Study of QoS-aware reliability transmission methods for edge computing networks in power distribution IoT

Yongchao Wang a, Dongdong Cai b, Yugui Nian c
State Grid Henan Electric Power Company Luoyang Power Supply Company
Luoyang, China
a yongchaowang@foxmail.com, b ddcai81@163.com, c nyg821@sina.com

Abstract. As the construction of IoT for power distribution advances, millions of power distribution equipment, electrical quantity sensors and condition sensors will be connected to the IoT network, thus generating huge amounts of heterogeneous power distribution data, whose collection, transmission and calculation will put great pressure on the communication channel and the master station storage and calculation system. To solve the above-mentioned distribution network problems, edge computing technology is introduced and an edge computing platform is deployed on the end-side to achieve lightweight data processing. The contradiction between the huge amount of data generated by IoT terminals and their QoS requirements, and the limited system resources, makes it imperative to study QoS-awareness transmission methods for edge computing networks.

Keywords: Distribution IoT, edge computing, QoS awareness, reliability.

1. Introduction

Distribution IoT is a new form of power network resulting from the deep fusion of traditional industrial technology and Internet of Things technology. Through the comprehensive interconnection, interoperability and interoperability of distribution network equipment, it can realize the comprehensive perception, data fusion and intelligent application of distribution network, meet the lean management needs of distribution network and support the rapid development of energy Internet, which is the distribution network in the new generation power system. As a new network computing paradigm, edge computing technology is the core link in the four-tier architecture of the distribution IoT, located between the terminal smart device and the cloud master, and is an open platform for data aggregation, computation and application integration. Edge computing enables node data to be processed in a timely and efficient manner near the source. Unlike cloud computing, this model does not require all data to be transferred to the data center, thus avoiding problems such as limited network bandwidth and data transfer delays, making it possible for data at the edge nodes to be processed more efficiently.

While edge computing solves the transmission challenges posed by exploding data volumes, it still suffers from the heterogeneity and reliable transmission of a large number of end nodes. As more and more end devices are connected to the wireless network, while the data they generate is no longer required to be transferred to a "distant" cloud in the edge computing network, somewhat slowing the network load, the number and unpredictability of the data transferred to the edge nodes still requires
competition for limited channel bandwidth resources. Also, most of the data generated by end devices is now real-time in nature and needs to be sent before a limited time is available. Traditional node-priority-based sending algorithms cannot guarantee that all real-time data is transmitted within the effective time, resulting in unstable system transmission performance and a high risk of reducing transmission success.

2. IoT QoS Assurance and Reliability Delivery Mechanisms

2.1. Traditional retransmission mechanisms
The simplest reliability transmission strategy. Re-transmission of failed packets, with high latency and wasted bandwidth resources.

2.2. Cluster-based transport protocols
This type of protocol divides the network into multiple clusters based on different factors, with one cluster header selected for each cluster. Other nodes send data to the cluster header, which manages and sends it uniformly. This type of transmission is only applicable to networks with stable node locations, where nodes in edge computing networks are highly mobile and the network state is complex, and the use of this type of transmission protocol leads to frequent failures of cluster headers and links, which is a fatal disadvantage for effective data transmission.

2.3. Collaborative communications transmission
This type of transmission mechanism uses relay nodes to help retransmit or forward data and improve network transmission reliability. Common transit node selection strategies are mainly energy-based and distance-based. However, such approaches are not applicable to edge computing networks due to a single consideration, which is affected by node data load, storage space, etc., generating high data latency and resulting in reduced network throughput and delivery success.

2.4. Transport mechanism based on ACO
Such algorithms typically construct heuristic colony algorithms based on adjacent node distance and residual energy to find the optimal transport path for nodes. This type of transmission mechanism is usually effective in reducing network energy consumption and balancing the load. However, it does not take into account the latency requirements for mixed transmission of multiple data streams, and the implementation is complex.

2.5. Reliable transmission mechanisms for emergency flow
This type of transmission mechanism uses congestion control and priority queue scheduling to meet the urgent data transmission needs and improve the user experience as much as possible. However, the failure to provide a guaranteed minimum transfer rate for low-priority data streams can result in "starvation" of low-priority data.

2.6. Energy-based transmission mechanisms
Energy-based transport protocols effectively reduce network energy consumption and extend network life. The mechanism requires global information from known networks, and the energy constraints of the IoT result in access to only local information. Moreover, such mechanisms do not take into account the different latency requirements of multiple types of data streams and are not suitable for edge computing networks with diverse node data.

3. Co-optimized synergistic transmission mechanisms for priority and equity
Priority and fairness are two key metrics for IoT device data transfer in edge computing systems. In order to ensure reliable transmission based on priority data and fairness in network resource allocation, and to meet the demand for real-time data latency transmission while avoiding starvation of low-priority
streams, this paper proposes a cooperative transmission mechanism that is jointly optimized by priority and fairness.

Node Classification. This paper divides all edge devices into two groups based on channel quality and determines which devices need transit nodes and which devices can communicate directly with edge servers. In particular, devices with poor channel quality are very likely to fail to transmit data and need to find the best relay nodes to assist in forwarding data and improve access success rates.

Transit node selection. After distinguishing the edge devices in the previous step, select the best relay node for the node that needs to relay the data. The choice of transit nodes has a significant impact on the transmission time and energy cost of the entire network. Therefore, in selecting the transit node, this paper integrates several more important factors. For example, available storage space, amount of data, location information, available energy, etc.

Priority setting. Prioritization of each data stream needs to be determined prior to data transmission, which is critical for bandwidth allocation and data stream scheduling. In this paper, the priority of each data stream is assigned based on its effective remaining time, device importance and throughput difference.

Bandwidth allocation and data flow scheduling. Data transmission is highly dependent on wireless channel bandwidth resources. Therefore, scheduling strategies and rate allocation algorithms are critical. In this paper, we calculate the optimal allocation factor for bandwidth allocation when we can make full use of channel resources, which improves the transmission success rate of packets and system throughput.

After all devices have determined the transmission path, they begin sending data to the edge server. Due to the effective transmission time of the data stream and the limited channel bandwidth resources, a scheduling strategy that combines priority with the shortest remaining time can effectively improve the transmission performance of the network system.

The packet waiting time cannot be taken into account when allocating bandwidth, so the packet transmission time should be:

$$T_{im} = D_{im}/r_i, \forall f_i, \forall i \in N$$

For each data stream, its transmission performance can be expressed as.

$$\varphi_i = 1/T_i$$

Among them, $T_i = \sum_{f_i} \sum_{s_{im}} s_{im} T_{im}$, is the duration of the data stream. In addition, this paper allocates bandwidth based on the principle of fairness, where the data streams are prioritized, i.e. the transmission performance metrics of the streams should be proportional to their priority.

$$\varphi_i \propto p_i, \forall f_i, \forall i \in N$$

$$r_i \propto p_i D_{f_i}, \forall f_i, \forall i \in N$$

Among them, $D_{f_i} = \sum_{m \in f_i} D_{im}$, is the size of the data stream.

$$r_i = a p_i D_{f_i}, \forall f_i, \forall i \in N$$

$a$ is allocation factor. It is related to the size of the channel capacity and it can be deduced that its optimal solution is:

$$a^* = B/\left\{ \sum_{i \in N} (h_i + 1)p_i D_{f_i} \right\}, \forall i \in N$$

After obtaining the optimal allocation factor, the size of the bandwidth allocated to each stream is calculated, and since the optimal allocation factor is calculated based on the size and priority of the data stream under the bandwidth constraint, it complies with the principle of fairness and makes full use of bandwidth resources. The scheduling mechanism in this paper takes into account each data stream and allows for an equitable distribution of data stream levels.
4. A hierarchical data transmission algorithm based on channel quality

In a heterogeneous wireless network coexistence environment, the number of data transfers between diverse devices faces multiple challenges such as spectrum limitation and wireless interference. The acquisition and utilization of wireless channel quality has proven to be one of the most effective means of sharing spectrum resources. A data transmission strategy based on channel quality can explore and exploit all potential data transmission opportunities to improve data transmission success and system throughput and maximize resource utilization.

1) Hq Rating.
If the current channel quality level is Hq, it means that the channel is of good quality and can guarantee a high data transmission success rate and robustness. It is possible to send node data directly, ensuring the accuracy of data information and positioning analysis results.

2) Mq rating.
If the current channel quality level is Mq, it means that the current channel quality is average and the throughput of the system is limited, and the data needs to be pre-processed to extract valid data before sending data. For indoor positioning data for smart security systems, multipath reflections and ambient noise can affect the CSI variation of subcarriers, and these disturbed subcarriers can negatively affect target positioning. Therefore, it is necessary to filter out those subcarriers that are highly affected by multipath and noise and extract only the CSI variation of the "clean" subcarriers for position estimation. Also, if the amount of data sent is too large, transmission errors or long delays can easily occur, resulting in degraded transmission quality.

3) Lq rating.
Channels at the Lq level are of very poor quality and require reasonable planning of data processing to ensure transmission performance. In this case, less data needs to be extracted to ensure the accuracy of the data transfer. Therefore, in case of poor channel quality, the subcarriers affected by the multipath effect are first removed in this paper with the same treatment as for Mq-rated channels. Beyond that, further treatment is needed. The data between the transceivers is transmitted via subcarriers, each pair of transceiver antennas containing multiple subcarriers. Due to the similarity of the subcarriers, there are many redundant subcarriers. Therefore, in order to mitigate the communication load and make full use of the bandwidth resources, after the nodes filter out the subcarriers affected by the multipath effect, PCA (principal component analysis) extraction method is used to remove the redundant data and identify and extract k subcarriers that are highly correlated with the positioning analysis.

After channel quality detection and classification, node data is pre-processed according to different levels of channel quality to fully utilize bandwidth resources and ensure system transmission quality. After determining the channel quality classification method and the data processing operations corresponding to the different levels of channel quality, the two operations are combined and applied to the data transmission mechanism.

First, the number of weak subcarriers in the current channel is measured at certain time intervals. The current channel quality class is determined based on the threshold values σ1 and σ2. The device that needs to send the data is then preprocessed to match the channel quality level of the data in the buffer based on the current channel quality level feedback. If the current channel quality level is Hq, the data is not processed. Finally, the sending device establishes a TCP connection with the target node, sends the pre-processed device data, and disconnects when the sending is complete.

Assuming that all data streams to be sent are of the same type. All nodes set to $G = \{1,2,3,...,i\}$, the flow is set to $F = \{f_1,f_2,...,f_i\}$, $\forall i \in G$, the size of each data stream is $V_i, \forall i \in G$, the current channel capacity is $C$.

This section uses a weighted allocation algorithm that allocates bandwidth to each data based on the amount of data sent in its stead. The binary variable $t_i$ is used to indicate whether the node is ready to send data.

$$t_i = \begin{cases} 1, & \text{Node i Data to be sent} \\ 0, & \text{other} \end{cases}$$

According to the weighted allocation principle, the bandwidth that can be allocated to each device is:
\[ r_i = \frac{C}{\sum t_i V_i} \]

After all nodes have processed the source data according to the current channel quality level, they send the data to the target node according to the allocated bandwidth. During the sending process, the data to be sent node first sends a connection request to the target node, and after the target node sends an ACK (acknowledge), the source node also sends an ACK message, and the connection is successful. Depending on the bandwidth rate allocated, the to-be-sent data node sends pre-processed data to the target node. If sending is successful, continue sending packets. If the send fails, retransmit the failed send packet again, up to three times per packet. After the count is sent, the source node sends a request to release the connection. The target node receives the request and first replies to confirm the known release connection request. After completing data reception, a consent release message is sent, the source node replies with a confirmation message, and the connection is disconnected. Each terminal sends the pre-processed data to the target node according to the above process.

Due to the high number of weak subcarriers that are prone to errors due to interference when the channel quality is poor, there are limited reliable subcarriers that can be utilized. If no hierarchical transmission policy is used, sending data directly based on the allocated bandwidth will result in higher bit error rate and packet loss, reducing transmission success rate and throughput. The hierarchical transmission algorithm proposed in this paper can extract highly reliable data, ensure transmission quality, improve data accuracy and transmission success rate.

5. Sum up
In recent years, the application of edge computing has become more and more widespread, providing an effective solution for the fast processing of massive data in the distribution IoT. However, with limited resources in edge computing network systems, how to ensure the efficient transmission of large amounts of data has become a hot issue of concern to researchers and major enterprises. In this paper, we propose a reliable QoS-aware transmission strategy for edge computing networks, firstly, we analyze the transmission needs of end nodes in edge networks, and then we model the data co-transmission for this problem, and propose a data co-transmitter and a transmission algorithm based on channel quality hierarchy for joint optimization of fairness and priority..

Acknowledgments
I would like to express my gratitude to all those who helped me during the writing of this article. I would also like to give special thanks to all my colleagues who have benefited from their work experience and guidance in building the foundation for the writing of this article.

References
[1] Lacanina P J M , Jaen A D L V , Ramos J L M . A new technique for short-term reliability assessment of transmission and distribution networks[C]// Probabilistic Methods Applied to Power Systems, 2006. PMAPS 2006. International Conference on. IEEE, 2006.
[2] Song Y , Yau S S , Yu R , et al. An Approach to QoS-based Task Distribution in Edge Computing Networks for IoT Applications[C]// IEEE International Conference on Edge Computing. IEEE, 2017.
[3] Macit M , Gungor V C , Tuna G . Comparison of QoS-aware single-path vs. multi-path routing protocols for image transmission in wireless multimedia sensor networks[J]. Ad Hoc Networks, 2014, 19(aug.):132-141.
[4] Nateghi R , Guikema S D , Wu Y , et al. Critical Assessment of the Foundations of Power Transmission and Distribution Reliability Metrics and Standards[J]. Risk Analysis, 2016, 36(1):4-15.
[5] Al-Zahid K K , Matsumoto M . A Dynamic QoS-Aware Transmission for Emergency Traffic in IP Networks[C]// Quality of Service in Multiservice Ip Networks, Third International Workshop, Qos-ip, Catania, Italy, February. Springer-Verlag, 2005.
[6] Vijendran A S, Thavamani S. ANALYSIS STUDY ON CACHING AND REPLICA PLACEMENT ALGORITHM FOR CONTENT DISTRIBUTION IN DISTRIBUTED COMPUTING NETWORKS[J]. International Journal of Peer to Peer Networks, 2012, 3(6):13-21.

[7] Xiao M, Chuang L, Han Z, et al. Energy-Aware Computation Offloading of IoT Sensors in Cloudlet-Based Mobile Edge Computing[J]. Sensors, 2018, 18(6):1945-.

[8] Ye Q, Wu M, Wang Y F. A Survey on QoS-Aware Secure Data Transmission in Multimedia Sensor Networks[J]. Applied Mechanics & Materials, 2014, 511-512:120-125.

[9] Siya X U, Lei W, Zhu L, et al. A QoS-Aware Packet Scheduling Mechanism in Cognitive Radio Networks for Smart Grid Applications[J]. 2016(2):68-78.

[10] Ye Q, Wu M, Wang Y. A QoS-aware Trust Model for Multipath Load Balancing in Multimedia Sensor Networks[J]. Journal of Software Engineering, 2015, 9(1):50-65.

[11] Zou C, Jin T, Chigan C, et al. QoS-aware distributed spectrum sharing for heterogeneous wireless cognitive networks[J]. Computer Networks, 2008, 52(4):864-878.

[12] Ye Q, Wu M, Wang Y. A QoS-aware Trust Model for Multipath Load Balancing in Multimedia Sensor Networks[J]. Journal of Software Engineering, 2015, 9(1):50-65.

[13] Pham K, Blasch E, Chen G. QoS-aware dynamic spectrum access for cognitive radio networks[J]. Proceedings of SPIE the International Society for Optical Engineering, 2013, 8739(5):511-521.

[14] Siya X, Lei W, Zhu L, et al. A QoS-aware packet scheduling mechanism in cognitive radio networks for smart grid applications[J]. Communications, China, 2016, 13(2):68-78.