Application of Wireless Return
Topology Planning Based on K-Means

Yang Weixia
Xi'an Technological University
School of computer science and engineering
Shaanxi 710021, Xi'an, China
e-mail:1195862787@qq.com

Li He
Xi'an Technological University
School of computer science and engineering
Shaanxi 710021, Xi'an, China
e-mail:1003294436@qq.com

Xu Fei
Xi'an Technological University
School of computer science and engineering
Shaanxi 710021, Xi'an,
e-mail:China29112462@qq.com

Chen Yuan
Xi'an Technological University
School of computer science and engineering
Shaanxi 710021, Xi'an, China
e-mail:1677622439@qq.com

Abstract—**With the development of communication technology and the general trend of interconnection and intercommunication, people's demand for high-quality information and communication is increasing day by day. Therefore, the intensive deployment of base stations has become the development trend of the construction of a new generation of communication network. In this paper, in order to solve the urban and rural electric scenarios such as traditional transport unreachable problem, on the basis of the operator site cost at the same time, keep the quality of signal transmission, give full play to the advantages of the traditional planning concept and artificial intelligence algorithm, integrated "pieces", "step by step optimization", "local optimal", "the results of reverse", such as ideas, set up a wireless back mathematical analysis model of network topology programming problem.**

The purpose of this paper is to reduce the cost and reduce the return path loss. Aiming at the lower cost problem, a local optimal model is constructed. First division, through the K - Means algorithm is divided into K, for each region is also based on K - Means algorithm is further split into its n tribal groups, limited each tribal group one and only one host station, recently to tribal group of perimeter centroid position of butterfly stand stand as a host, the rest of the site as a child, according to the type of tribal group of judge whether the number of sub station meet the qualification, if not satisfied, change the K value and the adjustable radius, respectively, until meet the qualification, the best solution is calculated. For the lower return path loss problem, if only the loss problem is considered, the loss result is obtained by changing the distance between the first jump and the second jump based on the model. Through the comparison of the results, the best scheme is obtained when only loss is considered.

Keywords-Wireless Return; Base Station Deployment; K-Means Algorithm; Piecemeal Optimization

I. THE PROBLEM BACKGROUND

With the rapid development of mobile communication network, various mobile terminal devices and applications affect every aspect of people's life. However, the current network information construction faces some problems: the acceleration of urban construction makes the urban environment more and more complex, which leads to the formation of many wireless signal black spots and weak coverage areas in densely populated urban areas; Some urban residents misunderstand the construction of base stations, believing that the harm of base stations into the community is serious, which makes the deployment of base stations significantly more difficult, and the phenomenon of station demolition is also increasing. Due to the difficulties in property coordination in many communities, the arrival rate of the last kilometer of transmission fiber deployment is low.

In literature [1], the deployment principles, coverage characteristics, station type selection and comparison with conventional micro-base stations are studied, and the actual deployment scheme of Relay micro-base station is proposed, and the effectiveness of the scheme is verified. Literature [2] focuses on analyzing the principle and characteristics of Relay technology, and deeply studies the deep coverage of urban areas and the wide coverage of rural areas and roads. Literature [3] studies the impact of Relay technology on the wireless network structure, wireless network planning and the development of wireless network planning tools. Literature [4] discussed whether the broadband wireless communication system

can be deployed under the condition of non-line-of-sight NLOS. Can you provide reliable high bandwidth connections? Is there a reliable high bandwidth NLOS solution available? The above literature mainly starts from the theory, while this paper mainly studies the base station deployment plan from the actual situation, establishes the model and conducts the feasibility verification, and finally achieves the goal of low cost and small path loss to improve the user experience.

II. MATHEMATICAL MODEL

A. The known conditions

The location distribution of candidate sites in this paper is known, and there are 1000 sites. Only the mutual location and topological relationship between sites are considered, that is, only the distance between the host station and the sub-station is considered. The integrated cost of various station types, including the integrated cost of the host station, the integrated cost of the sub-station, and the cost of satellite equipment are also known.

Total cost = host station cost * number of host stations + substation cost * number of substations + satellite cost * number of satellites

Average cost = total cost/number of sites in the region

The topological relationship between sites meets the following conditions:

1) The distance between the host station and the first-level sub-station is no more than 20km, and the distance between sub-stations is no more than 10km.

2) The site is divided into two types: RuraStar (one sector) and butterfly station (two sectors).

3) If it is the host station, the maximum number of sub-stations at level 1 in each sector is 4, and the total number of sub-stations does not exceed 6.

4) Regardless of the coverage direction of butterfly station sector;

5) The limit of microwave communication distance between host stations is 50km.

6) Wireless return connection is adopted between host station and sub-station and between sub-stations;

7) Each sub-station can only have two wireless back links at most, that is, the upstream and downstream links are unique;

8) The relation diagram between host station and substation is similar to the tree diagram.

9) There is only one path between any substation and the host station, and the number of hops is less than 3.

10) There is a one-to-many relationship between the satellite and the host station. A host station group with less than 8 satellites can share a satellite.

The wireless back propagation is affected by the NLOS scene, and the free space propagation is adopted to simplify the calculation.

The model estimates the path loss between sites, and the formula is as follows:

$$PL = 32.5 + 20*\lg（D）+20*\lg（F）$$

Where, PL stands for path loss, D is the distance between the two stations, the unit is km, F is the transmission frequency, the default is a constant, 900MHZ.

Average system loss = sum of all wireless return connection losses/number of wireless return connections
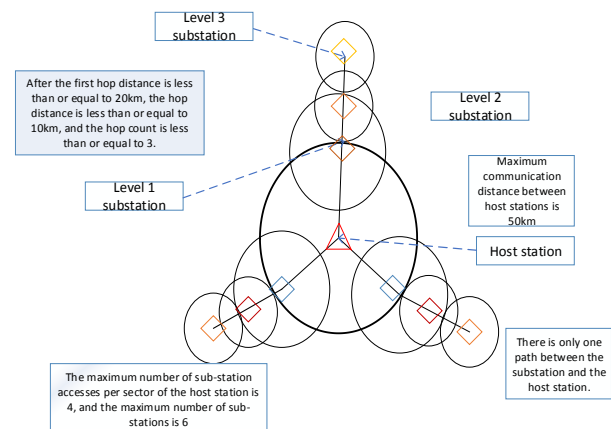

Figure 1.   Schematic diagram of connections between sites

B. Question assumptions

1) Suppose RRN(eRelay Remote Node) wireless transmission device as a substation;

2) It is assumed that DeNB, as the host station, can be divided into 1 to 3 host cells, covering different directions.

3) It is assumed that the effects of terrain blocking and ordinary mobile phone access blocking on the back transmission quality are not considered.

4) It was assumed that ReBTS interference was not considered.

5) The interference of adjacent base stations is not considered.

6) It is assumed that the first jump of cascade between base stations is not greater than 20km, and then it is not greater than 10km.

7) Assuming that the sector coverage direction of butterfly station is not taken into account, the maximum number of sub-stations is 12.

8) It is assumed that the maximum number of RuralStar stations is 4.

9) It is assumed that the host stations are connected by microwave and the maximum communication distance is 50KM.

10) Suppose that the host station and the sub-station and the sub-station are connected by wireless back transmission.

11) It is assumed that a substation can only have one host station, and there is only one path to the host station, and the path contains no more than 3 hops.

12) It is assumed that there can be no more than 2 wireless return connections between each substation.

13) Assuming that only one satellite of any host station is responsible for the back transmission, host stations connected by slices can share one satellite, and each satellite can only bear the back transmission data of 8 host stations.

14) It is assumed that the total number of host stations is unlimited.

15) Assuming that the maintenance costs of the host station, substation and satellite are not taken into account.

16) Assuming that other factors are not considered, the spherical model is transformed into a plane model.

17) It is assumed that the path consumption is estimated by the free space model without considering the influence of NLOS.

C. *Meaning of the symbols*

TABLE I.   NOTATION TABLE

symbols	Meaning
LOS	Line-of-sight transmission capability
ROI	Return on investment
NLOS	Non-line-of-sight transmission capability
RRN	The infinite return device of the infinite return scheme
RN	Relay station
UE	Ordinary mobile phone
PL	Path to the consumption
APL	Average path consumption
D	Site spacing
F	Transmission frequency
R	Radius of the earth
S	The distance between the spheres
$α_i$	I point longitude
$β_i$	I point the dimension
i	Base station identification
X_i	The host site
Y_i	The child site
Z_i	Satellite point
C	The overall cost
FD	The first jump distance
ND	After each jump distance
FX_i	Butterfly host station
RX_i	Star host station
WL	Microwave connections between host stations
WBL	Wireless return connection between host station and sub-station and between sub-stations
JS	The number of hops from a substation to a host station
Ceil	The function that rounds up

D. *The establishment and solution of the model*

1) Analysis of problems

In combination with table 2, the cost of satellite > host station costs > sub station costs, to achieve the overall cost of wireless return part of the station is lower, the first need to ensure the minimum number of satellites, that is, as far as possible 8 host stations share a satellite; Secondly, the number of host stations should be kept as small as possible. Since butterfly station has one more sector than RuraStar, the maximum access number of sub-stations is 12, covering a wider range, so butterfly station should be selected as the host station as far as possible. In real life, the deployment of the site is usually partitions, according to the "division", "local optimum, and the thought of" step by step optimization ", first by using K Means clustering algorithm, all the site is divided into K classes, each

kind of belonging to a satellite, then each category divided into n host station tribal groups, the mathematical model established between K, n, and cost, butterfly site closest to the center of mass as a host station, and then judge the location of the relationship between sub station with the host station whether meet the constraints of distance, the host station to station traverse connection relations, Select the connection path with the most extensive coverage, and get the scheme with lower cost.If a plane carrying a large number of passengers is replaced by a small passenger plane some passengers will overflow, and some passengers will not be able to register because of the number of passengers.

For path loss, only consider the child stand back part of the path loss, and not to consider path loss between host station, just meet the distance limit, so can be achieved by increasing the number of the host station to smaller path loss, when the host stand for most time, path loss minimum, but also will increase the cost of satellite transmission. If the effective distance of wireless return transmission is limited, that is, the distance between the substation and the host station and between the substations is limited, then in theory, a low path loss can be obtained. By means of the idea of "result inverse", the model is modified, and then the modified station model is deduced, and the lowest cost scheme, namely the optimal solution of the problem, is finally screened out.

2) Problem model establishment
a) Establishment of objective function

Total cost: number of host stations * host station cost + number of substations * substation cost + number of satellites * satellite cost.

The number of satellites is equal to Ceil (number of host stations /8), and Ceil () indicates the upward direction.

$$MinC = X \times 10 + Y \times 5 + Z \times 50 \qquad (1)$$

Where: C represents the total cost under the topology;

X represents the number of host stations;

Y represents the number of sub-stations;

Z is the number of satellites.

TABLE II.   SHOWS THE COSTS OF VARIOUS MODES OF TRANSMISSION IN WUSD

transport	cost
host station	10
child station	5
satellite	50

b) Establishment of constraint conditions:

- The distance of the first hop is 20km, and that of each subsequent hop is 10km.

- Site includes RuralStar and butterfly station two different station type; Among them, RuralStar contains 1 sector in total, and butterfly station contains 2 sectors in total. If the site is the host station, the maximum number of sub-stations at the first level of each sector is 4, and the maximum total number of sub-stations is 6. In order to simplify the problem, the sector coverage direction of butterfly station is not considered for the moment.

- Microwave connection is adopted between the host stations, and the maximum communication distance is 50km.

- Wireless return connection is adopted between host station and sub-station and between sub-stations.

- Each sub-station can only have two wireless back links at most.

- Any sub-station can only belong to one host station, and there is only one path to the host station, and the number of hops contained in the path is less than or equal to 3.

- Any host station has and only one satellite responsible for the back transmission. Host stations connected by slices can share the same satellite, but a satellite can only bear the back data of eight host stations at most.

- In a monolithic host station, there is no upper limit for the total number of host stations, i.e., the constraint conditions are as follows:

$$S.T \begin{cases} FD \leq 20 \\ ND \leq 10 \\ RX_i \leq 6(Y) \\ FX_i \leq 12(Y) \\ (X_i)WL(X_j) \leq 50KM \\ (Y_i)WBL(Y_j) \\ Y_{WBL} \leq 2 \\ (Y_i)JS(X_i) \leq 3 \\ Z = Ceil(X/8) \end{cases} \tag{2}$$

c) Establishment of the model

Based on the above analysis, a lower overall cost model with the highest priority is established as follows:

$$MinC = X \times 10 + Y \times 5 + Z \times 50$$

$$S.T \begin{cases} FD \leq 20 \\ ND \leq 10 \\ RX_i \leq 6(Y) \\ FX_i \leq 12(Y) \\ (X_i)WL(X_j) \leq 50KM \\ (Y_i)WBL(Y_j) \\ Y_{WBL} \leq 2 \\ (Y_i)JS(X_i) \leq 3 \\ Z = Ceil(X/8) \end{cases}$$

III. K-MEANS ALGORITHM

Clustering analysis is an important analysis method in data mining. Its goal is to divide the data set into several clusters, so that the similarity of data points within the same cluster is as large as possible, while the similarity of data points between different clusters is as small as possible. The study of Clustering Algorithms has a long history. Hartigan systematically discussed Clustering Algorithms in his monograph Clustering Algorithms as early as 1975. Since then, the academic circle has proposed a variety of clustering algorithms based on different ideas, mainly including the algorithm based on partition, the algorithm based on hierarchy, the algorithm based on density, the algorithm based on grid and the algorithm based on model. All these algorithms can achieve good clustering effect, among which the k-means algorithm based on partition is the one that is applied most and has a simple algorithm idea. By dealing with the difficult constraints, k-means algorithm makes the solution of the problem relatively easy. The algorithm has a good convergence, and the solution speed of the algorithm can also meet the requirements of real-time.

A. Processing flow of k-means algorithm:

1) For a randomly given set of 1000 sites, the samples are divided into K clusters according to the size of the distance between sites. Suppose the cluster is divided into (B_1, B_2 ,... B_k), then our goal is to minimize the squared error E:

$$E = \sum_{i=1}^{k} \sum_{x \in B_i} \|X - \mu_i\|_2^2 \tag{3}$$

Where μ_i is the mean vector of cluster B_i, called the center of mass, and the expression is:

$$\mu_i = \frac{1}{|B_i|} \sum_{x \in B_i} X \tag{4}$$

A suitable K value range of 10~25 was selected through cross validation. That is, k samples are randomly selected from 100 data sets as the initial k centroid vectors: { $\mu_1, \mu_2 ,... \mu_k$,}.

2) Take N as the maximum number of iterations. For N =1,2,3... N.

a) Classify the cluster B and initialize it as $B_t = \Phi$ t=1,2... K;

b) For I =1,2... M, calculate sample x_i and each centroid vector μ_1 (j=1,2... K) distance from: $d_{ij} = \|X_i - \mu_j\|_2^2$, mark x_i with the smallest d_{ij} as the corresponding category λ_i, and then update $B_{\lambda_i} = B_{\lambda_i} \cup \{X_i\}$;

c) For j = 1, 2,... K, recalculate the new center of mass $\mu_j = \frac{1}{|B_j|} \sum_{x \in B_j} X$ for all the sample points in B_j;

d) If all k centroid vectors do not change, go to step (3).

3) Output cluster division $C = \{C_1, C_2,... C_k\}$.

IV. SIMULATION EXPERIMENT

A. Cost modeling experiment

The specific steps for solving the model are as follows:

Step1: use k-means algorithm to conduct data aggregation and violent calculation on difficult constraints, and obtain the upper bound of the original problem and the next one, that is, the solution without considering other constraints will make the overall cost the lowest;

Step2: by solving the lower overall cost model with the highest priority, k-means algorithm is used again to conduct data aggregation for the classified subregions, and considering the solutions of other constraints, the number of host stations in each subregion is obtained;

Step3: if the obtained solution satisfies the optimal solution of the condition, stop the algorithm, add the number in each cell, and the obtained solution is the optimal solution of the problem, otherwise go to Step2.

Obtain the topology structure satisfying the constraint conditions through Step3. Take k=20 and the adjustable radius of the cell is 20km as an example, as shown in Figure 2. In the figure, the green "o" represents the host station, the blue "*" represents the substation, and the black line segment represents the connection relationship between stations.
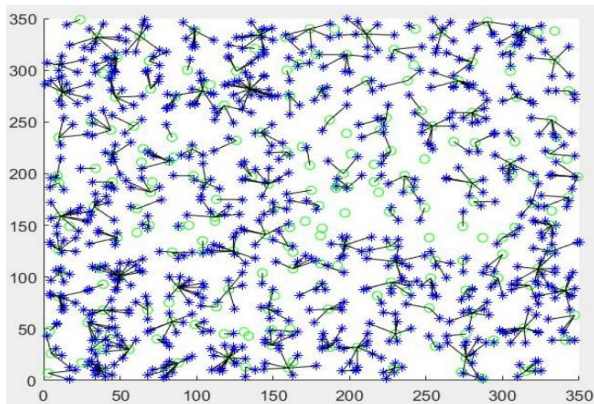

Figure 2.   Base station topology

Step4: obtain the minimum cost result output through the overall cost formula for the optimal solution obtained in each subregion.

Get the result of the lowest cost through Step4. Take k=20 as an example, and the results are shown in figure 5-6. The number of host stations is 222, the number of sub stations is 778 and the number of satellites is 28. The total cost is 7510 (WUSD) and the average cost is 7.5100 (W USD). The efficiency of the algorithm is less than 2 minutes, with strong convergence.


```
STEP1 is over.
Now,STEP2 is begining:calculate the topology of sites.
The Calculation has finished %0
...%5...%10...%15...%20...%25...%30...%35...%40...%45...%50...%56...%60...%6
STEP2 is over.
Now,STEP3 is begining:report the solution.
The datas of Posi and Graph are being generatinged %0
...%10...%20...%30...%40...%50...%60...%70...%80...%90...%100
The datas of Posi and Graph have been generatinged all over.
The datas of Posi and Graph have been saved.
Now, graphing the topology of these sites....
Topology-graphing is over.
Now, report the solution...As following:
    本方案共设置:
        宿主站222 (站),
        子站  778 (站),
        卫星  28 (个)。
    本方案未完成设置的站点数量为:   0 (站)。
    本方案的总体成本为:  7510 (W USD)。
    本方案的平均成本为: 7.510000e+00 (W USD)。
    本方案系统平均损耗为: 1.117896e+02 (dB)。
时间已过 103.787610 秒。
```

Figure 3.   Program run result

By contrast the chosen radii are 20, 25, and 30. Get the corresponding site distribution, number and total cost. See table 3.

TABLE III.   COST COMPARISON TABLE

Radius(km)	Host station	child station	satellite	Overall cost(W USD)
20	222	778	28	7510
25	136	858	17	6500
30	110	871	14	6155

By comparison, when the cell radius is 30, the overall cost is the lowest. The minimum cost is 6155 (W USD). The larger the radius is, the more points of untreated stations are. However, even if the radius is 30, there are no 6 stations connected by leakage, which can be ignored.

B. Path loss modeling experiment

According to the simulation result of cost, the connection relation and physical location between host station and sub station are obtained. If only the path loss of the back transmission part of the sub-station is taken into account, and the path loss between the host stations is not taken into account, it only needs to meet the distance limit, and a smaller path loss can be achieved by increasing the number of host stations. When X is the maximum, PL is the minimum, but it also increases the cost required by satellite transmission. If the first and second hop distances are limited, that is, the distance between the sub-station and the host station and between sub-stations is reduced to obtain a lower path loss. If there is only a level 1 sub-station, the number of sub-stations in a

single sector of the host station is required to be less than or equal to 4, and then the modified sub-station model is deduced in reverse. Finally, the scheme with the lowest cost is selected, that is, the optimal solution to the problem.

Although there is an influence of non-line-of-sight transmission capability (NLOS) in wireless return transmission, in order to simplify the problem, the free-space transmission model is adopted to estimate the path loss between stations. The formula is as follows:

$$PL = 32.5 + 20 \times \lg(D) + 20 \times \lg(F) \qquad (5)$$

Where, PL is the path loss and the distance between the two stations, D is km, F is the transmission frequency and the unit is MHz, 900MHz is adopted by default here.

The average system cost in APL is equal to the sum of the losses of all wireless back links/the number of wireless back links

It should be noted that the path loss only considers the return part of the sub-stations, and the microwave transmission is adopted between the host stations. The loss is not calculated as long as the distance limit is satisfied.

By comparison, the first jump/second jump distance combinations were selected as 20/10, 15/8 and 12/6, respectively. Get the corresponding number of sites and average loss. See table 4.

TABLE IV.   NUMBER OF SITES AND AVERAGE WASTAGE

First jump/second jump	Host station	child station	satellite	Average loss (dB)
20/10	222	778	28	111.79
15/8	385	610	49	109.44
12/6	588	410	74	107.30

Reducing the distance between the first jump and the second jump can reduce the average loss, so the 12/6 jump scheme is the best choice. However, the number of sites not connected to the topology increases, the signal coverage decreases, and the cost is very high.

To sum up, if only the cost is considered, the best solution is that when the adjustable constraint radius is 30km, the overall cost is the lowest, and the lowest cost is 6155 (W USD). If only loss is considered, k=30, the

first jump distance 12km and the second jump distance 6km are selected as the best scheme.

V.   SUMMARY

In this paper, in the process of achieving the goal, a local optimal model based on k-means algorithm is constructed. By traversal comparison of the site deployment of host station and sub-station under the global large clustering and local small clustering, the lowest cost sub-station scheme is screened out. The establishment of the model adopts the idea of "dividing the whole into zero" and applies the k-means algorithm. Due to the scalability and high efficiency of the algorithm itself, it can simply and quickly divide 1000 candidate sites into K parts for step-by-step solution, which significantly reduces the amount of calculation and improves the speed of operation. However, because the k value is predetermined, the selection of this k value is very difficult to estimate, and the one-time optimal programming cannot be realized. In addition, due to the clustering division of the overall site, the final solution of the model is always locally optimal, which may be slightly different from the overall optimal solution. Therefore, the initial value of k is limited in the modeling process. Let k be between 15 and 25, and solve different k values for many times, so as to obtain the relative optimal solution by comparison. After getting the lower cost of the substation scheme, considering the return path loss of the substation, the established local optimal model is modified from the perspective of reducing the loss. Considering the results, microwave connection is adopted between the host stations and no loss is calculated. The average loss of the system is related to the wireless return distance, so the algorithm efficiency is significantly improved by limiting the distance between stations. However, the limitation of this model lies in the increase of host station and the increase of overall cost. Although it is beneficial to the service quality of users, it increases the cycle of investment return, which is not in line with the original intention of operators. In order to take into account the interests of operators, follow-up can be adjusted according to the situation.

REFERENCES

[1]   Ma Liang . Wireless Relay transmission technology in the micro base station deployment strategy [J]. Journal of mobile communications, 2017, 41 (24) : 13 to 18.

[2]   Li xin, peng xiongen. Deployment and application of Relay technology in LTE network [J]. Mobile communications, 2016,40(11):32-35.

[3]   Deng shuifa. Research on LOS/NLOS communication environment identification and NLOS error elimination technology [D]. Southwest jiaotong university,2016.

[4]   Wang bo. Wideband lineless communication system in NLOS field scene without apparent distance [J]. Digital communication world,2017(02):55-57.

[5]   Liu sen. Analysis on the return rate of investment in enterprise cloud computing information [J]. China science and technology BBS,2014(06):83-87.

[6]   Zhang yong-jun, yao zhi-cong, wang jing, li hua-hua. Capacitor voltage balance control strategy of a cascade h-bridge static reactive power generator [J]. Chinese journal of electrical engineering,2014,34(27):4621-4628.

[7]   Wu huayi, liu bo, li dajun, ling nanyan. Research review on topological relations of spatial objects [J]. Journal of wuhan university (information science edition),2014,39(11):1269-1276.

[8]   Shi langyu, ma ling. Application of Relay wireless back transmission technology in td-lte networking scheme [J]. Telecommunications engineering technology and standardization, 2017, 30(2):41-44.

[9]   Xie J y, gao h c. selection algorithm based on statistical correlation and k-means to distinguish gene subset [J]. Journal of software, 2014(9):2050-2075. (in Chinese)

[10] Zuo l, he y g, li b, zhu y q, fang g f. research on path loss of passive uhf rfid system [J]. Acta phys. Sin,2013,62(14):150-157. (in Chinese)

[11] Jia h J, ding s f, shi z z. approximate weighted kernel k-means algorithm for solving large-scale spectral clustering [J]. Journal of software, 2015, 26(11):2836-2846. (in Chinese ）