Yet Another Parallel Hypothesis Search for Inverse Entailment

Hiroyuki Nishiyama and Hayato Ohwada

Faculty of Sci. and Tech.,
Tokyo University of Science

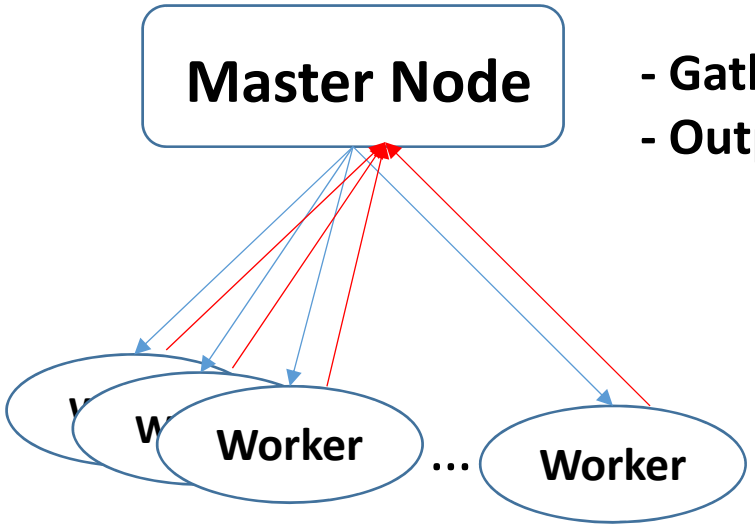
Background

Example of Distributed Computing Model: **MapReduce**

Map Step

Divide a Task
(input data)
into Subtasks

To Worker Nodes



Reduce Step

- Gather the results
- Output the gathered result

Each result

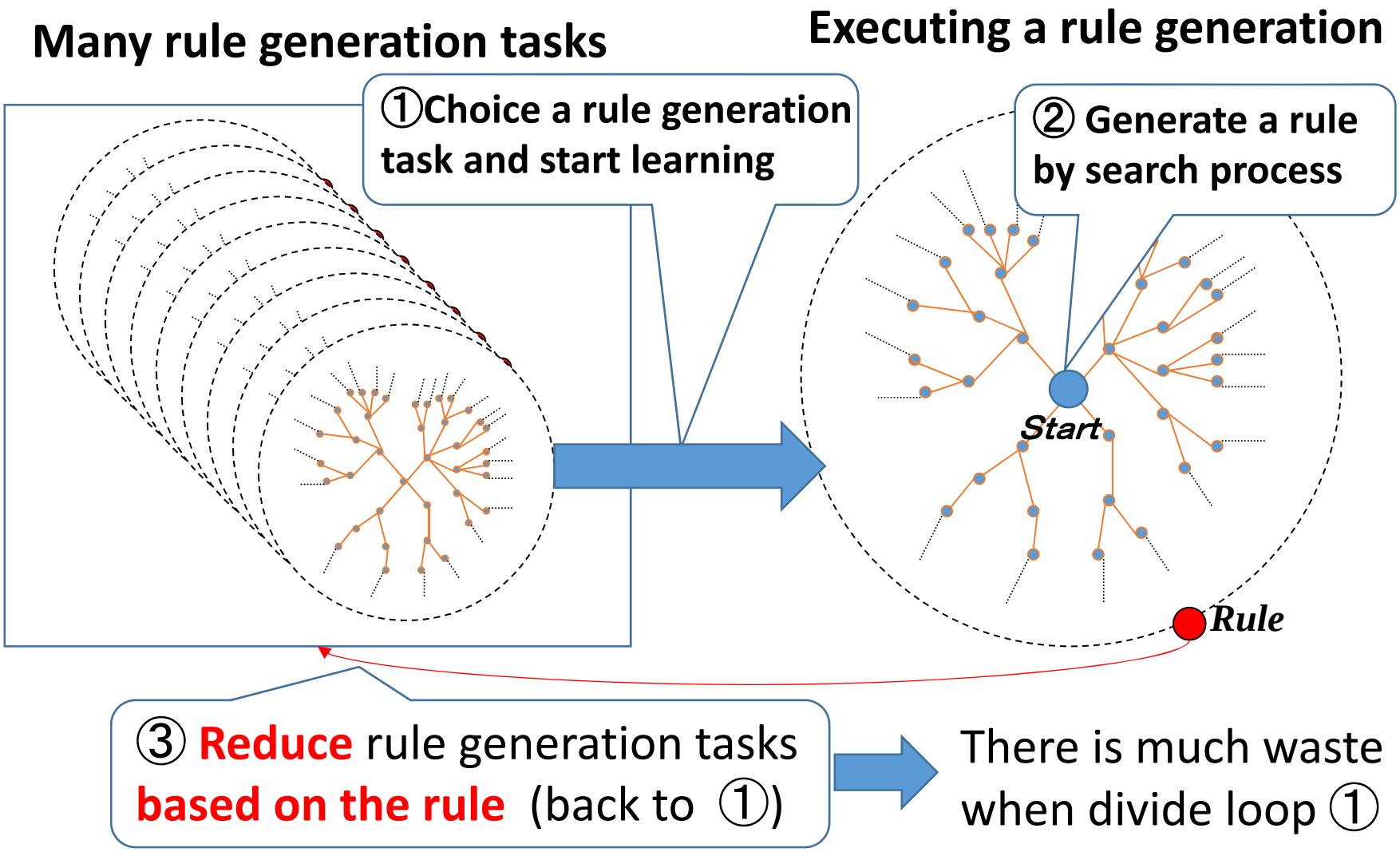
However…

MapReduce is effective when each subtask is full independent

Not suitable for Inductive Logic Programming

Flow of ILP（Inductive Logic Programming）

ILP has 2 repetition processing loops in learning



Our Purpose

Speed up ILP tool by parallel processing

We divide the search process for generate a rule

Design and Implement a Parallel Hypothesis Search System

- Divide the search area using Worker Nodes (CPUs)
- The Worker Node helps other Worker Nodes when finish own search task

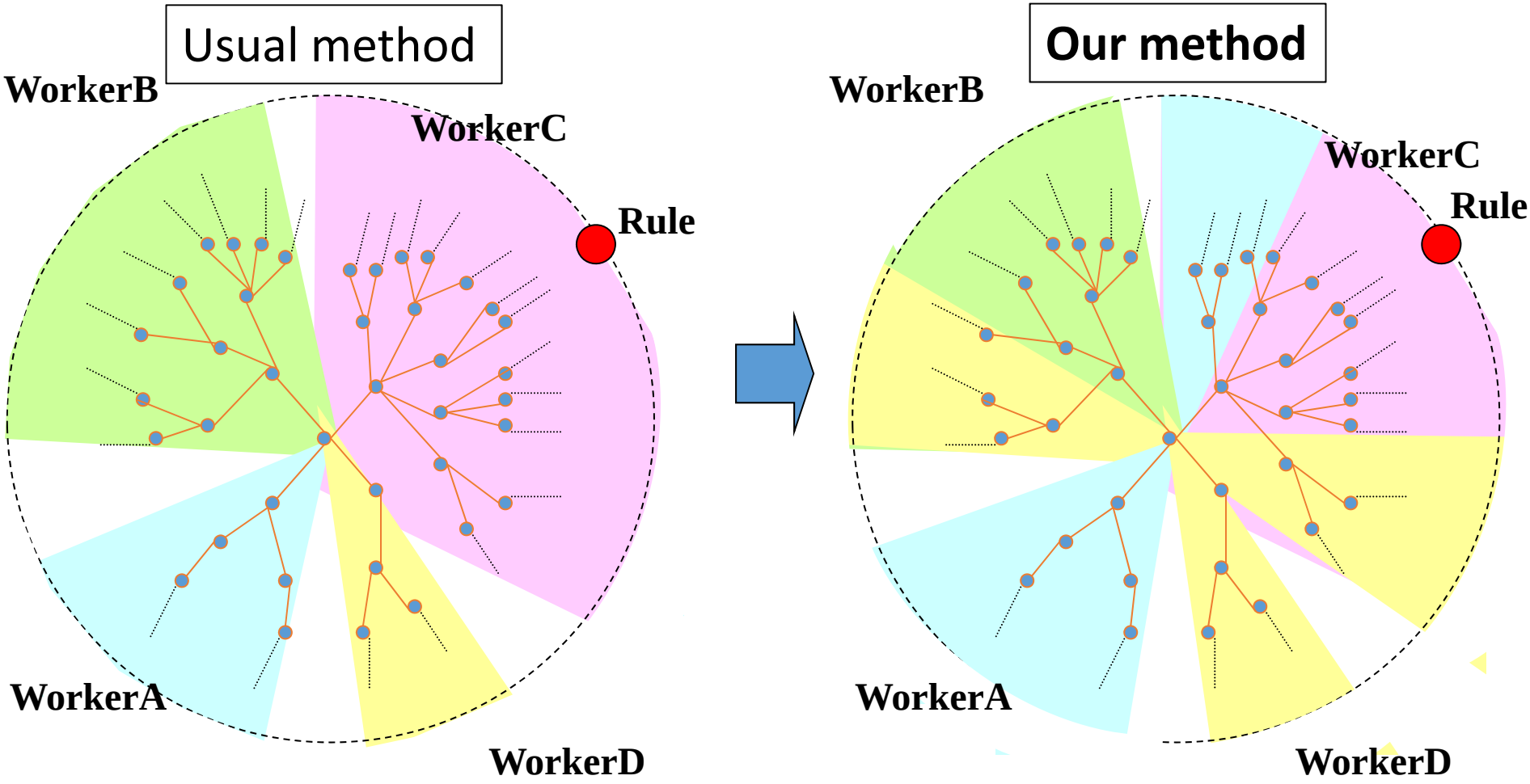
Realize effective parallel processing

Our Approach

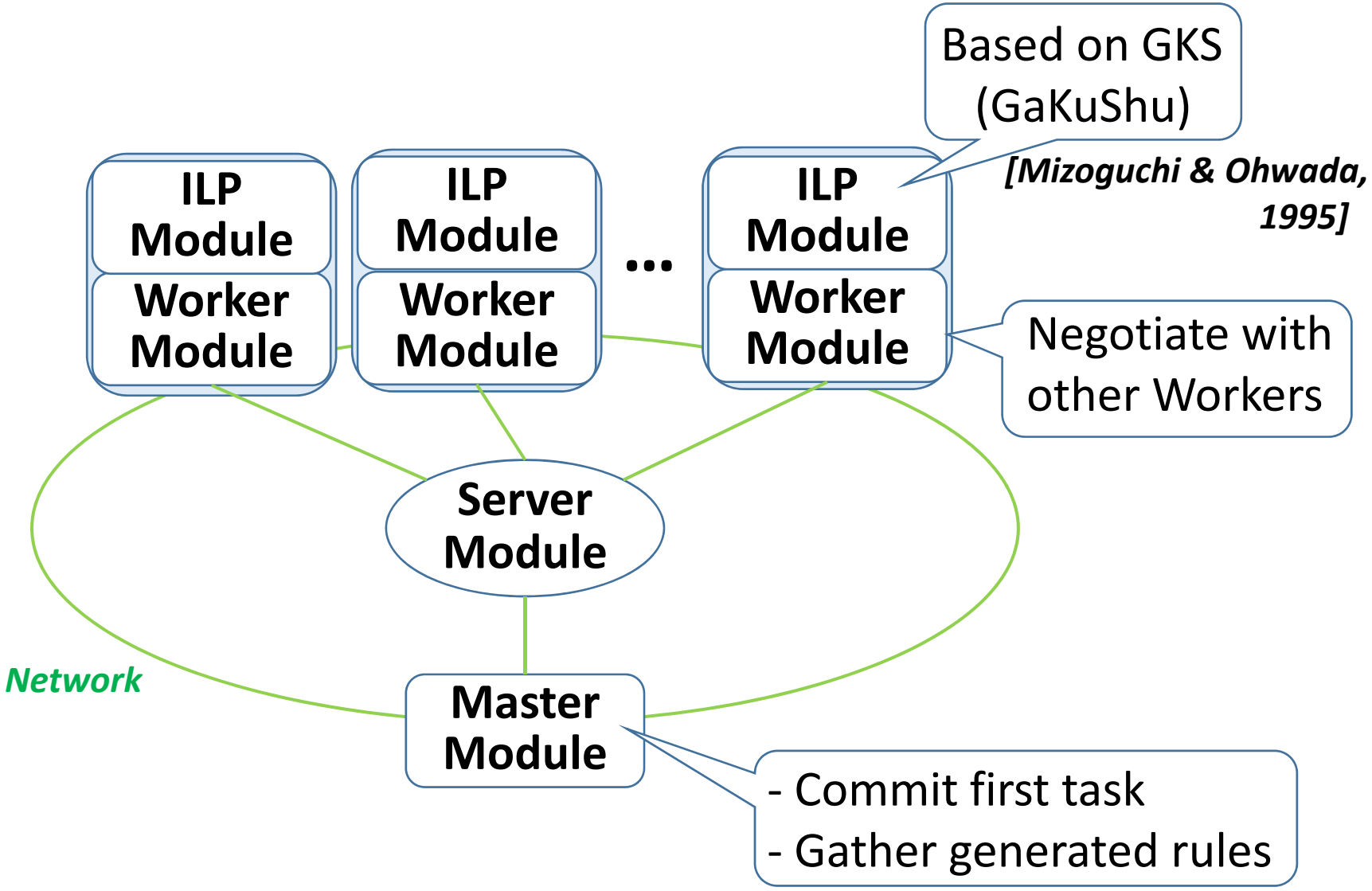
Usual method only assigns subtasks to workers

Our method

Each worker communicates with each other and requests (or accepts) a part of the subtask



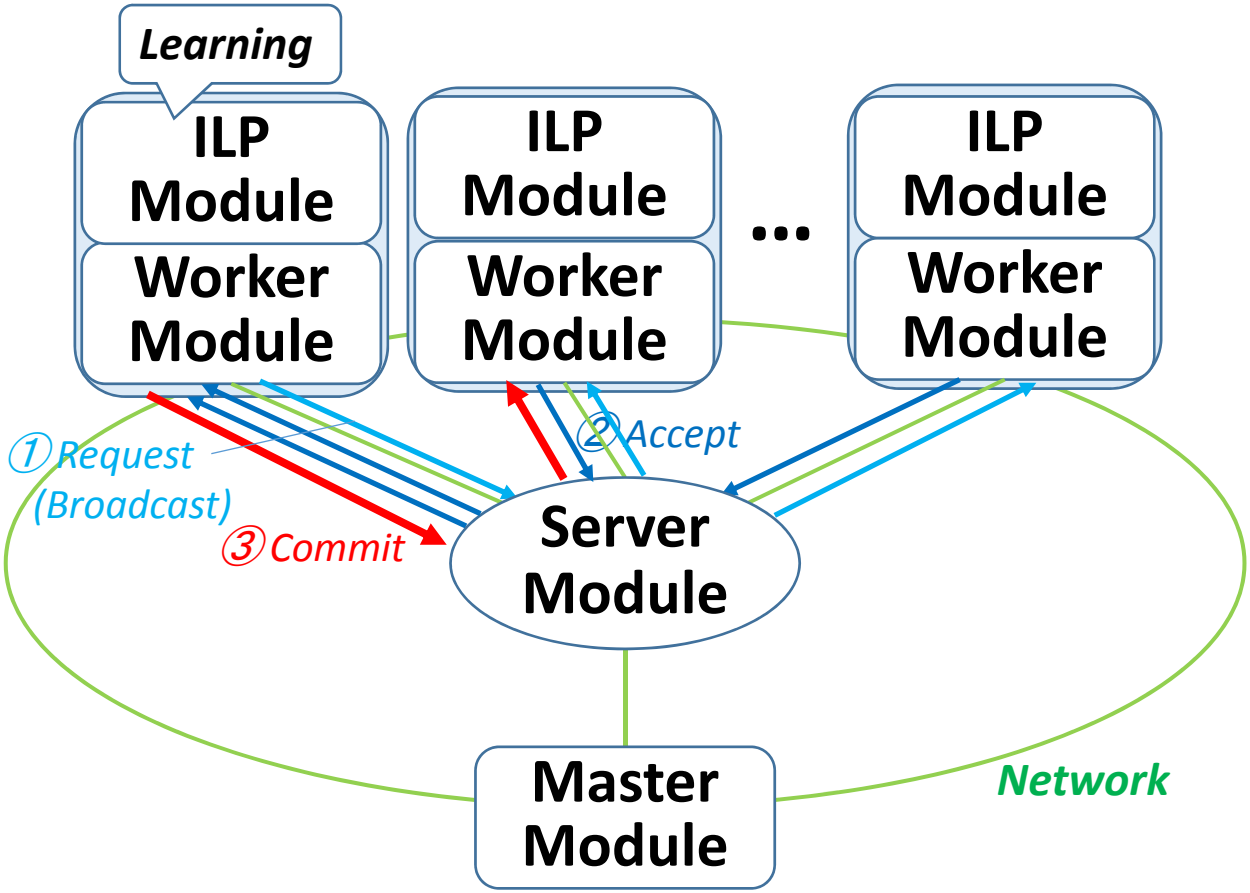
System Configuration



(All modules were implemented by Java language)

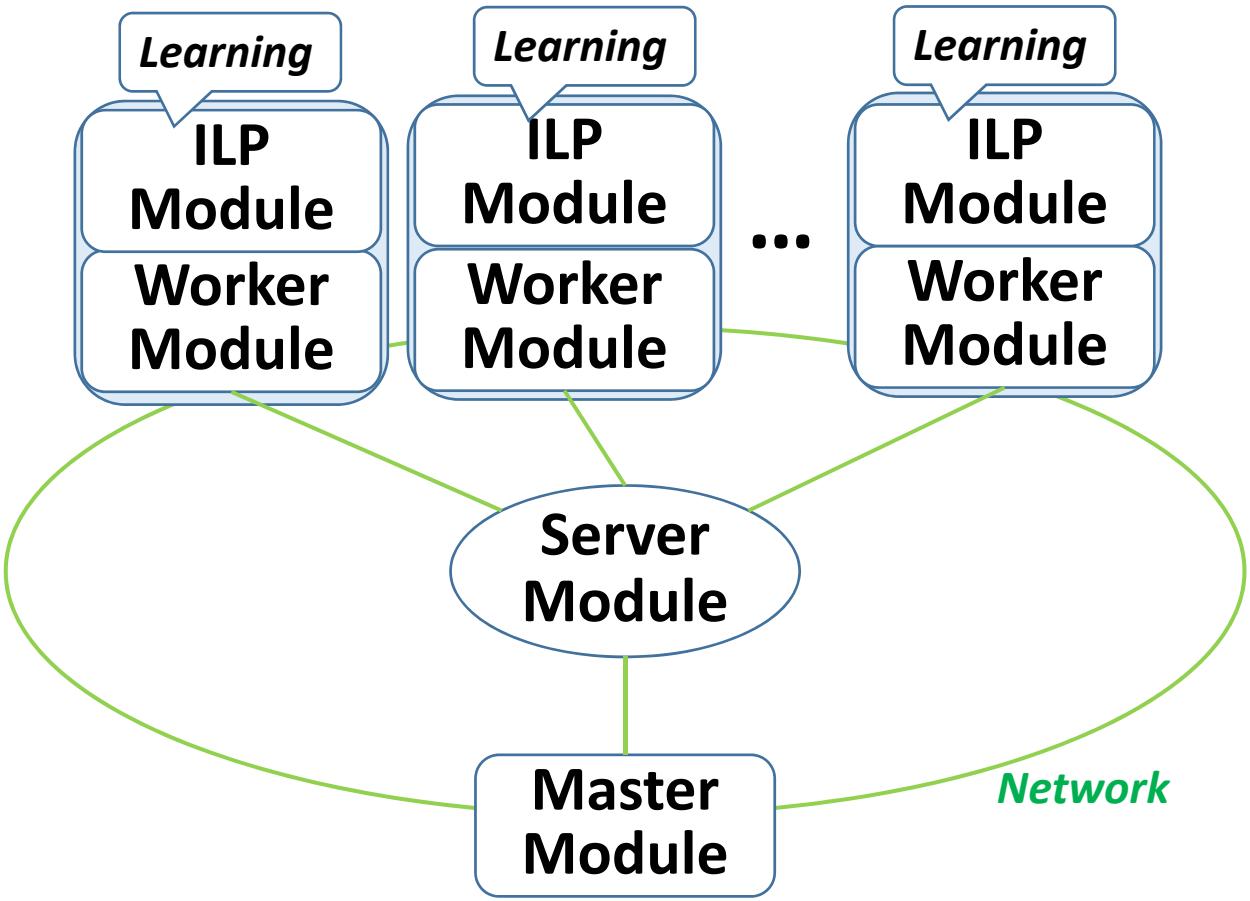
Flow of Negotiation to Divide a Task

① Learning worker **requests** a part of task to other all workers
② Non-learning worker **accepts** the requested task
③ Requesting worker **commits** to an accepting worker



State that all Workers are Assigned (Saturation of the Task)

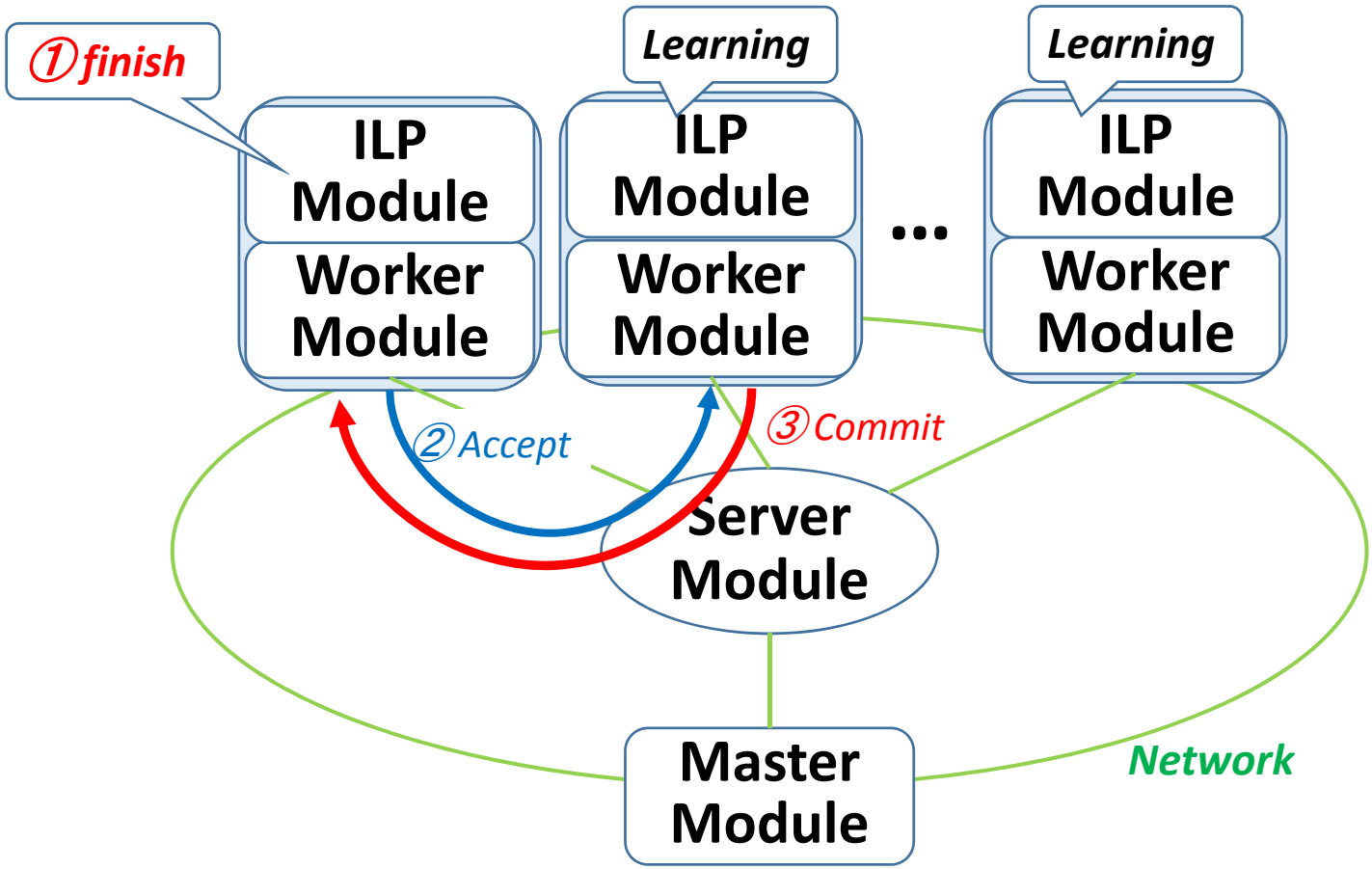
- All workers sent a request message to other all workers
- All workers received request messages from other all workers



State that all Workers are Assigned (Saturation of the Task)

When one worker's learning task is finished

The worker accepts and is committed



Experiment

Enviroment

2 PC (total 12 CPUs)
（Intel(R) Core(TM) i7-5820K CPU @ 3.30GHz 16.0GB 64bit）

Data

- **Drug Design for Human (Sample Problem)**

From “In Silico Screening of Zinc (II) Enzyme Inhibitors Using ILP”

[Ito, Ohwada et al., ILP2015] (Yesterday’s presentation)

Using 1 CPU: 724 sec.

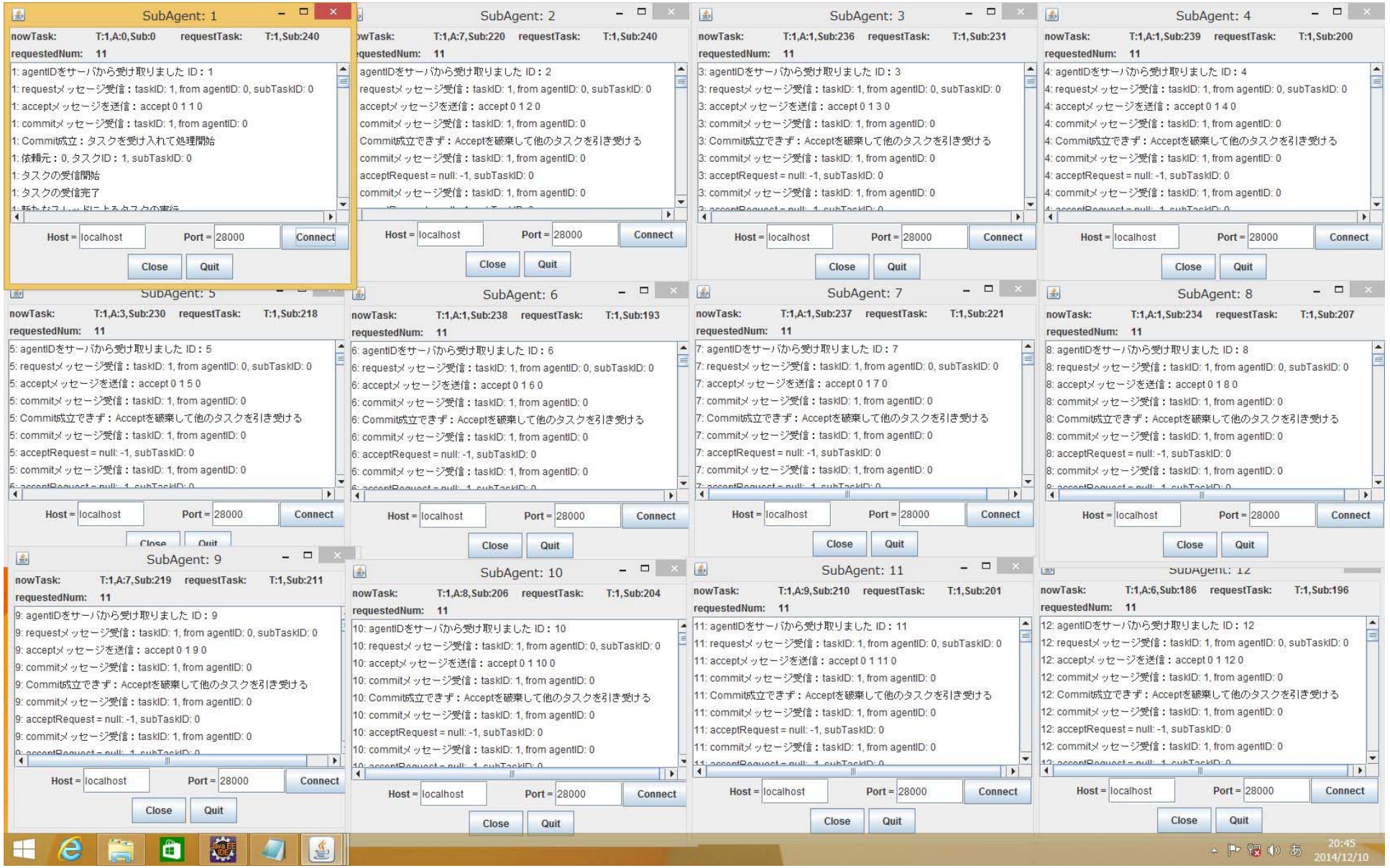
- **Drug Design for Plant (Large Scale Problem)**

From “Extracting the Common Structure of Compounds to Induce Plant Immunity Activation using ILP”

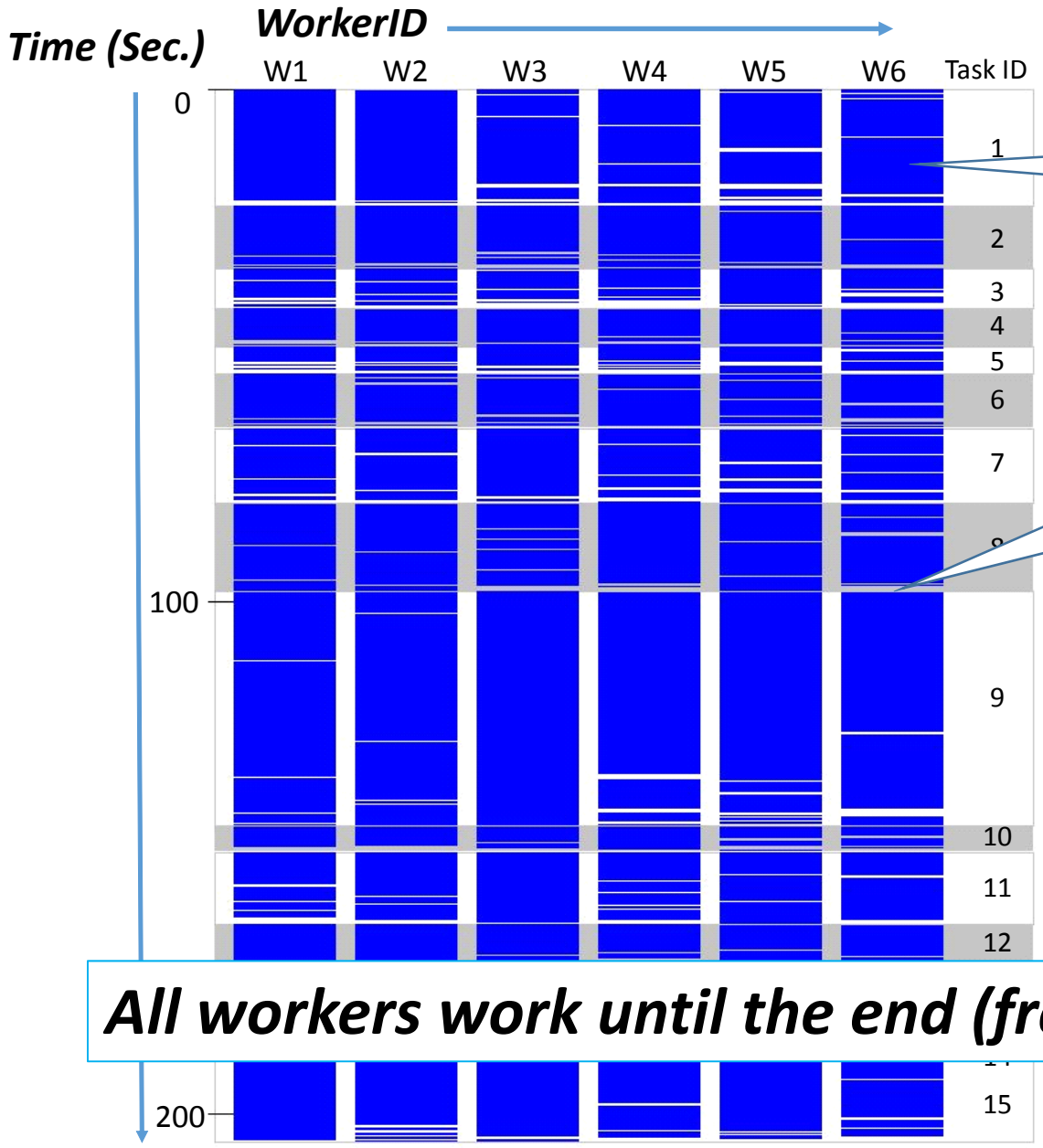
[Matsumoto, Ohwada et al., ILP2015] (Yesterday’s presentation)

Using 1 CPU: 56,372 sec. (15.66hours)

Main Display of Parallel Execution (12 workers)



Operational status of each worker in the parallel-processing experiment using six workers (6CPUs) for the Sample Problem



Blue area:
Learning

White area:
Communication or Waiting

Average learning time:

94.49 %

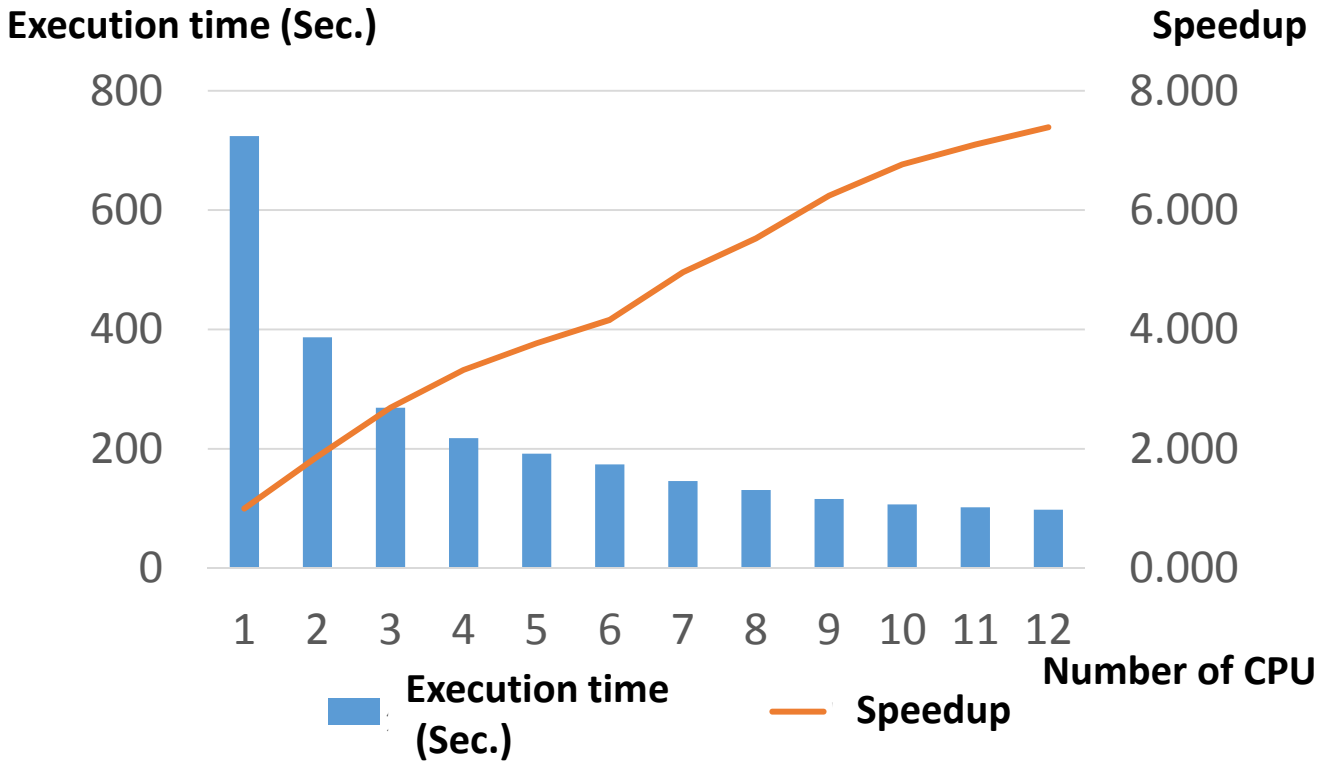
Average communication time:

3.07 %

All workers work until the end (free time is very short)

Experimental Results

Sample Problem(Drag Design for human) (12CPU: 6 CPU × 2 PC)



Large-scale problem (Drug Design for plant) (6CPU ×2 + 4CPU ×2)

Add PC

Number of CPU	Execution time (sec.)	Speedup
1	56372	1.000
12	5723	9.850
20	3563	15.821

Conclusion

Designed and Implemented a Parallel Hypothesis Search System

- Divide the search area using Worker Nodes (CPUs)
- The Worker Node helps other Worker Nodes when finish own search task

All workers work until the end (free time is very short)

Realize effective parallel processing