

Grid Computing Based Search Engine

Mr. Tejasvi Kattimani T R
Dept. Of ISE, GMIT

Mr. Veeragangadhara Swamy T M
Dept. Of ISE, GMIT

Mr. Amith Shekhar C
Dept. Of ISE, GMIT

Abstract— The idea intends to develop a “Grid Computing Based Search Engine” mechanism, where the processing of the Search Engine is distributed in the Gridlets in a network that performs search and consolidates the results to the Grid Server which is then forwarded to the System that issued the Search request.

Keywords— *Cryptography, Steganography, Confidential image, Stego image, Cover image.*

I. INTRODUCTION

Grid Computing is an enhanced version of distributed computing that involves coordinating and sharing computing, application, data, storage, or network resources across dynamic and geographically dispersed organizations. The goal of this paper is to achieve a developed system capable of searching vast amounts of geographically distributed information, which in many cases could not be searched effectively, if at all, by a centralized search.

By using a Search Engine with grid technology, the customer requirements can be satisfied by searching for the required data and retrieving the results in real time. At the same time, it improves the resource usage and provides efficient workload optimization. The system intends to develop the architectural design of a grid which can be used for the purposes of a search engine. This is an efficient method for searching a desired file and is faster when compared with other methods used for searching in a network. The search engine can then be used for searching a desired file in the network based on the file type.

II. SYSTEM DESIGN

Our idea is mainly for searching information, dispersed in the network. Our main aim is to build an architectural design of search engine which is cost effective and user friendly. Fig.1 shows the architecture. It shows how actually our search engine works. It also shows the interaction between the Grid Server, Client and the Gridlets. The main modules to be considered are i) Grid Server ii) Client iii) Gridlet iv) Search Engine.

Grid Server Module

The Grid Server must be started up initially. When the Grid Server module is executed the Server's User Interface is generated in which there is gridlet configuration option. The gridlets can be configured through the server module and gridlet configuration module or by directly editing the configuration detail file. This configuration file contains the domain names of the systems that have been chosen as gridlets.

The Grid server then waits for the search request from client and upon receiving the request, dispatches the search job to the selected gridlets. Upon receiving the search results from the gridlets, the Grid server has to aggregate the results and return it to the Client. In this module there are options for starting, stopping and quitting the Grid server. The user can stop the grid server or quit at his/her discretion.

Client Module

When this module is executed the Client User Interface is generated. First the users have to connect to the server. Then the user can type in the search string and the file type in the relevant text areas and then initiate the search. This search request will be sent to the Grid Server for processing and mean while the client waits for the result. In this module there are options for the client to connect and disconnect to the grid server so that user can connect or disconnect from the server at his/her discretion.

Gridlet Module

This module must be executed in all machines that are chosen as gridlets. This receives the Grid Server message “SEND CPU USAGE”. It then calculates its current system wide CPU usage and sends it to the grid server. If its CPU usage is less than 80%, it then receives the search request from the grid server.

Search Engine Module

This module searches for the file name and file type specified by the client in the gridlets and sends the results to the server. The search engine application must run in all the gridlets. After receiving the search job from the server, the gridlet has to search for the specified file.

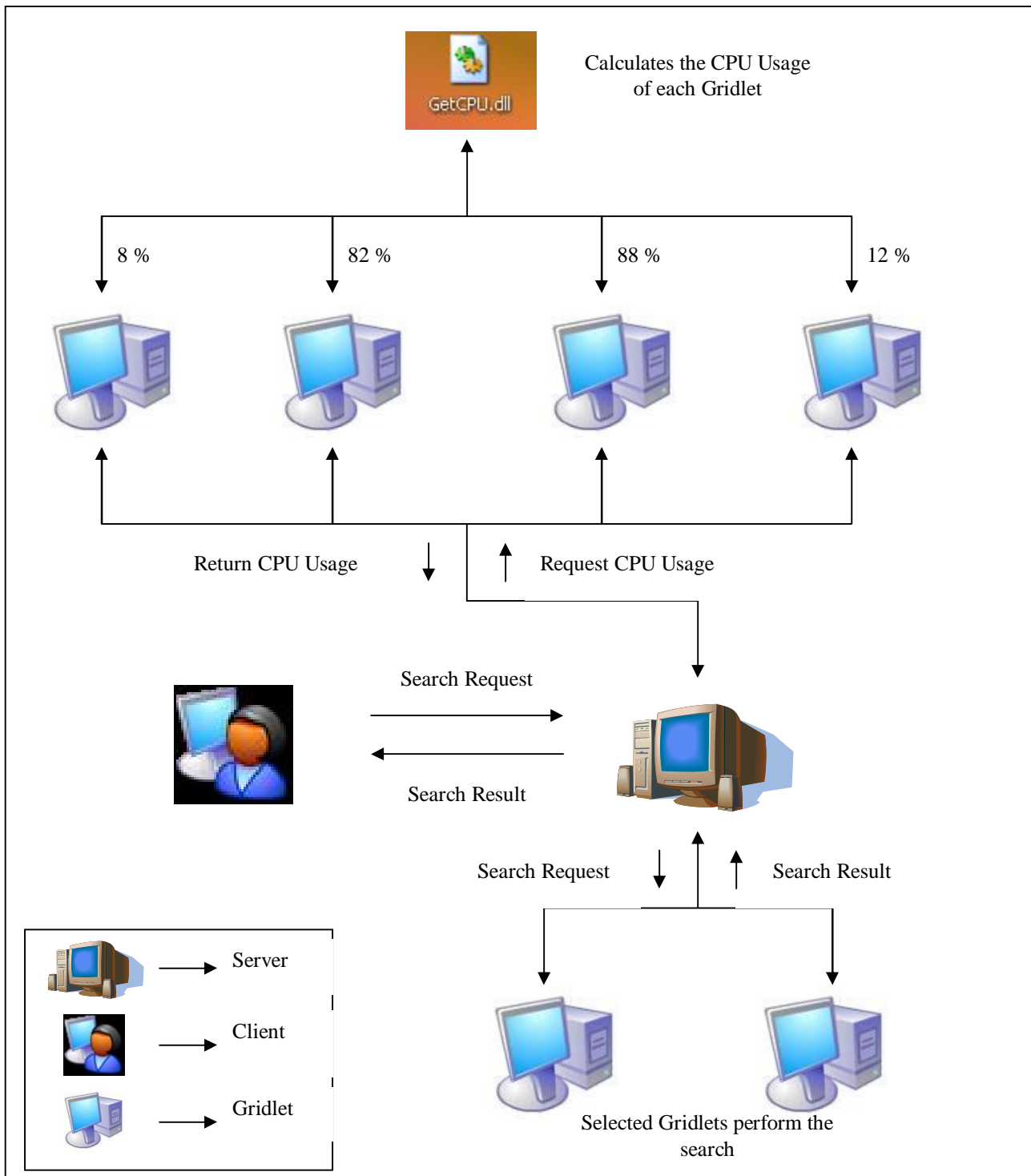


Fig. 1: Project Architecture

III. SERVICES

Services provided by the Grid server

- Receiving request from client
- Job Allocation

- Job Scheduling
- Job Dispatching
- Receiving results from gridlets
- Aggregation of the results
- Updation of log file
- Returning results to the client

The grid server first receives a search request from the client. It is responsible for allocating the job fragments to the various machines in the grid. In our concept the job is a search based on a file type. Job scheduling is done by the grid server by broadcasting a “SEND CPU USAGE” message to all the gridlets. It in turn receives the CPU usage of all the gridlets and based on that it does the job dispatching. Finally it gathers results from various gridlets, aggregates the results, updates the log file and forwards the results to the requested client.

Services provided by the Client

The client interface provides user interaction. It allows the user to get connected to the grid server and enter a desired file name and file type in the specified field. The user can then initiate the search by using appropriate buttons so that the search request is sent to the grid server. On receiving the results from the server, it displays the locations of the requested file in a systematic manner on the interface

Services provided by the Gridlet

The actual search is done by the gridlets. On receiving the “SEND CPU USAGE” message from the grid server, it invokes the .dll file GetCPU.dll which calculates the current system wide CPU usage. It then returns this to the server. Based on its CPU usage it receives the search request. On receiving the search request, it searches for the file and forwards the results to the grid server.

IV. DATA FLOW DIAGRAMS

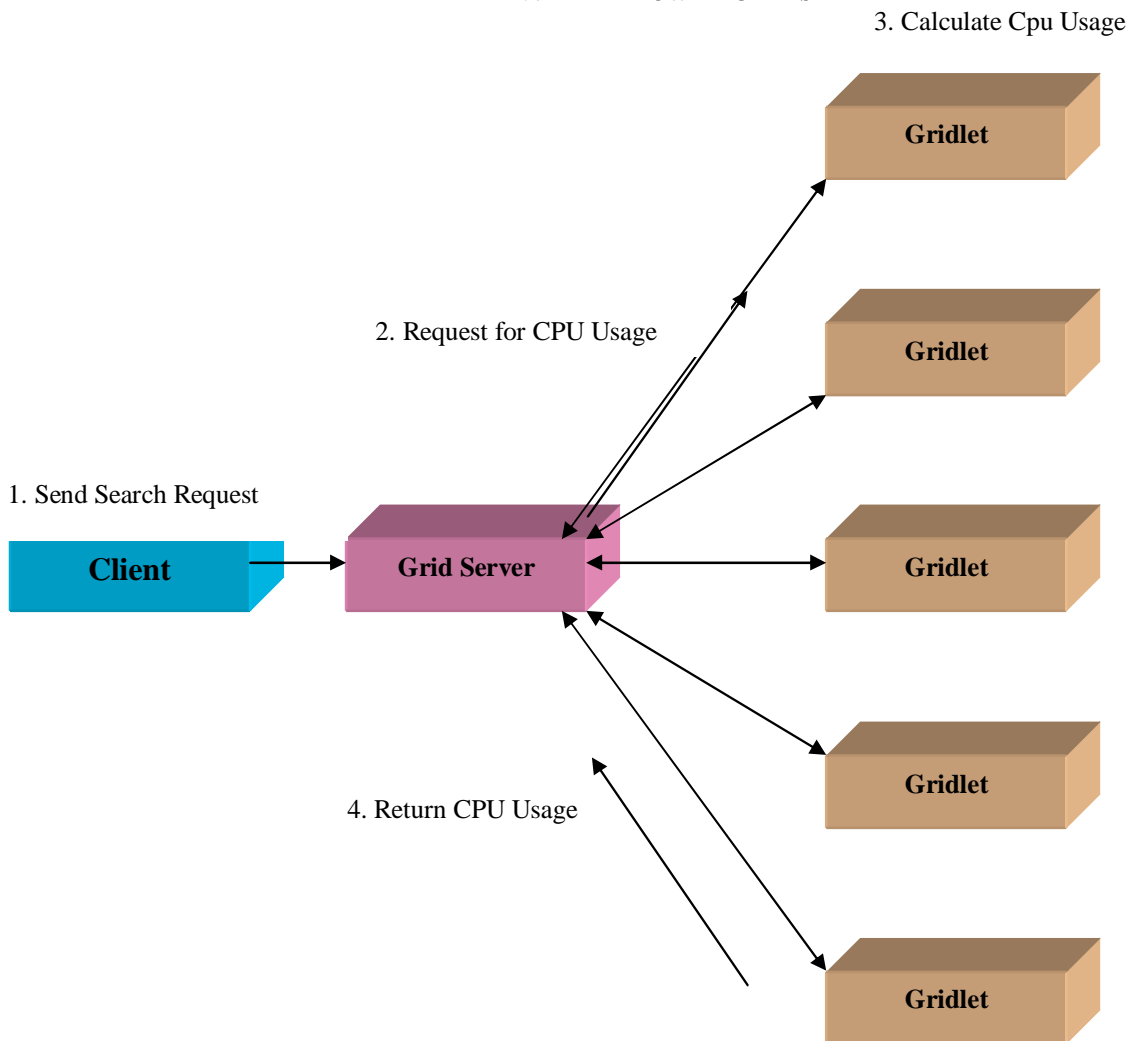


Fig.2. Data Flow Diagram

Fig.2 shows that the client sends request to the server; the server in turn requests the gridlets for their CPU Usage. The gridlets calculate their CPU Usage and returns it to the server.

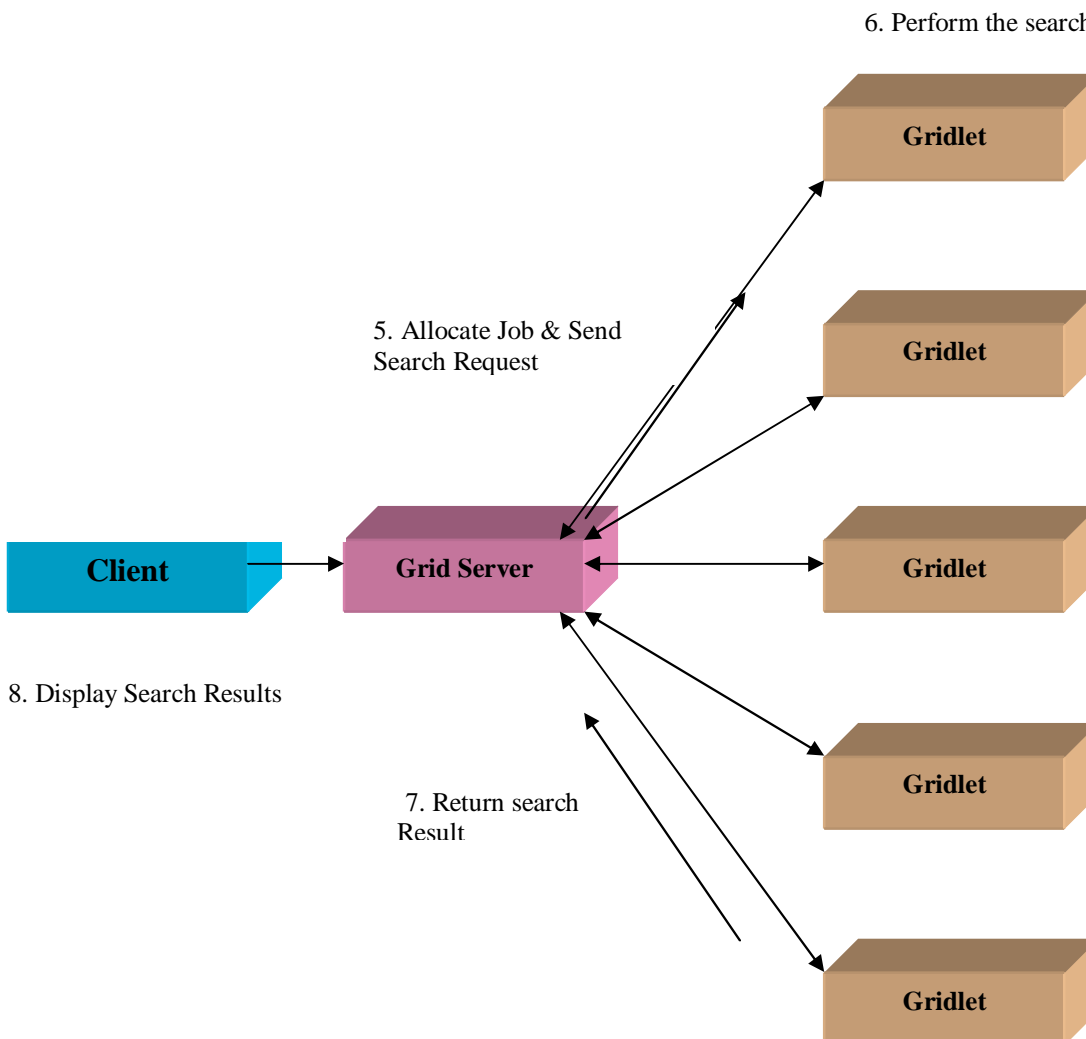


Fig.3. Data Flow Diagram

Fig. 3.shows that the server forwards the request to the selected gridlets; the gridlets in turn perform the search and returns the results to the server. It is then forwarded to the client.

V. IMPLEMENTATION DETAILS

The purpose of this paper is to give a brief description about the various implementation tools, like the languages, the modules created, and the details of each module implementation. The goal of the implementation phase is to translate the design of the system produced during the design phase into code in a given programming language. This involves the details of the methods and interfaces used and their implementation.

VI. CONCLUSION

We have presented the architectural design of a grid which can be used for the purpose of a search engine. The search engine is used for searching a file in the network based on file types. Some of the advantages are portability as it is built in Java. The developed system is capable of working in a dynamic environment with controlled and coordinated resource sharing and multithreading facility. The Architecture has been clearly defined to delegate the tasks [functionality of the Grid Server] for efficient results.



ACKNOWLEDGEMENT

Our most sincere thanks go to our advisor Dr. Yuvaraj B N, Professor, Dept. Of CSE, NIE, Mysore. We thank him for providing us an opportunity to work in the area of Cryptography and Steganography for securing confidential images. We thank him for consistent guidance, encouragement and support during initial development of this paper. He has been helping us to improve our English communication and writing skills.

REFERENCES

Books

- [1] Herbert Schildt. *The Complete Reference Java™ 2*. New Delhi: Tata McGraw-Hill Publishing Company Limited, Edition 2002.
- [2] Steven Holzner. *Java™ 1.2*. New Delhi: BPB Publications, Edition 1998.
- [3] Patrick Naughton. *The Java Handbook*. New Delhi: Tata McGraw-Hill Publishing Company Limited, Edition 2002.

Websites

- [http:// www.GridComputing.com](http://www.GridComputing.com)
- <http://www.globus.org/research/papers/anatomy.pdf>
- <http://www.gridforum.com>