Evolutionary Routing-Path Selection in Congested Communication Networks

Mario Köppen, Kaori Yoshida, Masato Tsuru, Yuji Oie
Kyushu Institute of Technology
Network Design and Research Center (NDRC)
680-4 Kawazu, Iizuka, Fukuoka, Japan
mkoeppen@ieee.org

Abstract—This paper proposes an evolutionary approach to the network traffic optimization under the constraint of congestion avoidance. The individuals of the evolving population directly represents a set of paths in a network, and corresponding crossover and mutation operators are provided. The optimization is a global one, i.e. it will not optimize the paths independently but also taking link sharing into account. To avoid the situation that the optimization will result in no traffic for some of the senders (which is also an element of the feasible space in congestion avoidance), we use the user fairness concept. A general approach to user fairness is also provided. The fitness of an individual (path set) is computed from the total traffic in the maxmin fairness state. Experiments on certain graph structures were performed. The results were compared with a path selection strategy based on single path evaluation only. The experiments for networks in the dimension 10 to 80 nodes demonstrate that an increase in performance of around 10% can be achieved in many cases, even with rather small population sizes and number of generations.

Index Terms—networking, network congestion, maxmin fairness, fairness relation, evolutionary algorithm

I. INTRODUCTION

Optimal routing of traffic in communication networks is still an important research issue. In comparison to other optimization problems, there are a number of specific issues. In general, fairness among users has to be taken into account. If modeling a communication network as a graph, each link can be assigned a maximum capacity. This is the maximum amount of (for example data-) traffic that can be routed via this link. Congestion occurs if the amount of incoming traffic exceeds this capacity. To avoid congestion, two approaches were taken into account so far: in the “greedy” congestion control approach, the entire traffic is routed among a path-topology (a set of individual paths) based on some criteria and the traffic on all incoming paths to the congested link is reduced by some rule-of-thumb (or more elaborated concepts[1][2][3][4]. In the global approach, an optimal point of the feasible space of the traffic of all links is selected. The first approach is more easy to realize in practice (it only needs computations at the routers, senders, and/or receivers) than the global approach. The global approach is at much higher computational cost, it needs a centralized and more complex network control, and it scales with the size of the network. However, more cost-efficient networking hardware, and larger computational power at the lower network layers are expected to make the global approach more feasible as well. Moreover, realizations of the global approach may help to justify the higher investment by learning about the expected gains in performance.

In this paper, we want to investigate the opportunities of the global approach. This needs clarification of the meaning of “optimal” selection of a feasible point. Seen as a graph problem, network traffic optimization is the assignment of traffic routes and amounts to each sender. However, using standard combinatorial or real optimization in this context has an important drawback. It cannot avoid that border points of the feasible space may be selected as optimal points. If the problem would be just to decide about cost-efficient network hardware, then there is no big problem if one of the components would not be needed at all. But in a communication network, a border point of the feasible space could mean that some users may not send traffic into the network at all - for the sake of having optimal deployment of network resources. To explain this more clearly, consider the following example.

In the so-called parking-slot scenario [5], a sender 1 is sharing links in a network with two other senders 2 and 3. With sender 2, she has to share link 1, and with sender 3 link 2. Both links have a maximum capacity $m$ for data transmission1. If the incoming traffic for a link exceeds this maximum capacity (congestion), the link distributes the overexcess traffic in a linear manner (the traffic of sender $i$ is denoted by $x_i$): If $x_1 + x_2 > m$, then $x_1$ can only transmit $m \times x_1/(x_1 + x_2)$, and similarly for senders 2 and 3. If we want to maximize the total throughput $T = x_1 + x_2 + x_3$ of this network, we have to consider whether links are getting congested or not. For simplicity, we will assume $x_3 = x_2 = x_3$. Then, if both links are congested, the throughput $T$ will be

$$T = m \frac{x_1}{x_1 + x_2} + 2m \frac{x_2}{x_1 + x_2} = m \left( 1 + \frac{x_2}{x_1 + x_2} \right)$$  \hspace{1cm} (1)$$

and since all $x_i \geq 0$, this expression attains its maximum $2m$ if $x_1 = 0$. However, if at least one of the links is not congested, the total throughput will be surely lower than $2m$, and the global optimum for the total throughput in this network is indeed $2m$ for sending with $x_1 = 0$ and $x_2 = x_3 = m$. This is a rather unfortunate situation, as optimizing the traffic gives

1Note that the traffic could not be constant at any time-scale in reality, and the sending rate at a given time-scale (i.e., averaging in a given unit time) should be considered. Hereinafter we just assume that the sending rate can be treated as constant for simplicity.
raise to a solution where sender 1 cannot transmit any data due to the sharing of network resources with more than one other senders. It cannot be avoided, as long as such a minimum share for sender 1 is part of the feasible space.

As a countermeasure, the concept of fairness has been introduced in the network optimization [6]. Instead of seeking a globally maximum of a certain measure, the allocation of traffic to all senders is compared. The goal is to avoid a situation, where some senders get higher access to resources for the price that other senders, who already share less resources get even lower access to the same resources. A number of formalizations have been provided, and a few algorithms exist to find such an equilibrium state in the feasible space.

In this paper, we want to employ fairness for the global approach to congestion control. Having a path assignment for network routing of all send traffic, the fairness points can be found by a standard algorithm, and the total traffic can be easily found then as well, simply by adding up the single traffics. Thus, the problem of optimizing the network traffic is shifted to the selection of paths through the network itself. Due to high complexity of this problem, we propose an evolutionary computation of this path selection. Using a variant of the Genetic Algorithm, where a path selections is directly represented by the individuals of the evolving population, we can use this metaheuristic to finally get a better path selection, taking path overlap better into account than considerations based on single path measures alone.

This paper is organized as follows. In section II, we have to introduce some basic notations and concepts, which will be needed later on. Then, section III introduces all components of the proposed approach. Section IV will present a number of experimental results and demonstrate the feasibility of the proposed approach.

II. PREREQUISITES

A. Communication networks as graphs

In the following, we are considering communication networks as graphs $G = (N, L)$, where $N$ is a set of nodes $N_i$ ($i = 1, \ldots, n_N$) and $L$ is a set of links $l_j$ ($j = 1, \ldots, n_L$). Alternatively, the notation $l_{kl}$ is used for a link connecting node $N_k$ and node $N_l$. Each link $l_i$ (or $l_{kl}$) has a capacity $c_i$ (or $c_{kl}$) assigned, which gives the maximum traffic that can be routed via this link.

A path $\Pi$ is a sequence of nodes, which are all connected via links. The first node in the sequence will be called sender (or source), and the last node receiver (or sink). Paths will be written as $\Pi = (N_{i_1}, N_{i_2}, \ldots, N_{i_p})$, where $p$ is the length of the path.

In the routing problem, we are considering a set of pairs $(S_{ik}, R_{jk})$ with $k = 1, \ldots, n_T$ of corresponding senders and receivers. Here sender $k$ sends from node $i_k$ to the receiver $j_k$ at node $j_k$.

B. Relations

We will also make use of the general notion of a relation and its maximum set. In set theory, given two sets $A$ and $B$, a relation $R(a, b)$ with $a \in A$ and $b \in B$ is specified as subset of $A \times B$. Thus, a relation is given as set of pairs $(a, b)$, and only the elements appearing in a pair together are seen as standing in relation $R$ to each other.

Each relation $R$ with $A = B$ has a maximum set (possibly empty) of elements from $A$, to which no other element in $A$ is in relation $R$ to:

$$M_R = \{a | \exists x \in A, (x, a) \in R\}$$  (2)

In particular we will consider the Pareto-dominance relation, where $A = R_n$ (n-dimensional Euclid Space). It is said that $x \in R_n$ Pareto-dominates $y \in R_n$ if for all $i = 1, \ldots, n$ $x_i \geq y_i$ and for at least one $j$ $x_j > y_j$ holds. Correspondingly, the Pareto-dominance relation can be defined for any subset of $R_n$, or with regard to smaller values. The Pareto-dominance is a general means for handling multi-objective optimization problems, thereby extending the notion of the maximum of a set of real numbers. The Pareto-dominance relation is transitive, but it is not complete. The maximum set of the Pareto-dominance relation is usually called Pareto front (or Pareto set).

III. COMPONENTS

A. Fairness Theory

As already said in the introduction, the purpose of using fairness is to avoid optimality of border points of the feasible space. In the past, a number of definitions have been provided for subsets of $R_n^+ \times R_n^+$ as feasible space:

1) Maxmin Fairness: here, an element $x$ of the feasible space is maxmin fair if for any other element $y$ and for each component $y_i$ of $y$, which is larger than the corresponding component $x_i$ of $x$ there is a component $x_j$ of $x$ which is (already) smaller than $x_i$ such that $x_j$ is smaller than this component $x_j$. In other words, the “price” for improving at one component of a maxmin fair point is always paid by another component, which has already less.

2) Proportional Fairness: Here, a point $x$ of feasible space is considered proportional fair, if for any other point $y$

$$\sum_{i=1, \ldots, n} \frac{y_i - x_i}{x_i} \leq 0$$  (3)

holds. This definition is not as strict as maxmin fairness, and is rather comparing the average gain of larger components with the average loss of smaller components.

3) Weighted $\alpha$–Fairness: This is a generalization of the proportional fairness, and is using the expression

$$\sum_{i=1, \ldots, n} w_i \frac{y_i - x_i}{x_i^\alpha} \leq 0$$  (4)

instead. For large exponents $\alpha$, this fairness converges to maxmin fairness.

So far, there has been no general concept of a fairness, where all these approaches appear as special cases. In order to gain some insight, we were basing fairness on a relation...
among points of the feasible space, and seeking the maximum set of this relation. In case there is only one point (we will see that this holds for convex feasible spaces) then this is the fair point (or fair state).

The formal way to get a generalized definition for fairness is using an unbounded everywhere non-convex function $P(x, y)$ (in case of two components, only to simplify notation here). For a given point $(x_0, y_0)$, we define a fairness relation in the following way:

1) Consider the surfaces $P(x, y) = P(x_0, y_0)$ and its tangent at the point $(x_0, y_0)$. We call it iso-average surfaces, and $P$ the generating averaging operator.

2) The intersection of the feasible space and the halfplane under the tangent not containing the surface is the subspace dominated by $(x_0, y_0)$ (see fig. 1). It is said that $(x_0, y_0)$ is more fair than any other solution $(x, y)$ within this subspace.

3) For this relation, we search the maximum set, i.e. the set of all points (solutions) in the feasible space, for which no other point of the feasible space is more fair. These are the best solutions in terms of $P$.

For example, if we choose $P(x, y) = xy$, then this procedure leads to the proportional fairness. If we choose other power means (with negative exponent), we get alpha-fairness. In general, if $f$ and $g$ are invertible and monotonic functions (somehow representing utility functions), then the procedure can start from an expression like $f(x)g(y)$, or $1/f(x) + 1/g(y)$, and thus, the concept of fairness can be generalized in a very flexible manner.

Having a generalized fairness, we can see general properties of fairness points directly. For space reasons we will not give the proofs here, but in fact all of them can be seen rather easily.

1) The fairness relation is not transitive and not complete.

2) The maximum set of the fairness relation is a subset of the Pareto set of the feasible space.

3) If the feasible space is convex (i.e. with $x$ and $y$ being elements of the feasible space, also $\alpha x + (1 - \alpha)y$ with $\alpha \in [0, 1]$ is an element of the feasible space) then the maximum set contains only one point.

B. Maxmin Path Selection

For selecting path candidates, we take the robustness principle into account and enforce the random selection of shorter paths. Given nodes $N_A$ and $N_B$, then $\Pi_{\text{min}}(N_A, N_B)$ will be a shortest paths in the network between these two nodes (if there are more than one path of the same minimum length, then one of them will be randomly selected). We select neighbouring nodes to the shortest path from sender to receiver to get a set of path candidates, and from these path candidates we select the one with the largest minimum capacity along the path.

Algorithm MPS: Maxmin Path Selection

1) Get $\Pi_{\text{min}}(S, R)$, the shortest path from sender to receiver.

2) Select $r$ times a relay node $P$, which is a neighbouring node to the shortest path. Compute the shortest path $\Pi_1$ from sender node to relay node, and the shortest path $\Pi_2$ from receiver node to relay node. Join $\Pi_1$ and $\Pi_2$ and remove loops to get path candidate $\Pi$.

3) Assign the minimum capacity among all links of a path candidate to the path, and select the path with the largest minimum capacity (or select one randomly, if more than one).

C. Fairness State of Network Traffic

Despite of having a larger selection of various fairness relations, there are no exact algorithms known so far to find the fairness point. In [7] it has been demonstrated how multi-objective evolutionary algorithms can be used to approximate fairness points. Only in case of maxmin fairness, such an algorithm is known, and we want to make use of this fact for focussing on the path selection problem.

The algorithm to find maxmin fair points in a network traffic problem is based on the idea of bottleneck links. We shortly want to recall the algorithm, known as Water Filling algorithm, here. It is assumed that the network traffic starts with 0 on all
paths and is increased equally, until one or more links become congested. Then, the capacity of this link is divided by the number of paths sharing this link, and this value is assigned as traffic to the senders, which use these paths. These senders are excluded from the following processing, the traffic amount reached so far is subtracted from all capacities for each path, and the procedure iterates, until no more sender remains.

It can be easily seen that by this procedure, an increase in one traffic for a sender would necessarily result in the decrease of an equal amount of traffic of another sender, or exceed the capacity of a link.

Algorithm FSM: Fairness State of multi-path selection

1) Set the remaining set to the set of all paths. Set the traffic for all sender \( s_i \) through their paths \( \Pi_i \) to 0.
2) While the remaining set is not empty, perform the following steps.
3) For all links \( l_i \) used by the remaining paths, get the number \( w_i \) of paths that pass through this link.
4) Find the minimum of \( m_i = c_i/w_i \) where \( c_i \) is the capacity of link \( l_i \).
5) Add \( m_i \) to the traffics for all senders through the links with minimal \( m_i \).
6) Remove all senders through links, for which \( m_i \) is minimal, from remaining paths.
7) Set new capacities of network links \( c_i \leftarrow c_i - m_i * w_i \).

**D. Evolutionary Computation**

As already said before, the goal of the proposed optimization procedure is to increase the network traffic for all senders, while avoiding network congestion. But this goal cannot be reached by maximizing the sum of network traffic, since this can result in no traffic for some senders at all. Therefore we are using the maxmin fairness state as an auxiliary objective, a general concept proposed and applied in [8]. This will definitely reduce the total network throughput that can be achieved, if neglecting the aspect of fair network resource sharing, but ensure that a non-zero (or better: non-least) amount of traffic is allocated to each user. The performance of a set of paths, connecting the senders with the receivers, then is evaluated by the sum of the traffic through the network for all paths.

We are using an evolutionary approach to handle the remaining problem: the selection of paths. The setting is close to a standard genetic algorithm, and in the following the structure of the algorithm will be described.

**Individual representation:** Each individual of the population corresponds to a complete set of paths \( \Pi_k \) \( (k = 1, 2, \ldots, n_T) \) from sender nodes \( S_{in} \) to receiver nodes \( R_{in} \). This representation is different from other evolutionary algorithms (using bit-strings, syntax trees, grammars, vectors etc.) but nevertheless straightforward. Other ways of representing could be taken into account (e.g. a parametric description of the path creation procedure), but we selected this approach due to its direct representation of link sharing (and thus more likely congestion).

**Fitness:** The fitness computation represents above-mentioned paradigm. An individual is a set of paths from senders to receivers. Given this, the maxmin fair state can be found by using the FSM algorithm. The sum of send traffic for all paths is the fitness of the individual path set.

**Initialization of population:** The population is initialized by application of the MPS algorithm, but keeping all path candidates and selecting randomly from them. One extra individual equals to the path selected at the end of the MPS algorithm to ensure that at least the performance of the greedy strategy will be achieved.

**Cross-over operator:** Using the direct representation by paths, also the cross-over can be handled in a straightforward manner. By tournament selection, from two pairs of individuals each the one with higher fitness is selected. A new path set is created by selecting for each sender-receiver pair either from the one or other individual in the ratio of the value of the parameter cross-over ratio.

**Mutation operator:** Mutation is realized similar to the MPS algorithm. For each path in the path set, and with probability \( \mu_{mut} \), a neighbouring node to some node of the path is selected. Then, the sender node is connected by one of the shortest paths with this relay node, and the relay node on one of the shortest paths with the receiver node. Loops are removed. Thus we can favor the selection of still short paths, to grant robustness of the approach.

**Parameter:** In order to make the approach applicable in practice, the dimension of the algorithm should be kept rather small (means smaller population sizes, less number of generations).

**Feasible space:** Despite of the fact that the individuals are representing path sets, the feasible space is defined from the allocation of traffic amounts to the links being traversed by these paths. It has to be ensured that the traffic does not exceed the capacities at the links. This is a convex space, since if two traffic amount vectors (having one component for each link) stay below maximum capacity, any convex combination will also stay below maximum capacity. Thus, there is only one point in the maximum set of the fairness relation, and it can be found by the FSM algorithm.

**Evolutionary algorithm:** After explaining all necessary steps, the evolutionary algorithm follows the processing of any other evolutionary algorithm: at first, a population of \( m \) individuals is created. Then, for a number \( g \) of generations (or if some other stopping condition is meet), the following is repeated. From the population at generation \( g \) (the parent generation) \( m \) children are created by using cross-over, and modified by mutation. From parent and children together, the fitness values are computed, and the \( m \) individuals with highest fitness constitute the population of generation \( (g + 1) \).

**IV. Results**

We have conducted experiments on graphs with increasing dimension \( d \). Especially two graph types have been used: the completely connected graph of \( n_N = d \) nodes, and the Helm graph with a ring of \( d \) nodes, having together \( n_N = 2d + 1 \).
nodes. The capacities have been set as i.i.d. integer random numbers from [1, 100]. For convenience, only even $d$ have been used, and there were $d/2$ sender, sending traffic from half of the ring (the ring spikes in case of Helm graph) to the other half of the nodes (ring spikes). Thus, the selection of multiple paths will necessarily produce sharing of links among different sender-receiver pairs.

The setting of the evolutionary algorithm with path encoding is given in table II. The population size and number of generations is kept rather small.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size of population</td>
<td>10</td>
</tr>
<tr>
<td>Number r of path candidates</td>
<td>10</td>
</tr>
<tr>
<td>Crossover ratio</td>
<td>0.4</td>
</tr>
<tr>
<td>Mutation probability</td>
<td>0.1</td>
</tr>
<tr>
<td>Generations</td>
<td>50</td>
</tr>
</tbody>
</table>

Table II
Parameters of the evolutionary algorithm used in the experiments.

We have compared the path selection by separate MPS algorithm for each path with the evolved path selection. Table III shows the average increase for 30 runs for the evolutionary approach for several values of $d$. In general, an increase in feasible traffic in the fair state of about 10% can be seen. The results for Helm graphs are slightly larger, which can be seen as a result of the larger number of nodes in this network. Also the increase of the result for larger $d$ shows that the larger number of available nodes (and thus alternative paths) is in advantage of the optimization procedure.

For larger $d$ the performance is decreasing, but not very strongly. Here, the limited number of generations (50) comes into account. For larger graphs, the evolution has not converged at this generation.

Table III
Comparison of evolved path selection to selection by MPS algorithm. The increase in total traffic in the maxmin fair state is given in %.

<table>
<thead>
<tr>
<th>dimension d</th>
<th>increase in %</th>
<th>Full Graph</th>
<th>Helm Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>4.77</td>
<td>15.33</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7.97</td>
<td>16.83</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>7.6</td>
<td>18.53</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>8.33</td>
<td>12.1</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>9.13</td>
<td>9.07</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>10.56</td>
<td>11.4</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>12.1</td>
<td>10.577</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>8.33</td>
<td>13.23</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>7.0</td>
<td>8.12</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>1.72</td>
<td>6.73</td>
<td></td>
</tr>
</tbody>
</table>

Table IV
Single path traffics in fairness state for the five paths in the example shown in Fig. 2.

V. CONCLUSION

In this paper, we have investigated an evolutionary concept for traffic optimization in data networks under the constraint of congestion avoidance. The concept is using user fairness to ensure that the traffic will be shared in a more average manner among all senders. Otherwise, it can happen that points in feasible space are selected, where some user will not be allowed to send into the network at all. The evolutionary algorithm uses a population of individuals, where each individual directly represents a set of paths. The fitness of an individual is the total traffic in the maxmin fair state of the network. Thus, the optimization is targeting the path selection, such that the corresponding maxmin fair state has highest total network traffic and also such that it is ensured that no user will not get zero traffic allocated.
The feasibility of the approach could be demonstrated on some graphs with dimension 10 to 80, and, compared to the “greedy” approach of selecting paths individually by selecting the path with maximal minimum capacity among the shortest paths from sender to receiver, an performance increase of about 10% could be achieved. This also demonstrates the potential of global approaches in network control optimization. We will evaluate the proposed approach in more realistic network topologies.

Future work will focus on further optimization by using other fairness concepts. Using maxmin fairness here had the advantage that an exact algorithm is known to compute the maxmin fair state from link capacities alone. For using other fairness relations, heuristic methods will be also needed to approximate the fairness states.

Moreover, the approach presented in this paper can be directly employed also for multi-path routing (simply by adding nodes at the sender source nodes with unlimited capacities, one for each multi-path of a sender). Thus, we may study this approach for multi-path routing, and take into account that a fairness situation may arise in this context as well (allowing some users to send via multiple paths at network fair state may increase or reduce the total traffic for this sender).

ACKNOWLEDGMENT

This work was partly supported by the Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research (S) (No. 18100001)

REFERENCES


