Multi-mode Dynamic Aggregation Strategy for Mobile Edge Resources

Weihong Wang\textsuperscript{a} Fuxiang Zhou\textsuperscript{b} Qian Feng\textsuperscript{c} Yuhui Cao\textsuperscript{d} Hongyan Lv\textsuperscript{e}
Hebei University of Economics and Business
\textsuperscript{a}wangwhs@163.com 18131112668, Hebei. China
\textsuperscript{b}zhouFx_08@163.com 18330167149, Hebei. China
\textsuperscript{c}15733100624@163.com 15733100624, Hebei. China
\textsuperscript{d}412503238@qq.com 18131112669, Hebei. China
\textsuperscript{e}981220135@qq.com 15227193192, Hebei. China

ABSTRACT: In view of the problem of network service delay and low resources utilization caused by frequent resources failure in mobile network service processing, this paper proposed a multi-mode dynamic aggregation strategy for mobile edge resources (MDDAS). Firstly, the definitions and expressions of mobile edge resources were given in the paper by the form of resource vectors, and a two-layer resource aggregation model was proposed. After that, the parameters and operators required for resources aggregation are defined. Then, the resource aggregation schemes and the construction of the resources view are described by applying definition formulas and calculation parameters. Finally, it is verified by experiments that the MDDAS achieves the purpose of reducing the processing delay of network services and improving the utilization of network resources to a certain extent.

CCS Concepts
- Computer systems organization → Architectures → Distributed architectures → Client-server architectures.

1. INTRODUCTION
In order to realize the timely and effective processing of mobile network services, the mobile computing model is moving from the center to the edge of the network. However, the dynamic, disorder, and local characteristics of the mobile network cause the effectiveness of the acquired mobile network resources to be seriously degraded, so that the mobile service processing delay is still long. In particular, the number of mobile edge resources has increased sharply, the problem of low utilization rate have become increasingly prominent. To this end, there is an urgent need for a resource aggregation method to facilitate the management and operation of mobile edge resources. In view of the fact that the existing resource aggregation methods are mostly applied to fixed networks, this paper proposes the MDDAS. It aims to improve the utilization of network mobile edge resources, facilitate the scheduling and sharing of resources, and reduce the processing delay of mobile services as much as possible.

Mobile edge resource aggregation is a method to improve the utilization of mobile network resources, balance network load, reduce the frequency of resource transmission in the network, facilitate resource management scheduling, and make scheduling decisions by the resource manager. The paper is organized as follows: In Section 2 of this paper, we describes relevant research in the field. In Section 3, the definitions and expressions of mobile edge resources were given in the paper by the form of resource vectors, and a two-layer resource aggregation model was proposed. In Section 4, the parameters and
operators required for resources aggregation are defined, and then the resource aggregation schemes and the construction of the resources view are described by applying definition formulas and calculation parameters. In Section 5, the evaluation indicators are defined, and the availability of the MDDAS is proved by experiments. Finally, the summary and outlook of this paper are given.

2. DEFINITION OF RESOURCES

At present, due to the time-varying characteristics of resources in mobile networks, there are few types of research on resources aggregation, and most of them are for the relatively stable bandwidth resources. For example, P.Sharma etc. proposed an intelligent bandwidth aggregation method based on the mobile community [1], which aggregates dynamic resources into virtual resource pools to realize resources sharing and multiplexing. Literature [2-6] constructed different architectures or systems for bandwidth aggregation in a multi-radio network coverage environment. These two methods of mobile bandwidth resources aggregation have reached the goal of mobile resources aggregation to a certain extent, which provides ideas for the research of this paper.

The idea of [1] is to aggregate all available resources into a virtual space and deploy resources support services from this virtual resource space when users need them. This method of virtual aggregation has valid verification in other network resources aggregation. Y.Chu, etc. proposed a virtual service resource aggregation method based on dynamic pruning for personalized requirements [7], which can quickly meet the needs of personalized applications. In a mixed environment, F.Huang and D.Li proposed a virtual machine server aggregation algorithm based on hierarchical clustering [8], which can effectively aggregate widely distributed, heterogeneous and autonomous resources on the Internet. H.Zhan and so on show a multi-user scalable resource aggregation service virtual view that provides users with a set of distributed and autonomous virtual compute nodes to choose from [9].

Literature [2-6] conveys the idea of resources classification or hierarchical aggregation, which is effectively validated in general network resources aggregation. X.Zheng et al. proposed an unstructured peer-to-peer network resource aggregation model [10]. The core of the model is the regional layered aggregation algorithm, which reduces the communication consumption in the unstructured peer-to-peer network to a certain extent. Z.Li, etc. proposed a topology-level aggregation mechanism based on node splitting [11], to achieve visualization of resources aggregation. Literature [12] proposed an algorithm based on maximal group edge cloud resources aggregation, which realized resources allocation in units of small resource blocks, so as to reduce the global maximum communication delay. G.Yuan proposed a networked resources sharing architecture for heterogeneous multi-port aggregation [13], which ensures that the terminal nodes in the heterogeneous converged network can always transmit data concurrently through two links. It makes full use of the transmission capacity of the terminal and the cooperation ability of different standard networks to provide users with higher speed services.

Therefore, this paper proposes the MDDAS. The method combines the layering and virtualization technologies of resources aggregation.

3. DEFINITION OF RESOURCES

The study considers a single resource individual as a mobile resource subject, and defines a mobile network with \( N \) mobile resource subjects. The network divides \( L \) domains into \( D_j \) \((j=1, 2, \cdots, L)\) according to the resource subject domain distribution. Each resource subject \( i \) \((i=1, 2, \cdots, N)\) has the characteristics of computing power \( C_i \), storage capacity \( S_i \), time limit \( T_i \), and bandwidth \( B_i \). The resource subject \( i \) publishes its resource information in the form of a vector \( R_i = (C_i, S_i, T_i, B_i, \cdots) \), where the free available resources are represented by \( FR_i = (FC_i, FSi, FT_i, FB_i, \cdots) \), \( FR_i \leq R_i \). Collecting all resource subject vectors in each domain \( D_j \) and publishing them to the intra-domain resources aggregation centre site in the form of information matrix \( M_j \) and free information matrix \( FM_j \), Where \( |D_j| \) represents the number of resource subjects in the domain. This paper selects four characteristics to perform modelling experiments.
In order to avoid the impact of aggregation resources failure on tasks execution, this paper established a stable resources view. By extracting the same or similar resources in different domains into \( k \) classes, the resources are re-aggregated in each region and published in the resources view in the form of the class matrix \( CM_k \) (the \( CM_k \) is defined in the aggregation scheme). The resource aggregation model is divided into two layers, as shown in Figure 1.

\[
M_j = \begin{bmatrix}
m_{j1} & m_{j2} & \cdots & m_{jn}
m_{j1} & m_{j2} & \cdots & m_{jn}
\vdots & \vdots & \ddots & \vdots
m_{j1} & m_{j2} & \cdots & m_{jn}
\end{bmatrix},
\]

\[
FM_j = \begin{bmatrix}
f_{m_{j1}} & f_{m_{j2}} & \cdots & f_{m_{jn}}
f_{m_{j1}} & f_{m_{j2}} & \cdots & f_{m_{jn}}
\vdots & \vdots & \ddots & \vdots
f_{m_{j1}} & f_{m_{j2}} & \cdots & f_{m_{jn}}
\end{bmatrix},
\]

\[
FR_j = \begin{bmatrix}
FR_{j1}
FR_{j2}
\vdots
FR_{jn}
\end{bmatrix}
\]

(1)

Fig. 1. Mobile edge resource aggregation view model.

4. RESOURCES AGGREGATION

4.1 Definition of Aggregation Parameters

This section gives definitions of resource information parameters and resources aggregation operation operators required in the MDDAS, and points out the role of each part of different scenarios.

4.1.1 Computing power

The measurement unit of the computing power of the resource subject is measured in millions of instructions per second. By controlling the minimum, maximum, and addition operations of the information matrix, the computing power of the \( D_j \) domain can be controlled.

\[
FC_j^* = \min_{o \in |P|} FC_j, \quad FC_j^* = \max_{o \in |P|} FC_j, \quad FC_j^* = \sum_{i=1}^{o \in |P|} FC_i \quad \forall FC_j^*\quad (2)
\]

The minimum operator can be used to obtain the minimum effective computing power of the domain \( D_j \), ensuring that the assigned tasks can be completed if the specified resource fails. The maximum operator is used to obtain the maximum free computing power of a resource subject, and the additive operator is used to obtain the total free computing power of the domain. When the tasks can be distributed to different objects, the addition operator is used to find the appropriate domain.

4.1.2 Storage capacity

In order to control the overall storage capacity of the resource subject in the domain, the storage capacity of the resource subject in the \( D_j \) domain is aggregated by using the minimum, maximum, and addition operation. When the data can only be stored in a single subject, the minimum or maximum operators is used to determine whether the domain has storage capability. When a tasks can be split and stored into different resource subjects, the add operator is used to determine whether the domain has storage capability.

\[
FS_j^* = \min_{o \in |P|} FS_j, \quad FS_j^* = \max_{o \in |P|} FS_j, \quad FS_j^* = \sum_{i=1}^{o \in |P|} FS_i \quad \forall FS_j^*\quad (3)
\]
4.1.3 Bandwidth
The bandwidth resources of different resources entities cannot be simply superimposed to determine the rate of a resource scheduling and tasks transmission. The proposed resource aggregation strategy only considers the minimum and maximum operations to aggregate bandwidth resources in the domain.

\[ FB^*_j = \min_{\alpha \in \mathcal{D}_j} FB_\alpha, \quad FB^*_j = \max_{\alpha \in \mathcal{D}_j} FB_\alpha \quad \forall FB^*_j \quad (4) \]

4.1.4 Time limit
The time limit \( T \) of the resource subject is expressed in the form of a time interval set. Considering the variability of time limit, the timeliness of the resource subject is estimated by the evaluation of the frequency of the resources main operation and the trajectory estimation. The validity period of the time is defined by the start time \( ST \) and the end time \( ET \), and a Boolean operation is used to find the time period.

\[ FT^*_j = \min_{\alpha \in \mathcal{D}_j} FT_\alpha, \quad FT^*_j = \max_{\alpha \in \mathcal{D}_j} FT_\alpha, \quad \forall FT^*_j \quad (5) \]

As shown in Figure 2, assuming that each row represents the free time \( T_i \) of a resource subject in the time period, the shortest effective time \( \min FT_i \) in the \( D_j \) domain is the latest starting time point of all resource subjects in the domain, and the ending point is all resources in the domain. The earliest end time point in the subject, as shown in Figure a. Then the longest effective time \( \max FT_i \) in the \( D_j \) domain is the earliest starting time point of all resource subjects in the domain, and the ending point is the latest ending time point of all resource subjects in the domain, as shown in Figure b.

Fig. 2. Time limit aggregation example diagram

In any case, the list of parameters and aggregation operators defined above is just a different option that can be used by the aggregation plan. Other parameters and operations are defined, depending on the application and the scheduling algorithm used.

4.2 Resource Aggregation Strategy
This paper analyses the possible forms of resource requirements, and proposes three aggregation strategies. Then, intelligent scheduling method is used to form a hybrid resource aggregation model.

4.2.1 Intra-domain Monotonic Aggregation (IDMA)
In the IDMA scheme, all resource subjects in each domain are aggregated into an information vector. The formulation is expressed as follows (6), in which the matrix \( M_j \) has \( |D_j| \) rows are reduced to an aggregate information matrix \( mdM_j \) with only one row. The IDMA is mainly designed for the distributed demand execution tasks in the domain. The IDMA strategy is designed for the distributed demand execution tasks. It aims to reduce the frequency of resources cross-domain transmission through single-point aggregation.

\[ M_j = \begin{bmatrix} m_{1j} & m_{2j} & \cdots & m_{nj} \\ \vdots & \vdots & \ddots & \vdots \\ m_{1\alpha_j} & m_{2\alpha_j} & \cdots & m_{n\alpha_j} \end{bmatrix} = \begin{bmatrix} R_{i1} \\ R_{i2} \\ \vdots \\ R_{i\alpha} \end{bmatrix} \Rightarrow mdM_j = [R'] = (C', S', B', T', \cdots) \quad (6) \]

4.2.2 Intra-domain Clustering Aggregation (IDCA)
The IDCA technology aims to reduce the time consumption of resources scheduling by facilitating intra-domain resources lookup based on single-point aggregation. Each \( D_j \) domain is segmented into \( h_j \leq |D_j| \)
Intra-domain clusters. Appropriate clustering method is applied to each resource subject, and the objects with approximate resource characteristics are combined to form an intra-domain cluster, so that the aggregated information vector better represents the resource subject characteristics in the domain.

\[
M_j = \begin{bmatrix}
m_{j11} & m_{j12} & \ldots & m_{j1p} \\
m_{j21} & m_{j22} & \ldots & m_{j2p} \\
\vdots & \vdots & \ddots & \vdots \\
m_{jN1} & m_{jN2} & \ldots & m_{jNp}
\end{bmatrix} = \begin{bmatrix} R_{j1} \\ R_{j2} \\ \vdots \\ R_{jN} \end{bmatrix} \Rightarrow \mathbf{c}\mathbf{d}M_j = \begin{bmatrix} R'_{j1} \\ R'_{j2} \\ \vdots \\ R'_{jN} \end{bmatrix}
\]  

(7)

4.2.3 Cross-domain Aggregation (CDA)

In order to avoid the impact of aggregation resources failure on tasks execution, this paper re-aggregates the resources matrix \( M_j \) in each domain. The resources with the same or similar characteristics in different domains are classified into k-type aggregates, and are published to the resources view in the form of the class matrix \( CM_k \).

\[
CM_k = (C_{R_1}, C_{R_2}, \ldots, C_{R_m})^T
\]  

(8)

Where \( C_{R_m} = \{ j \in L, i \in N | M_j, R_i \} \), \( M_jR_i \) indicates that the resource subject is the \( i \)-th resource subject of the \( j \)-th domain, \( m (m>1) \) represents the total number of resource subjects in the class. The class matrix is identified by the minimum resource object \( \min C_{R_m} \) in the matrix. The size of the resources in the class matrix does not exceed \( \Delta R \), and the size of \( \Delta R \) is determined by the characteristics of the user population. The size of the resources in the class matrix does not exceed \( \Delta R \), and the size of \( \Delta R \) is determined by the characteristics of the user population. \( CM_k \Rightarrow \min C_{R_m} = (\min C_m, \min S_m, \min T_m, \min B_m, \cdots) \).

4.3 Resource Aggregation View

This section describes the construction process of the mobile edge resource aggregation view through the process of logical evolution resources aggregation and scheduling.

| Logical evolution | Semantic definition |
|-------------------|---------------------|
| \( (\exists y \in D_j ) \Leftrightarrow (\exists R, P, DC_j) \) | \( R, P, DC_j \): The resource information in the resource body \( i \) is sent in the form of a resource vector \( R_i \) to the \( DC_j \) controller of the domain in which the resource body is located. |
| \( \exists R, P, DC_j \) \( \Longleftrightarrow \exists M_j, U \) \( \rightarrow (s_i|s_j) \rightarrow (s_i|s_j) \) | \( R, U, M_j \): The intra-domain resource vectors form the primary resource domain matrix \( M_j \). |
| \( \exists M_j, U \) \( \rightarrow \exists F, M_j \) \( = \exists \left( (\exists C_1 \cup \exists C_2 \cup \exists C_3) \cup \exists P \right) \) | \( M_j, U \) \( \rightarrow \exists F, M_j \) : The primary resource domain matrix is sent to the central controller \( F \). |
| \( \exists M_j, U \) \( \rightarrow \exists F, M_j \) \( \rightarrow \exists C_1 \) | \( M_j \) \( \rightarrow \exists F, M_j \) : The primary resource domain matrix forms some class resource matrix \( C_1 \) by class. |
| \( \exists R, P, DC_j \) \( \rightarrow \exists M_j, U \) \( \rightarrow \exists F, M_j \) \( \rightarrow \exists C_1 \) | \( \exists R, P, DC_j \) \( \rightarrow \exists M_j, U \) \( \rightarrow \exists F, M_j \) : The resource information in the resource body \( i \) is sent in the form of a resource vector \( R_i \) to the \( DC_j \) controller of the domain in which the resource body is located. |
| \( \exists y \) | \( y \): The number of row vectors in \( M_j \) is greater than 2. |
| \( \exists y \) | \( y \): The number of row vectors in \( M_j \) is greater than 2. |
| \( \exists y \) | \( y \): Indicates that there are resources belong to the cluster aggregation matrix in the domain that meet the requirements. |
| \( \exists y \) | \( y \): Indicates that there are resources belong to the cluster aggregation matrix in the domain that meet the requirements. |

5. EXPERIMENT AND ANALYSIS

To evaluate the effectiveness of MDDAS, Stretch Factor (SF) is defined as an indicator of scheduling efficiency. SF is the ratio of the delay of working with the MDDAS to working without the MDDAS.

\[
SF_{\text{aggregation strategy}} = \frac{\text{Delay}_{\text{aggregation strategy}}}{\text{Delay}_{\text{no aggregation}}}
\]  

(9)

When the SF \( \leq 1 \), the aggregation strategy is valid. If the SF > 1, it indicates that the aggregation scheme does not improve the impact of dynamic resources change on tasks scheduling. On the contrary, the smaller the SF, the more stable the view formed by the MDDAS.

Then, according to the mobile network layout, the experimental network environment is divided into 7 domains. According to the mobile network layout, the experimental network environment is divided
into 7 domains. In order to facilitate the analysis and comparison of the experimental results, 500 and 1000 resource nodes were dynamically distributed to the network. By changing the number of requirements, collecting the time when the MDDAS is applied to perform the tasks, and the remaining amount of available resources over time to analyse the experimental results.

Fig. 3. The SF value of different aggregation methods as the number of demand increases.

From the results shown in Fig 3, with the increase of resource demand, the MDDAS has significantly reduced the delay of network service processing. It also reflects that the more resource subjects, the more stable the resource aggregation view, and the service delay tends to decrease.

Fig. 4. Resource utilization of different aggregation methods in different scenarios

The graph on the left shows the average of resources utilization under different demand levels. Another statistic is the utilization of resources at various times during the network work process. The Figure 4 show that the MDDAS proposed in this paper can achieve full utilization of resources to a certain extent. The stable resources view basically realizes the difficulty of effectively improving resources mobility for network service execution.

6. CONCLUSION

Mobile edge resources aggregation is a method to improve the utilization of mobile network resources, balance network load, and reduce the delay of resources transmission in the network. This paper proposed the MDDAS. Firstly, the definitions and expressions of mobile edge resources were given by the form of resource vectors, and a two-layer resource aggregation model was proposed. After that, the resource aggregation schemes and the construction of the resources view are described. Finally, proved the availability of the MDDAS. To some extent, it improves the utilization of mobile network resources, and reduce the delay of the network. In the next work, we will further optimize the mobile edge resource aggregation algorithm, reduce the time consumption and energy consumption of resources aggregation as much as possible, and improve network work efficiency.

ACKNOWLEDGMENTS

This work was financially supported by the Ministry of Education's Funding for Returning Students Studying Abroad Projects of China (No. C2015003042), the Hebei Provincial Natural Science Foundation Youth Project (No. F2015207009), the Hebei University of Economics and Business Key Research Project (No. 2014KYY09 & 2017KYZ11), the Key Teaching Research project of Hebei University of Economics and Business (No.2018JYZ06), and the Education technology Foundation of the Ministry of Education (No. 2017A01020).
REFERENCES
[1] Sharma P, Lee S J, Brassil J, et al. Distributed Channel Monitoring for Wireless Bandwidth Aggregation[J]. Lecture Notes in Computer Science, 2004, 3042:345-356.
[2] Chebrolu K, Rao R. Bandwidth aggregation for real-time applications in heterogeneous wireless networks[J]. IEEE Transactions on Mobile Computing, 2006, 5(4):388-403.
[3] Sharma P, Sharma P, Lee S, et al. Handheld Routers: Intelligent Bandwidth Aggregation for Mobile Collaborative Communities[C]// International Conference on Broadband Networks. IEEE Computer Society, 2004.
[4] Habak K, Harras K A, Youssef M. What Goes Around Comes Around: Mobile Bandwidth Sharing and Aggregation[C]// IEEE International Conference on Mobile Ad Hoc & Sensor Systems. IEEE Computer Society, 2015.
[5] Habak K, Harras K, Youssef M. OSCAR: A Deployable Adaptive Mobile Bandwidth Sharing and Aggregation System[C]// International Conference on Mobile & Ubiquitous Systems: Computing. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014.
[6] K. Pandi, H. Charaf, K. Fekete, Network bandwidth aggregation based mobile resources management, 2013.
[7] Chu H, Wang XZ, Wang ZJ, et al. Virtual Service Resource Integration Method for Personalized Requirements [J]. Chinese Journal of Computers, 2011, 34(12): 2370-2380.
[8] Huang F, Li D. A Virtual Machine Server Aggregation Algorithm Based on Hierarchical Clustering [J]. 2009.
[9] Zhan H, Kang L, Liu L, et al. Scalable Resource Aggregation Service of an ErlangOTP PaaS Platform[C]// IEEE International Symposium on Service Oriented System Engineering. IEEE, 2013.
[10] Xiao-Jian Z, Tong L I, Tie-Wei F U. Unstructured Peer-to-Peer Network Resource Aggregation Model [J]. Value Engineering, 2014.
[11] Li Z, Ji L, Huang R. Network Hierarchical Compression Algorithm Based on Splitting-Fusion of Non-Important Nodes, Journal of Computer-Aided Design & Computer Graphics, 2018.
[12] Zhu J, Wu J, Yan X, Edge Cloud Clustering Algorithm Based on Maximal Clique [J]. Computer Science, 2018, 45(4): 60-65.
[13] Yuan Q, Research and implementation of networked resources sharing mechanism for heterogeneous multi-port aggregation, Diss, 2017.