (\beta_1 + \beta_3)xs(k-1)\}$. Thus $p_1^{(k-1)} \in W_1$ a contradiction. Consequently this case cannot occur. $\square$.

Remark: Since orbits that begin below the line l_1 blow up we will not discuss these here.

6.3 Diamond Route Distribution with decreased h_S

Using the notation introduced in Section 6.2 we consider $\nu > 1$. This has the effect of increasing the weight on links 2 and 3 with a compensating decrease of weight on link 5. The discussion in Section 6.2 shows that the corresponding entropy h_s is decreased. In contrast with the previous case, the algorithm fails to converge and in fact the root $\gamma(k)$ of (17) diverges as $k \to \infty$. This section details a proof of this fact. For a particular sub-case (see Case 1a) we assume that $2c_2 > c_5$.

We assume initial values of $\mathbf{p}$, such that $p_i^{(0)} \in W_i$ $i = 1, \cdots, 5$. The subsequent dynamics at least intially can therefore be described by equations (48)-(49) Let the xs values for which $\Delta d_j^{(k)} = 0$ $j = 1, 2, 3$ be denoted by xs_j^*. The relation (55) is now,

$$xs_1^* < xs^* < xs_3^* \tag{56}$$

Let τ_1 be the line in the $\mathbf{d}$ plane given by $\bar{x} = \bar{x}_1^*$, where $\bar{x}_1^*$ is the mean cost corresponding to xs_1^*. The line τ_3 is defined analogously. The line $\tau_d : \bar{x} = \bar{x}^*$ lies between τ_1 and τ_3 so we will define $\mathcal{C}$ to be the region bounded by these lines , the d_3 axis , and the line $\{(d_1, d_3)|d_3 = d_1\}$. If $d^{(0)}$ is an initial point whose corresponding xs value is $xs(0) < xs_1^*$, then d_1 and d_3 will decrease. On the other hand if the corresponding xs value satisfies $xs(0) > xs_3^*$, then d_1, and d_3 will increase (see Figure 7). Orbits either enter the region $\mathcal{C}$ or they intersect or approach the d_1 or d_3 axes. We describe the fate of orbits that enter $\mathcal{C}$. The behavior of $\bar{x}^{(k)}$ is governed by (54). The contraction property of T implies that $\mathbf{d}^{(k)}$ approaches the line τ_d and once the orbit reaches the line it moves along it. By (56) see that d_1 increases and d_3 decreases so it is clear that $\mathbf{d}^{(k)}$ approaches the line $d_1 = d_3$. Thus γ^k must diverge.

Suppose now that an orbit's initial position is located above the line l_0 (see equation (51) and Figure 7). In this situation, at least initially $xs(k) \le xs_1^* < xs_3^*$, therefore d_1 and d_3 are decreasing and orbits approach the d_3 axis. We have $d_3^{(0)} \ge \delta$ where $(0, \delta)$ is the intersection of l_0 with the d_3 axis. Thus the equation (48) will cease to be valid for some k for the first time. Since $d_1^{k-1} = p_1^{(k-1)} + p_3^{(k-1)}$, we can suppose either $p_1^{(k-1)} \notin W_1$ or $p_3^{(k-1)} \notin W_3$. One can verify after some algebra, an inequality that will be used in the arguments that appear below. Recall that xsp_i^* is the value of xs for which $\Delta p_i^{(k)} = p_i^{(k+1)} - p_i^{(k)} = 0$, when $p_i^{(k)} \in W_i$. For $\nu > 1$, $xsp_3^* = xsp_2^* < xsp_1^* < xsp_5^*$.

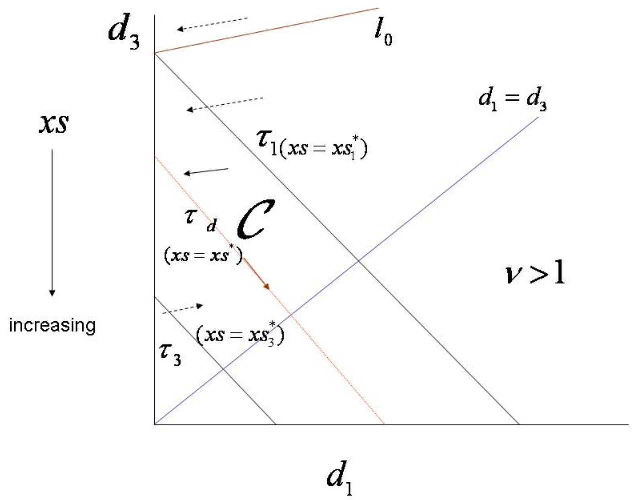

Figure 7: (d_1, d_3) plane for $h_s < h_D(c)$

Case 1: First assume that $p_1^{(k-1)} \notin W_1$ and thus $p_1^{(k)} = 0$.

Case 1a If $p_2^{(k)} = p_3^{(k)} \in W_2$ and $p_5^{(k)} \in W_5$, an equation for $\bar{x}^{(k+1)}$ can be obtained in terms of $p_2^{(k+1)}$ and $p_5^{(k+1)}$ of the form $\bar{x}^{(k+1)} = T\bar{x}^{(k)}$ analogous to (32). T is a contraction with a fixed point $\bar{x}_{2,5}^*$ such that $xs_{2,5}^* = \left(\frac{w}{\bar{x}_{2,5}^*}\right)^{1/2}$, and $xs_{2,5}^* = (1-\mathfrak{f})xsp_2 + \mathfrak{f}xsp_5$ with $\mathfrak{f} = \frac{\beta_3^2}{(\beta_3^2 + (1-\beta_3)\beta_2)}$. Now suppose $xs_{2,5}^* < xsp_1^*$. Then $\bar{x}^{(k)} = \beta_1 p_3^{(k)} + \beta_2 p_2^{(k)} + \beta_3 p_5^{(k)}$ will move closer to the value of $\bar{x}$ corresponding to $xs_{2,5}^*$. Thus $xs(k+1) < xsp_1$ and therefore $p_1^{(k+1)} = 0$. We could repeat this argument so that by induction this implies then that $p_1^{(l)} = 0 \;\; l \geq k$. Now $d_3^{(l)} \geq d_1^{(l)}$ and $p_1^{(l)} = 0 = p_4^{(l)}$, therefore $p^{(l)} \geq p_2^{(l)}$

In the p_2, p_5 plane, $xs(k) \to xs_{2,5}^*$ and since $xsp_2^* < xs_{2,5}^* < xsp_1^* < xsp_5^*$, we must have p_5 decreasing and p_2 increasing. Thus $(p_2^{(k)}, p_5^{(k)})$ approaches the line $p_2 = p_5$ as $k \to \infty$. This means that we approach the line $d_1 = d_2 = d_3$. The following sufficient condition for $xs_{2,5}^* < xsp_1$ is derived in the Appendix and is satisfied when $\kappa = \frac{c_2}{c_5} > 1/2$ and $\nu > 1$.

$$\left\{\frac{1}{\kappa + 1 + (1-\nu)\kappa} - \frac{\mathfrak{f}}{1 + 2(1-\nu)\kappa}\right\} > 0 \tag{57}$$

$$\tag{58}$$

The preceeding reasoning assumed $p_2^{(k)} \in W_2$. The convergence of $(p_2^{(l)}, p_5^{(l)})$ to the line $\bar{x} = xs_{2,5}^*$ and the fact that $xs_{2,5}^* > xsp_2$ shows that $p_2^{(l)} \in W_2$ for $l \geq k$.

Case 1b: Suppose $p_3^{(k)} \notin W_3$. Then $xs(k) \leq xsp_3^*$ and since $xsp_3^* < xsp_1^* < xsp_5^*$, both $d_1^{(k)}$ and $d_3^{(k)}$ decrease. Therefore even though $p_3^{(k+1)}$ is 0, $xs(k)$ increases. xs continues to increase until $xs(l) \geq xsp_3^*$, for some $l \geq k$. If $xs(l) = xsp_3^*$ then note that $p_5^{(l)}$ decreases, and hence $\bar{x}^{(l)}$ decreases. Therefore xs will continue to increase and p_3 eventually becomes positive and we are back in the previous case where $p_3^{(k)} = p_2^{(k)} \in W_2$. If we have $p_5^{(k)} \notin W_5$, then $p_1^{(k)} = 0$ implies that $d_1^{(k+1)} = p_3^{(k+1)} \geq d_3^{(k+1)} = p_5^{(k+1)} = 0$. So we have crossed the line $d_1 = d_3$ causing γ to diverge.
Case 2: Finally let us consider the case when for the first time , $p_3^{(k-1)} \notin W_3$. Then $xs(k-1) \leq xsp_3^* < xsp_1^* < xsp_5^*$ so d_1 and d_3 decrease, causing $xs(k)$ to increase. This increase continues until $xs(l) \geq xsp_3$, for some $l \geq k$.
Case 2a: If $p_1^{(k)} \in W_1$, then xs continues to increase and eventually, $p_3^{(l)} \in W_3$ for some $l \geq k$ and is increasing.One can check that $xsp_3^* = xs_1^* < xsp_1^* < xsp_5^*$. If $xs(l) \leq xsp_1$, then $(d_1^{(l)}, d_3^{(l)}) \in \mathcal{C}$ and we can apply the results of the discussion of orbits in that region to this case. If $xs(l) > xsp_1$ and $p_5^{(k)} \in W_5$ then, equations (48)-(49) can be used

to show that $(d_1^{(l)}, d_3^{(l)})$ is either in or will eventually enter $\mathcal{C}$. If $p_5^{(k)} \notin W_5$ then $d_1^{(k+1)} = d_3^{(k+1)}$, i.e. we are at the boundary of the feasible region so $\gamma^{(k+1)} = \infty$.

Case 2b Note that $xsp_3^* < xs_{2,5}^* < xsp_1^*$. If $p_1^{(k)} \notin W_1$ then xs increases until for some $l \geq k$, $p_3^{(l)} > 0$. As in Case 1a, the dynamics of $d_1^{(l)}, d_3^{(l)}$ are determined by $p_3^{(l)} = p_2^{(l)}$ and $p_5^{(l)}$ as long as $xs(l) < xsp_1^*$ with $l > k$. We can repeat the arguments of Case 1a to show that in fact $xs(k) \to xs_{2,5}^*$.

7 UTILITY-ENTROPY TRADEOFF

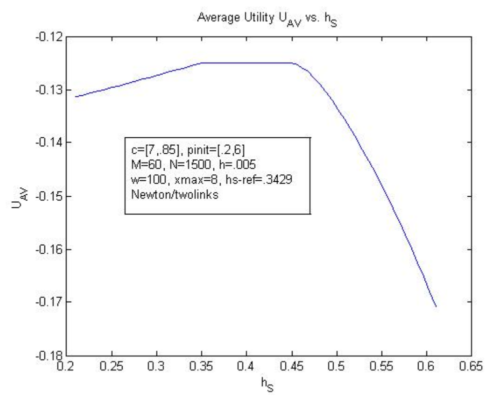

Figure 8: Twolinks Utility vs. entropy

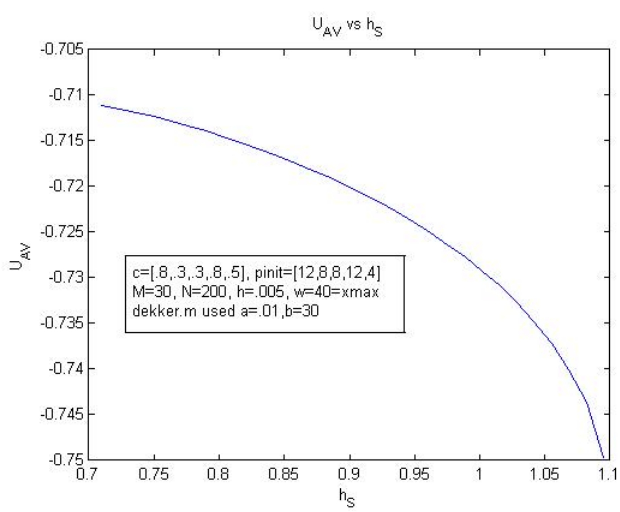

Figure 9: DIAMOND Utility vs. entropy

The route distribution entropy h_s is a measure of the degree of path diversity of an allocation scheme that uses that distribution.The tradeoff between path diversity and utility is not well understood in general but our approach allows us to frame the issue in relation to a reference entropy that defines an intrinsic allocation defined by the link capacities and topology of a given network. Using the critical entropy value as a reference, the effect of decreasing or increasing path diversity beyond this allocation can be assessed. This is illustrated in our examples by plotting the utility as a function of h_s. In each case, this was done by computing the average utility over a finite time interval for each h_s for a range of values that contained the critical entropy value. Figure 8 shows the computation for TwoLinks and Figure 9 shows the corresponding computation for Diamond. For values of h_s less than the critical value, a time interval was chosen so that orbits were still within the feasible region. Increasing the entropy from $h_s = 0$, the value for a single path allocation scheme, the iterations did not converge until a critical entropy value was reached.

Inspection of the curves shows that a network using an allocation with larger entropy can increase utility by decreasing the entropy (that is, increasing the fraction on less expensive links) up until the critical value is reached. The differing topologies of TwoLinks and Diamond affect what happens when the allocation entropy is less than the critical value. In TwoLinks the utility cannot be increased by decreasing the entropy. In Diamond, the utility is increased until the value $h_s = \log 2$. Note in contrast to [9] the utility does not reach a maximum at the critical entropy value. We could not stably iterate equations (38) -(42) for values $h_s < \log 2$. We find that the iterations fail to converge for any $h_s < h_D(c)$. A possible reason for this comes from a visualization of the route allocation distributions for 3 routes (Figure 10). As discussed in Section 4, each $\beta = (\beta_1, \beta_2, \beta_3)$, such that $\beta_i \geq 0 \ i = 1, 2, 3$, $\beta_1 + \beta_2 + \beta_3 = 1$, can be associated with a point in the equilateral triangle with altitude height 1. The transformation from the simplex in 3-space to this triangle is smooth and invertible. The region defined by the constraints of the optimization problem i.e. inequalities (11)-(13) is convex and bounded by the level curve for h_s. Level curves for various values of h_s are mapped to the curves in the triangle. When $h_s = \log 2$, the curve is tangent to the boundary of the triangle and therefore the level curve is tangent to the boundary of the simplex. The points of tangency are singular points for the entropy function (i.e. points where the derivative of h_s as a function of β fail to exist). Once $h_s < \log 2$, the curve becomes disconnected. It is interesting to note that the source of the trouble comes from the implicit equation. Indeed there are two solutions of the implicit equation for each h_s and when $h_s < \log 2$, the values of $\gamma_(k)$ oscillate between these values introducing sustained oscillations into the system that resemble the route flapping phenomenon discussed in [12] and other work. For such h_s note that we are effectively operating with only two of the three links, so that the Diamond topology is equivalent at this stage to the Twolinks topology.

8 CONCLUSION

In this paper, we presented a convex network optimization problem for simultaneously controlling congestion and route allocation. The set of possible route allocations that can be used by a source s is constrained by placing a lower bound on the route distribution entropy h_s. As a first step towards understanding the issues involved in any protocol based on this model we derived a dual iteration scheme and analyzed the dynamics for two simple topologies. Sufficient conditions insuring that the equilibria of the iteration scheme are solutions of the optimization problem are given. We also discuss the limits of the algorithm. Specifically in our examples we identify parameter values h_s where the iterates eventually leave the region of feasibility. For such parameter values some other approach-using subgradient methods may prove useful.

Another key issue here is the trade-off between network utility and h_s, a parameter that measures path diversity. After making mild assumptions about the capacity of the links in each example we found a critical value of the entropy. The average utility can always be increased if the system has an allocation with route distribution entropy h_s that is larger than the critical value. Figure 8 shows how the average utility for the TwoLinks topology plateaus so that its maximum value is achieved at or near the value at the critical entropy $h_T(c)$ similar to the results of the simulation in [9]. However Figure 9 shows that the average utility for the Diamond topology continues to increase for $h_s < h_D(c)$. A natural lower bound of $\log 2$ is reached, corresponding to the point where one of the 3 routes is eliminated and the topology is effectively equivalent to TwoLinks. What we can therefore say in this case is that the utility value at $h_D(c)$ is the largest practicably achievable one given a protocol based on dual iteration.

In summary, we see that one cannot increase the utility by choosing a path distribution with entropy less than the reference entropy without producing infeasible solutions. For each of the examples, the link capacities and the topology introduce inherent limits to how closely the route allocation can approach the single path specification scheme.

9 APPENDIX

9.1 Constant $\beta^{(k)}$ in TwoLinks

In this section we will show that

Theorem 9.1 *Let* $\frac{c_1}{c_2} = \lambda$ *and* $\frac{\beta_1^{(hs)}}{\beta_2^{hs}} = \lambda + \mu$ *with* $\lambda > 1$ *and* $\lambda + \mu \geq 0$. *Then*

$$\beta_i^{(k)} \equiv \beta_i^{(hs)} \ i = 1, 2, \ k \geq 0 \tag{59}$$

where $\mathsf{H}(\beta^{(hs)}) = -(\beta_1^{(hs)} \log \beta_1^{(hs)} + \beta_2^{(hs)} \log \beta_2^{(hs)}) = hs$.

Proof: For simplicity we will drop the hs notation from $\beta_i^{(hs)}$ $i = 1, 2$. Given an entropy value $hs < \log 2$ there are exactly two associated probability distributions, either $\{\beta, 1-\beta\}$ or $\{1-\beta, \beta\}$. If we define the ratio $\rho = \frac{\beta_1}{\beta_2}$, then the distributions can be expressed as either $\{\frac{\rho}{1+\rho}, \frac{1}{1+\rho}\}$ or $\{\frac{1}{1+\rho}, \frac{\rho}{1+\rho}\}$. If $p_1^{(k)} < p_2^{(k)}$ then the choice of a positive root of (17) implies that $\beta_1^{(k)} = \frac{\rho}{1+\rho}$ where $\rho < 1$. If on the other hand, $p_1^{(k)} > p_2^{(k)}$ then $\beta_1^{(k)} = \frac{1}{1+\rho}$. $\square$

Note that the condition $hs < \log 2$, rules out the possibility that $p_1^{(k)} = p_2^{(k)}$ for any finite k since the solution of (17) would then be ∞. These facts lead to the observation that $\beta_i^{(k)}$, $i = 1, 2$ remains constant in k if and only if $p_2^{(k)} - p_1^{(k)}$ does not change sign. In fact a change in sign leads to a jump discontinuity in the value of β_i, $i = 1, 2$.

9.2 Diamond route distributions and their entropy

There are 3 routes in the diamond topology. Using planar geometry one can show that every distribution in the simplex of distributions on 3 routes $\mathsf{S} = \left\{ \hat{\beta} = (\beta_1, \beta_2, \beta_3) \mid \quad \beta_i \geq 0 \quad i = 1, 2, 3 \ \beta_1 + \beta_2 + \beta_3 = 1 \right\}$, can be represented by a point in an equiliateral triangle with an atltitude of height 1. This result is known as Viviani's theorem. If we position the leftmost vertex of the base of the triangle at the origin (as shown in Figure 10), then the distribution $(\beta_1, \beta_2, \beta_3) \in \mathsf{S}$ corresponds to coordinates (x, y) in the triangle through $\Psi : \mathcal{R}^2 \to \mathsf{S}$ given by,

$$\beta_1 = y \ , \ \beta_2 = \frac{\sqrt{3}x - y}{2} \ , \ \beta_3 = 1 - \left(\frac{y + \sqrt(3)x}{2} \right) \tag{60}$$

The entropy function $\mathsf{H}(\beta_1, \beta_2, \beta_3)$, defines a function of x, y through (60) and $h(x, y) = \mathsf{H}(\Psi(x, y))$. The distribution with $\beta_1 = \beta_2 = \beta_3 = 1/3$, is the image of $(1/\sqrt(3), 1/3)$. A simple computation shows that the latter is a non-singular critical point of h. Hence nearby level curves are nearly ellipses and other level curves for smaller values of h_s that do not touch another critical point must have the same topological type and so they are simple closed curves [10]. Since Ψ is affine and H is convex, $h = H \circ \Psi$ is convex and therefore quasi-convex. The region enclosed by a level of h is therefore convex. At $\log 2$, the level curve has several critical points that are the images of the intersection of the black curve in Figure 10 with the sides of the triangle. The topological type of the level curve therefore changes.

9.3 Suffcent condition that $x^*_{2,5} < xsp^*_1$

We present a proof of the inequality (58) that appeared in Section 6.3.

Lemma 9.2 *Let xsp^*_1 be the value of xs for which $\Delta p_1^{(k)} = 0$ in equation (38). Further let, $xs^*_{2,5}$ be the value of xs for which $\Delta \bar{x}^{(k)} = 0$ restricted to the (p_2, p_5) plane. Then $xs^*_{2,5} < xsp^*_1$ if and only if inequality (58) is satisfied .*

Proof:
We have,

$$xsp_1 = \frac{c_3 + c_5}{c_3 + c_5 + (1-\nu)(c_2 + c_3)} \ , \quad xsp_2 = xsp_3 = \frac{c_2 + c_3 + c_5}{\nu} \tag{61}$$

$$xs^*_{2,5} = (1 - \mathfrak{f})xsp_2 + \mathfrak{f}xsp_5 \tag{62}$$

$$\mathfrak{f} = \frac{\beta_3^2}{\beta_3^2 + (1-\beta_3)\beta_2} \tag{63}$$

$$= \frac{(c_5 + 2(1-\nu)c_2)^2}{[2(\nu c_2)^2 + (c_5 + 2(1-\nu)c_2)^2]} \tag{64}$$

Writing $xs^*_{2,5} - xsp1 = (xsp2 - xsp1) - (xsp2 - xs^*_{2,5})$, after some algebra it can be seen that

$$xs^*_{2,5} - xsp_1 = \left(\frac{1-\nu}{\nu} \right) \left[\frac{1}{c_3 + c_5 + (1-\nu)c_3} - \frac{\mathfrak{f}}{c_5 + (1-\nu)(c_2 + c_3)} \right] (c_2 + c_3 + c_5) \tag{65}$$

The expression in the square brackets of equation (65) is the left hand side of inequality (58). Setting $\kappa = \frac{c_2}{c_5} = \frac{c_3}{c_5}$ it can be seen that choosing $\kappa > 1/2$ is sufficient when $\mathfrak{f}$ is rewritten in terms of κ and the inequality is verified by calculation. $\square$

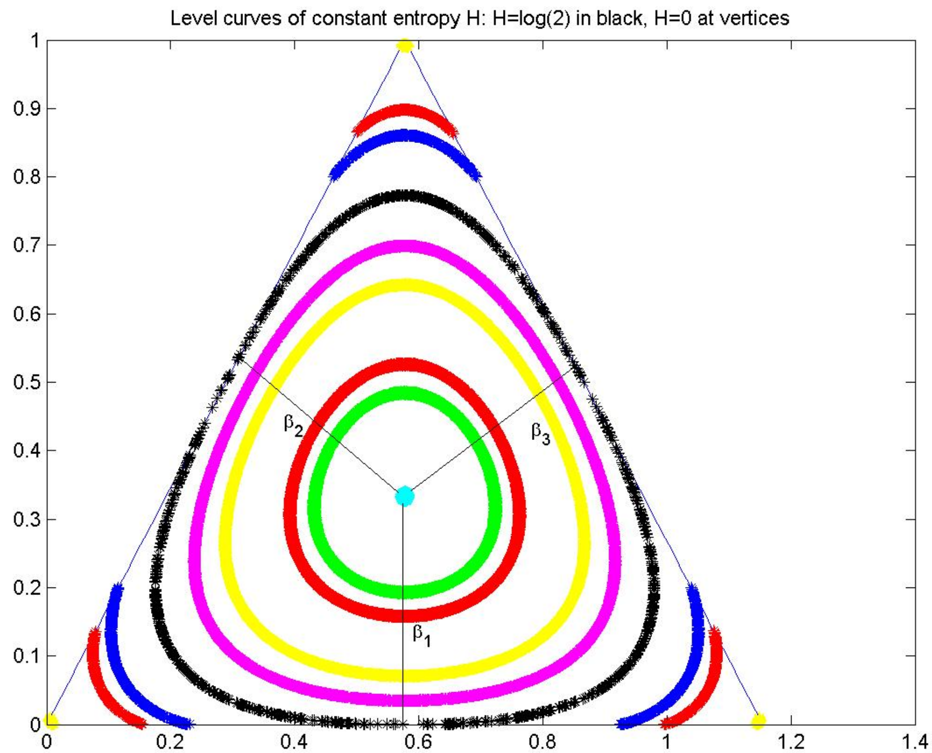

Figure 10: Diamond route allocations

References

[1] D.P. Bertsekas, J. N.Tsitsiklis "Parallel and Distributed Computation: Numerical Methods", Prentice Hall, Englewood Cliffs, N. J. 1989

[2] S.Boyd, L. Vandenberghe "Convex Optimization", Cambridge University Press, Cambridge UK, 2004

[3] M.Chiang, S.H. Low, A.R. Calderbank, J.C. Doyle "Layering as optimization decomposition: A mathematical theory of network architectures", Proceedings of the IEEE, Vol. 95, no. 1, pp. 252-312, 2007

[4] J. He, M. Bresler, M. Chiang, J. Rexford "Towards Robust Multi-Layer Traffic Engineering: Optimization of Congestion Control and Routing" IEEE Journal on Selected Areas in Communications, Vol. 25, No. 5, pp.868-880, 2007

[5] F.P.Kelly, T. Voice "Stability of end-to-end algorithms for joint routing and rate control" Computer Communication Review, Vol. 35, No. 2, pp. 5-12, 2005

[6] X. Lin, N. Shroff, R. Srikant "On the Connection-Level Stability of Congestion-Controlled Communication Networks", IEEE Trnasactions on Information Theory, Vol. 54, No. 5, 2008

[7] F.P. Kelly, A.K.Maulloo, D.K.H. Tan "Rate control in communication networks: shadow prices, proportaion fairness and stability." Journal of the Operational Research Society, Vol. 49, pp. 237-252, 1998.

[8] S.H. Low, D. E. Lapsley "Optimization Flow Control I: Basic Algorithm and Convergence" IEEE/ACM Transactions on Networking Vol. 7, No. 6, pp.861-874, 1999
available at http://netlab.caltech.edu/pub.

[9] V. Marbukh, S. Klink "Decentralized Control of Large Scale Networks as a Game with Local Interactions:Cross Layer TCP/IP Optimization" ValueTools07

[10] Y. Matsumoto "An Introduction to Morse Theory" Translations of Mathematical Monographs, Vol. 208, American Mathematical Society, 2002

[11] A. Peressini, F.Sullivan, J.Uhl "The Mathematics of Nonlinear Programming", Springer Verlag, New York, N.Y. , 1988

[12] J.Wang, L. Li, S.H. Low, J.C.Doyle "Cross-Layer Optimization in TCP/IP networks" IEEE Transactions on Networking, Vol. 13, No. 3, pp. 568-582, 2005