Data Fusion Algorithm of Power Line Carrier Communication Based on Minimum Spanning Tree Theory

Hai Lu¹, Ziyang He²*, Haibo Pen²*, Ting Yang²*

¹Electric Power Research Institute, Yunnan Power Grid Co., Ltd., Kunming, 650217, China
²School of Electrical and Information Engineering, Tianjin University, Tianjin, 300072, China
*Corresponding author’s e-mail: penhaibo@tju.edu.cn, heziyang@tju.edu.cn, yangting@tju.edu.cn

Abstract: With the expansion of the clean energy photovoltaic power generation systems, the requirements of communication and data exchange capacity become higher and higher. The status information of each photovoltaic unit should be collected. However, when the number of PV units is too large or the collection frequency is too high, it will generate huge amount of data to the communication system. If each monitoring terminals transmits the data to the master station independently, the huge traffic flows occupy bandwidth and reduce the communication efficiency, which also affect the real-time monitoring of the photovoltaic system. To solve this problem, this paper proposed a novel data fusion algorithm for power line carrier communication based on minimum spanning tree transmission path. The data fusion packet format is also presented. Finally, the effectiveness of the proposed algorithm is proved by comparing the amount of data brought by the fusion algorithm and the non-fusion algorithm.

1. Introduction

Power line carrier (PLC) communication technology is widely used in the local communication method of the photovoltaic power generation system. In this deployment, each carrier module is a communication node. The monitoring information is transmitted to the concentrator through the power line, and then sent to the master station. With the construction of ubiquitous power Internet of things, the system requires higher frequency of data collection for photovoltaic power generation. With the expansion of the photovoltaic system scale, frequent information interaction between high-density nodes is bound to bring massive data and consume a lot of communication resources. Therefore, how to reduce the amount of data in the network and transmit data reliably and efficiently is the core problem of the whole ubiquitous power Internet of things’ access layer. To solve this problem, this paper proposes a power line carrier communication algorithm based on minimum spanning tree routing for network data fusion. By establishing the minimum spanning tree route, the data format of DL/T645-2007 protocol is used for network fusion at the path sink node, so as to realize effective compression of mass data transmission, save communication transmission resources and improve transmission capability of power line carrier access network.
2. Data Volume Analysis of Power Line Carrier Communication

The power industry standard DL/T645 protocol specifies the frame format for power line carrier transmission, as shown in Figure 1:

Each frame consists of seven fields: frame start character, address field, control code, data length field, data field, frame information longitudinal check code and frame terminator. Each field is composed of several bytes. \( L \) is the number of bytes in the data field, \( L \leq 200 \) when reading data, and \( L \leq 50 \) when writing data. However, in the actual electricity data acquisition system, the number of bytes required by reading the energy measurement data is generally only 4 bytes, which is far less than the reserved 50 bytes and does not occupy the maximum length of the data domain. For a distributed photovoltaic power generation system, the photovoltaic data to be collected includes active power, reactive power, voltage and current. If each data is a 32-bit floating-point number occupying 4 bytes, there are only 16 bytes in total. If each node sends metering information separately, the actual transmission contains a large number of empty bytes and duplicate frames. When there are a large number of nodes and frequent information interaction in the network, it will waste the precious communication resources and reduce network performance.

The specific distributed photovoltaic power generation and load advanced measurement nodes are mostly connected to the concentrator by diffraction structure. Therefore, network data fusion technology can be adopted to fuse the transmitted data in network fusion, so as to reduce the total amount of transmitted data and improve the performance of network transmission on the basis of guaranteeing the integrity of data transmission. For the total positive active power transmitted by power line carrier communication, the amount of data transmitted from an independent node to the master node is 16 bytes. If 100 nodes in the network transmit data to the master node separately, the amount of data transmitted by the whole communication network is 1600 bytes. When the frequency of acquisition is high, the system will generate a huge amount of data. When monitoring system collects information from each node in the photovoltaic system, it can be known from the specified frame field that each node only has a different data field and address field, so there is certain redundant information in the data frame of each node.

3. Data Fusion Based On Power Line Carrier Communication

The traditional data fusion model is applied to the calculation of the communication information system in the network: the sink node collects the originally collected information of all connected nodes and integrates it into a data packet [1]. This model does not consider the number of fusion nodes and the number of data packets generated in \([t_0, t] \) time period. Obviously, this model has certain limitations and does not conform to the application of ubiquitous power Internet of things. In the actual system, each node collects different real-time information, such as the active and reactive power output, instantaneous voltage and current of photovoltaic unit, and the power quality harmonic information on the user side. Whether the heterogeneous original data can be completely compressed into a packet needs to be determined according to the actual data content.
This paper improves the model and establishes a more general information fusion model as shown in Figure 2. The mapping relationship between the number of output packets and the number of input packets is as follows:

\[
\bar{X} = \sum_{i=1}^{n} m_i (x_i - c_i) + c
\]  

(1)

Where c is overhead transmission data, and its value is different according to the fusion type (lossy fusion and lossless fusion). \((x_i - c_i)\) is the length of the information segment of the input packets, \(m_i\) is the data compression rate, and the value of \(m_i\) is determined by the entropy value of the data to be fused [2-3]. Because information fusion does not increase the load of each input packet, usually \(m_i \leq 1\). Here the "pan-sensor node" can be a single node or a node cluster.

**Definition:** when the concept of the signal source is extended to a set of random events -- information source, the information entropy can describe some uncertainty of the set of random events.

All the activities that lead to the increase or decrease of the certainty, organization, regularity or order of random time set can be measured by the change of information entropy.

According to the data fusion model, the total length of bytes output by power line carrier communication at the fusion node is:

\[
X(x_i) = \sum_{i=1}^{n} x_i + s + a
\]  

(2)

Where \(n\) is the number of acquisition nodes associated with the fusion node. \(x_i\) is the total number of bytes occupied by the data and address of each independent node, and \(s\) is the total number of bytes occupied by the data and address of the fusion node, and \(a\) is the number of bytes occupied by the fixed format of the data frame, generally \(a = 7\).

When writing data, if the number of bytes occupied satisfies equation (3), the data and address of \(n\) nodes can be fused to the fused node for data fusion and transmission.

\[
\sum_{i=1}^{n} x_i + s \leq 50
\]  

(3)

For example, when the number of data and address bytes of the independent node to be uploaded is \(x_i\) bytes, the data on \(n\) carrier modules can be fused.

\[
n = \left\lfloor \frac{50 - s}{x_i} \right\rfloor
\]  

(4)

The data fusion model collects the data information of all subordinate nodes at the fusion node and integrates the core information locally through unpacking/packaging operations [4-5]. The data on each independent node is fused to the fusion node for centralized transmission, and each independent node should have its own address field and data field, in which the address field occupies 6 bytes, while the number of bytes occupied by the data field is generally 4 bytes depending on the type of information collected. Figure 3 shows the data frame format transmitted at the fusion node.
Here, the data fusion efficiency is analyzed: suppose that the monitoring photovoltaic grid contains 100 acquisition nodes, and active power of each photovoltaic panel needs to be collected. If each node separately transmits the data to the master node, the data transmitted by the whole communication system is 1600 bytes. When applying the data fusion method proposed in this paper, the data and address of each independent node take up 10 bytes. A fusion node can fuse up to 4 subordinate nodes nearby, so there are 20 fusion nodes in the system to transmit the fusion data to the master node. Therefore, the minimum amount of data to be uploaded by the whole communication system is 1100 bytes, and the amount of data is reduced to 68.75% as compared without using the fusion method.

### 4. Minimum Spanning Tree Routing Algorithm for Data Fusion

The power line carrier communication network including m communication nodes can be expressed as a directed graph \( G(V,E) \), where \( V = M \cup \{s\} \) is the collection of \( m+1 \) nodes, \( s \) is the root node. Each communication node needs to establish a connection with the root node via the communication link, and there are multiple transmission paths between each node and the root node. Finding the shortest effective path between the root node and each node in the network topology will reduce the attenuation of high-frequency signals in the transmission process of power line carrier communication and improve the transmission success rate [6]. Therefore, the data modeling of this problem is to establish the minimum spanning tree from \( m \) communication nodes to the root node \( s \). A given network \( G = (V,E,W) \), let \( T = (V,E') \) be a support tree of \( G \), order:

\[
\text{w}(T) = \sum_{e \in E'} \text{w}(e)
\]  

(5)

\( w(T) \) is the weight of \( T \). The spanning tree with the least weight in \( G \) is the connected tree with the smallest \( w(T) \).

![Figure 4](image_url)

Figure 4. Minimum spanning tree path

Figure 4 shows the route established based on the minimum spanning tree algorithm. Where CCO
is a central coordinator in broadband carrier communication network, PCO is a proxy coordinator, STA is a station.

Algorithm steps:

Step1. Let \( v \) be any vertex of \( V \), \( S_0 = \{v\}, E_0 = \emptyset, k = 0 \).

Step2. If \( S_k = V \), the algorithm ends, and the graph with \( S_k \) as vertex set and \( E_k \) as edge set is the minimum spanning tree of \( G \); Otherwise, execute step 3.

Step3. Construct \( [S_k, S] \), if \( [S_k, S] = \emptyset \), \( G \) is disconnected and terminated. Otherwise, calculate \( w(e_k) = \min_{e \in [S_k, S]} \langle e \rangle \), \( e_k = v_k v_k' \), \( v_k, v_k' \in S_k \), let \( S_{k+1} = S_k \cup \{v_k'\}, E_{k+1} = E_k \cup \{e_k\}, k = k+1 \), execute step 2.

According to the above data fusion transmission model in the communication network, when the network establishes the minimum spanning tree, it can establish transmission link from each node in the network to the master node, and data fusion is done once data reaches each fusion node. In the process of information transmission, the photovoltaic data collected by the master station is firstly modulated to the high-frequency carrier, and the attenuation degree of the high-frequency signal is related to the distance of the wire. Therefore, when constructing the minimum spanning tree, distance between two communication nodes is taken as the weight of each edge, and each communication node in the communication network is connected through the above steps to construct a minimum spanning tree.

5. Performance Analysis

By comparing the data transmission amount of this method and traditional method, the proposed data fusion algorithm is proved to compress data effectively.

For the single-phase connection mode, all nodes are connected to the same phase, so data from all nodes in the network can be merged with data from other nodes of the same phase. The amount of data using fused and unfused transmissions is shown in Figure 5(a). It can be seen from the figure that for a large-scale photovoltaic area with hundreds of nodes, the data transmitted by the data fusion method is smaller than that generated without data fusion. Especially when the number of nodes is large, the data amount is greatly reduced.

For the three-phase connection mode, under the same network scale, it is assumed that the acquisition nodes in the network are evenly distributed on each phase, and the amount of data transmitted by the two methods on each phase is shown in Figure 5(b). It can be seen that for the same network scale, the average distribution of nodes in the network on three phases can reduce the amount of data generated by power line carrier communication in each phase. In addition, the amount of data transferred with data fusion is significantly lower than that reference network without data fusion.
6. Conclusion
In this paper, the concept of data fusion is introduced into the power line carrier communication in photovoltaic system data acquisition. By using the minimum spanning tree transmission path algorithm, the data on each node is fused at the fusion node according to the fusion mechanism. The fusion node transmits the data to the master node via the minimum spanning tree path established, and then the master node transmits the data to the master station. The master station can collect information from each communication node with significantly reduced amount of data transferred via power line carrier communication owing to the proposed algorithm in this paper. By comparing the data generated by the fusion method with the data generated by traditional method, it is verified that the method proposed in this paper can effectively reduce the traffic in the network, realize the saving of network resources, and improve the transmission efficiency of power line carrier communication.

Acknowledgments
This work was sponsored by the National Key Research and Development Program of China (2017YFE0132100), National Natural Science Foundation of China (61971305, 61571324), China Postdoctoral Science Foundation (2019M651037) and Natural Science Foundation of Tianjin (19JQNJ06000)

Reference
[1] Jia, L. (2019) Research on Low Energy Transfer Algorithms for Data Fusion in Industrial Internet of Things. Electronic world., 06: 55-56.
[2] Kang, J., Zuo, X Z., Tang L. W., et al. (2010) Data Fusion Technology in Wireless Sensor Networks. computer science., 37(04):31-35+58.
[3] Wu Y., Li Y., Shi L. (2019) "Efficient Linear Sensor Fusion Over Multiple Lossy Channels With Local Observability.". IEEE Control Systems Letters., 3(3):721-726.
[4] Ehatisham-Ul-Haq M., et al. (2019) "Robust Human Activity Recognition Using Multimodal Feature-Level Fusion.". IEEE Access., 7(5):60736-60751.
[5] Pai, H., Han, Y. (2018) "Power-Efficient Direct-Voting Assurance for Data Fusion in Wireless Sensor Networks,". IEEE Transactions on Computers., 57(2):261-273.
[6] Liao, F B., Zhang, W M. (2017) Non-uniform Clustering Routing Protocol Based on Minimum Spanning Tree. Journal of Sensing Technology., 30(09):1412-1416.