Divide and Share: A New Approach for Optimizing Backup Resource Allocation in LTE Mobile Networks Backhaul

VENMANI Daniel Philip*, Yvon GOURHANT*, Djamal ZEGHLACHE†
*Orange Labs, France Telecom R&D, Lannion, France
† Institut Mines-Télécom, TELECOM SudParis & CNRS, SAMOVAR / UMR 5157, Evry, France

Abstract—In this paper we analyze the link capacity requirements for microwave backhaul architecture when two different Mobile Network Operators (MNOs) decide to “divide and share” their primary resource (working path) with each other as an alternative for investing in a backup path to reduce cost investments. To examine and to develop practicable performance bounds on resource sharing, we make an estimation of the resource utilization and derive integer linear programming (ILP) counterparts. Given the complexities of solving ILP, we propose heuristic-based resource provisioning algorithm which allows MNOs to share their primary resource with (an)other MNO(s), without having to sacrifice their own traffic requirements. Illustrative numerical results show the effectiveness of our resource provisioning approach in terms of network resource utilization and connection blocking probability.

Keywords—4G-LTE; Capacity Provisioning; Infrastructure Sharing; Operating Point Equilibria; Theory and Modeling.

I. INTRODUCTION

Mobile Network Operators (MNOs) are challenged by the rapid bandwidth increase due to the bandwidth-intensive applications and the bursty traffic behaviors. As a result, MNOs decided to adopt backup-path sharing as one of the possible remedies, i.e. to allow backup paths associated with disjoint working paths to share bandwidth [1]-[8]. Our claim is that this backup resources provisioning also comes with an additional cost. This cost is extraordinarily large especially in emerging economies such as the Sub-Saharan African economies. That said, in this work, we analyze the link capacity requirements for microwave backhaul architecture when two different Mobile Network Operators (MNOs) decide to “divide and share” their primary resource (working path) as an alternative for investing in a backup path. Our present work here is a continuation of our previous works [9]-[11] in which we have proposed, designed, and evaluated our design by comparing it with existing approaches analytically. Therefore, in this work, we exclusively focus towards the “optimum configuration choice” for capacity allocation between MNOs that are agreeing to share their primary resource (working path), so that the overall bandwidth reservation for the backup path would be minimal. By doing this, we could minimize the total cost for additional backup resource. Within this context, we propose and evaluate:

- An approach that allows MNOs to optimize backhaul link bandwidth resources to share with one another, deciding the degree of sharing without jeopardizing their own availability requirements.
- A systematic optimistic methodology to efficiently define the capacity bounds (the upper bound and the lower bound) for each connection going through a shared MNOs’ backhaul network in its ability to offer a high quality of experience (QoE) for end-users.

II. DIVIDE AND SHARE OPTIMIZATION

A. ILP based Optimization for Link Bandwidth Allocation through Sharing

Formal Notations: The physical backhaul network topology is represented as a weighted bidirectional graph \( G = (V, E, A, \lambda) \) where \( V \) is the set of nodes, \( E \) represents the set of all microwave links between source \( s \) and destination \( d \), each link with end points \((i, j)\). \( A: E \to (0, 1) \) is called point-wise availability, instantaneous availability, or transient availability between source \( s \) and destination \( d \) (where \( 0, 1 \) denotes the set of positive real numbers between 0 and 1). Thus, we denote the availability of each microwave link \((i, j)\) by \( A_{ij} \). \( \lambda: E \to \mathbb{Z}^+ \) indicates the available bandwidth between source \( s \) and destination \( d \) (where \( \mathbb{Z}^+ \) denotes the set of positive integers) and similarly we denote the bandwidth of microwave link \((i, j)\) by \( \lambda_{ij} \). According to our proposal, since two MNOs share the total link bandwidth \( \lambda_{ij} \), \( \mathbb{P}_{ij} \) represents the share of bandwidth for the donor and \( B_{ij} \) represents the backup bandwidth that can be utilized by the recipient on \( \lambda_{ij} \). Let \( t = (s, d, SLA) \) be an incoming connection request on the link \((i, j)\). Now, the \( k \)-th connection request for the donor is denoted as \( C_{ijk,D} \) and for the recipient is denoted as \( C_{ijk,R} \). The set of all connections \( \{t_1, t_2, ... t_k, ... t_n\} \) on \( \mathbb{P}_{ij} \) is denoted by flow \( S \) and flow \( W \) represents the set of all connections on \( B_{ij} \). This causes \( \mathbb{P}_{ij} \) equal to \( \lambda_{ij} \) when \( B_{ij} \) is not consumed. Furthermore, we define \( b \) as the Blocking Ratio (BR), assuming that not every connection request from the recipient can be always accommodated by the donor (due to various factors such as abrupt spike in traffic load, unstable traffic characteristics on the donor’s backhaul and/or unacceptable QoS request by the recipient). Therefore, the number of connections accepted successfully by the donor for the recipient’s request is:

\[
BR_{Recipient} = b \cdot \sum_{W} \sum_{ijk} C_{ijk,R}
\]

(1)

Since the additional capacity required by the donor to support the recipient traffic contributes significantly to its cost, we classify a hypothetical description to sustain availability...
requirements cost effectively, i.e. given that $A_{ij}$ as the availability of the microwave link $(i,j)$, we associate the cost $C_{ij}$ as a function of $A_{ij}$. This is formally written as,

$$C_{ij} = -\log A_{ij}$$ (2)

We denote $-\log A_{ij}$ as a constant $\alpha_{ij}$. Therefore,

$$C_{ij} = \alpha_{ij}$$ (3)

Variables: $P_{C_{ij}^{D,K,D}}^{C_{ij}^{K,R}}$ and $B_{C_{ij}^{K,R}}^{C_{ij}^{K,R}}$ are the variables, since this optimization problem concerns with allocating the backup path bandwidth on the primary path bandwidth. $P_{C_{ij}^{K,R}}^{C_{ij}^{K,R}} = 1$ if the flow $S = \{t_1, t_2, ..., t_n\}$ successfully traverses through the link $(i,j)$; else $P_{C_{ij}^{K,R}}^{C_{ij}^{K,R}} = 0$. $B_{C_{ij}^{K,R}}^{C_{ij}^{K,R}} = 1$ if the flow $W$ is successfully re-routed through the link $(i,j)$; else $B_{C_{ij}^{K,R}}^{C_{ij}^{K,R}} = 0$.

Objective:

$$\text{Min} \left( \left( \sum_{W} B_{C_{ij}^{K,R}}^{C_{ij}^{K,R}} \cdot \varepsilon \right) + \left( \sum_{S} P_{C_{ij}^{K,R}}^{C_{ij}^{K,R}} \cdot \alpha_{ij} \right) \right), \forall i, j \in E$$ (4)

subject to:

- Flow constraints.
  $$\sum_{j \in E} P_{C_{ij}^{K,R}}^{C_{ij}^{K,R}} - \sum_{i \in E} P_{C_{ij}^{K,R}}^{C_{ij}^{K,R}} = \begin{cases} 1, & i = s \\
-1, & i = d \\
0, & \text{otherwise} \end{cases} \forall S$$ (5)

- Demand constraints.
  $$\sum_{j \in E} P_{C_{ij}^{K,R}}^{C_{ij}^{K,R}} = D_{ij}^{C_{ij}^{K,R}}, \forall i, j \in E$$ (7)

- Availability constraint.
  $$\sum_{j \in E} P_{C_{ij}^{K,R}}^{C_{ij}^{K,R}} \cdot \alpha_{ij} \leq \alpha_{D}, \forall S$$ (9)

- Bandwidth constraints.
  Primary Bandwidth constraints for the donor
  $$\sum_{j \in E} P_{C_{ij}^{K,R}}^{C_{ij}^{K,R}} \leq \lambda_{ij}, \forall S$$ (10)

Backup Bandwidth constraints for the recipient
  $$\sum_{j \in E} B_{C_{ij}^{K,R}}^{C_{ij}^{K,R}} \leq P_{ij}, \forall W$$ (11)

where, $P_{C_{ij}^{K,R}}^{C_{ij}^{K,R}}, \lambda_{ij} \in (0,1)$, such that $ij \in E$

The objective function illustrated above in (4) tries to maximize the availability of the bandwidth reserved for the flows of the donor. At the same time, it minimizes the bandwidth resource utilization for the flows of the recipient, where $\varepsilon$ is a small value that we have assigned arbitrarily, so that satisfying the “required traffic demand” of the donor is of higher priority, MNOs decide and define what their “required traffic demand” is. It could be satisfying a set of flows consisting of high revenue generating premium customers and/or delay sensitive voice call users etc. Equation (5) gives the flow balance for the primary path for the set of flows $S$ and (6) gives flow balance for the backup path for the set of flows $W$. In the rest of paper, we will refer to (5) as the primary flow constraint and (6) as the backup flow constraint. By (7), we ensure that there is enough bandwidth resource for the donor to accommodate its own required traffic demand ($D_{ij}^{C_{ij}^{K,R}}$) for the set of connections $S$ before it allows the recipient to share. Only when (7) is fulfilled, the recipient is permitted to take hold of its part of the bandwidth resource, if and only if the demand of the recipient is no greater than the donor. This is ensured by (8). Furthermore, in (9) we impose that the availability of a connection on the primary path should not be downgraded than the minimum required availability to satisfy the “required traffic demand” of the donor. Now, considering the problem of dividing and sharing the bandwidth resource between the MNOs as a means for backup, in typical real world situations, the bandwidth that is reserved for the connections of the recipient as the backup path are not utilized as much as the bandwidth reserved for the donor as the primary path. Therefore, the above defined optimization problem in (4) is relaxed by decoupling the (backup) bandwidth reserved for the recipient ($B_{ij}^{C_{ij}^{K,R}}$) on the link $(i,j)$ from the (working) bandwidth reserved for the donor ($P_{ij}^{C_{ij}^{K,R}}$) on the same link $(i,j)$ as in (10) and (11). In (10), we assume that initially all of the bandwidth on link $(i,j)$ (i.e. $P_{ij}^{C_{ij}^{K,R}}, \lambda_{ij}$) is used for allocating the (working) bandwidth for the donor. A general static connection-provisioning problem can also consider optimizing the bandwidth on the primary path to avoid over-utilizing or congesting links. (11) ensures that the (backup) bandwidth which is reserved for the recipient does not exceed the (working) bandwidth reserved for the donor. To obtain a linear program, we relax the last constraint to $P_{C_{ij}^{K,R}}^{C_{ij}^{K,R}}, B_{C_{ij}^{K,R}}^{C_{ij}^{K,R}} \in (0,1)$ in (11). Therefore, the optimal solution to a LP relaxation of an ILP problem gives us a bound on the optimal objective function value. For maximization problems, the optimal relaxed objective function value is an upper bound on the optimal objective value. For minimization problems, the optimal relaxed objective function value is a lower bound on the optimal objective value.

III. HEURISTICS BASED APPROACH FOR LINK BANDWIDTH ALLOCATION THROUGH SHARING

A. Algorithm for Link Bandwidth Allocation through Sharing

The above ILP formulations of section II can be solved only for smaller node sizes. The complexity of an ILP formulation is the product of the number of equations by the number of variables, such as our case too. Due to the
complexity that ILP solvers face while solving even mid-sized networks with a large number of binary variables, in what follows next, we resort to heuristics. Here, we propose an algorithm that we term as Divide and Share by Infrastructure Sharing (DASIS) Algorithm. The specificity of this algorithm is to allow (backhaul) link bandwidth sharing among MNOs excluding the need for an additional backup path, in addition to be able to satisfy the “required traffic demand” for the donor. The idea of our algorithm is that the bandwidth resource of the donor is re-evaluated constantly, so that the existing traffic demand is not jeopardized. During this re-evaluation, the donor’s network resources are constantly computed for the (backup) bandwidth reservation to determine if it is efficient enough to maximize the network resource utilization subject to: Availability constraint, SLA constraint and Bandwidth Constraint. Algorithm 1 describes this. To interpret the algorithm below easily, we assume that the working path of MNO A (W1) becomes unavailable and the traffic is re-routed through working path of MNO A (W1).

Algorithm 1 Divide and Share by Infrastructure Sharing (DASIS) Algorithm

Input: Network topology \( G = (V, E, A) \); Connection request \( C_{i,j,k,R} \)

Output: New backup-path \( W \) with a backup bandwidth \( B_{i,j} \) is established satisfying the connection \( C_{i,j,k,R} \)’s constraints; or refuse \( C_{i,j,k,R} \) if no such path(s) is found.

Step 1) Compute the minimum required bandwidth \( P_{i,j} \) on \( \lambda_{i,j} \) from \( s \) to \( d \) on \( W1 \) for the donor.

Step 2) Set the initial backup bandwidth \( B_{i,j} \) on \( \lambda_{i,j} \) for the recipient as null.

Step 3) Compute the backup bandwidth \( B_{i,j} \) that has to be reserved on \( W1 \) from \( s \) to \( d \) on \( W1 \) based on the following three constraints:

Step 3.a) Availability constraints: Compute the availability of \( W1 \) and check if the \( k \)-th connection request \( C_{i,j,k,R} \) is satisfied according to \( 0 < B_{C_{i,j,k,R}} \leq B_{i,j} \) such that the set of flows \( W \) utilizing \( B_{i,j} \) of the recipient are re-routed successfully and go to Step 3.b. Refuse connection \( C_{i,j,k,R} \) if availability constraints are not satisfied and/or path \( W1 \) is not found.

Step 3.b) SLA constraints: If \( B_{i,j} \leq SLA \) agreed between the donor and the recipient, set \( W1 \) as the working path for the recipient and go to Step 3.c. Refuse connection \( C_{i,j,k,R} \) if SLA constraints are not satisfied and/or path \( W1 \) is not found.

Step 3.c) Bandwidth constraints: Now check whether the bandwidth availability requirement for the donor is also still satisfied, i.e. \( B_{i,j} \leq B_{C_{i,j,k,R}} \leq P_{i,j} \) and go to Step 4. Refuse connection \( C_{i,j,k,R} \) if bandwidth constraints are not satisfied and/or path \( W1 \) is not found.

Step 4) Connection \( C_{i,j,k,R} \) is accepted and a new backup-path \( W \) with a backup bandwidth \( B_{i,j} \) is established for the recipient on the donor’s backhaul.

Step 5) Update network resource information accordingly so that the total available bandwidth is divided and shared appropriately.

B. Description of the Algorithm

From the previous discussions in section II, it is made clear that the requested bandwidth for the backup path should be allocated along the primary path. Thus for any link with end points \((i,j)\), the amount of bandwidth reserved for the donor is the sum of the requested bandwidth of individual connections in flow \( S \). This is denoted as:

\[ P_{i,j} = \sum_{S \in F_{i,j}} B_{C_{i,j,k,R}} \]  (13)

where, \( B_{C_{i,j,k,R}} \) denotes requested bandwidth of an individual connection \( C_{i,j,k,R} \) of the donor. Thus, we first determine the minimal bandwidth for the donor which can satisfy the “required traffic demand”. To do this, we first set an initial high value, and start to decrease this value one by one until all the connections can be set up in the optimization. The process is repeated until some connections cannot be set up. This value is fixed as the value in previous loop. Here, no connections are blocked since all the connections can be carried satisfactorily. In particular, across any link \((i,j)\), we allocate the minimum required bandwidth for the set of flows \( S \) on the primary path through which the minimum demand \( B_{i,j} \) is satisfied. We define this as the Minimum Required Primary path Bandwidth (MRPB). Formally it is written as,

\[ P_{MRPB} = \min_{i,j \in E} \sum_{S \in F_{i,j}} B_{C_{i,j,k,R}} \]  (14)

At the same time, we enforce a lower bound on bandwidth allocation for backup path since allocating more bandwidth for the backup may interrupt the bandwidth reserved for the primary path. For this, we consider the baseline approach in which no backup path bandwidth is allocated initially. We define this as the Zero Backup path Bandwidth (ZBB).

\[ B_{i,j}^{ZBB} = 0 \]  (15)

So the goal of the backup path bandwidth allocation algorithm is to determine \( B_{i,j} \), the amount of bandwidth that needs to be reserved on link \((i,j)\) for the backup path across this link. Similar to primary path bandwidth allocation, a naive approach for backup path bandwidth allocation is to reserve the requested bandwidth of each flow along the backup path. We define this as the Minimum Required Backup path Bandwidth (MRBB) across link \((i,j)\) which is the sum of the requested bandwidth of those flows whose backup paths use this link. Formally it is given as,

\[ B_{i,j}^{MRBB} = \sum_{S \in F_{i,j}} B_{C_{i,j,k,R}} \]  (16)

Finally, we enforce the maximum bound for the bandwidth allocation for the backup path on this link. This bound is necessary so that the bandwidth reserved for the backup neither affects the availability requirements nor ignores the SLA constraints nor affects the bandwidth for the “required traffic demand” of the donor. We define this as the Maximum Backup path Bandwidth Allocation (MBBA).

\[ B_{i,j}^{MBBA} = \max_{i,j \in E} \sum_{S \in F_{i,j}} B_{C_{i,j,k,R}} \]  (17)

Therefore, using Algorithm 1, the degree of sharing is controlled stringently such that network resources are shared and utilized more intelligently and efficiently, nevertheless to say that the flow connections of the donor are sustained to be able to meet the demand requirements. In practice, we do not execute these three constraints step by step but mix them together. Thus, the adaptation of our approach enables to cautiously control the total capacity allocated to the backup paths without affecting the bandwidth requirements for the
The backup capacity optimized indirectly depends upon the availability of the primary path, which directly depends upon the cost $C_{ij}$. This encourages MNOs to divide and share their resources to reduce the total cost of setting up an entire backhaul network.

IV. ILLUSTRATIVE NUMERICAL EVALUATION

A. Performance Comparison between ILP and Heuristics.

1) Bandwidth Utilization versus Node Size: Network resource utilization efficiency is defined as the ratio of the bandwidth reserved for the backup connections to the bandwidth reserved for primary connections [15]. It is a measure that exhibits the additional resource overhead (RO) utilized for backup. Better backup-sharing optimization is achieved when the resource overhead is lower. Thereby, the objective here to estimate the usage of the backup path bandwidth of the recipient on the primary path of the donor.

![Figure 1. Resource Overhead versus node size.](image)

Fig. 1 illustrates the RO for the ILP-based and the heuristics-based approaches for different node sizes. From the results, it can be noticed that, as the node size increases, the resource overhead value decreases because of our approach. What this means to us, is that, by our approach, we could possibly achieve less resource overhead, meaning that, the occupancy of the backup bandwidth on the primary bandwidth gradually decreases for large network sizes, which gives more freedom for the donor to provision more resources for their own demands. Furthermore, it can also be seen that the RO due the ILP-based approach has higher values than those of the heuristics-based approach due to computational complexity.

2) Blocking Probability: The blocking probability is a measure of the number of connection requests rejected against the total number of connection requests. The main measure here is to compare the performance of the ILP-based approach and the heuristic-based approach. The number of connections provisioned is calculated by simulating at least 10,000 connection requests, under dynamic traffic. Fig. 2 demonstrates the number of connections that are provisioned with ILP and heuristic based approaches for different node sizes. We find that a connection is less likely to be blocked by in ILP than in heuristics. Our justification for this is that in ILP routing is not limited by candidate routes, since our heuristics are based on shortest path routing. Also, from Fig. 6, we could observe that ILP has overall lower blocking probability compared to heuristics-based approach.

![Figure 2. Blocking probability versus node size.](image)

3) Cost versus Node Size: As we observe from the results, as the node size increases, the cost for provisioning additional backup resource decreases for the heuristics as well the ILP. It can also be seen that the performance of heuristics is almost the same as compared to the ILP for smaller network sizes (for $N=5$, $N=6$ etc) while the performance improves and gets better than those of ILP for larger networks ($N=14$, $N=22$ etc). What is illustrous here, this efficiency that is achieved is without any additional cost at all for backup resource.

![Figure 3. Cost versus node size.](image)

4) Overall Performance: Overall, we observe that the performance of resource sharing by the heuristic and the ILP approaches improves as the size of the network increases. From our experience, we reason out that ILP solvers have a difficult time solving our model which has binary variables in abundance. On a conclusive basis, we could say that the ILP-based approach necessitates more resources (link bandwidth) to produce similar performances as that of the heuristics.

V. CONCLUDING DISCUSSIONS

As discussed in this paper, we have presented a novel resource sharing framework which can cost-effectively provide protection services by guaranteeing the service demands without jeopardizing minimum required demands for the MNOs. True, there is a trade-off between resource utilization, availability and the cost here. Nevertheless, since the backup capacity of another MNO is not used under normal no-failure conditions (except by low priority pre-emptible traffic), the objective of minimizing restoration capacity overhead in the network translates to higher network utilization. Moreover, our approach here is to define which type of traffic needs to be protected all the time, and which can have a lower level of protection, thus allowing the MNOs to protect and improve the availability of their revenue generating services to ensure high-quality, uninterrupted user experience, and increase link capacity to offer more data services.
REFERENCES