Multi-Agent Coalition Formation Tactic Based on Grid

Qi Xin
College of Computer and Communication Engineering, China University of Petroleum, Dongying Shandong, China
Email: qixin@hdpu.edu.cn

Chen Jianjian
College of Computer and Communication Engineering, China University of Petroleum, Dongying Shandong, China
Email: chenjj129@163.com

***Abstract*—Because of the dynamic and complex grid environment, the speed and stability of grid data transfer can't be guaranteed, which has become the "bottleneck" that restricts grid applications,Replica is also very hard to incarnate advantage. Multi-Agent, with virtue of saving network bandwidth, realizing load and uninstall, increasing application's, and affording , can solve problems in the transfer process primely. However, they are the key factors that multi-replica transfer task is transformed into multi-agent coalition formation, and the coalition formation tactic.This paper analyzes and researches the relationship between replica transmission and Multi-Agent coalition, puts forward a high efficient coalition formation tactic based on linear programming optimization methods of mathematical theory.**

***Index Terms*—grid;multi-agent; coalition formation tactic;linear programming**

I. INTRODUCTION

Data grid meets the demand for data-intensive tasks with good data sharing and collaboration capabilities, such as high-energy physics, climate modeling and so on. However, because of the dynamic and complex grid environment, node failures and unexpected changes in network occur frequently. So the speed and stability of grid data transfer can't be guaranteed, and it has become the "bottleneck" that restricts grid applications [1].

With introduction of Multi-Agent technology, it brings the virtue of saving network bandwidth, realizing load and uninstall, increasing application's robustness, and affording platform independence and so on, which can solves problems in the transmission process primely; namely, multi-replica transmission task is transformed into multi-agent coalition formation [2]. The resolved thought is shown in figure 1.

As illustrated in figure 1: user submits task request R, system locates the replica set A that data replicas the R requested are composed of; and then, adopting a certain coalition formation tactic to construct Multi-Agent coalition and leaguer set $AR = \{ar_1, ar_2, ..., ar_n\}$, and meanwhile, assigns the transmission task R to the every replica node; at last, completes the transmission task R via negotiation mechanism.

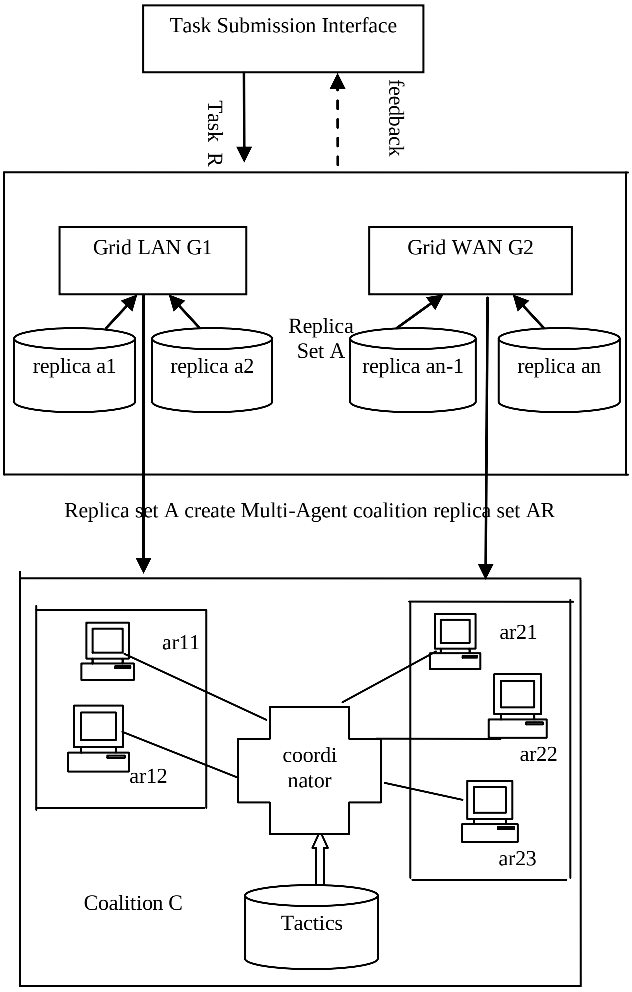

Figure 1. Multi-agent coalition engaging in transmission task

However, Multi-Agent technology brings a great deal of challenge, moreover, the most important question is

Fund Project: the project of CNPC supported by the Innovation Fund for young and middle-aged (07E1024)

doi:10.4304/jsw.5.1.36-43

the question of Multi-Agent coalition formation, which this paper focuses on. The rest of the paper proceeds as follows: In section 2, we analyze the key question for the process of Multi-Agent coalition formation in grid environment; And in section 3 we discuss the related work; Section 4 is the emphases in this paper, which introduces Multi-Agent coalition formation tactic; In section 5 we provide experiment result and analysis; and section 6 is conclusion.

II. DESCRIPTION OF MULTI-AGENT COALITION FORMATION

A. Definition of Multi-agent Coalition Structure

Taking charge of Multi-Replica transmission task, Multi-Agent coalition is defined as:

$$C=\langle AR,T,T_{max},V(c),U_c\rangle$$

Where:

1.Replica node set AR:

all of the replica agent in grid forms one set A: $A=\{a_1,a_2,\ldots,a_n\}, AR\subset A$.

Each Agent a_i (1<=i<=n) has its capacity set:

$$B_{ai}=\{b_1^{ai},b_2^{ai},\ldots,b_k^{ai}\}$$

Where b_j^a is measurement of capacity for Agent a_i completing a certain special task. Here, it mainly indicates the capacity of transferring data (Approximately, it is considered as agent node's static and dynamic information, such as OS, CPU, System Framework, Host's Location, or available HD, available Memory, Load, and bandwidth capacity and so on);

Faith set among agents is defined by ally set in this coalition structure. And then, Agent a_i ally set is defined as follows:

Definition one: Ally Set S_{ai}: describes symbiosis among agents in gird environment. S_{ai} is the set, including the agent with the times of cooperation between Agent a_i and the other agents exceeding a certain frequency (Let R_m be the frequency).

Where R_{ij} denotes cooperation degree between Agent a_i and Agent a_j; S_{ai} indicates the Agent ai ally set. And, when $R_{ij}>R_m$,then $S_{ai}\leftarrow\delta\sigma\ \{S_{ai}\cup a_j\}\wedge S_{aj}\leftarrow\{S_{aj}\cup a_i\}$. Every agent in AR set has respective ally set. Forming coalition set S by calculating R_{ij},simplifies replica set in forming coalition out and away.

2.Task Set T: The task set Coalition taking charge of, denotes m individual unresolved transmission task: $T=\{t_1,t_2,\ldots,t_m\}$;

3.Stated Time T_{max}: $T_{max}=\{t_{max1},t_{max2},\ldots,t_{max\,m}\}$, represents task's time limit.

4.Coalition Value V(c): The coalition C value is the value gained after the leaguer in coalition jointly completing a certain task, which is defined as utility:

Definition two: Utility denotes the profit gained by replica arj completing a certain task tx, which is defined as follows:

$\forall t_{ij}<t_{max\,i}$ (1<=x<=m), where $p_i(t_x)$ is the utility gained by completing task tx; let α be utility gene; function F( ) has the ability to measure stated time and node's capacity.

So, coalition value of coalition C is simply defined as:

$$V(c)=\sum p_i(t_x)-F1(C)\ (1<=x<=m)$$

Where F1(C) denotes the cost of coalition formation.

5. $U_c=\{u_1,u_2,\ldots,u_k\}$, indicates that V(c) is assigned to the leaguer in coalition.

B. Description of Issue for Multi-agent Coalition Formation

Assignment of multi-replica transmission task R is transformed into multi-agent coalition formation. Namely, the process of forming coalition, assigns multi-replica transmission task R dynamically, and the R is deployed to every agent node, while the coalition C is formed.

Usually, on the precondition of transmission time is less than stated time T_{max} and no value is more than coalition value V(c), the issue of coalition formation seeks one coalition structure CS satisfied with stability, CS={C1,C2,C3,…,Cg}, which is one of set AR's combination.

And then, using mathematic method describes and analyzes the above questions:

Definition of Set Variables:

Task Set: $T=\{t_1,t_2,\ldots,t_m\}, m>=1\wedge m\in Z^+$

Replica Node Set: $A=\{a_1,a_2,\ldots,a_n\}, n>1\wedge n\in Z^+$

Coalition Formation Replica Node Set: $AR=\{ar_1,ar_2,\ldots,ar_k\}$, $n>k>1\wedge k\in Z^+\wedge AR\subset A$

Stated Time:

$$T_{max}=\{t_{max1},t_{max2},\ldots,t_{maxm}\}, m>=1\wedge m\in Z^+$$

Utility gained by Coalition Nodes Completing Task: $Profit=\{p_1,p_2,\ldots,p_k\}$,

Description of Coalition Formation:

Suppose that the file is cut into m blocks $(t_1,t_2,\ldots,t_m)$, the m tasks are assigned to n replica nodes and gains the file data. And now, in the replica set A, each agent's ally set forms coalition formation set S, from which k nodes selected constitutes AR; and according to combinatorics, there are $CN=C_S^k$ combination modes.

Time matrix T_{matrix} is used to denote Transfer time for the task set T. where the number of rows m indicates m tasks; the number of columns denotes k replica nodes.

$$T_{matrix}=\begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1k} \\ t_{21} & t_{22} & \cdots & t_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ t_{m1} & t_{m2} & \cdots & t_{mk} \end{bmatrix}$$

Where t_{ij} $(1<=i<=m\wedge 1<=i<=k)$ refers as the time requested replica node j obtaining file task i;

Utility matrix P_{matrix} indicates utility value set for every task t_i (1<=i<=m), and where every replica node arj (1<=j<=k) will gain corresponding "reward" p_{ij}

(1<=x<=k) after completing stated transmission task. Where the number of rows m indicates m tasks; the number of columns denotes k replica nodes.

$$P_{matrix} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ p_{m1} & p_{m2} & \cdots & p_{mk} \end{bmatrix}$$

Where p_{ij} (1<=i<=m $\wedge$ 1<=i<=k) refers as the utility value gained by replica node j completing task i;

Similarly, cost matrix $Cost_{matrix}$ indicates cost set, and where replica ar_j (1<=j<=k) has to pay the cost $Cost_c$(1<=c<=k) in order to completing the t_i (1<=i<=m). Where the number of rows m indicates m tasks; the number of columns denotes k replica nodes.

$$Cost_{matrix} = \begin{bmatrix} cost_{11} & cost_{12} & \cdots & cost_{1k} \\ cost_{21} & cost_{22} & \cdots & cost_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ cost_{m1} & cost_{m2} & \cdots & cost_{mk} \end{bmatrix}$$

Let $cost_{ij}$(1<=i<=m $\wedge$ 1<=i<=k) be the cost value after replica node arj completing task t_i.

Each combination form one coalition structure CS, moreover, the optimal combination is the coalition structure CS_i satisfied with condition as follows:

$$\forall\, t_{ij}<t_{maxi} \wedge Max(\Sigma\, p_{ij})$$
$$\wedge\, Min(\Sigma\, cost_{ij}),1<=i<=m \wedge 1<=i<=k$$

Coalition structure C meeting the above condition, gets close to the optimal coalition structure,and has the maximal coalition value V(c). So, its objective function is defined as:

$$V(C) = Max(\sum p_{ij} - \sum \cos t_{ij})$$

Constrained Condition: $\forall t_{ij} < t_{\max i}$ (1)

And where 1<=i<=m,1<=i<=k

III. RELATED WORK

A. Research of Coalition Formation Method

Over the past few years, Multi-Agent systems with selfish agent become more and more important, and moreover, coalition formation is the key question for multi-agent systems.

More and more researchers take agent coalition structure formation arithmetic into account[3]:

Sandholm et al [4]have proved that they can gain the agent coalition structure whose utility value is $1/\lceil n/2 \rceil$ at least after in search of L1,L2, Ln in coalition structure chart. Using dynamic programming method to search Agent coalition.

Dang et al[5] can also get secondly optimal agent coalition structure.

Rothkopf et al can resolve the question of optimal coalition structure formation by using dynamic programming to solve the optimal combination auction.

Hu shanli et al[6] bring forward one kind of any-time coalition structure formation arithmetic through interlayer search. Using step-search method that the maximal coalition's potential is not less than $\lceil n(q-1)/4 \rceil$, where q from q=$\lceil (n+1)/4 \rceil$ to 2, Jennings et al can obtain the coalition structure which utility value is not less than 1/(2q-1) for the optimal utility. Hu Shanli et al study the optimal coalition structure formation through searching layers.

Liu Jinglei et al search the optimal coalition structure by adopting dynamic programming.

Boella[7] et al reduce coalition structure's search space based on the protocol of using capacity structure to assign task.

But, there are Sum(n)= $\sum_{i=1}^{n} S(n,i)$ coalition structure for n agent, where S(n, i) is the second Striling, Sum(n) is increased by exponential times.

Therefore, although arithmetic brought forward by Sandholm and Jennings et al are sure of gaining the optimal solution, its searching capacity is very gigantic. And although reduce searching capacity, arithmetic brought forward by Dang et al only assure that it get secondly optimal agent coalition structure. Dynamic programming of Rothkopf et al, include a lot of sub-issue repeat computing. Liu Jinglei et al have resolved repeat computing question.

B. Research in Optimization Method

Optimization methods (also known as operations research methods) formed over the past decades, which mainly use mathematical method to study all kinds of system's optimization approach, provides decision maker with basis of scientific decision.

Linear programming is a theory-integrity, method-maturity, application-extension and very important branch in optimization methods[8]. Linear programming mainly transforms practical problems into solving maximum and minimum of linear objective function with constraints for a set of linear inequality or equality; its most major solution is simplex method.

Nonlinear programming is the extension to linear programming, which doesn't contain constraints in the nonlinear programming model, and its typical methods include steepest descent method and quasi-Newton methods.

Constrained optimization method, mainly discusses optimal condition of constrained optimization problems, and namely, points that optimal solution of objective function and constraint function for optimal problems should meet these conditions such as necessary condition, sufficient condition and necessary sufficient condition.

Multi-objective optimization method, is that considers many aspects for practical problems, and at the same time, makes it possible for multiple objectives achieving the optimal objectives synchronously.

In 1951, dynamic programming brought forward mathematician R Bellman at al, is one kind of methods that can resolve optimization problems in the multistage decision process.

IV. DESCRIPTION OF MULTI-AGENT COALITION FORMATION TACTIC

At the beginning of coalition formation, we should initialize some variables:

All of R_{ij} ($i,j=1,2,\ldots,n;i=j \vee i \neq j$) in system should be initialized as R_m, and namely, every replica agent's ally set could contain all of agents, which can guarantee that the original system's candidate set includes all agents;

Users initialize the task and stated time for system: ($i,j=1,2,\ldots,n;i=j \vee i \neq j$). Where T is initialized as logic description of files requested; according to the description of files, users bring forward approximate T_{max};

Coalition utility value: V(c)=0.

Concrete process is divided into steps as follows:

Step 1 Create and Adjust Coalition Candidate Set S

(1) Every replica agent should exchange its ally set S_{ai} and capacity set B_{ai};

(2) Create coalition candidate set S: $S=S_{a1} \cup S_{a2} \cup \ldots \cup S_{ak}$

(3) Adjust coalition candidate set S

Predict completion time of forming coalition, if $T_j>t_{maxj} \wedge |S|<|A|$, then we should extend coalition candidate set S, and the concrete arithmetic is shown as follows:

```
FOR(aj∈{A-S})
{
   IF(Tj>tmaxj)
   {
//finding the aj with maximum Rjj from the left set
      S=S∪{aj};
      A=A-{aj};
//predict Tj;
   }
}
```

Step 2 Obtain Utility Matrix P_{matrix} and Cost Matrix $Cost_{matrix}$

(1) Utility Matrix: P_{matrix}

$\forall p_{ij}$ In matrix denotes the utility value obtained by agent ar_j in coalition C completing task t_i. Depended on the definition of coalition C's structure, p_{ij} can obtained utility value through $p_j(t_i)=\alpha *F(ar_j,t_{maxi})$.

Function F( ) involves two variables ar_j and t_{maxi}, now let us analyze the factors affecting function F( ) in these variables respectively.

There are tow part mainly affect The capacity set B_{arj} of agent ar_j: static factor and dynamic factor.

(a) Static Factor

CPU Processing Ability: ar_{jcpu}--For transmission performance of agent ar_j influence weight is w_{cpu};

Memory Size: $ar_{jmemory}$--For transmission performance of agent ar_j influence weight is $w_{mememory}$;

Network Interface Bandwidth: $ar_{jinterfaceband}$--For transmission performance of agent ar_j influence weight is $w_{interfaceband}$;

Host: Site: ar_{jsite}=0/1 1 stands for host and destination node are in the same LAN; but 0 is not -- For transmission performance of agent ar_j influence weight is w_{site};

(b) Dynamic Factor

Network Load: ar_{jload}-- For transmission performance of agent arj influence weight is w_{load};

Available Network Bandwidth: $ar_{javailband}$ ----- For transmission performance of agent ar_j influence weight is $w_{availband}$;

Where $w_{cpu}+w_{memory}+w_{interfaceband}+w_{site}+w_{load}+w_{availband}=1$.

Integrating Static and dynamic factors of agent ar_j, and t_{amxi}, we define function F( ) as:

$$F(ar_j,t_{maxi})=\delta\times(ar_{jcpu}\times w_{cpu}+ar_{jmemory}\times w_{memory}+ar_{jinterfaceband}\times w_{interfaceband}+ar_{jsite}\times w_{site}+ar_{jload}\times w_{load}+ar_{javailband}\times w_{availband})+\varepsilon\times(t_{maxi}-t_{ij}) \quad (2)$$

Where

$$V_j=(ar_{jcpu}\times w_{cpu}+ar_{jmemory}\times w_{memory}+ar_{jinterfaceband}\times w_{interfaceband}+ar_{jsite}\times w_{site}+ar_{jload}\times w_{load}+ar_{javailband}\times w_{availband});$$

δ and ε are the influencing factors of agent ar_j and stated time t_{maxi}; where δ influences transmission quality more, and transmission time more for ε; therefore, they are paradoxical , and the relationship is: $\delta+\varepsilon=1$. Their values should be fixed based on the task's demand, if transmission task requires real time more, ε should be higher; if requires transmission quality more, δ should be higher.

After obtaining the Function F( ), the Multiplication of α and F( ) can get utility value including t_{ij}: $\alpha*F(ar_j,t_{maxi}) =p_j(t_i)$.

(2) Cost Matrix: $Cost_{matrix}$

When the agent ar_j assigns and performs transmission task t_i, it mainly pays the cost, which includes that: earlier task assignment cost $Cost_{applyij}$ and performance cost $Cost_{runij}$. And then we will introduce them:

We suppose that earlier task assignment cost is changeless,and that is constant. Then,the difference among cost mainly is different from performance cost $Cost_{runij}$,and static and dynamic factors of agent ar_j are also the key factors for $Cost_{runij}$. However, there is difference: the weight of each influencing factor is different from the weight in F( ), and even has a cost gene β,which also determine cost:

$$Cost_{runij}=\beta\times V_j$$

So:

$$Cost_{ij}=Cost_{applyij}+Cost_{runij} \quad (3)$$

Step 3 Using linear programming to resolve optimal coalition C

According to formulas (1)(2)(3), we can define the objective function that gains the maximum coalition value as follows:

$$V(c)=Max[\sum_{1\le i\le m}\sum_{1\le j\le k}(\alpha\times(\delta-\beta)\times V_j+\alpha\times\varepsilon\times(t_{maxi}-t_{ij})-Cost_{applyij})]$$

$$\text{Constrained Condition: } \forall a_{ij}*t_{ij}<t_{maxi} \quad (4)$$

Next, we will resolve the objective function by using linear programming, and the concrete steps as follows:

Step One: Constitute Linear Programming Mathematical Model:

$$V(c)=Min\{-[\sum_{1\le i\le m}\sum_{1\le j\le k}(\alpha\times(\delta-\beta)\times V_j+\alpha\times\varepsilon\times(t_{maxi}-t_{ij})-Cost_{applyij})]\}$$

$$S.t: \forall a_{ij}*t_{ij}<t_{maxi}$$

Where $\alpha,\delta,\beta,V_j,\varepsilon,t_{maxi},a_{ij}$ and so on are known variables.

Step Two: With the introduction of relaxation factors $t_{xy}(m<x<m*k,k<y<m*k)$, we can transform the above model into standard linear programming:

$$V(c)=Min\{-[\sum_{1\le i\le m}\sum_{1\le j\le k}(\alpha\times(\delta-\beta)\times V_j+\alpha\times\varepsilon\times(t_{maxi}-t_{ij})-Cost_{applyij})]\}$$

$$s.t.\ a_{ij}*t_{ij}+txy=t_{maxi}$$
$$t_{ij}>0$$

Step Three: Firstly, using simplex method, we should obtain basic feasible solutions T satisfied with constrained condition:

Supposed:

$B=(P_1,P_2,\ldots,P_m)$ is the feasible solutions that column vectors of coefficient for basis variable $T_m=(t_1,t_2,\ldots,t_m)$ are constituted

$N=(P_{m+1},P_{m+2},\ldots,P_k)$ is the matrix that column vectors of coefficient for non basis variable $Tk=(t_{m+1},t_{m+2},\ldots,t_k)$ are constituted;

So the constrained equation: $a_{ij}*t_{ij}+t_{xy}=tmaxi$ can be expressed as $BN(T_m, T_k)^{-1}=T_{max}$;

Then, we can get the basis solutions: $T=(B^{-1}T_{max},0)^{-1}$

If basis solution is not existed, we can conclude that there is something wrong with constrained condition, turn step six;

Step Four: From the starting point of the basis solution, based on optimal and feasible conditions, we import non basis variable to replace a certain basis variable, and find another better feasible solution of the objective function value:

Let basic feasible solution T, finding the objective function value V(c) to $\{-[\sum_{1\le i\le m}\sum_{1\le j\le k}(\alpha\times(\delta-\beta)\times V_j+\alpha\times\varepsilon\times(t_{maxi}-t_{ij})-Cost_{applyij})]\}$.

And then, let Tm, replacing the objective function with non basis variable to the objective function, and that is:

$V(c)=\sigma*T_k$

Where $\sigma=(\sigma_{m+1},\sigma_{m+2},\ldots,\sigma_k)$ indicates test vector of non basis variable, whose component is called as test number. If each test number is less than 0, namely, $\sigma_s<=0$, well then, the current feasible solution is the optimal solution.

If σ includes test number more than 0, namely $\sigma>=0$, well then, the current feasible solution is not the optimal solution, and now we are required to search one new feasible solution, and the concrete practice is:

(1) Using maximum increase principle, confirming one exchange variable from non basis variable whose test number is positive, namely $\max\{\sigma_s\}$, we transform the non basis variable into basis variable (increase the value from 0 to positive number).

(2) Using minimum ratio principle, confirming one exchange variable from the primary basis variable, namely $\min\{\{B^{-1}T_{max}\}/B^{-1}P_x\}$, we transform the basis variable into non basis variable (increase the value from positive number to 0).

Step Five: Iterate as step four, we find the optimal solution satisfied with corresponding test number, and that is the optimal solution t* for problems.

Step Six: The process of iteration is end

V. EXPERIMENT RESULT

A.. Analysis and Comparison of Coalition Formation Tactic

(1) Time Complexity Analysis

N-P problems have been proved that m tasks are assigned to k replica nodes and gains the file data. Therefore, using the above method can also gain the optimal task assignment's solution approximatively. Well, then, we will analyze the time complexity through formation tactic's steps.

Step one: Creating candidate set S, whose time complexity is O(n), for the primary candidate set S including the most of replica nodes which have a ability to complete the task, and adjusting candidate set S, whose time complexity is O(1), So, the time complexity in step one is O(n)=O(n)+O(1);

Step Two: When calculating utility matrix P_{matrix} and cost matrix $Cost_{matrix}$, which are m*k matrix, the time complexity needed is O(2*m*k);

Step Three: Each iteration's time complexity is O(m2+k2+2*(k-m-1)) in the process of using optimization method, and suppose the number of iteration is Z, then, the time complexity is O(Z(m2+k2+2*(k-m-1))) for this process.

According to the above time complexity analysis, for the whole coalition formation tactic,the time complexity is:

O(n)+O(2*m*k)+O(Z*(m2+k2+2*(k-m-1)))

In the earlier stage, system's efficiency is low, however, with the renewal for ally set, it will reduce the time complexity greatly, and improve formation efficiency.

(2) Space Complexity Analysis

In order to test the coalition formation tactic proposed this paper, we realize coalition formation tactic through using java. We carry out 20 times experiments in number, and then average the searched space, moreover, compare the tactic proposed this paper with the other arithmetic and tactics, and the searched space required every arithmetic is shown in figure 2.

Depending on the time and space complexity analysis, our coalition formation tactic have good excellent performance.

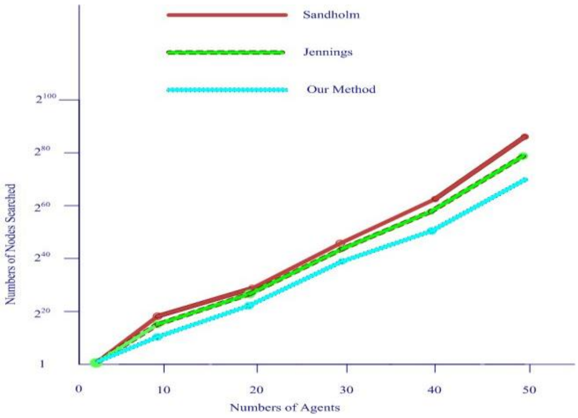

Figure 2. Space complexity analysis

B. Simulation Test Multi-agent Coalition

In order to test the performance of this multi-agent coalition, we need to realize the coalition formation tactic through java. For fairness, this program can generate experimental data randomly, nodes' performance data, experimental environment topological graph, transmission task R, and node's location and so on, and carry out 4 times test, and the result of experiment is shown in table 1:

As shown in table 1, the multi-agent coalition formed, whose leaguer meet the principle of excellent performance and the nodes in the same LAN in preference. Thus, compared to the other transmission models, the coalition like this will get higher transmission efficiency.

Experiment Topological Graph is shown in figure 3:

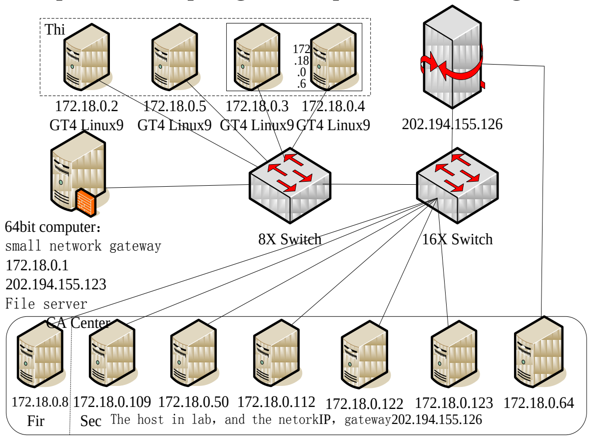

Figure 3. Experiment topological graph

Analysis:

(1) Because of the cost for replica location, coalition's transmission speed is lower than any other mode, when transferring small files, the size such as 50M;

(2) When the size of file is more than 100M, coalition will perform excellently for transferring big file, and improve transmission speed greatly. However, with continuing strengthening of file, the transmission speed will descend while the size of file is 1G, which may be caused by the huge task assigned.

TABLE I.
MULTI-AGENT COALITION TEST RESULT

NO	replica Nodes	Node Performance Ranking(from big to small)	Experiment Topological Graph	Task	Block	CS
1	No : 1-15	No3 → No1 No12 → No6 No8 → No10 No7 → No14 No4 → No9 No11 → No2 No5 → No15 No13	**G11:** No3 No7 No11 No15 No9 **G12:** No1 No13 No2 No14 No5 No8 No10 **G21:** G11 G12 No1 No4 No6	100M , D ∈ G11	10	No3 No7 No12 No6 No8 No10 No13 No4 No9 No15
2	No : 1-9	No1 → No5 No7 → No6 No8 → No3 No4 → No2 No9	**G11:** No2 No3 No6 **G21:** No1 No4 No5 No7 No8 No9	50M, D ∈ G21	5	No1 No5 No7 No6 No8 No3
3	No : 1-20	No1 → No13 No6 → No10 No18 → No7 No9 → No14 No4 → No19 No11 → No2 No5 → No15 No20 → No3 No8 → No17 No12 → No1 6	**G11:** No10 No14 No9 No8 No16 **G12:** No1 No6 No12 No11 No3 No17 **G13:** No 13 No5 No15 No4 No20 **G21:** No19 No18 No2 No7	150M , D ∈ G13	15	No13 No4 No5 No15 No20 No1 No6 No10 No18 No7 No9 No14 No4 No19 No11
4	No : 1-8	No2 → No4 No8 → No5 No3 → No1 No7 → No6	**G11:** No1 No3 No4 **G21:** No2 No6 No5 No7 No8	6M, D ∈ G11	2	No4 No3

VI. CONCLUSION

This paper has discussed that the problem of data transmission is transformed into the problem for multi-agent coalition formation, and applied the linear programming to the multi-agent coalition formation through the research of coalition formation tactics and optimization methods. As shown in experiment, the tactic

The Result of Experiment is shown in figure 4:

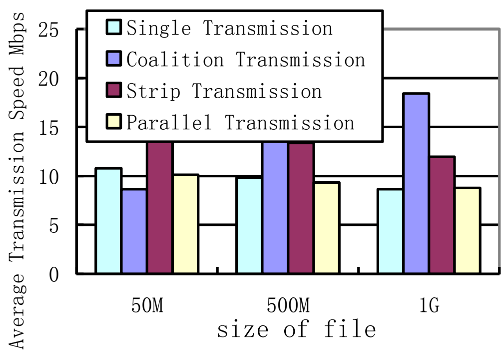

Figure 4. Comparison between coalition structure and other transfer modes

we proposed is able to resolve data transmission's task assignment excellently, and moreover, by analyzing this tactic, the efficiency of the tactic is better. So, the coalition formation tactic will have active significance to research of data transmission in the future.

APPENDIX A DATA GRID SIMULATION TOOLS OPTORSIM

The role of the grid simulator is a grid simulation environment, the people in this simulation environment is to examine various issues, such as the feasibility and performance issues. Through the configuration parameters can be more realistically simulate the actual situation in the various scenes, so that the authenticity of simulation results is more realistic; through the analysis of the results on the simulator, researchers can design continuous improvement. At present, there are many grid simulator[9],Including:Bricks,MicroGrid,SimGrid,ChicSim,EDGSim,GirdNet,OptorSim and so on.

It is a Grid simulator implemented by Java based on time , the user only needs is specified in the configuration file of the Grid topology and work list before the start of the simulation. Resources of different scheduling algorithm and agent scheduling algorithm can be setup parameters in one parameter configuration file, at the same time other parameters can also be specified , such as file transfer mode of the task, initialization of the document distribution, processing time and the degree of network congestion . After the simulation, there will be some statistical data output.

a. Compared with other simulation tools

Data Grid simulation tools is developed by the European Data Grid project team to verify the simulation of the copy of the data grid management technology . by setting the environment configuration file , describes the data grid system of environmental features, including topology, network bandwidth, system operating characteristics and job scheduling strategies; through the implement of a copy of the specific management strategies can simulate the data grid environment of the a copy of the creation, positioning, selection and management process consistency, and statistical data throughout the grid system in a number of performance parameters, including all the total execution time of the operating system, the copy number of the system is running, the visit times of the system for each document, operations execution time, the consumption of storage resources on the site, the times of remote access to the document on the site. The simulated environment can be very detailed description of the work process of the entire data grid management system in copy management system and get a number of performance parameters of copy management strategy in data grid environment.

TABLE 2 CONTRAST WITH OPTORSIM AND OTHER GRID SIMULATOR

Simulator	Visit Feature from users	Simulation in Grid		Information Services	
		Calculate	data	Focus	Distribution
ChicagoSim	√	√	√		√
EDGSim	√	√	√		√
GridSim		√		√	
GridNet			√		√
OptorSim	√	√	√	√	

Table 2 shows the results contrast with OptorSim and other Grid Simulator in different points of view. The analysis process is as follows:

(a) Simulation of Grid users: Simulation is mainly the user's access features (such as the arrival rate) to access the grid, and submit the job. Supporting for user's access features makes the simulation is closer to the real grid environment. The access features of individual users may have no law, but the overall access features probablly accord with a certain law. ("√" represents that the simulator has the characteristic, the same below). All the results show that the simulator has user's access characteristics.

(b) Grid virtual simulation: Most of the simulators use virtual applications, only a small number can use the real grid applications. Real grid application can make experimental results closer to the real situation, but it also needs more time to complete an application, that is because the application is true, but resources of the Grid is a virtual simulation of the limited physical resources, through the Theoretical calculation to take into consideration the time spent, so it is more suitable for a large number of simulation experiments.

(c) Besides the features listed in Table 6-1, optorSim also simulates job scheduling, resource management, data management, network environment in data grid, as it is the simulator mainly for European Data Grid project, so it has its own inadequacies. We will discuss the modification for it in detail below, in order to make it closer to the real data grid under virtual environment.

Therefore, the use of OptorSim simulator can simulate the data transfer model MATM process based on Multi-Agent in data grid.

b. Test configuration files

OptorSim has three configuration files: grid configuration file, job configuration files and Simulation Parameter File.

(a) grid configuration file

The grid configuration file describes the component of each site (CE number, CE site that is its computing power, SE number and its storage capacity) and the grid topology (network connections between two grid sites and the effective network bandwidth between two sites).

The grid configuration file designed is as shown in Figure5, as too much data, this figure gives only the GR1 configuration file, and Figure 6 shows the complete data grid topology.

In the grid configuration file, the first column represents computing power, Numerical, the larger the number, the higher the Computing capability; said that the calculation of the number of units the greater the computing power of the stronger; the second column as the number of memory cells, if zero, it means the grid site has no memory cell, it can only be used for calculating; The third is a memory cell storage, unit for the MB; the remaining as the effective bandwidth between sites for the Rf. Because this paper used the grid-based regional system model, therefore the network bandwidth belong to the same regional grid between site for the 10-100MB/s, and between different regions of the grid for the site 0-10MB / s. In the matrix of the storage network bandwidth, site with the site itself, the network bandwidth is zero. In addition, the choice of computing power and storage capacity of a strong in the region is as a copy of the server.

1 1 10000 0 100 100 100 100 100 100 100 100 100

1 1 10000 100 0 100 100 100 100 100 100 100 100

1 1 10000 100 100 0 100 100 100 100 100 100 100

2 1 20000 100 100 100 100 100 0 100 100 100 100

3 1 30000 100 100 100 100 100 100 100 100 100 0

Figure 5. Grid Configuration File

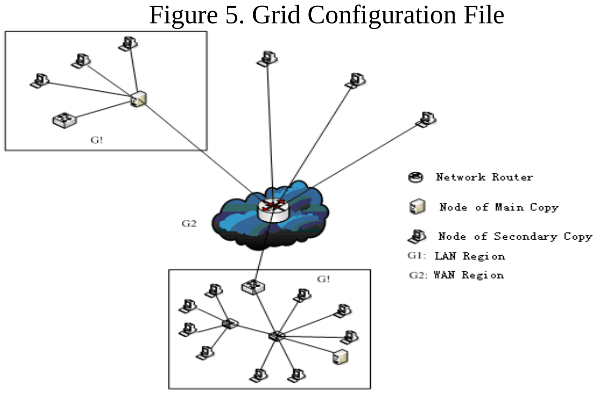

Figure 6. Chart of Grid Topology

(b) Job Configuration File

Job Configuration File,It contains the information of simulated job,which include the form of File Table(logic file name,size,Index No.), the form of Job Table(the logical file name that the job need to access), etc.

For a file, need for both logic name and physical file name, logical file name is the file's abstract name, it has nothing to do with the location of the file and the number of copies; but the physical file name refers to the actual location of the file, that is the same as the logic of the file name may have several physical file names, corresponding to multiple copies of this file.

In OptorSim,Job Configuration File has been built completely before the simulator run, including File Table, Job Table,Scheduling Table of the calculation unit, and the Job Option.

(c) Simulation Parameters File

Users can set a variety of the simulation parameters of the data grid with the Simulation Parameters File, such as grid configuration file, job configuration file, the selection of the simulation parameters file's path, job scheduling algorithm (including the research of coalition formation strategy),the design of the consultation mechanism, etc.

ACKNOWLEDGMENT

This research is supported by the Innovation Fund for young and middle-aged of CNPC under Grant No. 07E1024.

REFERENCE

[1] W. Allcock, A. Chervenak, I. Foster,C. Kesselman,C. Salisbury,S. Tuecke. The Data Grid:Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets. Journal of Network and Computer Applications

[2] J.M. Bradshaw. An introduction to software Agents. In J.M. Bradshaw, editor, Software Agents, pages3-46.A AAI Press/TheMITPress,1997

[3] Zlotkin Gilad, Rosenschein Jeffrey S. One,Two, Many: Coalition in Multi-agent Systems. Castelfranchi C, muller J,eds. From Reaction to Cognition, Lectuer Notes in Artificial Inteligence. Vol 95 7[M]. Berlin: Springer, 1993.96-110

[4] T Sandholm, K Larson, M Andersson, et a1. Coalition structure generation with worst case guarantees. Artificial Intelligence, 1999

[5] Nicholas R Jennings, Viet Dung Dang. Generation coalition structures with finite bound from optimal guarantees. AAMAS'04, New York, 2004

[6] Hu Shanli, Shi Chunyi. An anytime coalition structure generation algorithm. Journal of Software, 2001, 12(5):729—734(in Chinese)

[7] Guido Boella, Luigi Sauro, Leendert van der Torre. Reducing coalition structures via agreement specification. AAMAS'05, Utrecht, Netherlands, 2005

[8] F. H. Clarke, Optimization and Nonsmooth Analysis, John Wiley & Sons, New York, 1983

[9] WILLIAM H B, DAVID G C, LUIGI C, et al. OptorSim-agrid simulator for studying dynamic data replicationstrategies[J]. International Journal of High PerformanceComputing Applications, 2003, 10(3): 256-268

Qi Xin, Female, born in WuQiao County , HeBei Province, China. Date of birth: December 1978. Graduated from the China University of Petroleum (East China), Computer Application Technology, Master's degree. Who is currently teaching in the College of Computer and Communication Engineering, China University of Petroleum (East China), DongYing City in Shandong Province, China. The main research fields are grid computing, parallel architecture.

Chen JianJian, Male, born in LinyYi City, Shandong Province,China. Date of birth: January 1984. College of Computer and Communication Engineering Master's in Reading, China University of Petroleum (East China),DongYing City, Shandong Province, China. The main research direction is grid computing.