MODIFICATION OF HIERARCHICAL AGGLOMERATIVE CLUSTERING WITH COMPLETE LINKAGE FOR OPTIMAL USAGE WITH PASSIVE OPTICAL NETWORKS

Tomáš Pehnelt, Pavel Lafata*, Marek Nevosad

Department of Telecommunication Engineering, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, Prague 6, Czech Republic

* corresponding author: lafatpav@fel.cvut.cz

Abstract. Today, passive optical networks (PONs) represent a modern solution for high-speed FTTx subscriber lines and networks. Due to the fact that optical fibres and cables are deployed in last-mile network segments. Unfortunately, the costs of trenching and installation of optical cables in cities are still very high today. However, the application of mathematical methods and algorithms for designing optimum network solutions and topologies might significantly decrease the overall capital expenditures (CAPEX) of the whole process. This article introduces an efficient method based on agglomerative clustering for an optimization of deployment of PONs in practice. The paper describes its application and proposes the necessary modifications of the clustering method for the environment of PONs. The proposed algorithm uses the technique to cluster and connect the adjacent optical terminals and splitters of PONs and to calculate the optimum locations of these units in order to achieve the most optimum network solution and to minimize the summary capital expenditures.

Keywords: passive optical network; network optimization; topology optimization; algorithms; hierarchical clustering; optical splitter placement.

1. Introduction

Nowadays, the demands for transmission speed and overall transmission capacity of access networks are rapidly increasing [1, 2]. One of the most promising network solutions for building modern FTTx lines and networks are passive optical networks (PONs) [3]. A typical PON consists of one Optical Line Termination (OLT), which acts as a central unit, and Optical Network Terminals (ONTs) or Units (ONUs), which are located at the end-point users and connect these users to the PON itself [3, 4]. The Optical Distribution Network (ODN) is then composed of all optical elements between OLT and all ONTs including optical fibres, patch cords, connectors, and especially passive optical splitters [3, 4]. Since all these components are typically passive (no power consumption, no management), no optical signal regeneration, routing nor switching is possible between OLT and ONTs, therefore, the ODN must always be properly designed and optimized [3, 4]. However, there are also economical aspects and criteria, which need to be addressed during the process of planning of the ODN and its topology in practice [5, 6]. Today, the costs of deployment of optical fibres and cables, especially the trenching costs and the installation costs, are still very high. The last aspect, which needs to be addressed during the designing and planning of realistic ODNs and PONs in practice, are the locations and placements of all optical components in real networks and situations [7, 8]. Generally, the ODNs are built mainly in cities, suburbs, residential areas, industrial and commercial objects, because the optical fibres are mostly trenched or installed along the roads, pavements [9], in drains and subways, using existing street lights, telephone poles, etc. It is also necessary to adequately place passive optical splitters and the rest of the network elements. It is evident that the designing of passive ODN in order to meet all the network criteria for specific PON version is an important process, which needs specific calculations and designing methods [10]. There are various existing techniques described in numerous scientific papers focused on the optimization and planning of PON topologies and networks, the most important ones will be briefly discussed in the following section.

This paper, however, is focused especially on clustering of ONTs in order to obtain optimum locations of passive optical splitters and optimum network topology. Through performing a correct cluster analysis of ONT units and by designing their proper clusters, the optimum placement of passive splitters can be achieved and the overall network topology planned. The solution described in this paper is based on a hierarchical agglomerative clustering algorithm with implemented modifications and innovations. The proposed algorithm is primarily focused on the number of ONTs in formed clusters in order to optimize the splitting ratio of splitters. As the passive splitters greatly increase the summary attenuation in the PONs, the optimization of their positions and splitting ratios also play an important role. The proposed algorithm was also tested through typical realistic scenarios and the results are included in this paper as well.

2. LITERATURE REVIEW

This section brings a short review and comparison of existing and proposed techniques and algorithms for designing optimum PON topologies and networks.

Various existing studies have investigated methods for the optimization of network deployment including optical network and PON deployment. One of the very first methods was to apply a heuristic algorithm together with a minimum spanning tree algorithm in order to optimize the fibre length in the PON deployment, as presented in [11] and further improved in [12]. However, this solution does not use real map data and the computation time of heuristic method is also quite long, therefore, large scenarios must be divided into smaller ones, as illustrated in [8]. Another approach is to use ant-colony algorithms, as described in [13], and their combination with a metaheuristic method and updated techniques in [14]. In [14] and [15], an ant-colony method is used to find optimum locations of passive splitters to solve the optimization task of passive optical networks. However, the complex solution of the whole network topology is not provided, and moreover, no further optimization criteria are implemented for a multi-parametric optimization process. Similarly, [16] and [17] propose the application of genetic algorithms and genetic oriented techniques, however, the optimization is oriented towards optimum optical topology without the possibility of CAPEX minimization option or attenuation minimization decision.

In order to offer a complex PON optimization, [9] and [18] provide the combination of various algorithm techniques and methods to solve this problem. In [9], all end-users in the designed PON topology are first separated into clusters, then the clusters are connected to the OLT using splitters. Similarly, in [18], the hierarchical cascading technique of passive splitters is presented together with the application of clustering algorithms. In [19] and [20], surveys dealing with the recent PON evolution and innovations can be obtained. Nevertheless, none of the above referenced papers deals with the optimization of splitting ratios and with the clustering of the ONT units based on their optimum number in order to obtain an optimized splitting ratio solution. The solution presented here is partially based on initial ideas already presented in [21], however, this paper is focused exclusively on the optimization of splitting ratio performed through a hierarchical agglomerative clustering technique and on proposing an innovative algorithm.

3. PROPOSED HIERARCHICAL AGGLOMERATIVE CLUSTERING ALGORITHM

The main contribution of this paper is the designing of an innovative hierarchical agglomerative clustering algorithm with a modified complete linkage. These connections are updated throughout the running of an algorithm resulting in updating clusters of ONT units in order to optimize the splitting ratios of passive splitters in these clusters.

A passive optical splitter is a key component of PONs, providing splitting of optical signals in downstream direction towards all its outputs, while it combines the optical signals together in the upstream direction. The internal structure of each passive splitter is composed of basic Y-junctions, while each Y-junction forms a passive splitter with a ratio 1:2. By cascading these junctions, a resulting splitting ratio can be achieved. Due to that, the splitting ratios of passive splitters are typically powers of 2, e.g., 1 : 2, 1 : 4, 1 : 8, 1 : 16, 1 : 32, 1 : 64, 1 : 128 and 1 : 256. Due to that, the optimum number of ONT units in each cluster should be close to the power of 2 if possible. The following section brings the description and proposal of an innovative algorithm technique. The basic hierarchical agglomerative clustering algorithm is modified and amended with a complete linkage functionality.

3.1. DESCRIPTION OF EXISTING HIERARCHICAL AGGLOMERATIVE CLUSTERING METHOD

The hierarchical agglomerative clustering algorithm is a hierarchical method of a gradual merging of existing clusters until the desired number of clusters is reached. There are several methods of the process, which clusters should be merged, these methods are called linkage criterion. This criterion is used to determine the distance between two clusters. The following list contains some of the best-known criterions, the list is far from complete, as there are many existing types of linkage criterions:

Complete linkage. It uses the maximum possible distance between two clusters, i.e., distance between two farthest points in two sets.

Single linkage. in contrast to the complete linkage, single linkage uses the smallest distance between two clusters.

Centroid. This linkage uses distances between two centres of the two clusters.

Average linkage. It uses the average distance between each point in one cluster to each point in the other cluster.

Another important parameter for hierarchical clustering is the metric. This means, which distance function will be used for evaluating the distance between points, most commonly used are Euclidean or Hamming. The algorithm for the hierarchical agglomerative clustering starts with all points being single clusters each. In each step, the clusters with the smallest distance according to the selected linkage and metric criterions are merged. This process continues until the initial limit of number of clusters is reached.

3.2. Modified linkage for hierarchical agglomerative clustering algorithm

This section introduces the main contribution of this paper. Our goal was to optimize the resulting numbers of ONT units in all clusters, so that the number of units would be close to the number of ports of passive optical splitter (splitting ratio). Since the splitting ratio is usually the power of 2, the optimal number of ONT units including a reserve in the form of one port for potential updates and topology modifications in each cluster should be as close as possible to the nearest higher power of 2. The whole algorithm is described in the following flowchart in Figure 1.
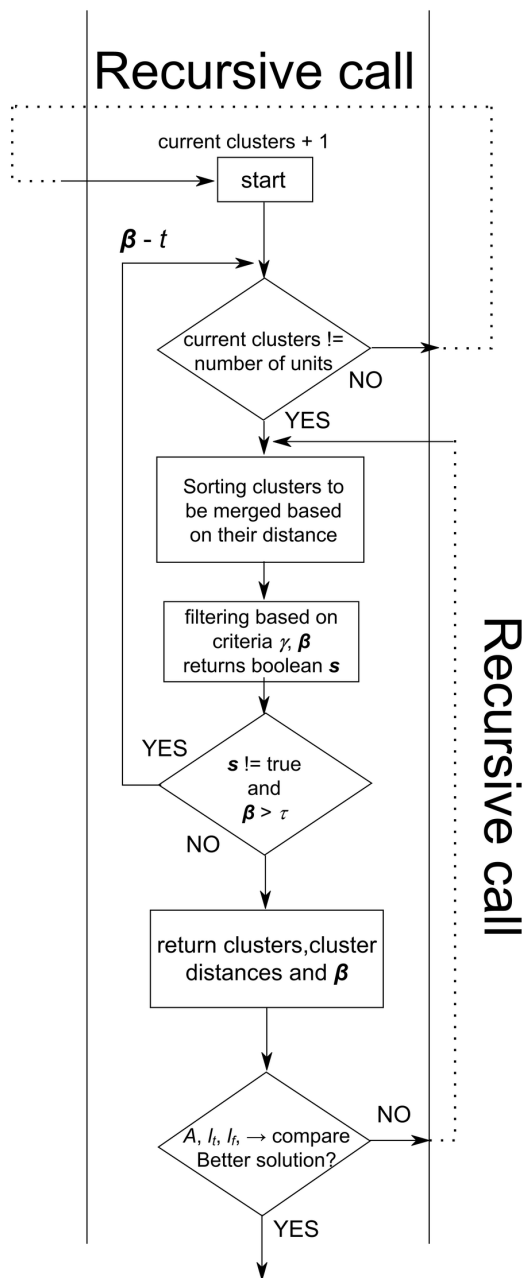

Figure 1. Recursive hierarchical agglomerative clustering.

The modified linkage for hierarchical agglomerative clustering algorithm in Figure 1 is recursive so that it is possible to have a feedback allowing for changing the parameters in the previous step. This includes mainly the parameters s and β, which will be described later.

The first step of the algorithm is the decision, whether the current number of clusters corresponds to the total number of points to merge. The algorithm starts with the number of clusters corresponding with the number of ONT units in the entire PON, the recursive function is called with the current number of clusters plus one. This process stops when the current number of clusters is the same as the number of ONU units to merge. When this number of clusters is reached, the main body of an algorithm begins. First, all possible cluster merges are sorted based on the distance between two farthest points in clusters to be merged (complete linkage). The list of these merges is further filtered, based on the metrics calculated through the following set of equations. In the first recursive step, the value β is set to 1 (meaning that the merging probability is 100 %). The threshold value τ is calculated in each recursive step as

$$\tau = (1 - \beta) \cdot 20. \tag{1}$$

The filtering process returns a boolean value s, which holds the result of the filtering, if no value is left after the filtering, the s is false, otherwise, it is true. If the filtering was not efficient and s holds false, then the algorithm returns, decrements the value β by t and again checks the value of current clusters. This mechanism returns the algorithm to its beginning until β reaches the threshold value τ or s holds the false value. The decision of merging two clusters (filtering process) is constrained by the threshold γ calculated as

$$\gamma = \mu(\tau + 1), \tag{2}$$

where μ is the minimum distance between two points in the whole dataset. Constraining the number of elements in the list according to the threshold τ as well as according to the number of units in merged cluster so that they only contain allowed amount of ONT units plus the reserve based on the bounds. This lower bound (lb) and upper bound (ub) are calculated as

$$ub = 2^n - 1, \tag{3}$$

where n is the nearest greater power of 2 and the number of units in a cluster; and

$$lb = \lfloor \beta(2^n - 1) \rfloor - 1, \tag{4}$$

where β is an acceptance criterion from the input of the process and the part $2^n - 1$ represents the greatest smaller power of two to the number of ONT units in the cluster.

Finally, the result of an algorithm (number of clusters, number of ONT units in clusters, cluster distances and β) is then used to calculate the resulting
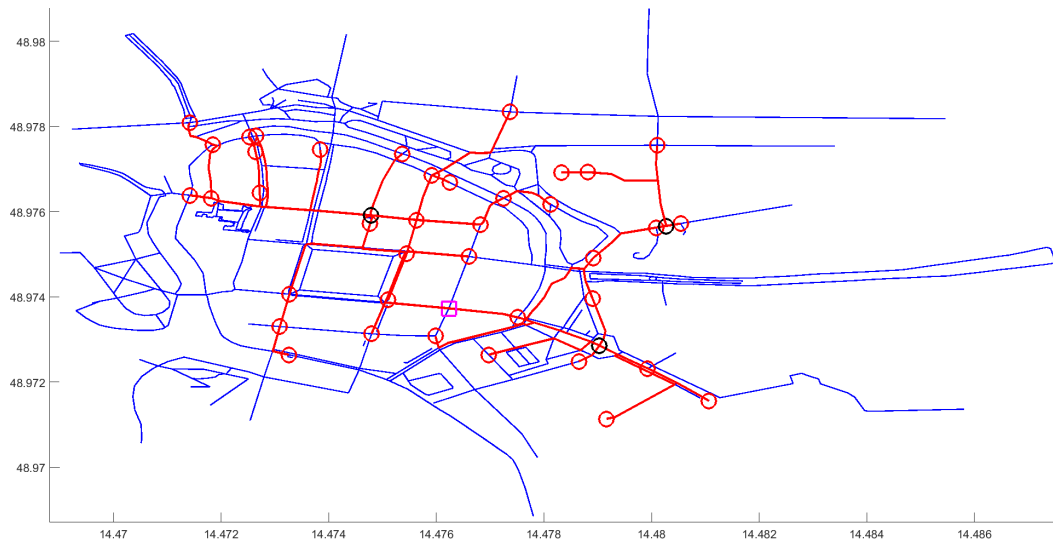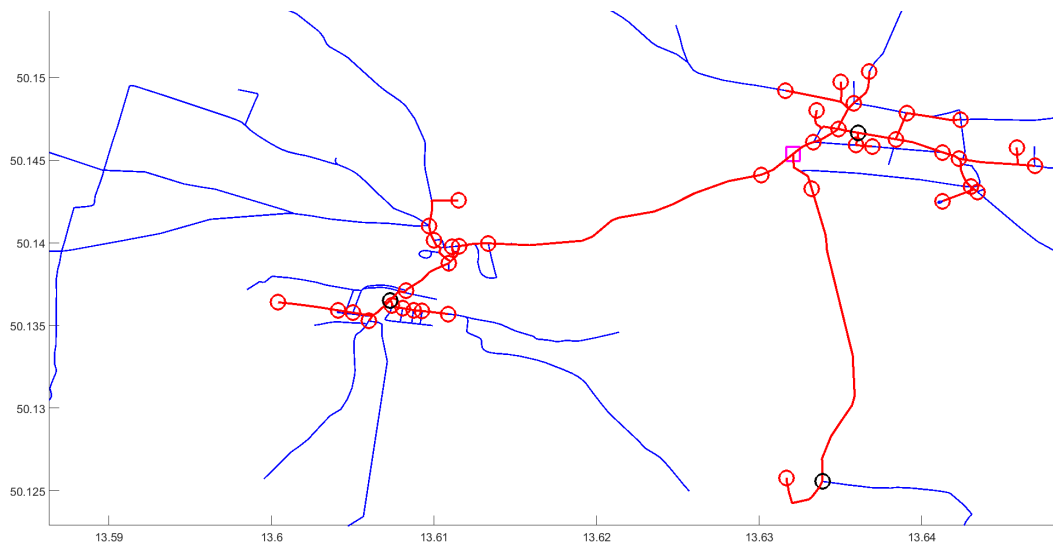
FIGURE 2. An example with the city center.

FIGURE 3. Rural area with 3 villages.

attenuation of the topology A, the summary length of optical fibres l_f and the total trenching distance l_t. Based on these values, the solutions obtained during the recursive runs of proposed algorithm can easily be compared and the best one identified. Once the solution with the desired values A, l_f, l_t is reached, or the best solution based on their comparison is achieved, the algorithm ends. Both the threshold value τ as well as parameter t influence the results of the algorithm. The parameter t represents a granularity of the algorithm steps and through numerous testing and calculations, its value was set to 0.05.

4. Results and discussion

The following section contains a discussion of the results and their comparison with the existing hierarchical agglomerative clustering algorithm. For the following comparison, the realistic data and maps from OpenStreetMaps.org were exported and two different examples were selected. The first one in Figure 2 illus-

trates a typical situation in the historical city centre (urban area), while Figure 3 contains a typical scenario for a rural area with 3 villages. In both figures, the position of Optical Line Termination (OLT) is illustrated using a violet square, the positions of passive splitters are marked using black rounds, while ONT units are illustrated by red circles. All paths (roads, streets, walkways, etc.) are represented by blue lines, while the resulting optical topology (optical fibres, cables) by red lines. These topologies were obtained by previously presented methods and algorithms described in a previous article [21] with the implementation of the presented innovative hierarchical agglomerative clustering with a complete linkage. The PON topology designing presented in [21] is based on innovative techniques using the Dijkstra's algorithm, metrics and proposed hierarchical clustering method with the complete optimum linkage described in this paper.

In order to provide a comparison with existing methods for different scenarios and topologies, the proposed

Hierarchical Agglomerative Clustering	Area type	Summary fiber length, l_{f} [m]	Number of ONTs in clusters	Maximal attenuation, $A_{\max}$ [dB]	Total trenching distance, l_{t} [m]
Proposed	city center	13405	7, 30, 3	23.1118	4590
Existing [21]		13410	11, 29	22.8004	4592
Existing [5]		14236	10, 30	22.8561	4604
Existing [22]		13951	11, 29	22.8004	4592
Proposed	urban	10511	14, 26	21.4530	3241
Existing [21]		10512	3, 5, 6, 26	21.4627	3253
Existing [5]		10705	4, 4, 7, 25	21.8909	3372
Existing [22]		10532	3, 5, 6, 26	21.4627	3253
Proposed	rural	22115	10, 30	22.2105	10474
Existing [21]		22115	10, 30	22.2105	10474
Existing [5]		22287	11, 29	22.3992	10532
Existing [22]		23563	7, 10, 23	23.0035	10701
Proposed	urban	14490	5, 13, 13, 9	24.8365	4605
Existing [21]		14270	5, 7, 15, 13	26.6512	4570
Existing [5]		14963	6, 6, 12, 10, 6	27.3008	4638
Existing [22]		14806	8, 14, 15, 3	27.0850	4623
Proposed	rural	18933	15, 22, 3	24.9055	4866
Existing [21]		22808	27, 13	32.5624	4952
Existing [5]		24073	14, 5, 11, 10	26.0898	4913
Existing [22]		23223	4, 15, 21	25.7300	4920
Proposed	city center	11181	11, 6, 11, 12	21.5039	5015
Existing [21]		11207	4, 8, 11, 8, 9	24.4557	5117
Existing [5]		11205	10, 7, 12, 11	26.0046	5095
Existing [22]		11340	5, 8, 8, 10, 9	24.3980	5109

TABLE 1. Comparison of existing and proposed methods.

algorithm and methods were used for several situations including rural, urban and city areas and were compared with the existing methods [5, 21, 22]. Based on these results, a statistical processing and comparison was performed, and its results are presented within the following Table 1. Rural areas represent typical village regions, urban maps illustrate typical city parts containing block of flats and houses, while centre areas consist of historical city centres. The existing methods presented in [5, 21, 22] are based on existing agglomerative clustering techniques, while the proposed method presented within this paper is referred to as Proposed in Table 1. The summary length of optical fibres l_{f} and the total trenching distance l_{t} [m] are obtained directly as the results of the algorithm in Figure 1. The maximum attenuation $A_{\max}$ [dB] is calculated as the maximum of attenuation A between the central OLT unit and the ONT units in the designed topology. Generally, the attenuation A in the PON can be calculated as

$$A = \sum A_{\mathrm{S}} + \sum n_{\mathrm{pe}}A_{\mathrm{pr}} + \sum \alpha l_{\mathrm{total}} + A_{\mathrm{res}} \quad [\mathrm{dB}], \tag{5}$$

where the sum A_{S} [dB] represents the sum of the attenuation of all passive splitters located between the OLT unit and the selected ONT unit, the sum $n_{\mathrm{pe}}A_{\mathrm{pe}}$ [dB]

represents the sum of the attenuation of n_{pe} passive elements each with an attenuation A_{pe} (connectors, splices, passive filters, etc.), the sum $\alpha l_{\mathrm{total}}$ represents the summary attenuation of optical fibres with an attenuation factor α [dB/km] and length l_{total} [km] and A_{res} [dB] stands for the attenuation reserve, which covers the attenuation compensation of aging, temperature fluctuations, deformations of optical fibres, etc. The attenuation of a symmetrical splitter A_{S} can be calculated according to

$$A_{\mathrm{S}} = 10\log_1 0N + A_{\mathrm{U}}\log_2 N + A_{\mathrm{pe}} \quad [\mathrm{dB}], \tag{6}$$

where N is the number of output ports of a splitter ($1 : N$ is its splitting ratio), A_{U} [dB] represents the uniformity of one single Y-junction cascaded in the internal structure of a splitter and A_{C} [dB] is the additional attenuation of connectors or splices used to connect the splitter in the PON topology. In all calculations, A_{U} is considered as 0.3 dB. In (5), the attenuation factor α is considered as 0.4 dB/km meeting the ITU-T G.652D optical fibre type, A_{res} as 1 dB and A_{pe} in (5) and (6) as 0.4 dB, which represents an attenuation of one standard optical connector of a type SC/PC or SC/APC.

Based on the results in Table 1, it is evident that the proposed algorithm provided the best solution in

almost all scenarios. The only situation, in which the existing method [21] provided a better result, was one of the urban scenarios. Evidently, the proposed hierarchical agglomerative clustering method with modifications presented in §3.2 performed the lowest summary fibre length, total trenching distance and the maximum attenuation in almost all the situations. Due to that, the CAPEX of this solution should be the lowest compared with all existing methods. Moreover, the proposed algorithm provided the best solution for all three different scenario types (city centre, urban, rural), therefore, it can be used for any of these situations.

5. Conclusion

This paper is focused on planning and designing an optimum network topology of passive optical networks by using advanced algorithms and techniques. The article presents an innovative hierarchical agglomerative clustering algorithm with a modified complete linkage ability. The algorithm is presented by a flowchart in §3.2 (Figure 1), while §4 contains its comparison with several existing methods and techniques. These results were obtained for various different scenarios and situations based on realistic map data obtained from OpenStreetMap.org and processed in Matlab environment. The performance of all methods and algorithms is compared through elementary output parameters and results, especially the maximum attenuation, summary fibre length and the total trenching distance. These parameters form the resulting capital expenditures (CAPEX) when building the network in practice. It is evident that the presented algorithm always provided the best solution, except for a single scenario. The presented algorithm provides the best results for all three typical types of scenarios in practice, the city centre area, the urban area and the rural area. The algorithm will be further evaluated and modified within the following research. For example, the comparison in Table 1 was performed for a fixed number of ONT units in a network, therefore, the following step is to implement the variable number of units in network topology.

References

[1] ARÉVALO, G. V., GAUDINO, R.: A techno-economic network planning tool for PON deployment including protection strategies. 19th International Conference on Transparent Optical Networks (2017). DOI:10.1109/ICTON.2017.8025122

[2] AKGUN, T., ÜNVERDI, N. O.: FTTX analysis and applications. 24th Signal Processing and Communication Application Conference (2016). DOI:10.1109/SIU.2016.7496220

[3] LAM, C., F.: Passive Optical Networks: Principles and Practice. Academic Press of Elsevier Inc., Burlington, USA (2007). ISBN 0-12-373853-9.

[4] NESSET, D.: PON roadmap. IEEE/OSA Journal of Optical Communications and Networking (2017). DOI:10.1364/JOCN.9.000A71

[5] PAPAEFTHIMIOU, K., TEFERA, Y., MIHAYLOV, D., GUTIERREZ, J. M. and JENSEN, M.: Algorithmic PON/P2P FTTH access network design for CAPEX minimization. 21st Telecommunications Forum Telfor (2013). DOI:10.1109/TELFOR.2013.6716199

[6] DAVIM, J. P., PINTO, A. N.: CAPEX Model for PON Technology. 2nd International Conference on Evolving Internet (2010). DOI:10.1109/INTERNET.2010.41

[7] CRAMER, S. and KAMPOURIDIS, M.: Optimising the deployment of fibre optics using Guided Local Search. IEEE Congress on Evolutionary Computation (2015). DOI:10.1109/CEC.2015.7256973

[8] ARÉVALO, G. V., SIERRA, J. E., HINCAPIÉ, R. C., GAUDINO, R.: A novel algorithm for PON optimal deployment over real city maps and large number of users. 18th Italian National Conference on Photonic (2016).

[9] LI, J., and SHEN, G.: Cost Minimization Planning for Greenfield Passive Optical Networks. IEEE/OSA Journal of Optical Communications and Networking (2009). DOI:10.1364/JOCN.1.000017

[10] MITCSENKO, A., PAKSY, G., CINKLER, T.: Topology design and capex estimation for passive optical networks. 6th International Conference on Broadband Communications, Networks (2009). DOI:10.4108/ICST.BROADNETS2009.7245

[11] KHAN, S. U.: Heuristics-based PON deployment. IEEE Communications Letters (2005). DOI:10.1109/LCOMM.2005.1506723

[12] LAKIC, B., HAJDUCZENIA, M.: On optimized Passive Optical Network (PON) deployment. Second International Conference on Access Networks & Workshops (2007). DOI:10.1109/ACCESSNETS.2007.4447124

[13] XIONG, W., WU, C., WU, L., GUO, X., CHEN, Y. and XIE, M.: Ant colony optimization for PON network design. IEEE 3rd International Conference on Communication Software and Networks (2011). DOI:10.1109/ICCSN.2011.6013738

[14] HU, W., WU, K., PING SHUM, P., ZHEDULEV, N., SOCI, C.: Using nonlinear optical networks for optimization: primer of the ant colony algorithm. IEEE Conference on Lasers and Electro-Optics (CLEO), (2014).

[15] ZU, Y., YANG, W., WANG, Y., SHAO, L.: The routing algorithm of multi-layer multi-domain intelligent optical networks based on ant colony optimization. 15th IEEE International Conference on Communication Technology (2013). DOI:10.1109/ICCT.2013.6820399

[16] LIU, Z., JAEKEL, A., BANDYOPADHYAY, S.: A genetic algorithm for optimization of logical topologies in optical networks. International Parallel and Distributed Processing Symposium (2001). DOI:10.1109/IPDPS.2002.1016609

[17] VILLALBA, T., ROSSI, S., MOKARZEL, M., SALVADOR, M., NETO, A., CESAR, A., ROMERO, M., ROCHA, M.: Design of Passive Optical Networks Using Genetic Algorithm. SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference (2009). DOI:10.1109/IMOC.2009.5427496

[18] LIN, B., LIN, L., and HO, P. H.: Cascaded splitter topology optimization in LRPONs. IEEE International Conference on Communications (2012). DOI:10.1109/ICC.2012.6364216

[19] BUTT, R. A., HASUNAH, M. S., IDRUS, S. M., REHMAN, S. U.: Evolution of Access Network from Copper to PON – Current Status. ARPN Journal of Engineering and Applied Sciences (2015).

[20] BUTT, R., IDRUS, S., ZUL, N., ASHRAF, M.: A Survey of Energy Conservation Schemes for Present and Next Generation Passive Optical Networks. Journal of Communications (2018). DOI:13.10.12720/jcm.13.3.129-138

[21] PEHNELT, T., and LAFATA, P.: Optimizing of Passive Optical Network Deployment Using Algorithm with Metrics. ADVANCES IN ELECTRICAL AND ELECTRONIC ENGINEERING, (2017). DOI:10.15598/aeee.v15i5.2285

[22] OLSEN, D. A. Closing the loop on a complete linkage hierarchical clustering method. 11th International Conference on Informatics in Control, Automation and Robotics, (2014). DOI:10.5220/0005058902960303