Optimal Relation Between the Flow and Latency of a Given Network

Optimization Theory and Applications

Eduardo Costa, 2019-82118
Professor: Insoon Yang

ABSTRACT

There are numerous articles that focus on optimizing the throughput of a network. Most of these articles use non-convex algorithms, and the few that use convex optimization, do not take into account the latency of the network's links. This project suggests a simple yet reliant approach to optimize this problem, resorting to convex optimization and finding the optimal network flow for given latency values. It is now left to try this approach in a real-world network and verify that the results hold.

INTRODUCTION

The objective of this project is to find the optimal ratio between the flow rates and the flow latency of a network, given its routing matrix and the link's capacity vector. To do so, both routing matrices and capacity vectors are randomly computer (within certain parameters) in order to simulate a real-world network, with different speeds and variable links. Then, both utility and latency functions are determined, in order to formulate the problem into a convex one.
With the problem formulated into a convex problem, **cvx** is run in order to optimize it, generating a regression of the utility in function of the latency.
A simple example of a network can be found in figure 1.

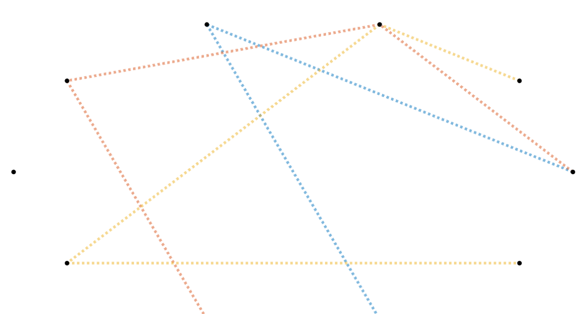

Figure 1: Graph of a simple network with 10 links (Black nodes) and 3 flow routes (coloured edges)

NETWORK PARAMETERS DEFINITION

Any network is made up by different devices, such as network interfaces, hubs, switches, routers and so on. These devices communicate between themselves, establishing links and then transmitting information on them. This transmission can be interpreted as a flow, and multiple flows can traverse the same link at the same time.

Therefore, any given network has **L** links, and **F** flows. However, a link between any two interfaces is not perfect and has an associated positive maximum capacity $\mathbf{C_i}$. Each flow passes over a fixed set of links, having its own flow route. Also, each flow has a non-negative flow rate $\mathbf{f_j}$ associated to it.
It is needed to identify which flows traverse which links (i.e. the links that assemble the different flow routes). To do so, it is used a routing matrix $\mathbf{R} \in \mathbb{R}^{\mathrm{LxF}}$, defined as

$$R_{\mathrm{ij}} = \begin{cases} 1, & \text{if flow } j \text{ traverses link } i, \\ 0, & \text{otherwise.} \end{cases}$$

The traffic on a given link **i**, represented by $\mathbf{t_i}$ is given by the sum of all the flow rates correspondent to the flows that traverse that link. Therefore, the link traffic vector $\mathbf{t} \in \mathbb{R}^{\mathrm{L}}$, can be defined as

$$t = Rf,$$

where $\mathbf{f} \in \mathbb{R}^{\mathrm{F}}$ represents the vector of flow rates.

It is now needed to define the latency of the network. The first step is to define the average queuing delay of any packet on a given link **i**, that is given by

$$d_{\mathrm{i}} = \frac{1}{c_{\mathrm{i}} - t_{\mathrm{i}}} = \frac{1}{c_{\mathrm{i}} - R_{\mathrm{i}} f},$$

assuming that no packets are dropped from the queue. Note that if $\mathbf{c_i} = \mathbf{R_i f}, \mathbf{d_i} = \infty$, which means that the packets that traverse the link **i** will never reach their destination. The total latency of a given flow $\mathbf{l_j}$ is given by the sum of the delays of the links traversed by it, $\mathbf{d_{ij}}$. This can be expressed by

$$l = R^{\intercal} d,$$

where $\mathbf{l} \in \mathbb{R}^{\mathrm{F}}$ is the vector that contains all the flow latencies, and $\mathbf{d} \in \mathbb{R}^{\mathrm{L}}$ is the vector that contains the delays of each link.

With this, the maximum flow latency of the network **L** is defined as

$$L = max\{l\} = max\{l_1, ..., l_{\mathrm{F}}\}.$$

FORMALIZING THE PROBLEM

The next course of action is to express this as a convex optimization problem. The first step is to define the utility function of this problem, which is given by

$$U(f) = \sum_{j=1}^{F} log(f_{\mathrm{j}}),$$

where the logarithm was applied individually to each flow rate so that the utility function is concave.

Focusing first on maximizing the utility function and ignoring the latency, it is still needed to take into account some constraints. As said initially, the flow rates must always be non-negative and the traffic of a given link cannot be superior to its capacity, i.e.

$$Rf \preceq C,$$

where $\mathbf{C} \in \mathbb{R}^{\mathrm{L}}$ is the vector that contains the capacities of the links.

This conditions can all be expressed in a single optimization problem

$$\begin{array}{ll} \underset{f}{\text{maximize}} & \displaystyle\sum_{j=1}^{F} log(f_{\mathrm{j}}) \\ \text{subject to} & Rf \preceq C, \\ & f \succeq 0. \end{array}$$

Next the utility function is ignored, and the maximum flow latency is minimized. Since the link delays increase with traffic, having a zero flow on the network will minimize the flow latency, which makes

$$l = R^{\intercal} d = R^{\intercal} \left(\frac{1}{c_1}, ..., \frac{1}{c_{\mathrm{L}}} \right).$$

This gives that

$$L_{\min} = max \left(R^{\intercal} \left(\frac{1}{c_1}, ..., \frac{1}{c_{\mathrm{L}}} \right) \right),$$

where max is the maximum over the entries of the vector $R^{\intercal} \left(\frac{1}{c_1}, ..., \frac{1}{c_{\mathrm{L}}} \right)$.

With both maximum achievable utility and minimum achievable latency determined, it is possible to find the optimal relation between them, maximizing the utility and minimizing the latency. Therefore, the final problem takes the form

$$\begin{array}{ll} \underset{f}{\text{maximize}} & \sum_{j=1}^{F} log(f_{\mathrm{j}}) \\ \text{subject to} & \sum_{i=1}^{L} \frac{R_{\mathrm{ij}}}{c_{\mathrm{i}} - R_{\mathrm{i}}{}^{\intercal} f} \leq L, \quad j = 1, ..., F, \\ & Rf \preceq C, \\ & f \succeq 0, \end{array}$$

for a range of values of $\mathbf{L} \geq \mathbf{L_{min}}$.

If $\mathbf{Rf} \preceq \mathbf{C}$ is true, then $\mathbf{C_i} - \mathbf{R_i}^{\intercal}\mathbf{f} \geq 0$ is also true for $\mathbf{i} = 1, ..., \mathbf{L}$. This means that the constraints $\sum_{\mathbf{i}=1}^{\mathbf{L}} \frac{\mathbf{R_{ij}}}{\mathbf{c_i} - \mathbf{R_i}^{\intercal}\mathbf{f}} \leq \mathbf{L}$, $\mathbf{j} = 1, \ldots , \mathbf{F}$ are convex.
Therefore, this is a convex optimization problem.

RESULTS

The problem is now optimized using **cvx**, a Matlab-based modeling system for convex optimization. The network is created with 100 links and 30 flows, and the utility function is optimized for 25 different values of latency. The trade-off curve between the utility and the latency of the network can be found in figure 2.

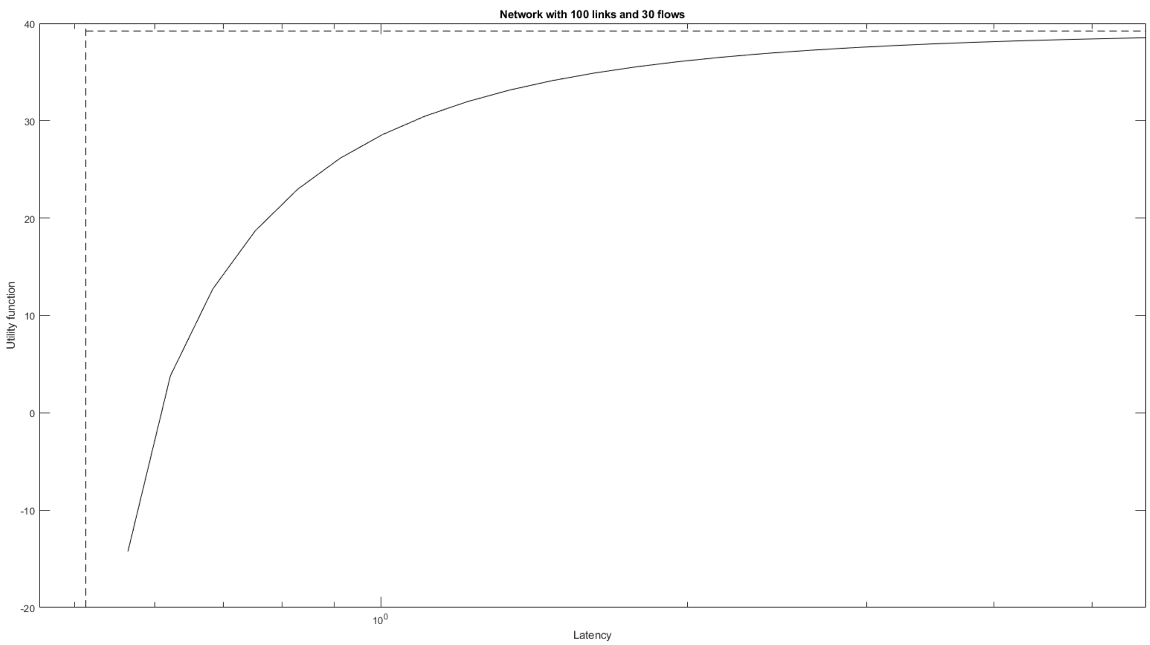

Figure 2: Solid line: Optimal trade-off between the flow and latency of the network. Dashed lines: Minimum latency (vertical) and maximum utility function (horizontal)

DISCUSSION AND CONCLUSIONS

The obtained trade-off between both flow and latency of the network plotted in figure 2 is as expected. For the minimum value of latency, the utility function will tend to minus infinity, which corresponds to a zero flow rate. For an infinite value of latency, the flow rates are maximum, which corresponds to the maximum utility function.

REFERENCES

[**1**] Michael Zargham, Alejandro Ribeiro, Ali Jadbabaie, Accelerated Dual Descent for Constrained Convex Network Flow Optimization: 52nd IEEE Conference on Decision and Control, December 10-13, 2013. Florence, Italy
[**2**] Alireza Nazemi, Farahnaz Omidi, A Capable Neural Network Model for Solving the Maximum Flow Problem: Journal of Computational and Applied Mathematics, Volume 236, Issue 14, August 2012, Pages 3498-3513