Joint control and routing optimizations on overlay networks with an OSP budget constraint

Project Report CMPSCI 691ee

Bruno Ribeiro

I. MOTIVATION AND RELATED WORK

BGP policies and/or global routing optimization create a set of end-to-end Internet routes that are not optimal for a given source-destination pair. Overlay networks are usually seen as a way to find better alternative routes than the one provided by the underlay [1], [2]. The idea is to use an overlay path instead of the underlay one. A weakness of this approach is no guaranteed good performance over some network scenarios. Suppose an overlay path has recurrent short high-loss periods but otherwise it is loss-free. The afore mentioned best-path strategy will have a hard time to achieve good performance in this scenario.

Another rationale against the best-path approach is the optimization problem it is solving. In the current Internet, routing is optimized independently of the congestion control mechanism even though both have the same high-level objective (to maximize user throughput given some measure of fairness). The first optimization that takes place is the routing optimization. After all routes are optimized given their traffic demands, these computed routes become inputs to the congestion control optimization problem. This can be shown to be sub-optimal [3]. Multi-path TCP was proposed to overcome these weaknesses [3]. Its main advantage is to jointly optimize routing and congestion and consequently overlay and underlay routes are optimized together. The total flow throughput is the sum of the overlay and underlay throughputs. But in a realistic scenario the overlay bandwidth is not for free and there is a monetary price to be paid for such improvement.

Users can pay the overlay according to the amount of traffic the overlay is routing for them. In overall this is a good way to control the expenditure of the overlay network: transfer all the cost to the users. But this strategy has one important fallacy. Suppose an user is infected by a worm or a e-mail spam virus. It is clear that mixing user payments by usage and software security holes is not a good idea.

In a scenario where there is a single administrative entity running the overlay network, an alternative payment method can be used. This method seems to be more in tune with the way real enterprises work: monthly fixed payments. A single administrative entity or Overlay Service Provider (OSP) estimates the number of users of its network, sets a bound on the maximum budget expenditure on the overlay and formulates how much each user should pay for right to use this overlay network. As the OSP would like to provide good and reliable service over all user network scenarios, Multi-path TCP seems to be the best choice for the task. Clearly there is a deployment issue as both source and destination hosts should implement

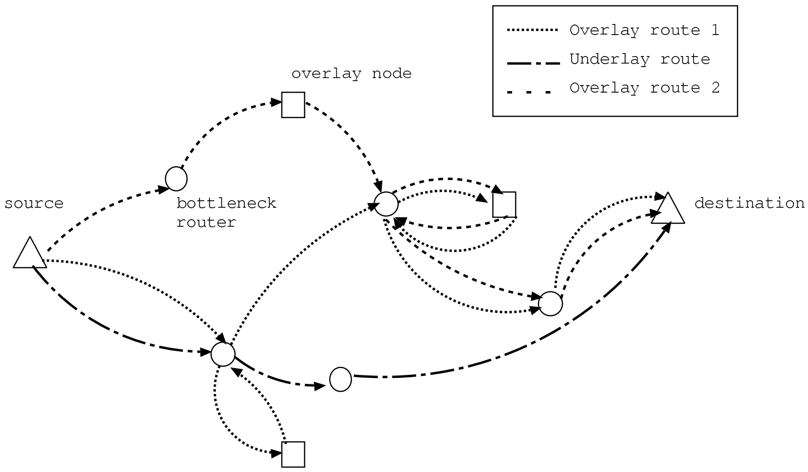

Fig. 1. Multi-path TCP topology example.

Multi-path TCP. Our hope is that in the future reliability and performance of some content sites will be vital and thus a sizeable amount of people would be willing to make these changes to their protocol stack to have better connectivity. Multi-path TCP has the advantage to work smoothly with normal TCP and there is no change to the underline Internet architecture.

To make the mathematical formulation simpler, we will assume all flows to be long lived and that the network reaches its steady state (no recurrent link failures or performance degradation [e.g.: environmental noise in wireless links]), altough these are not necessary conditions for the protocol to work.

II. Problem Introduction

The schematic of how Multi-path TCP works for a source/destination pair is depicted in Figure 1. The source has, besides its underlay path, a number of overlay paths to choose from. Notice that we are only representing bottleneck routers. Users are not necessarily multi-homed (see [4] for more details about multi-homed users). The technical details of the next two sections closely follows the notation and formulation given in [5].

Let S be a set of OSP source-destination pairs and J a set of bottleneck routers and overlay nodes these source-destination pairs have access to. Define $J' \subseteq J$ to be a set containing only overlay nodes and R to be the set of all available routes between source-destination pairs in S.

Let U_s be the utility function of user s and $C_j(y_j)$ be the congestion pricing at router or overlay node j given a traffic demand y_j. The monetary price paid by an overlay node $j' \in J'$ routing traffic $y_{j'}$ is given by the function $f_{j'}(y_{j'})$ and we assume that $f_{j'}$ is differentiable such that $g_{j'}(x_r) = \partial f_{j'} / \partial x_r$. Let B be OSP's maximum monetary expenditure on bandwidth.

Our system optimization problem can be written as

$$\text{maximize} \quad \sum_{s\in S} U_s(\sum_{h\in s} x_h) - \sum_{j\in J} C_j(\sum_{s\in S}\sum_{h\in s:j\in h} x_h) \tag{1}$$

subject to

$$\begin{aligned}&\sum_{j'\in J'} f_{j'}(\sum_{s\in S}\sum_{h\in s:j'\in h} x_h) \le B \qquad \text{and}\\ &x_h \ge 0, \quad \forall h \in R\end{aligned} \tag{2}$$

The KKT conditions state that the maximum of (1) subject to (2) is given by

$$\begin{aligned}\partial U_s(\sum_{h\in s} x_h)/\partial x_r - &\sum_{j\in r} \partial C_j(\sum_{s\in S}\sum_{h\in s:j\in h} x_h)/\partial x_r\\ &- \gamma \sum_{j'\in J'} \partial f_{j'}(\sum_{s\in S}\sum_{h\in s:j'\in h} x_h)/\partial x_r = 0,\end{aligned} \tag{3}$$

with

$$\begin{aligned}\gamma(\sum_{j'\in J'} f_{j'}(\sum_{s\in S}\sum_{h\in s:j'\in h} x_h) - B) &= 0\\ \alpha &\ge 0\end{aligned}$$

where γ is a Lagrange multiplier. Consider the following fluid rate control system:

$$\frac{d}{dt}x_r(t) = \kappa_r x_r(t)\left(1 - \frac{\lambda_r(t)}{\partial U_s(\sum_{h\in s} x_h(t-T_h))/\partial x_r}\right)^+_{x_r(t)} \tag{4}$$

where

$$\lambda_r(t) = \sum_{j\in r} p_j(\sum_{h:j\in h} x_h(t - T_{jh} - T_{hj})) + \gamma \sum_{j'\in ov(r)} g_{j'}(\sum_{h\in s:j'\in h} x_h(t - T_{j'h} - T_{hj'})). \tag{5}$$

Theorem 1: If vector $x = (x_r, r \in R)$ solves the optimization problem given by equation (1) subject to equation (2) then x is an equilibrium point of the system (4,5).

Proof: In equilibrium $x_r(t) = x_r$. Substituting $x_r(t) = x_r$ in (5) we have

$$\lambda_r(t) = \sum_{j\in r} p_j(\sum_{h:j\in h} x_h) + \gamma \sum_{j'\in ov(r)} g_{j'}(\sum_{h\in s:j'\in h} x_h)$$

and then (4) becomes

$$\frac{d}{dt}x_r(t) = \kappa_r x_r\left(1 - \frac{\sum_{j\in r} p_j(\sum_{h:j\in h} x_h) + \gamma\sum_{j'\in ov(r)} g_{j'}(\sum_{h\in s:j'\in h} x_h)}{\partial U_s(\sum_{h\in s} x_h)/\partial x_r}\right)^+_{x_r}. \tag{6}$$

From (3) we have

$$\partial U_s(\sum_{h\in s} x_h)/\partial x_r = \sum_{j\in r} p_j(\sum_{s\in S}\sum_{h\in s:j\in h} x_h) + \gamma \sum_{j'\in ov(r)} g_{j'}(\sum_{s\in S}\sum_{h\in s:j'\in h} x_h). \tag{7}$$

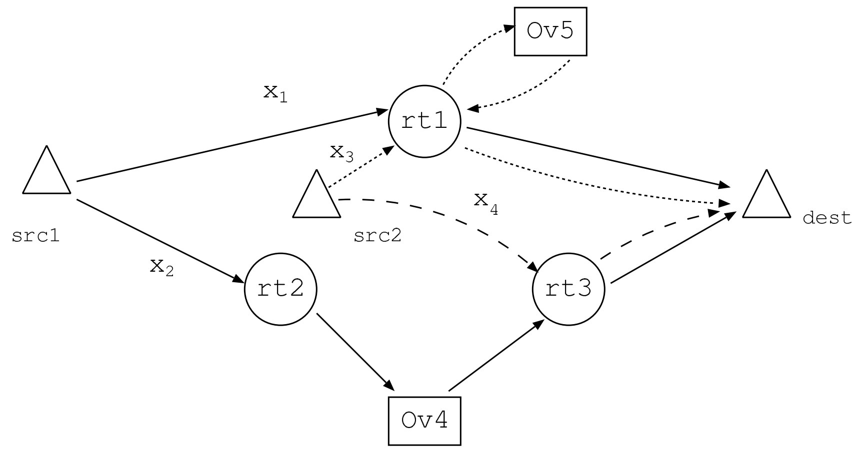

Fig. 2. Network example

Combining (6) with (7) we have

$$\frac{d}{dt}x_r(t) = \kappa_r x_r \left(1 - \frac{\partial U_s(\sum_{h\in s} x_h)/\partial x_r}{\partial U_s(\sum_{h\in s} x_h)/\partial x_r}\right)^+_{x_r(t)} = 0,$$

which is clearly a stationary point of the system (4,5). ■

Notice that the difference between the above fluid rate controller and the one given in [5] is in the formulation of function $\lambda_r(t)$. Function $\lambda_r(t)$ is the price marking rate. The fluid rate controller has a pricing marking scheme related to the network congestion and the budget constraint. The Lagrange multiplier is given by

$$\gamma = \frac{\partial U_s(\sum_{h\in s} x_h)/\partial x_r - \sum_{j\in r} p_j(\sum_{s\in S}\sum_{h\in s:j\in h} x_h)}{\sum_{j'\in ov(r)} g_{j'}(\sum_{s\in S}\sum_{h\in s:j'\in h} x_h)}, \quad r \in R : \exists j' \in J', j' \in r \tag{8}$$

or $\gamma = 0$ (which means that (2) is a loose bound, i.e., $\partial U_s(\sum_{h\in s} x_h)/\partial x_r = \sum_{j\in r} p_j(\sum_{s\in S}\sum_{h\in s:j\in h} x_h) \quad \forall r \in R$).

The term γ, from equation (8), can be interpreted as the steady state (optimal point) change in the utility $(\partial U_s(\sum_{h\in s} x_h)/\partial x_r - \sum_{j\in r} p_j(\sum_{s\in S}\sum_{h\in s:j\in h} x_h))$ over the money cost incurred by a change in the rate x_r $(\sum_{j'\in ov(r)} g_{j'}(\sum_{s\in S}\sum_{h\in s:j'\in h} x_h))$. This term, γ, is a constant that balances the increase in money cost with the respective increase in the utility minus congestion at the optimal rate allocation.

III. Example

We are now ready to look at a simple example. Take the network of Figure 2. RT1, RT2, RT3 are **bottleneck** routers, OV4 and OV5 are overlay nodes and SRC1 and SRC2 are sources that send traffic to destination DEST. Flows are represented by arrows labeled x_1, x_2, x_3 and x_4. Flows x_2 and x_3 use overlay routes and flows x_1 and x_4 use the available underlay path. Let $U_1(x_1+x_2) = (x_1+x_2)^{1-\alpha}/(1-\alpha)$ be the utility function of the

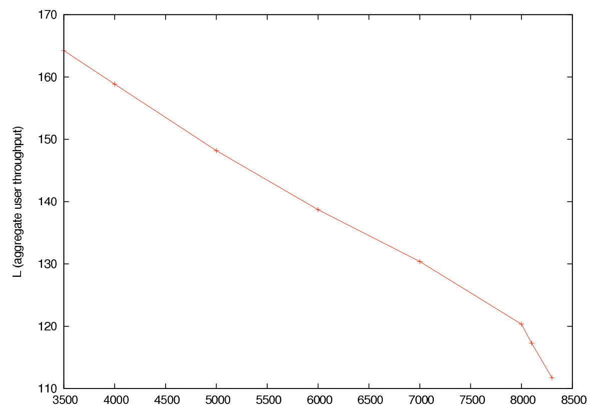

Fig. 3. Evolution of total aggregate user throughput (L) with overlay budget constraint (B).

source destination pair SRC1-DEST and $U_2(x_3+x_4) = (x_3+x_4)^{1-\alpha}/(1-\alpha)$ be utility function of the source destination pair SRC2-DEST. Let C_j be the outgoing link capacity of router/overlay j. Congestion pricing is given by $p_1(x_1+x_3) = ((x_1+2x_3)/C_1)^\beta/(\beta+1)$ at router RT1, $p_5(x_3) = (x_3/C_5)^\beta/(\beta+1)$ at overlay node Ov5 and so on. Outgoing link capacities are $C_1 = C_3 = 155$Mbps, $C_2 = 45$Mbps, $C_4 = C_5 = 100$Mbps. The money-traffic pricing at overlay node Ov4 is given by

$$f_4(x_2) = \begin{cases} \$1,000 & \text{if } x_2 < 1.5\text{Mbps} \\ \$1,000 + \$100 \times (x_2 - 1.5 \times 10^6)/10^6 & \text{otherwise} \end{cases}$$

and money pricing at overlay node Ov5 is given by

$$f_5(x_3) = \begin{cases} \$2,000 & \text{if } x_1 + x_3 < 3\text{Mbps} \\ \$2,000 + \$20 \times (x_3 - 3 \times 10^6)/10^6 & \text{otherwise} \end{cases}.$$

The total budget available to be spent on the overlay network is B dollars, i.e., $f_5(x_1 + x_3) + f_4(x_2) \leq B$.

Let $\beta = 30$, $\alpha = 2$ and $B = \$5,000$. Under this scenario, rates $x_1 \approx 79$Mbps, $x_2 \approx 17$Mbps, $x_3 \approx 27$Mbps and $x_4 \approx 23$Mbps are a solution to equation (1) subject to (2). Notice that for SRC1 the bottleneck of flow x_4 (C_2) is smaller than capacity the bottleneck of flow x_3 (C_1), SRC1 decides to send more traffic through the overlay. Applying the results to the Lagrange multiplier for both source/destination pairs SRC1/DEST and SRC2/DEST, equation (8), we find them to be equal $\gamma \approx 7 \times 10^{-5}$ and equation (6) shows the controller system to be in steady state.

Figure 3 shows the evolution of the the aggregate user rate ($L = \sum_{i=1}^{4} x_i$) of the example against an overlay budget B. There is a funny behavior in the system: The aggregated throughput goes down with the increase in the budget. To understand this behavior, let's look at the *per* user throughput. Figure 4 shows a graph of $L_1 = x_1 + x_2$ and $L_2 = x_3 + x_4$ against B. From the graph we can see that the increase in the overlay budget B makes the user throughput "more fair" until the budget is a little over \$8000. After that, SRC2 starts to get a higher throughput than SRC1. Two factors could have contributed to this fact. The first

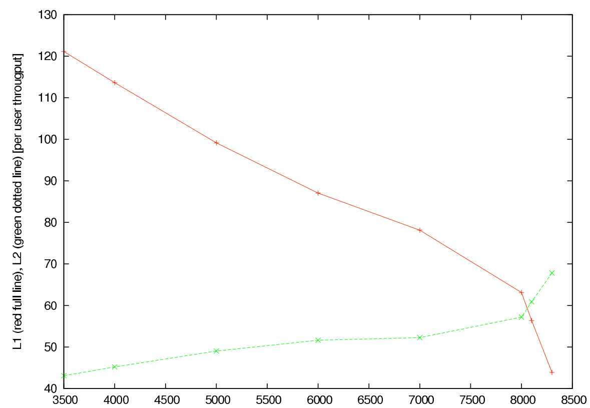

Fig. 4. Evolution of *per* user throughput (L_1 and L_2) with overlay budget constraint (B).

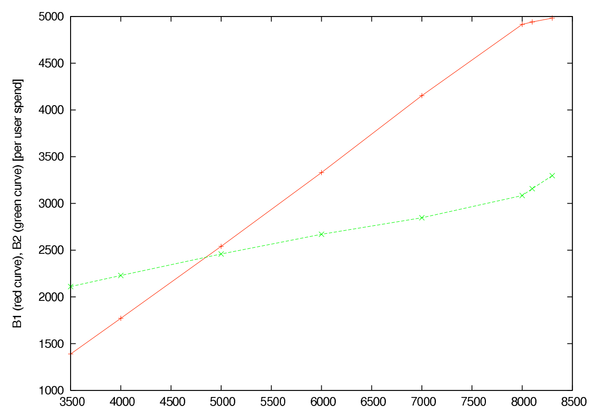

Fig. 5. Evolution of *per* user expenditure (B_1 and B_2) with overlay budget constraint (B).

is the traffic loop between the bottleneck router RT1 and the overlay node OV5. The second is the parameter $\alpha = 2$, which might not be adequate for to describe fairness in a Multi-path TCP scenario.

Another interesting metric is the *per* user expenditure (the amount of budget consumed by each user). Let $B_1 = f_4(x_2)$ and $B_2 = f_5(x_3)$, i.e., B_i money cost of routing SRCi user's traffic. Figure 5 plots B_1 and B_2 against B. Notice that user SRC1 expenditure rate is much higher than user SRC2 even though user SRC1 is losing throughput as user SRC2 increases its rate.

IV. Future work

- Computing γ to be used in the controller without solving the optimization problem is still an important open issue.

- The next step after computing γ is to analyze the dynamics of the system and its convergence.

Note: The software used for optimization problem showed itself somewhat unstable for $\alpha = 10$. It sometimes gives solutions that don't aggree with $dx_r(t)/dt = 0$ or sometimes it returns no solution at all. This instability should be investigated further.

REFERENCES

[1] S. Savage, A. Collins, E. Hoffman, J. Snell, and T. E. Anderson, "The end-to-end effects of internet path selection.," in *SIGCOMM*, pp. 289–299, 1999.

[2] D. G. Andersen, *Resilient Overlay Networks*. PhD thesis, Massachusetts Institute of Technology, 2001.

[3] H. Han, S. Shakkottai, C. Hollot, R. Srikant, and D. Towsley, ""overlay tcp for multi-path routing and congestion control."," in *Presented at the ENS-INRIA ARC-TCP Workshop*, Nov. 2003.

[4] D. K. Goldenberg, L. Qiuy, H. Xie, Y. R. Yang, and Y. Zhang, "Optimizing cost and performance for multihoming," *SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 4, pp. 79–92, 2004.

[5] F. Kelly and 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.