Seminar

Optimization in Communications and Signal Processing

Ferienakademie 2009, Sarntal

Optimal Cost Multicast over Coded Packet Networks

Stephan M. Günther

Abstract — **Consider a situation where one source wants to transmit the same information to multiple receivers. Such a problem is referred to as *single source multicast*. However, in case of routed networks it is very hard to find the optimal distribution path from a sender to a given set of receivers. Employing network coding allows packets leaving forwarding devices to be not only replicas but arbitrary causal mappings of incoming packets. This extension to basic capabilities leads to interesting simplifications regarding the calculation of optimal paths for distributing multicast traffic. This report gives a short introduction to problems regarding multicast in routed networks and introduces network coding. Afterwards the task of finding minimum cost paths for distribution of multicast traffic is stated as optimization problem. A polynomial time algorithm for static multicast (situations where membership of multicast groups does not change) first discussed by Lun et al. [LuRa06] is presented. Finally the problem is adapted to consider convex cost functions and elastic rates.**

1 Introduction

A wide range of media which used to be sent over broadcast channels in the past is distributed over data networks today. Examples are entertainment services like IPTV and internet based radio channels. The requirement of such a service may be given by a rate which the network must provide. Additionally, objectives are modeled by an abstract cost term, e.g. any combination of monetary cost to transmit a given amount of data over a link, the reliability of a link or the transmission delay. The goal is to find a path from the sender to the receivers which supports the demanded rate and concurrently minimizes the costs. One possible way to accomplish this is to find a minimum spanning tree (MST), which interconnects a sender with all receivers and concurrently satisfies the rate demanded by the given service. Note, that this MST does in general not contain all nodes of a network but only the subset of nodes necessary to establish a suitable communication path between the sender and all receivers. This would result in an optimal-cost multicast for routed networks. However, calculating an MST on a subset of nodes (even without considering the limited bandwidth of links) equates to the Steiner tree problem which is known to be $\mathcal{NP}$-complete. Employing network coding allows message forwarding devices (e.g. routers) not only to replicate incoming packets but to send packets which are arbitrary causal combinations of incoming packets. This was first described by Ahlswede et al. [AhCa00] in 2000. It is interesting that the problem of finding an optimal distribution path becomes solvable in polynomial time in this case.
The remainder of this report is organized as follows: Chapter 2 first considers dedicated unicast connections between the sender and each receiver which leads to a *multi-commodity flow problem*. Afterwards the problems of multicast in routed networks are discussed. Finally network coding is introduced and the resulting possibilities are explained by means of an example. In Chapter 3 the task of minimizing costs for the multi-commodity flow problem and multicast in coded packet networks is stated as an optimization problem. The latter is very similar to the multi-commodity flow problem and is discussed in more detail in Chapter 4 where approaches to solve it for certain types of cost functions are given.

2 Multicast in routed and coded packet networks

Throughout this report only single-source multicast is being discussed. This refers to situations, where exactly one sender wants to transfer the same data to a set of receivers. A communication network is represented by a weighted directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ where $\mathcal{N}$ denotes a set of nodes and $\mathcal{A}$ a set of arcs. Devices like sender, receivers and forwarding devices are each represented by a node $u \in \mathcal{N}$ while the arcs model the link between two nodes. For simplicity we assume lossless wireline links only. An arc can be written as tuple (u, v) where $u, v \in \mathcal{N}$ are its start and end nodes respectively. Each arc has a capacity c_{uv} which specifies the maximum rate at which data can be sent over it. To make the examples of this chapter more illustrative we further assume some kind of costs per link which are incurred by its usage, e.g. the monetary cost to send data over this link. The costs are denoted by a weight a_{uv} for each arc $(u, v) \in \mathcal{A}$. Figure

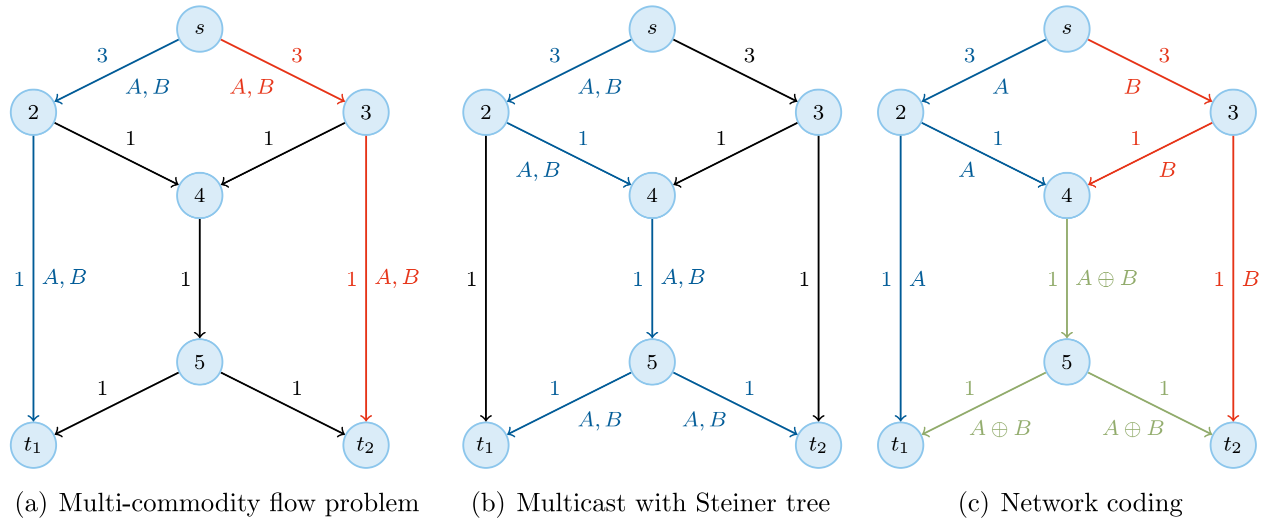


Figure 1: Comparison between routed and coded packet networks.

1 shows an example of such a network. The numbers beneath each arc represent the costs.

2.1 Dedicated unicast connections in routed networks

A trivial approach of distributing the same information from a sender $s \in \mathcal{N}$ to a set of receivers $t \in \mathcal{T} \subseteq \mathcal{N}$ consists of establishing unicast connections between the sender and each receiver. The same information is therefore transmitted $|\mathcal{T}|$ times. This results in a special case of the *multi-commodity flow problem* which models a situation, where different goods (commodities) are to be sent from a source to a specific receiver (also called *sinks*). In case of communication networks the goods are some kind of information flows. Figure 1(a) gives an example of this situation. The source node s wants to transmit two messages A, B to two sinks t_1, t_2. Consequently there are two flows, one for transmitting A, B to t_1 and the other one to transmit the same information to t_2. The fact that both flows are identical (both messages to both receivers) is not demanded by the multi-commodity flow problem, but is a result of the attempt to model a multicast scenario. This also implies that the source sends the same information multiple times. Note that a flow can be uniquely identified by the receiver t since there is only a single source. The objective is to find minimum-cost paths to both sinks. Assuming sufficient bandwidth of all links involved, this can easily be accomplished by calculating the shortest path tree (SPT) rooted at s with well known and efficient algorithms, e.g. Dijkstra [CoLe01]. This results in the optimal paths to both sinks individually. Note that this is not necessarily the lowest cost path for distributing the information to both nodes simultaneously as we will see in Section 2.2. Due to the symmetry of the network given in this example the shortest path to t_1 is $(s, 2, t_1)$ at cost of 4 while the shortest path to t_2 is $(s, 3, t_2)$ at the same cost. Transmitting both flows therefore results in a total cost of 16. If the link capacities are considered it is no longer possible to handle the flows independently since they might compete for resources on shared links. For example, arc $(4, 5)$ in Figure 1(a) becomes such a shared link if the costs of $(2, t_1)$ and $(3, t_2)$ are increased. Assuming a link capacity of one message per unit time for all links the rate demands of the flows are no longer satisfied by the shared link. However, there still exist algorithms to solve the problem in polynomial time, e.g. combining the successive shortest path algorithm with the capacity scaling algorithm (see [KoVy05] pp. 199).

2.2 Multicast in routed networks

As we have seen dedicated connections do not always yield an optimal way of distributing the same data to multiple receivers. If we do no longer demand dedicated unicast connections, the sending node s does not necessarily have to send each information $|\mathcal{T}|$ times. Therefore, lower costs might be achieved if intermediate nodes duplicate incoming traffic which would otherwise be sent multiple times by s. Given the situation in Figure 1(b) data is sent to node 5 only once. Afterwards node 5 duplicates the messages and sends a copy to each of the sinks. Such nodes form the root of a new subtree generating possibly many copies of incoming packets and are called *rendezvous points*. Assuming sufficient capacity of all links involved this results in a total cost of 14 in comparison to 16 as for optimal solution of the the multi-commodity flow problem. For routed networks with multicast ability the optimal distribution path is a Steiner tree. This corresponds to the subgraph in Figure 1(b). Depending on the weights, the Steiner tree may degenerate to individual shortest paths. This situation would arise if the costs for arc $(4, 5)$ were increased to a value of at least 2. The main problem here is to find the Steiner Tree which is not solvable in polynomial time, even without capacity constraints, and therefore considered to be very hard for large networks [HwRi92]. For this reason Steiner trees are not calculated in real world scenarios. Instead heuristics and specialized multicast routing protocols are used to find suboptimal but acceptable distribution paths in affordable time which is beyond the scope of this report.

2.3 Multicast in coded packet networks

Both the multi-commodity flow problem and multicast in routed networks are either not optimal or computational not affordable. However, there is no reason to restrict nodes of a network to replication of incoming packets. This is where *network coding* comes in: It allows nodes to send arbitrary causal combinations of incoming packets. For example consider a situation as depicted in Figure 1(c). The source s wants to send again messages A, B to both sinks. The messages A and B are sent directly to sinks t_1 and t_2 respectively using the shortest path. However, nodes 2 and 3 duplicate the incoming packets and send a copy also to node 4. This one makes use of its new abilities and calculates the bitwise XOR $A \oplus B$ which is sent to both sinks. Since each sink has already received one of the messages they can both recover the missing one. For example sink t_1 can recover $B = A \oplus (A \oplus B)$. The total cost for sending both messages in this example is 13 and therefore even lower than in the optimal case for routed networks. The costs improved by $18,75\%$ compared to unicast connections and by $7,14\%$ compared to the multicast scenario without network coding. The improvement greatly depends on the characteristics of the network. For example making the arc $(4, 5)$ arbitrary expensive results in dedicated unicast connections as optimal solution. In contrast by making the arcs $(2, t_1)$ and $(3, t_2)$ arbitrary expensive would result in a Steiner tree as optimal solution. Both cases require only trivial coding of packets $(X \oplus 0)$ but are also feasible solutions. In any case network coding finds the optimal solution for the given situation and simultaneously stays solvable in polynomial time. These two qualities make network coding interesting for real world applications.

3 Formulation of multicast as optimization problem

The situations described so far can be expressed in a more formal way. Let a non-negative value z_{uv} denote the total rate at which data is injected into arc $(u, v) \in \mathcal{A}$ and $\boldsymbol{z}$ the vector containg the rates for all links in some given order. The costs depend on the rate vector and can therefore be expressed by means of a *cost function*

$$f(\boldsymbol{z}) : \ \mathbb{R}^{|\mathcal{A}|} \rightarrow \mathbb{R}^+ \ := \ [0, \infty) \tag{1}$$

which increases monotonically in $\boldsymbol{z}$. All links have a capacity c_{uv}. Therefore, the total rate z_{uv} for a given link can never be greater than its capacity. This relation is also called *capacity constraint*:

$$c_{uv} \geq z_{uv}, \quad \forall (u, v) \in \mathcal{A}. \tag{2}$$

The relation between the total rate z_{uv} and the non-negative fraction $x_{uv}^{(t)}$ of this rate by flow $t \in \mathcal{T}$ depends on the network's capabilites, e.g. if coding is allowed. In any case the overall rate of a link must at least surpass the partial rates:

$$z_{uv} \geq x_{uv}^{(t)}, \quad \forall t \in \mathcal{T}, \ \forall (u, v) \in \mathcal{A}. \tag{3}$$

Together with the capacity constraint this describes a competitive situation between flows because available capacity must be shared in some way. Therefore, (3) is also called *coupling constraint* since it ties the flows together. Furthermore we assumed that information is generated only by a single source $s \in \mathcal{N}$ and consumed only by sink nodes $t \in \mathcal{T}$. All intermediate nodes only forward information. This is described by the *flow conservation constraint*

$$\sum_{\{v|(u,v) \in \mathcal{A}\}} x_{uv}^{(t)} - \sum_{\{v|(v,u) \in \mathcal{A}\}} x_{vu}^{(t)} = \sigma_u^{(t)}, \quad \forall u \in \mathcal{N}, \ \forall t \in \mathcal{T}, \tag{4}$$

where

$$\sigma_u := \begin{cases} R & \text{if } u = s, \\ -R & \text{if } u = t, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Here, R denotes a positive fixed rate which is demanded by the service. In case of dedicated unicast connections this is quite easy to see. Information for flow $t \in \mathcal{T}$ is generated by s at rate R and received by sinks $t \in \mathcal{T}$ at the same rate. All intermediate nodes send the same amount of messages as they receive. In case of multicast in routed networks it is possible that a single message belongs to two flows wherefore the number of messages entering and leaving a node may differ. This is the case at node 5 in Figure 1(b). It receives two messages A and B only once but both messages leave node 5 two times. This is taken into account in (4) since the message entering node 5 belong both to flow t_1 and t_2. The same argumentation holds for network coding. We will now state the optimization problem of minimizing $f(\boldsymbol{z})$ in compliance with these constraints for the multi-commodity flow problem and for multicast in coded packet networks. In the former case this results in the minimum-cost flow problem, a special variant of the multi-commodity flow problem. The latter case turns out to be very similar. The optimization problem for multicast in routed networks is omitted because it can not easily be written in a compact form.

3.1 Minimum-cost flow problem

We assume a routed network adhering to the flow conservation constraint. The capacity of each link is assumed to be finite. Multiple flows $t \in \mathcal{T}$ carrying identical information are to be sent from one source to multiple sinks. The partial rates $x_{uv}^{(t)}$ of each flow sent through a link are assumed to be non-negative. The problem of minimizing the costs given by a general cost function $f(\boldsymbol{z})$ with respect to the rate vector $\boldsymbol{z}$ is given by

$$
\begin{aligned}
&\text{minimize} \quad f(\boldsymbol{z}) && && \qquad (6)\\
&\text{subject to } c_{uv} \geq z_{uv} && \forall (u,v) \in \mathcal{A}, && \textit{(capacity constraint)}\\
&\qquad z_{uv} \geq \sum_{t \in \mathcal{T}} x_{uv}^{(t)}, \quad x_{uv}^{(t)} \geq 0 && \forall (u,v) \in \mathcal{A},\ \forall t \in \mathcal{T}, && \textit{(coupling constraint)}\\
&\sum_{\{v|(u,v)\in\mathcal{A}\}} x_{uv}^{(t)} - \sum_{\{v|(v,u)\in\mathcal{A}\}} x_{vu}^{(t)} = \sigma_u^{(t)} && \forall u \in \mathcal{N},\ \forall t \in \mathcal{T}. && \textit{(flow conservation)}
\end{aligned}
$$

The total rate of a link is bounded above as given by the capacity constraint. At the same time it is bounded below by the sum of the non-negative partial rates per flow. The optimal rate vector $\boldsymbol{z}^*$ for this optimization problem therefore lies within a bounded polyhedron in the positive orthant[1]. Since the cost function $f(\boldsymbol{z})$ is monotonically increasing the components z_{uv}^* must be minimal for an optimal solution $\boldsymbol{z}^*$. This is fulfilled iff the coupling constraint holds with equality which is why it is called an *active constraint*. The result of this optimization problem is given by the optimal rates z_{uv}^* and $x_{uv}^{(t)*}$ for all $(u,v) \in \mathcal{A}$ and $t \in \mathcal{T}$. These describe a set of individual shortest paths from the source to the sinks considering the capacity constraint and the competition for available resources in form of bandwidth between the flows.

[1] The positive orthant of an n-dimensional coordinate system is given by the section, where all components of an n-dimensional vector are non negative.

3.2 Single source multicast in coded packet networks

For this case we assume a network with the ability to code outgoing packets as described in Section 2.3. The other conditions are the same as in Section 3.1. The only modification compared to the multi-commodity flow problem affects the coupling constraint. The total rates z_{uv} are no longer bounded below by the sum of the partial rates for each flow. Since coding is allowed, messages for different flows may be combined resulting in a lower total rate. The optimization problem looks very similar:

$$\begin{aligned} &\text{minimize} && f(\boldsymbol{z}) && && (7)\\ &\text{subject to} && c_{uv} \geq z_{uv} && \forall (u,v) \in \mathcal{A}, && \textit{(capacity constraint)}\\ & && z_{uv} \geq x_{uv}^{(t)} \geq 0 && \forall (u,v) \in \mathcal{A},\ \forall t \in \mathcal{T}, && \textit{(coupling constraint)}\\ & && \sum_{\{v|(u,v)\in\mathcal{A}\}} x_{uv}^{(t)} - \sum_{\{v|(v,u)\in\mathcal{A}\}} x_{vu}^{(t)} = \sigma_u^{(t)} && \forall u \in \mathcal{N},\ \forall t \in \mathcal{T}. && \textit{(flow conservation)} \end{aligned}$$

The coupling constraint is active again. The optimal rate vector $\boldsymbol{z}^*$ still lies in the positive orthant of a bounded polyhedron, solely the bounds have changed. According to the new coupling constraint the total rate z_{uv} for a link $(u,v) \in \mathcal{A}$ has to surpass the partial rates $x_{uv}^{(t)}$ for all flows and is therefore limited by a new lower bound. This means, that solutions for the problem given in (6) are also feasible solutions for the new problem. However, there might exist better solutions now since the new constraint is less restrictive. The optimal rate vector $\boldsymbol{z}^*$ is now given, iff all its components fulfill the following equality:

$$z_{uv} := \|x_{uv}^{(t)}\|_{\infty,(t)} = \max_{t\in\mathcal{T}}\{x_{uv}^{(t)}\}, \quad \forall (u,v) \in \mathcal{A}. \tag{8}$$

Since we did not specify the cost function $f(\boldsymbol{z})$ we cannot make a statement about the complexity of this problem. The next chapter will assume two specific types of cost functions which allows to solve (7) in polynomial time.

4 Solving selected problems for coded packet networks

Based on the general formulation of minimum-cost multicast over coded packet networks as given in (7) we will now make additional assumptions with respect to the cost function. Depending on its special characteristics some of the constraints are weakened yielding the possibility to solve the problem more efficiently or in a distributed manner. Two of these cost functions are discussed in the following sections and a polynomial time solvable distributed algorithm first described by Lun et al. [LuRa06] is presented.

4.1 Linear, separable cost and separable constraints

Reconsider the situation discussed in Section 2.3 and illustrated by Figure 1(c). We assign a positive, fixed cost per unit rate a_{uv} to each link $(u,v) \in \mathcal{A}$. This can be expressed by a cost function $f(\boldsymbol{z})$ given as weighted sum of the total rates z_{uv} times the costs a_{uv} per link:

$$f(\boldsymbol{z}) := \sum_{(u,v)\in\mathcal{A}} a_{uv} z_{uv}. \tag{9}$$

According to this definition the cost function is linear and monotonically increasing in $\boldsymbol{z}$. Therefore it takes its minimum iff the total rates z_{uv} are minimal for all $(u,v) \in \mathcal{A}$. This is fulfilled iff the coupling constraint wherefore (8) holds. By taking this characteristic of the cost function into account we can restate general optimization problem:

$$\begin{aligned}
&\text{minimize} \quad f(\boldsymbol{z}) && \quad (10)\\
&\text{subject to } c_{uv} \geq x_{uv}^{(t)} \geq 0 && \forall (u,v) \in \mathcal{A},\ \forall t \in \mathcal{T},\\
&\qquad z_{uv} \geq x_{uv}^{(t)} && \forall (u,v) \in \mathcal{A},\ \forall t \in \mathcal{T},\\
&\qquad \sum_{\{v|(u,v)\in\mathcal{A}\}} x_{uv}^{(t)} - \sum_{\{v|(v,u)\in\mathcal{A}\}} x_{vu}^{(t)} = \sigma_u^{(t)} && \forall u \in \mathcal{N},\ \forall t \in \mathcal{T}.
\end{aligned}$$

Note, that z_{uv} is no longer explicitly bounded above while negative values are explicitly forbidden as solution to this problem due to the coupling between the total rates and the partial rates per flow. We refer to (10) as the primal problem. Now we dualize the coupling constraints $z_{uv} \geq x_{uv}^{(t)}$ and state the Lagrangian:

$$\begin{aligned}
L(\boldsymbol{x},\boldsymbol{z},\boldsymbol{\lambda}) &= f(\boldsymbol{z}) + \sum_{(u,v)\in\mathcal{A}} \sum_{t\in\mathcal{T}} \lambda_{uv}^{(t)} (x_{uv}^{(t)} - z_{uv})\\
&= \sum_{(u,v)\in\mathcal{A}} a_{uv} z_{uv} + \sum_{(u,v)\in\mathcal{A}} \sum_{t\in\mathcal{T}} \lambda_{uv}^{(t)} (x_{uv}^{(t)} - z_{uv})\\
&= \sum_{(u,v)\in\mathcal{A}} \left(a_{uv} - \sum_{t\in\mathcal{T}} \lambda_{uv}^{(t)} \right) z_{uv} + \sum_{(u,v)\in\mathcal{A}} \sum_{t\in\mathcal{T}} \lambda_{uv}^{(t)} x_{uv}^{(t)}.
\end{aligned} \tag{11}$$

The remaining flow conservation and capacity constraints can by abbreviated by means of a set $\mathcal{X}$ containing all vectors $\boldsymbol{x}$ adhering to these constraints. The dual function is now obtained by minimizing the Lagrangian with respect to $\lambda_{uv}^{(t)} \geq 0$ for all $(u,v) \in \mathcal{A}$ and $t \in \mathcal{T}$. Since the coupling constraint has been dualized and z_{uv} is no longer bounded explicitly, it can be assumed as unbounded for the minimization:

$$\Theta(\boldsymbol{\lambda}) := \inf_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{z}} L(\boldsymbol{x},\boldsymbol{z},\boldsymbol{\lambda}) = \begin{cases} \min\limits_{\boldsymbol{x}\in\mathcal{X}} \sum\limits_{(u,v)\in\mathcal{A}} \sum\limits_{t\in\mathcal{T}} \lambda_{uv}^{(t)} x_{uv}^{(t)} & \text{if } \sum\limits_{t\in\mathcal{T}} \lambda_{uv}^{(t)} = a_{uv},\\ -\infty & \text{otherwise.} \end{cases} \tag{12}$$

The dual problem now consists of maximizing the dual function which in general gives a lower bound of the primal optimal solution. Therefore, the dual problem yields no meaningful solution for $\Theta(\boldsymbol{\lambda}) \to -\infty$ although this would still be a valid lower bound for $f(\boldsymbol{z})$. To avoid this trivial bound, we demand a new constraint regarding the Lagrangian multipliers. This eliminates the total rates from the dual function:

$$\begin{aligned} &\text{maximize} \;\; \Theta(\boldsymbol{\lambda}) = \min_{\boldsymbol{x}\in\mathcal{X}} \sum_{t\in\mathcal{T}} \sum_{(u,v)\in\mathcal{A}} \lambda_{uv}^{(t)} x_{uv}^{(t)}, \\ &\text{subject to} \;\; \sum_{t\in\mathcal{T}} \lambda_{uv}^{(t)} = a_{uv},\;\; \lambda_{uv}^{(t)} \geq 0 \quad \forall (u,v)\in\mathcal{A},\;\; \forall t\in\mathcal{T}. \end{aligned} \tag{13}$$

Consequently, the optimal solution for the dual problem is given by $\max_{\boldsymbol{\lambda}} \Theta(\boldsymbol{\lambda}) \leq f(\boldsymbol{z}^*)$. In the scope of our problem this is also the optimal primal solution since strong duality holds. To see this we can rewrite the constraints which have been dualized as constraint functions:

$$g_{uv}^{(t)}(z_{uv}) := x_{uv}^{(t)} - z_{uv} \leq 0, \quad \forall (u,v)\in\mathcal{A},\;\; \forall t\in\mathcal{T}. \tag{14}$$

Obviously all constraint functions $g_{uv}^{(t)}$ are affine wherefore Slater's condition[2] reduces to primal feasibility and guarantees strong duality ([BoVa08], pp. 226 – 227).

To point out the advantage of solving the dual problem instead of the primal one we rewrite (13) as follows:

$$\text{maximize} \;\; \Theta(\boldsymbol{\lambda}) = \sum_{t\in\mathcal{T}} \xi^{(t)} \tag{15}$$

$$\text{with} \qquad \xi^{(t)} := \min_{\boldsymbol{x}^{(t)}\in\mathcal{X}^{(t)}} \sum_{(u,v)\in\mathcal{A}} \lambda_{uv}^{(t)} x_{uv}^{(t)}, \tag{16}$$

$$\text{subject to} \;\; \sum_{t\in\mathcal{T}} \lambda_{uv}^{(t)} = a_{uv},\;\; \lambda_{uv}^{(t)} \geq 0 \quad \forall (u,v)\in\mathcal{A},\;\; \forall t\in\mathcal{T}.$$

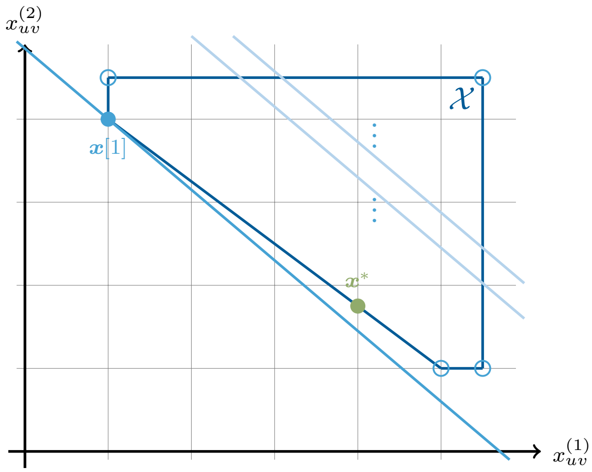
The dual problem (15) can be solved by maximizing it for each flow individually. In turn $\xi^{(t)}$ can be determined by minimizing (16) for all arcs of each node. The coupling between flows is broken which allows for distributed algorithms. Such an algorithm has been described by [LuRa06] and works roughly as follows:

1. Choose a feasible start vector $\boldsymbol{\lambda}[0]$ which fulfills the constraints imposed on $\lambda_{uv}^{(t)}$.
2. Obtain the optimal vector $\boldsymbol{x}[n]$ for the given $\boldsymbol{\lambda}[n]$ by solving (16).
3. Determine a better $\boldsymbol{\lambda}[n+1]$ to maximize (15) with respect to $\boldsymbol{x}[n]$.
4. Repeat steps 2) and 3) until some stop criterion is reached.

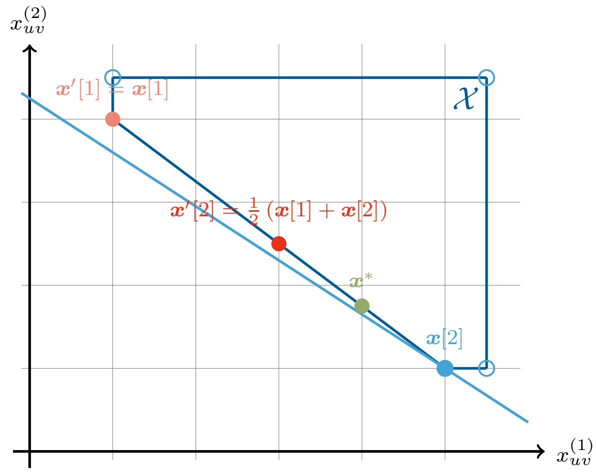
Subproblem (16) can thereby be solved in a variety of ways, e.g. ϵ-relaxation which is not covered here. To update $\boldsymbol{\lambda}[n]$ in step 3) we employ subgradient optimization since (16) is not differentiable. Furthermore the components of the updated vector $\boldsymbol{\lambda}[n+1]$ must comply with the constraints posed in $\lambda_{uv}^{(t)}$. Let

$$\Lambda_{uv} := \left\{ \boldsymbol{\lambda}_{uv} \;\middle|\; \sum_{t\in\mathcal{T}} \lambda_{uv}^{(t)} = a_{uv}, \;\; \lambda_{uv}^{(t)} \geq 0, \;\; \forall t\in\mathcal{T} \right\} \tag{17}$$

[2]Slater's condition is a constraint qualification guaranteeing equality of primal and dual optimal solutions. The case of exclusively affine constraint functions f_i is a special case of Slater's condition where $f_i \leq 0 \;\forall i$ is sufficient for strong duality.



(a) $\boldsymbol{x}'[1] = \mu_1[1]\boldsymbol{x}[1] = \boldsymbol{x}[1]$



(b) $\boldsymbol{x}'[2] = \mu_1[2]\boldsymbol{x}[1] + \mu_2[2]\boldsymbol{x}[2] = \frac{1}{2}(\boldsymbol{x}[1] + \boldsymbol{x}[2])$

Figure 2: A suitable convex combination of primal feasible points $\boldsymbol{x}$ as obtained by the subgradient optimization converges to optimal primal solution $\boldsymbol{x}^*$.

be the $|\mathcal{T}|$-dimensional simplex which contains all feasible $\boldsymbol{\lambda}_{uv} := [\lambda_{uv}^{(1)}, \ldots, \lambda_{uv}^{(|\mathcal{T}|)}]^T$ for a specific arc $(u, v) \in \mathcal{A}$. We can now determine a value $\boldsymbol{\lambda}_{uv}[n+1]$ by first executing one subgradient step of step size $h[n]$. This yields a $\tilde{\boldsymbol{\lambda}}_{uv}$ which is not necessarily feasible. Afterwards we can choose the nearest feasible value $\boldsymbol{\lambda}_{uv}$ according to the Euclidean norm:

$$\begin{aligned} \boldsymbol{\lambda}_{uv}[n+1] &:= \underset{\boldsymbol{\lambda}_{uv} \in \Lambda_{uv}}{\arg\min} \|\boldsymbol{\lambda}_{uv} - (\boldsymbol{\lambda}_{uv}[n] + h[n]\boldsymbol{x}_{uv}[n])\|^2 \\ &= \underset{\boldsymbol{\lambda}_{uv} \in \Lambda_{uv}}{\arg\min} \sum_{t \in \mathcal{T}} (\lambda_{uv}^{(t)} - (\lambda_{uv}^{(t)}[n] + h[n]x_{uv}^{(t)}[n]))^2 \end{aligned} \tag{18}$$

Equation (18) describes an Euclidean projection of $\tilde{\boldsymbol{\lambda}}_{uv}$ onto the feasible set Λ_{uv}. This can be solved by means of another optimization problem as discussed in [LuRa06].

Now that we have the optimal dual solution of the $(n+1)$-st step of the optimization we still have to recover an appropriate primal solution. Since the problem as given by (16) is not differentiable we use subgradients which are not unique and therefore yield no optimal or unique solution for the primal problem. However, an optimal primal solution can be recovered as described by Sherali and Choi [ShCh96]. The idea is, that the primal optimal solution is a convex combination of the dual optimal solutions of multiple optimization steps. Consider the example given in Figure 2. Note that the given set $\mathcal{X}$ is only used to exemplify the primal recovery and has no relation to the actual appearance as defined by the constraints. An optimal recovered solution $\boldsymbol{x}^*$ for the primal problem may be located on the boundary of a convex set. This primal optimum is a convex combination of two extremal points $\boldsymbol{p}_1, \boldsymbol{p}_2 \in \mathcal{X}$. Each optimization step n gives a primal feasible result $\boldsymbol{x}[n] \in \{\boldsymbol{p}_1, \boldsymbol{p}_2\}$. Let $\boldsymbol{x}'[n]$ denote a suitable convex combination in the n-th iteration:

$$\begin{aligned}
\boldsymbol{x}'[1]: \quad &= \mu_1[1]\boldsymbol{x}[1] = \boldsymbol{x}[1] \\
\boldsymbol{x}'[2] \quad &= \mu_1[2]\boldsymbol{x}[1] + \mu_2[2]\boldsymbol{x}[2] \\
&\vdots \\
\boldsymbol{x}'[n] \quad &= \sum_{l=1}^{n} \mu_l[n]\boldsymbol{x}[l] \\
\lim_{n\to\infty} \boldsymbol{x}'[n] &= \boldsymbol{x}^*
\end{aligned}$$

An acceptable option for the convex weights is $\mu_l[n] := 1/n$. This corresponds to an averaging of the primal feasible points with respect to their frequency of occurrence. A more general definition and the proof of convergence can be found in [LuRa06].

4.2 Convex, separable cost and separable constraints

Now consider a convex, monotonically increasing cost function. Such cost functions can be used to model the delay of a link as costs. As the rate, at which data is being injected, is increased the delay also increases. Let

$$f(\boldsymbol{z}) = \sum_{(u,v)\in\mathcal{A}} f_{uv}(z_{uv}) \tag{19}$$

be a new cost function where f_{uv} is convex and monotonically increasing, e.g.

$$f_{uv}(z_{uv}) := \frac{z_{uv}}{c_{uv} - z_{uv}}, \; \forall (u,v) \in \mathcal{A}. \tag{20}$$

Given this kind of cost function, we can remove the capacity constraint posed on the total rate z_{uv} per link since the costs to transmit data at rates near the maximum capacity of a link gets arbitrary expensive. An example is given in Figure 3 for a network with two arcs. Thus, we can restate the general optimization problem given in (7):

$$\begin{aligned}
\text{minimize} \quad & f(\boldsymbol{z}) := \sum_{(u,v)\in\mathcal{A}} f_{uv}(z_{uv}) && && (21)\\
\text{subject to} \quad & x_{uv}^{(t)} \geq 0 && \forall (u,v) \in \mathcal{A},\; \forall t \in \mathcal{T}, \\
& z_{uv} \geq x_{uv}^{(t)} && \forall (u,v) \in \mathcal{A},\; \forall t \in \mathcal{T}, \\
& \sum_{\{v|(u,v)\in\mathcal{A}\}} x_{uv}^{(t)} - \sum_{\{v|(v,u)\in\mathcal{A}\}} x_{vu}^{(t)} = \sigma_u^{(t)} && \forall u \in \mathcal{N},\; \forall t \in \mathcal{T}.
\end{aligned}$$

Given this new formulation we note that the coupling constraint is still active. Hence, the cost function takes it minimum for $z_{uv} := \|x_{uv}^{(t)}\|_{\infty,(t)}$. This allows us to write the coupling constraint

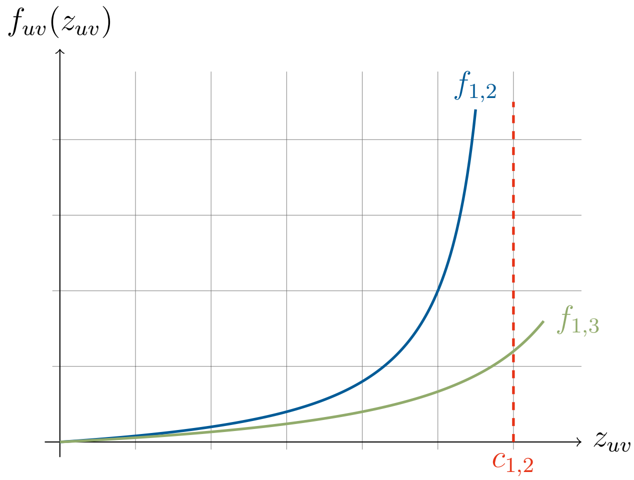


Figure 3: Cost function $f_{uv}(z_{uv})$ as given by (20).

as a number of equality constraints:

$$\begin{aligned}
\text{minimize} \quad & f(\boldsymbol{z}) := \sum_{(u,v)\in\mathcal{A}} f_{uv}(z_{uv}) && (22)\\
\text{subject to} \quad & x_{uv}^{(t)} \geq 0 && \forall (u,v)\in\mathcal{A},\ \forall t\in\mathcal{T},\\
& z_{uv} = \|x_{uv}^{(t)}\|_{\infty,(t)} && \forall (u,v)\in\mathcal{A},\\
& \sum_{\{v|(u,v)\in\mathcal{A}\}} x_{uv}^{(t)} - \sum_{\{v|(v,u)\in\mathcal{A}\}} x_{vu}^{(t)} = \sigma_u^{(t)}, && \forall u\in\mathcal{N},\ \forall t\in\mathcal{T}.
\end{aligned}$$

The lack of differentiability of $\|\cdot\|_\infty$ may pose problems when solving (22), e.g. finding the infimum of the Lagrangian involves differentiation of the objective function. However, we can approximate the maximum norm by the l-norm for l sufficiently large. This gives a differentiable and arbitrary close approximation:

$$\|x_{uv}^{(t)}\|_{\infty,(t)} = \lim_{l\to\infty}\left(\sum_{t\in\mathcal{T}}\left(x_{uv}^{(t)}\right)^l\right)^{1/l} \tag{23}$$

Problem (22) can finally be written as

$$\begin{aligned}
\text{minimize} \quad & f(\boldsymbol{z}') := \sum_{(u,v)\in\mathcal{A}} f_{uv}(\boldsymbol{z}'_{uv}) && (24)\\
\text{subject to} \quad & x_{uv}^{(t)} \geq 0 && \forall (u,v)\in\mathcal{A},\ \forall t\in\mathcal{T},\\
& z'_{uv} = \left(\sum_{t\in\mathcal{T}}\left(x_{uv}^{(t)}\right)^l\right)^{1/l} && \forall (u,v)\in\mathcal{A},\\
& \sum_{\{v|(u,v)\in\mathcal{A}\}} x_{uv}^{(t)} - \sum_{\{v|(v,u)\in\mathcal{A}\}} x_{vu}^{(t)} = \sigma_u^{(t)} && \forall u\in\mathcal{N},\ \forall t\in\mathcal{T}.
\end{aligned}$$

This is now a continuously differentiable and convex multi-commodity flow problem. To state the Lagrangian we dualize both primal feasibility and flow conservation constraints. This results in a total of $|\mathcal{T}| \cdot |\mathcal{N}|$ inequality constraints and another $|\mathcal{T}| \cdot |\mathcal{A}|$ equality constraints. The coupling constraint is considered by inserting the approximation of the infinity norm directly into the cost function:

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = &\sum_{(u,v)\in\mathcal{A}} f_{uv} \left(\sum_{t\in\mathcal{T}} \left(x_{uv}^{(t)} \right)^l \right)^{1/l} - \sum_{(u,v)\in\mathcal{A}} \lambda_{uv}^{(t)} x_{uv}^{(t)} \\ &+ \sum_{i\in\mathcal{N}} \sum_{t\in\mathcal{T}} \mu_i^{(t)} \left(\sum_{\{v|(u,v)\in\mathcal{A}\}} x_{uv}^{(t)} + \sum_{\{v|(v,u)\in\mathcal{A}\}} x_{vu}^{(t)} - \sigma_i^{(t)} \right) \\ = &\sum_{(u,v)\in\mathcal{A}} f_{uv} \left(\sum_{t\in\mathcal{T}} \left(x_{uv}^{(t)} \right)^l \right)^{1/l} + \sum_{(u,v)\in\mathcal{A}} \lambda_{uv}^{(t)} g_{uv}^{(t)} + \sum_{i\in\mathcal{N}} \sum_{t\in\mathcal{T}} \mu_i^{(t)} h_i^{(t)}. \end{aligned} \tag{25}$$

Equation (25) abbreviates the inequality constraints by $g_{uv}^{(t)}$ and the equality constraints by $h_i^{(t)}$. Since the primal problem is convex, the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for optimality for a set of points. If such a set $\boldsymbol{x}'$, $\boldsymbol{\lambda}'$, $\boldsymbol{\mu}'$ satisfy the KKT conditions

$$\begin{aligned} g_{uv}^{(t)} &\leq 0 \quad \forall (u,v) \in \mathcal{A},\ \forall t \in \mathcal{T} && \textit{(primal inequality constraints)} \\ h_i^{(t)} &= 0 \quad \forall i \in \mathcal{N},\ \forall t \in \mathcal{T} && \textit{(primal equality constraints)} \\ \lambda_{uv}^{(t)} &\geq 0 \quad \forall (u,v) \in \mathcal{A},\ \forall t \in \mathcal{T} && \textit{(dual feasibility)} \\ \lambda_{uv}^{(t)} g_{uv}^{(t)} &= 0 \quad \forall (u,v) \in \mathcal{A},\ \forall t \in \mathcal{T} && \textit{(complementary slackness)} \\ \nabla L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= 0 \end{aligned} \tag{26}$$

then $\boldsymbol{x}'$ is primal optimal and $\boldsymbol{\lambda}'$, $\boldsymbol{\mu}'$ are dual optimal ([BoVa08], p. 244).

4.3 Elastic rate demand

So far we have assumed a fixed rate R per flow. We will now modify the convex problem given in (24) to allow for elastic rates. To do so we introduce a utility function

$$U(\boldsymbol{x}, R) := U_r(R) - \sum_{(u,v)\in\mathcal{A}} f_{uv} \left(\sum_{t\in\mathcal{T}} \left(x_{uv}^{(t)} \right)^l \right)^{1/l}. \tag{27}$$

The utility describes the tradeoff between sending at a rate higher than R and the costs needed to transmit at this rate. It is reasonable to assume $U_r(R)$ as concave since the slope of utility in general flattens with increasing rate. Since $-f(\boldsymbol{z}')$ is also concave and concavity is preserved under sums, the utility function $U(\boldsymbol{x}, R)$ is also concave. The new optimization problem consists of maximizing the utility or conversely minimizing the negative utility:

$$\begin{aligned} &\text{maximize} && U(\boldsymbol{x}, R) && && (28)\\ &\text{subject to} && R \geq 0,\ x_{uv}^{(t)} \geq 0 && \forall (u,v) \in \mathcal{A},\ \forall t \in \mathcal{T}, \\ & && \sum_{\{v|(u,v)\in\mathcal{A}\}} x_{uv}^{(t)} - \sum_{\{v|(v,u)\in\mathcal{A}\}} x_{vu}^{(t)} = \sigma_u^{(t)} && \forall u \in \mathcal{N},\ \forall t \in \mathcal{T}. \end{aligned}$$

Again we can state the Lagrangian by dualizing all inequality and equality constraints:

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= U(\boldsymbol{x}, R) - \sum_{(u,v)\in\mathcal{A}} \lambda_{uv}^{(t)} x_{uv}^{(t)} - \lambda_R R \\ &\quad + \sum_{i\in\mathcal{N}} \sum_{t\in\mathcal{T}} \mu_i^{(t)} \left(\sum_{\{v|(u,v)\in\mathcal{A}\}} x_{uv}^{(t)} + \sum_{\{v|(v,u)\in\mathcal{A}\}} x_{vu}^{(t)} - \sigma_i^{(t)} \right) \\ &= U(\boldsymbol{x}, R) + \sum_{(u,v)\in\mathcal{A}} \lambda_{uv}^{(t)} g_{uv}^{(t)} + \lambda_R R - \sum_{i\in\mathcal{N}} \sum_{t\in\mathcal{T}} \mu_i^{(t)} h_i^{(t)}. \end{aligned} \tag{29}$$

The Lagrangian is very similar to the one given in (29) except for the slightly different objective function and the additional constraints on the rate demand. Maximizing the concave utility function is a convex problem again wherefore the KKT conditions are necessary and sufficient for optimality of a solution. Consequently, a set of points $\boldsymbol{x}'$, $\boldsymbol{\lambda}'$, $\boldsymbol{\mu}'$, which satisfies the KKT conditions given in (26) and the additional conditions $\lambda_R \geq 0,\ R \geq 0$, is primal and dual optimal.

5 Conclusion

We saw different approaches to accommodate single source multicast both in routed and coded packet networks and how to describe these problems in a formal way. Depending on the type of objective function we were able to modify the constraints to achieve a problem easier to solve. By dualization of the coupling constraint for example we were able to split the global optimization problem in local subproblems solvable by each node individually. Although the convex problem introduced in Section 4.2 could have been solved in a similar distributed way as the linear problem, we abandoned the distributed approach and showed how to avoid problems of primal recovery by reformulating the problem as a continuous differentiable one. The convexity ensured strong duality wherefore solving the dual problem guarantees a primal optimal solution. Finally we added support for varying rate demands of the sink nodes to the convex problem which turned out to be only a slight addition to the original problem.

References

[LuRa06] Desmond S. Lun, Niranjan Ratnakar, Muriel Medard, Ralf Koetter, David R. Karger, Tracey Ho, Ebad Ahmed and Fang Zhao: *Minimum-Cost Multicast Over Coded Packet Networks*, IEEE Transactions on information theory, vol. 52, no. 6, pp. 2608 – 2623, 2006.

[AhCa00] Rudolf Ahlswede, Ning Cai, Shuo-Yen Robert Li and Raymond W. Yeung: *Network Information Flow*, IEEE Transactions on information theory, vol. 46, no. 4, pp. 1204 – 1216, 2000.

[ShCh96] H.D. Sherali and G. Choi: *Recovery of primal solutions when using subgradient optimization mehtods to solve Lagrangian duals of linear programs*, Oper. Res. Lett., vol. 19, pp. 105 – 113, 1996.

[OuMa00] A. Ouorou, P.Mathey and J.P. Vial: *A survey of algorithms for convex multicommodity flow problems*, Manage. Sci., vol. 46, no. 1, pp. 126 – 147, 2000.

[CoLe01] T.H. Cormen, C.E. Leiserson, R.L. Rivest and C. Stein: *Introduction to algorithms*, MIT Press, Cambridge, Ma, 2001.

[BeHo95] D.P. Bertsekas, M.L. Homer, D.A. Logan, S.D. Patek, S.D.: *Nonlinear programming*, Athena scientific, 1995.

[KoVy05] B.H. Korte, and J. Vygen: *Combinatorial optimization: theory and algorithms*, Springer Verlag, 2005.

[HwRi92] F. Hwang, D.S. Richards and P. Winter: *The Steiner tree problem*, North Holland, 1992.

[BoVa08] Stephen Boyd and Lieven Vandenberghe: *Convex Optimization*, 6th edition, Cambridge University Press, 2008.