Demonstration of Precise Planning of Broadband Access Network based on Mining Traffic Trends and Demands from Hybrid Data Sources

Hui Li[1*], Xianyi Guo[1], Tianshun Zhan[1], Wu Jia[2], Yudan Su[2], Guangsheng Yang[1], Jinglei Sun[1], Yan Shao[2], Yuefeng Ji[3], Guangquan Wang[2]

1. Beijing Laboratory of Advanced Information Networks, Beijing University of Posts and Telecommunications(BUPT), Beijing 100876,China
2. Network Technology Research Institute, China Unicom, Beijing 100048, China
3. State Key Laboratory of Information Photonics and Optical Communications, BUPT, Beijing 100876, China
Author e-mail address: lihui@bupt.edu.cn

Abstract: We demonstrate a carrying capability evaluation system, which can evaluate and predict the access network capacity and efficiency by extracting detail network status and trends from hybrid data sources based on machine learning.

1. Introduction

Driven by the broadband China strategy, China's broadband access network is developing rapidly. By June 2019, the number of FTTH/O users in China has reached 396 million. Compared with five years ago, the average download rate of fixed broadband has increased by more than six times, but its traffic tariff has dropped by more than 90% [1]. At the same time, Higher speed PONs are emerging, such 25G PON, 50G PON and even 100G PON, which will provide ample bandwidth to support applications like 5G transport. [2]. How to introduce new access technology at the right time and right place has become a problem that operators must consider.

In the development of PON standard, the compatibility of loss budget has been carefully considered, most of the existing ODN can support consecutive PON technique evolution, so that the research of access network planning and design mainly focuses on the service level at present [3]. There are two approaches of optimization of PON, one is to analyze user behavior and aggregates business users and residential users with complementary traffic time characteristics to improve network efficiency, and the other one is to use mixed integer linear programming to obtain the best topology configuration in different cases for mix of residential and business subscribers [4] [5]. In the two methods, the prerequisites for successful optimization are detailed information of business requirement, traffic performance information and network distribution, which usually distribute in different systems. Another motivation is from the development of new broadband service such as Cloud VR, which make operators need to know more about the access network traffic feature to assess feasibility of the new services. As the evolution of data collection capabilities of network equipment, massive network data analysis needs will be inevitably produced for the access network, which provide a great platform to combine the emerging data mining tools with the access network analysis needs. In this demo, we developed a carrying capacity evaluation system (CCES) of broadband access network to integrate this information, which has imported 32-week operation data and related resource allocation data of a city access network (including more than 100 thousand PON interfaces). The CCES explores the use case of machine learning tools to support the construction, planning, management and analysis needs of access network, and constructs the basement for the application of emerging data analysis tools in the access network analysis.

2. Demonstration Overview

2.1 Architecture and workflow of the system

The CCES belongs to Operation Support System (OSS), and it is composed of application layer, service layer and data layer as Fig. 1 shows.

(1) ***Data Layer****:* This layer collects data from hybrid data sources, such as IP Network Management System (IP NMS), PON Network Management System (PON NMS), Resource Management System (RMS), Authentication Authorization Accounting (AAA) system, IPTV service system, Speed Measurement System (SMS), Element management system (EMS). After checking and linking, those data are stored in relational database with masking if necessary according to different scenarios. Long-term collection is adopted with regular intervals to get the change trend. The main function of this layer is to build an optimized data storage structure so as to better serve the data analysis function of the service layer. The data storage in this layer often does not follow the strict data paradigm design, and there will be some redundancy in data storage to improve the performance of data analysis. When necessary, data can be decomposed or merged by rules in some necessary occasions.

(2) ***Service Layer:*** Data association and calculation will be executed in this layer. Traditional evaluation analysis includes statistical analysis, such as basic summary and regression analysis, but for data characteristics, such as

user behavior profile and traffic trend analysis, it needs to adopt data mining and machine learning. These processing methods are integrated into modular service functions for the upper application layer to call. This layer requires high computing ability, which often involves large-scale computing activities such as big data flow computing or distributed computing.

(3) ***Application Layer:*** This layer provides evaluation reports according to different evaluation tasks, by calling corresponding service module in the Service Layer. It displays the network performance data and various business indicators after the calculation and processing, and describes the network status in the form of charts. Unlike traditional dedicated network management system or performance monitor system, the CCES focuses on the graphical presentation of the overall network and different levels of detail information. Before forming the final report, CCES users can invoke different analysis tools to aggregate data, mining and discovering the characteristics of subscribers' behavior and network structure. Thanks to the separation from the working system, users can analyze data and adjust the network structure more freely, without worrying about the impact on the operating networks.

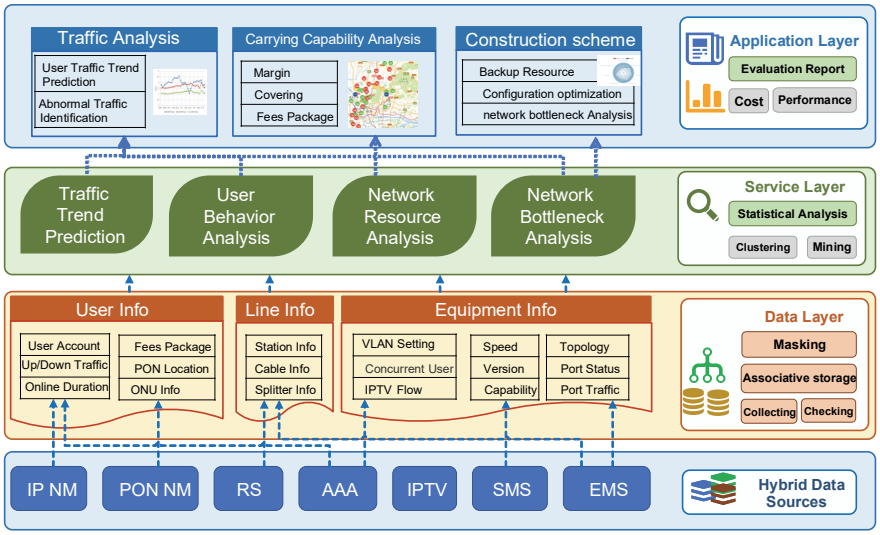

Fig. 1. The Architecture of Carrying Capacity Evaluation System.

In order to illustrate the function of the system, we present two cases as follows.

2.2 Traffic growth trend evaluation based on cluster analysis

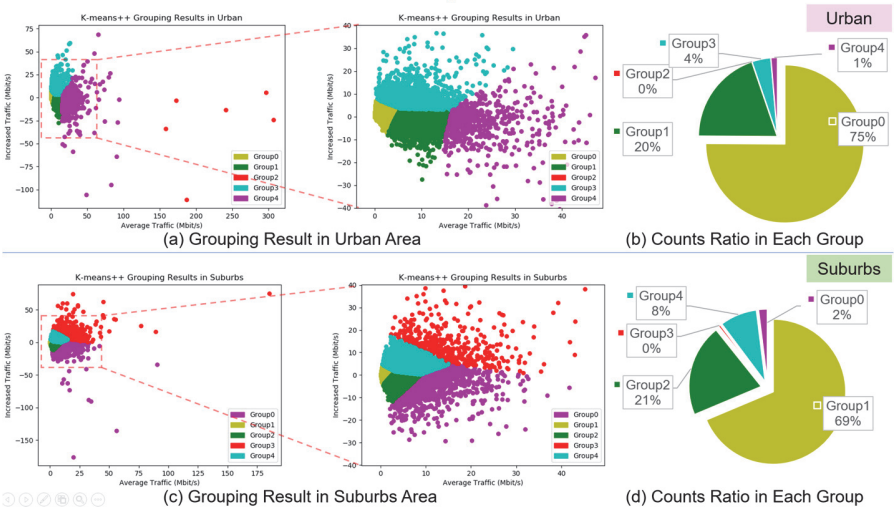

Fig. 2. Clustering analysis of upstream traffic in urban and suburbs. (Data migrate due to business reasons)

We collected 32 weeks traffic data of more than 100 thousand PON interfaces in a big city, and obtained the average traffic and traffic growth of each PON interface through regression analysis. Different from the traditional statistics, we then introduce the kmeans clustering algorithm[6] to analyze the characteristics of upstream traffic data, and cluster

the upstream traffic data of PON interface in urban and suburbs areas respectively like Fig. 2 shows (the downstream traffic can also be analyzed by similar methods)[7]. we can get the number distribution of PON interfaces in different groups and the group ID of each PON interface. Based on these detail information, operators can identify PON interfaces with large traffic fluctuation (such as Group2 in Fig2 (a), which includs only 6 PON), and can also judge the traffic pressure of the network from different levels.

2.3 Mining potential Gigabit access subscribers

Subscribers who upgrade to higher speed packages can bring more revenue. CCES can screen out high traffic subscribers according to their traffic histroy, and then locate and check their PON interfaces and the OLTs' info. If the OLT interfaces are already 10G PON, the subscribers can easily upgrade to higher speed without raise the central office cost. The specific process is shown in Fig. 3. Network designers can also set upgrade rules, and find those PONs need to transfer to higher speed PONs.

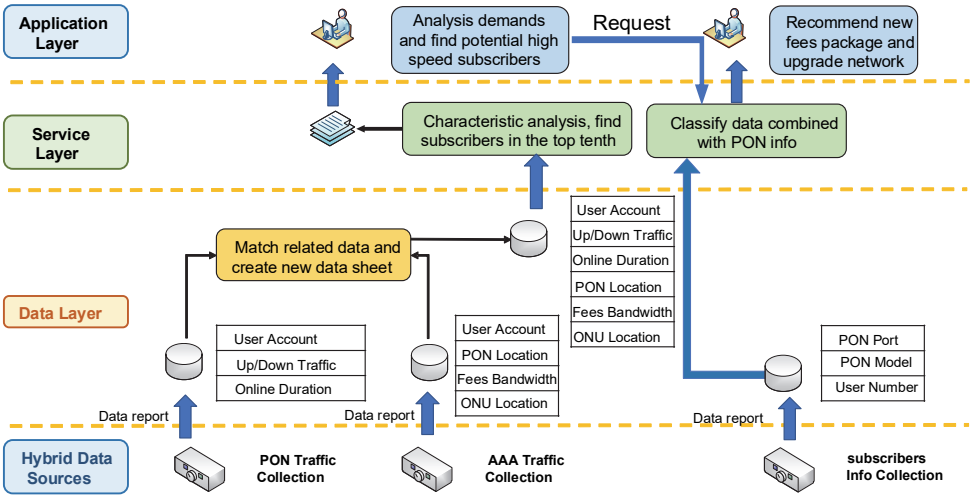

Fig. 3. The process to find potential Gigabit access subscribers

3. Innovation

This demonstration integrates data from different systems such as network management system, resource management system and IPTV system, modularizes functions such as data preprocessing and masking, and enables designers to input data analysis rules into the system. It classifies devices and users of access network through classification training, obtains traffic growth trend through regression analysis. User can locate the precise PON interfaces which need to update, and get more economical and forward-looking network construction plan.

4. OFC Relevance

In this demonstration, we propose an access network capacity assessment tool to solve the problem of hybrid data sources. Its unique data integration and analysis capabilities can provide equipment vendors and network operators with suggestions for equipment improvement and network construction. These are very important for OFC users. For interactive demonstration, we have prepared several specific demonstration cases, which can show the specific examples of user analysis and network hierarchical expansion. We hope that this demonstration can attract more developers to provide more functional modules of service layer, and we hope that with the help of this data analysis system, we can promote the standardization of network operation and maintenance data, so as to realize the interconnection of operation and maintenance data in the future.

References

[1] "The 44th statistical report on the development of Internet in China", http://www.cnnic.net.cn/ [acc Nov 14th, 2019]
[2] E. Harstead, *et al*., "Technology Roadmap for Time-Division Multiplexed Passive Optical Networks (TDM PONs)". JLT, 37(2): p. 657-664, 2019
[3] V. Houtsma, *et al*., "Optical Strategies for Economical Next Generation 50 and 100G PON", Proc OFC 2019 M2B1
[4] L., Shi, *et al*, "Energy-Efficient Long-Reach Passive Optical Network: A Network Planning Approach Based on User Behaviors," IEEE Syst. J., v4, no.4, pp. 449-457, 2010
[5] R., Sanchez, *et al,* "Network planning for dual residential- business exploitation of next-generation passive optical networks to provide symmetrical 1 Gb/s services," *JOCN*, vol. 8, no. 4, pp. 249-262, 2016.
[6] G. Varoquaux , *et al*, "K-Means clustering", https://www.lfd.uci.edu/~gohlke/pythonlibs/ [acc Nov 14th, 2019]
[7] S. Jinglei, *et al,* "Analysis of Access Network Requirements Based on Similar Service Feature Values Clustering", AOPC2019.