Automatic Node Insertion in a Multi-Technology Transmission Network

Luis Fontes Amorim da Cunha
Instituto Superior Técnico
Universidade de Lisboa
Email: luisfontesamorimcunha@gmail.com

Abstract—Telecom operators face serious challenges with their transmission network configuration and operation to ensure good Quality of Service (QoS). As a result, mobile operators are concentrating in increasing efforts to create tools and procedures that are intended not only to assist radio engineers in the process of maintaining and optimizing networks mobile, but also to turn the appropriate network more autonomous. It will be unavoidable to position and build investment in new infrastructures for the uncovered areas, or to provide greater capacity in places with high traffic growth or where emerging new radio technologies are initially deployed.

This thesis aims to develop an algorithm that optimizes the planning of a base station for a real telecom operator, taking into account the structure of the existing telecommunication network (locations, connections and traffic flow), maximizing QoS and minimizing Operational Expenditure (OPEX) and Capital Expenditure (CAPEX). This algorithm is intended to be incorporated into a network planning and monitoring platform in order to provide an automation process for introducing a base station. This process is intended to facilitate the operator’s choice when parameterizing technology, available bandwidth, geographic locations of transmission nodes and the cost associated with each transmission technology.

To create the structure of the network, Python library called NetworkX was used, which allows the creation, manipulation and study of the dynamics of a structure and functions of complex networks. This library enable a computational method to calculate processing priorities based on a strict priority scheduling, as well as the application of the Dijkstra package, which is found in this library. The function, imposed in Dijkstra, meets the objective of the thesis, which is to find the solution that presents the best possible QoS, and for this, the function takes into consideration the latency (delay) in each path.

I. INTRODUCTION

As the complexity of mobile networks grow due to the increasing performance requirements, it becomes harder for the mobile network operators not only to maintain, but also to optimize the performance of those networks. Telecom operators face serious problems related to the configuration and operation of their transmission networks, to guarantee the desired QoS. As a result, mobile network operators are targeting the creation of tools and procedures that aim, not only to assist radio engineers in the process of maintaining and optimizing the mobile networks, but also making the network itself more autonomous.

To meet these requirements, it will be inevitable to position and build new base station (BS) in areas without coverage, or to provide larger capacities to areas that need to face the traffic growth, or to deploy emergent new radio technologies.

This thesis aims to develop an algorithm that optimizes or plans a BS location for a real operator, taking into account an existing telecommunications network structure (locations, connections and traffic), maximizing QoS and minimizing OPEX and CAPEX. This tool is intended to be incorporated into a network planning and monitoring platform, in order to provide an automated process for new BS location, facilitating the operators work.

In order to achieve these results, it is essential to have a platform that contains this information. Hence, a database was used with real Mobile Network Operator (MNO) information.

This work is organized as follows. Section II presents the applied data pre-processing steps, how the platform contains the network information and explains its structure. Section III introduces the proposed algorithms, as well as the necessary background knowledge to develop the concerning methodology. Section IV includes the obtained results through the application of the proposed algorithm and the necessary uses cases that demonstrate the different software outputs. Section V a summary of the work carried out in this thesis is presented and some conclusions are drawn.

II. PLATFORM DEVELOPMENT

A. Network Structure

In order to enable the construction of a complex telecommunication structure, Python programming language was used, since offers various libraries. These libraries are algorithms collection to simplify the process of programming and removing the need of reprogramming the most used. The most important library to highlight, is the NetworkX, [I]. This is a package for the creation, manipulation, and study of the structure (graphs), dynamics, and functions of complex networks. In the graph, the vertex or node represents a location, that can be a data center, router or base station, and an edge represents a connection between two locations (that can be physical or logical), that can be linked by two different technologies: optical fiber or microwave.

Both elements of the graph have characterizing. These attributes help to redirect the network traffic data to the high level nodes. A real database was used that contains all the necessary information to characterize network elements.
It’s important to mention that each location can have more than one router, meaning more capacity in that location. Once again, that data was appealed on database. JavaScript, HTML language and the Openstreetmap API was used for a building a representation of the network. Thus, the notion of geographical location is clearer to interpret and explicit, as can be inspected on Figure 1.

![Figure 1: Overview of the network.](image)

B. Network Hierarchy

The network structure considers a hierarchy split according to its role and potentialities. So, it is organized in three different types of layers or levels.

- Transport or Core;
- Aggregation;
- Access.

Its differences are in the implemented equipment. The higher the level of the hierarchy the better the performance, intelligence and capacity of the equipment. Each router is associated with a maximum capacity (bits per second), a number of ports and a respective capacity for each port.

1) Location Attributes: In addition to increase demands for lower latency, data centers share a unique challenge in that the majority of data center traffic stays within the data center as workload processing is distributed across multiple compute nodes.

This graph component is associated with the following information:

- Location name;
- Location coordinates - longitude and latitude;
- Network type - Transport, Aggregation or Access;
- Bit rate Available;
- Load - percentage of use;
- Free ports;
- Priority queue;
- Environment rating - dense urban, urban or rural;
- Status - in service, turned off or in planning.

These attributes help to redirect the traffic data through the network.

2) Connection Attributes: Similarly to the nodes, the edges diverge depending on the technology used or on the medium in which traffic flows (fiber optics and microwaves), and also on their type of connection, which can be physical (Associated with one type of technology and can support the logic) or logical (Multiplexing virtual circuits supported by physicals and depend on them. If the physicals, for some reason, are turned off, the logic will also be turned off).

This graph component is related to the following data:

- Source and target - where and where are you going;
- Network type - Transport, Aggregation or Access;
- Technology - Optical fiber or microwave;
- Connection Identification;
- Bandwidth Available;
- Load - percentage of use;
- Support - which channels or services support;
- Priority queue;
- Status - in service, turned off or in planning.

It is important to label the edge with a connection identification because between two nodes there may be more than one link. It is also critical to know which technology is used because each one is associated with different transmission media and bit rates, causing different latencies. These attributes also help to redirect the network traffic data to the high level nodes.

C. Priority queue

Priority queuing is crucial to calculate system latencies. Different types of traffic need different QoS standards. In order to guarantee acceptable delay boundaries to delay-sensitive traffic (such as voice/video), several scheduling schemes – for switches and routers – have been proposed and analyzed, each with their own specific algorithmic and computational complexity. The algorithm that sets the worst case scenario is the strict priority scheduling. With this scheduling, as long as delay-sensitive (or high-priority) packets are present in the queuing system, high priority traffic is served. Delay-insensitive packets can thus only be transmitted when no delay-sensitive traffic is present in the system. As already mentioned, this is the most drastic way to meet the QoS constraints of delay-sensitive traffic, but also the easiest to implement.

A higher priority is typically assigned to a more important task. This is why the traffic in higher priority queues is always scheduled ahead of the traffic in the lower priority queues. This algorithm may carry bandwidth starvation for lower priority traffic, when there is an excessive amount of high-priority traffic. Such regulation might be that no traffic should be dropped no matter how congested the network might be.

D. Environment Rating

In order to classify a node according to its environment (dense urban, urban, suburban or rural), the Python library was again used. A package called Geocoding which receives geographic coordinates as input, returns the parish, district and the country corresponding to those coordinates. Then the
population density per parish was looked up \((3)\) and subsequently the environment classified according to the Portuguese communications regulatory authority, Autoridade Nacional de Comunicações (ANACOM) was applied.

The overcomed work described on this Subsection was very important due to the lack of Keys Performance Indicator (KPI) data that would be discussing on the next Subsection.

E. Traffic Modeling

This Section has four phases:

- Load the existing KPI;
- Classify BS without KPI data according to its environment;
- Load the dispersion nodes (node through which you can go more than one way);
- Generate traffic to higher nodes and edges.

KPI are referent to the different parameters values which evaluate network performance, used by the operators as a measure to know how good or bad his services are. The network has three major components (Core, Aggregation and Access), as already mentioned, and each of these components have their own KPI which affect the performance of the network as a whole.

The database contained several KPIs, total volume and throughput from uplink and downlink, from different dates and hours and different technology (2G and 4G) for each location. Analyzing the existing KPIs, was possible to identify the busy hour and baseline hour, where the traffic volume is higher and its opposite, respectively. One of the biggest issues was the lack of information from the KPI to the hierarchy’s highest nodes (Aggregation and Core), links and some BS (access nodes, lowest hierarchical nodes), which would add a major obstacle in the algorithm development. Only had KPI for some access nodes. In order to overcome this, it was necessary to associate each available KPI to the corresponding node and consequent to the environmental classification, mentioned above. Therefore, BS (Access nodes) that do not had these data were assigned with KPI from similar nodes according to their environmental classification. After associating KPI with access nodes, KPI were broken down into services, considering Ericsson mobility report, \((4)\), that measure the traffic impact of different application categories.

<table>
<thead>
<tr>
<th>Services</th>
<th>2G</th>
<th>4G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice</td>
<td>60</td>
<td>10</td>
</tr>
<tr>
<td>HTTP</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>Download</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Video-Streaming</td>
<td>--</td>
<td>60</td>
</tr>
</tbody>
</table>

Table I: KPI Division, \((4)\)

Afterwards, phase 3, KPI for dispersion nodes were computed, exemplified in more detail in Figure 2.

Analyzing the Figure 2, the KPI of dispersion nodes are the sum of nodes KPI of lower/equal hierarchy. It can be conclude that node 4, 6 and 7 will be the dispersion nodes. The Equation 1 shows the equations to calculate the dispersion nodes, taking Figure 2 into account.

\[
\begin{align*}
\text{node1} + \text{node2} &= \text{node3} \\
\text{b} + \text{c} &= \text{e} \\
\text{a} + \text{d} &= \text{f} \\
\text{node3} &= \text{node4} \\
\text{node5} &= \text{node6}
\end{align*}
\]

On the last phase, for flooding this data to other highest nodes and links, data transmission between these and dispersion nodes was performed randomly. Allowing the KPI data to be fairly and evenly distributed across the network. The KPI distribution was benefited from the algorithm known as Dijkstra, already implemented in NetworkX package, being used the “bidirectional” version for performance reasons, \((5)\), between dispersion nodes.

For these four phases was used the busy hour and base line hour traffic to flood KPIs data to all edges and nodes. Being possible to analyze and visualize the network with different loads at different times of day, as shown in the following Figures 3 and 4 respectively.

As can be seen in Figure 3, representing traffic at busy hour, it has a higher load than Figure 4. Can conclude that, there are more reddish and orange connections on first Figure than the second. That is, the links and nodes are more loaded with services.

F. Morphology

For a node insertion tool, it is important to access the terrain morphology and optical fiber ducts. This data is crucial for inserting a new node since it is necessary to know if the node
III. BASE STATION INSERTION ALGORITHM

As mentioned in the Section I, the algorithm aims to automate the insertion of a BS in a telecommunications network. This computing demands several parameters to be checked. It is necessary to check if there are fiber optic conducts and/or any LoS that connects this new node to the network. Moreover, it is essential to verify whether there are some network nodes that can hold with incoming traffic, generated by the new base station.

The Base Station Insertion algorithm brings some optimization in location planning, which informs which location supports the new and costs associated with each technology (optical fiber or microwave).

Therefore, the need to enter a new Base Station location may be due to:

- Zone with lack of coverage;
- Zone that needs increased capacity because it is congested;
- Zone that needs capacity to improve QoS;
- Implementation of new locations due to new application technology.

Next, the flowchart of the proposed algorithm for problem optimization will be presented.

A. Fiber Ducts Search algorithm

The algorithm looks for manholes in the ducts with a maximum distance of 100 meters from the coordinates that were given as input of the new location, considering that the costs provided by the Celfinet Operations department for the opening of the road are limited to that distance. If it finds one, the algorithm will check if this conduit connects with other(s) network location(s), ordering by fiber optic length. If does not found will report an alarm, and do not present a solution for this transmission technology.

In the possibility to connects with other network location, it is necessary to validate the node on which it supports the new one, i.e., it is essential to verify that traffic flows to the Core, satisfying certain requirements. For this it was suggested another algorithm exposed in Subsection III-C Validation algorithm. Subsequent to validation, if it verifies, the costs associated with the fiber optic connection, will be reported.

B. Line of Sight Search algorithm

This algorithm generates a list in ascending order of geographical distance from all locations that have LoS with a maximum radius of 25 km referring to the coordinates that...
were given as input to the new location. This maximum distance was again provided by Celfinet Operations department.

To have LoS it is imposed that the Fresnel ellipsoid is not obstructed by more than 40%.\(^6\) The frequency used for Fresnel ellipsoid calculation is in the C band, the most common frequency range for telecommunication communication.

The calculation for LoS takes into account the curvature of the Earth to block long distance transmissions. The algorithm considers the Spherical Earth model, making the necessary adjustments to the heights of the transmitter and receiver antennas.

If there is no LoS for a network node, the algorithm reports the situation and do not present a solution for this transmission technology. In the circumstances of existing it is essential to validate the node containing LoS, if the expected traffic associated to the new location reaches the Core fulfilling its requirements. To verify these, an algorithm exposed in the Subsection III-C is suggested, Validation algorithm. Then if validation takes place, the costs associated with the microwave connection will be reported.

C. Validation algorithm

As discussed earlier in Subsection III-A and III-B the goal is to provide the cheapest solution for each technology, so the existent node that connects to the new may differ between each technology. However, validation is performed the same way.

For the already existent and connecting node it is essential to verify some check-list, namely:

- Verify that the support node has port for the new connection;
- Validate that the port capacity supports the new connection, that is, if the port capacity is greater than the whole throughput of the new node (\(i.e., \) sum of all services bit rates);
- Find a path to the Core that supports all traffic associated with the new connection (Core Routing);
- Confirm that the Core node can flow the services to another Core node;
- Inspect if the final latencies of each service are within the standards.

The suggested check-list is done through the Core, because it is the top of the network hierarchy and because it is in a ring topology. This routing will be explained below.

Core Routing, uses an algorithm already mentioned in Section II-B applied for the purpose of generating network traffic. It is explored here for the purpose of routing all services that the new node contains in priority order. This routing takes into account this order, so that services with the highest priority, that is, the most sensitive to latency or with lower default values, are processed first. The Core nodes are not chosen randomly, but by jump order, \(i.e.,\) the amount of links traveled to reach this hierarchy. The path chosen for each service uses the Dijkstra algorithm that takes into account the following cost function, see Equation 2

\[
\text{cost} = \frac{1}{Bw_{\text{Available}}} * w_{Bw} + \text{Latency} * w_{La} + \text{Hierarchy}. \tag{2}
\]

The bandwidth and latency weights \((w_{Bw} \text{ and } w_{La})\), in Equation 2 were preset to 10% and 90%, respectively, aiming the thesis objective to give more importance to the path with the lowest possible latency, to maximize the QoS. As such, bandwidth will serve as a tiebreaker, that is, if there are two paths with the same latency, the algorithm will choose the one with the most available bandwidth. If one connections node state is non operating or still planned, it will have high cost (infinity), so that it won’t be considered. If connections have unavailable bandwidth (available bandwidth equal to zero), they will be handled equally. Since Core Routing is an upward hierarchical path, if the path has downward links it gives an additional cost and is less likely to be chosen.

By analyzing the cost function from Equation 2 it is possible to verify that, the higher the latency or the lower the available bandwidth, the higher the cost. This means a lower probability of being selected.

After routing is performed from the new node to the nearest Core node with a final check is performed whether final latency is within the service’s maximum delay value, see Table II to come across a good QoS.

Table II: Maximum delay per service.

<table>
<thead>
<tr>
<th>Service</th>
<th>Maximum delay [ms]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice</td>
<td>100</td>
</tr>
<tr>
<td>Streaming</td>
<td>150</td>
</tr>
<tr>
<td>HTTP</td>
<td>300</td>
</tr>
<tr>
<td>Download</td>
<td>300</td>
</tr>
</tbody>
</table>

On the next Subsection will be presented how the delays were computed, III-D

D. Delays computation

As mentioned in the Subsection II-C QoS is an important measure and it is related to the latency of each service. A service transmission may be affected by the intervention of two graph components (nodes and edges) as a function of its priority. Also states that there are several complex computational algorithms taking priority into account, referencing the strict priority scheduling.

1) Node delay computation: Node inherent latency is related to processing the volume of each service and the wait time to be served. As mentioned already, the node obeys a priority order per door, that is, it will first serve the highest service and so on. If there are two services of equal importance, the order will take into account First-In, First-Out (FIFO). When answering a service, the node dispose all its full capacity per door to process it and so on to the next. When the node serves the next service, latency is affected by the previous one.
2) Edge delay computation: The implicit latency to the edge is related to the volume processing of each service, the waiting time to be answered and its transmission. Like the node, and as stated in Subsection II-C, the edge also has a priority list. But its processing is distinct. Unlike the node, the edge instead of using its full capacity uses the throughput associated with each service taking into account the available edge bandwidth. In the event that the available bandwidth is greater than the service’s throughput, the service use the requested throughput, serving the next service by comparing the next service’s throughput and the remaining bandwidth. If the available bandwidth is lower, only available bandwidth is used until the first service being processed goes out. If the packet is not fully processed and already has available bandwidth for the requested throughput, use it. Therefore, more than one service can be processed at the same time. There is one more edge-related delay, which is service propagation. This depends on the type of technology associated with the edge, i.e., whether the transmission medium is fiber optic or microwave, that is as an attribute of the edge as already referenced in Subsection II-B2. Transmission media are associated with different transmission speeds, translating to different latencies.

E. CAPEX and OPEX

It was provided by the Celinet Operations department, the values associated with each operation and for each transmission technology. This translate to the costs associated with each planning being as realistic as possible.

The costs related to the optical fiber solution are associated:
- The required fiber optic length;
- Performing single mode fiber optic fusion on cable already terminated on join;
- Mobilization of a team;
- Fiber optic cable installation;
- Excavation and construction of ditches;
- Construction of manholes.

The associated cost, analogous with microwave solutions, depends on the throughput of all traffic flowing over the new link.

IV. USE CASES

In this Section, the result are organized into use cases. The main purpose is to show different outcomes of the proposed algorithm presented earlier. For that, in each use case is explained a brief context situation regarding the need for insertion and planning of a new BS.

It should be noted that the results consider the busy hour, since it is the time when the network has the highest data traffic, which translates to the worst quality of service and congested connection, justifying the planning of a new BS and therefore the usefulness of the proposed algorithm.

A. High Load Area

The area shown in Figure 6 during the busy hour is a high capacity (loaded) area. It can be observed that the transmissions links are mostly orange, which is expectable in an urban area.

Adding one more BS, in the zone identified in Figure 6 with a purple circle, will increase capacity in that area, which will relieve existing connections. To achieve this, the proposed algorithm was used after being provided the necessary inputs. Then, the algorithm uses the database containing the received KPIs from the existing BSs in that area, and uses them for the new BS. Those found, the volume and throughput for each service are presented in Table III.

Table III: Services that will be associated with the new BS.

<table>
<thead>
<tr>
<th>Services</th>
<th>Volume [bits]</th>
<th>Throughput [kbits/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice</td>
<td>606.0</td>
<td>37.62</td>
</tr>
<tr>
<td>HTTP</td>
<td>454.5</td>
<td>18.51</td>
</tr>
<tr>
<td>Download</td>
<td>202.0</td>
<td>6.17</td>
</tr>
</tbody>
</table>

The algorithm presents two possible solutions: the optical fiber solution and then the microwave solution.

As shown in Table IV, the latencies of each service, from the path taken to the Core in each solution, meet the default values for each service in the Subsection III-C Table II.

Table IV: Latencies associated with both transmission solution

<table>
<thead>
<tr>
<th>Services</th>
<th>Optical Fiber</th>
<th>Microwave</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice</td>
<td>49.1</td>
<td>32.8</td>
</tr>
<tr>
<td>HTTP</td>
<td>73.7</td>
<td>49.1</td>
</tr>
<tr>
<td>Download</td>
<td>98.3</td>
<td>65.5</td>
</tr>
</tbody>
</table>

It is noticeable that the latencies associated with the microwave solution are lower than the latencies associated with the optical fiber solution. This happens because the nodes supporting the new inserted BS are different for both solutions. As stated earlier, the Line of Sight algorithm tests the geographically closer node’ locations, while the Fiber Ducts Search algorithm tests nodes that will require less optical fiber cable. In addition, the path taken to the Core in the microwave solution travels fewer edges than the optical fiber solution.
The Core routing fiber optical and Microwave scenarios are presented in Figure, 7 and 8.

![Figure 7: Core Routing of fiber optical solution.](image)

![Figure 8: Core Routing of microwave solution.](image)

The uncoun optical fiber solution has a cost 3.5 times higher than the microwave solution. The new microwave connection has a length of 363 meters, while the optical fiber connection has a length of 200.80 meters.

In this use case was simulated the same new BS on a base line hour and as expected, the traffic flowing on the new location is smaller than on busy hour and translate on minors latencies.

B. Zone with lack of coverage

As stated in the Sector III, adding a BS is a solution when the network faces lack of coverage in a certain area. Figure 9 shows the lack of coverage (zone with the blue dots) at the frequency of 2600 MHz. This frequency is the most used for increasing capacity purposes. Regarding coverage, there are other frequencies that guarantee it in that area.

![Figure 9: Area with lacking coverage for the 2600 MHz frequency.](image)

However, to improve capacity, QoS and to ensure longer continuity at the 2600 MHz frequency, a new BS is proposed in that area, marked in Figure 10.

![Figure 10: Area where the new base station will be inserted.](image)

Initially, the algorithm relies on the database to gather the collected KPI in that area, through the existing locations where are described in the Table IV.

### Table V: Services that will be associated with the new location.

<table>
<thead>
<tr>
<th>Services</th>
<th>Volume [bits]</th>
<th>Throughput [kbits/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice</td>
<td>260.3</td>
<td>16.65</td>
</tr>
<tr>
<td>Streaming</td>
<td>26620.2</td>
<td>7048.29</td>
</tr>
<tr>
<td>HTTP</td>
<td>5324.0</td>
<td>149.81</td>
</tr>
<tr>
<td>Download</td>
<td>86.8</td>
<td>2.77</td>
</tr>
</tbody>
</table>

The Fiber Ducts Search algorithm indicates that the new BS is not near to fiber optic ducts. That said, it has no solutions for this technology.

On the other hand, the LoS algorithm presents a workable microwave solution with the new connection and path taken by routed services to the core, shown in Figure 11.

One can see from the following Figure 12 there is LoS between the new location and the base station that supports this new site.

The latencies given in Table VI satisfy the default values found in the Subsector III-C Table III for each service.
V. CONCLUSIONS

The main purpose of this thesis was to develop an algorithm that will allow telecom operators to improve network configuration and transmission, in order to assure good QoS. The algorithm is able to add a new BS by defining the best transmission solution, dependent on the used technology, available bandwidth, geographical locations of the transmission nodes while keeping good QoS at the minimum cost.

For all scenarios, the results obtained from the simulations, provide the best possible links for each transmission solution associated with the lowest cost, i.e., the locations supporting the new BS had LoS or optical fiber cable size under budget. The presented latencies for each scenario, portray a good QoS for each service. The different latencies are due to the volume size and throughput associated with each service, given that the available bandwidth on each link was sufficient. Regarding the latency associated with each node, it is negligible, as it uses its full capacity per port to process one service at a time, translating into very low latency.

In conclusion, the purpose of the thesis was achieved by showing the importance of an automated and fast software tool for defining new BS, and how operators can benefit from it, without doing the usual manual process.

Table VI: Latencies associated with microwave solution.

<table>
<thead>
<tr>
<th>Services</th>
<th>Latencies [m.s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice</td>
<td>47.0</td>
</tr>
<tr>
<td>Streaming</td>
<td>11.9</td>
</tr>
<tr>
<td>HTTP</td>
<td>107.2</td>
</tr>
<tr>
<td>Download</td>
<td>99.7</td>
</tr>
</tbody>
</table>

REFERENCES


